Title

Alzheimer's amyloid-β and the disordered structural ensemble characterized using molecular dynamics and NMR spectroscopy

Author

Ball, Katherine Aurelia

Publication Date

2013

Peer reviewed|Thesis/dissertation

Alzheimer's amyloid-β and the disordered structural ensemble characterized using molecular dynamics and NMR spectroscopy

By

Katherine Aurelia Ball

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Teresa Head-Gordon, Chair

Professor David Wemmer

Professor Berend Smit

Professor Jhih-wei Chu

Spring 2013

Alzheimer's amyloid-β and the disordered structural ensemble characterized using molecular dynamics and NMR spectroscopy

by

Katherine Aurelia Ball

Abstract

Alzheimer's amyloid-β and the disordered structural ensemble characterized using molecular dynamics and NMR spectroscopy

by

Katherine Aurelia Ball

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor Teresa Head-Gordon, Chair

We used simulations and NMR experiments to investigate the diverse structure of amyloid-β (Aβ) peptide in the soluble non-aggregated form in order to better understand this peptide's role in Alzheimer's disease. Because amyloid-β is intrinsically disordered in its monomeric state, the combination of molecular dynamics simulation and NMR spectroscopy was crucial to determining the individual conformations that make up the amyloid-β structural ensemble. Initially we focused on amyloid-β 1-42 (Aβ42), which is the most toxic form of amyloid-β. We collected homonuclear Nuclear Overhauser Effect (NOE) data on the peptide, and used extensive molecular dynamics simulations to characterize its conformational ensemble. We found that the conformational ensemble of Aβ42 is extremely heterogeneous. However, it also contains many structured populations with long-range NOE contacts. This is in contrast to Aβ21-30, an amyloid-β fragment. Aβ21-30 is mostly extended and unstructured, with no long-range NOEs measured. Next we characterized Aβ40, another common form of amyloid-β, which is less toxic and aggregation prone than Aβ42. Again we saw many long-range NOEs and structured conformations in the Aβ40 ensemble, but the most populated conformations for Aβ40 and Aβ42 were quite different. From our simulations we had seen that Aβ42 adopts a β-turn and β-strand, which together form the most common long-range interaction of the peptide, and that this turn is consistent with the same bend and β-strand segment seen in the aggregated form of the peptide. Aβ40 also adopts many different long-range β-strand conformations, however, none of them are similar to the fibril-like turn and β-strand seen in the Aβ42 ensemble. This is one possible explanation for the greater aggregation rate and toxicity of Aβ42.

Amyloid-β presents a difficult case for characterizing an intrinsically disordered disease protein because it contains many structured conformations within its ensemble. We therefore decided to examine the effectiveness of different computational methods for determining the conformational ensemble of this intrinsically disordered protein. We compared the knowledge-based approach to our *de novo* molecular dynamics approach. The knowledge-based approach randomly generates an ensemble and refines it to fit the NMR data. The *de novo* molecular dynamics approach, on the other hand, uses no experimental information to form the amyloid-β ensemble. In both methods, we compare the simulated ensemble to the experimental data after it is created. We found that the knowledge-based approach is highly dependent on the starting pool

of structures that it refines, and that a randomly generated pool does not contain structured conformations which are able to fit the NMR data. We also found that certain types of NMR data, like J-coupling constants and NOEs, do a much better job of distinguishing between vastly different ensembles than other types of NMR data like chemical shifts, which are calculated to be the same for both unstructured and heterogeneous structured ensembles. We did find that the knowledge-based approach was useful for further refining the molecular dynamics simulation ensemble to give a better fit to the NMR data. This refinement yielded a slightly different picture of the Aβ40 and Aβ42 monomer conformational ensembles. The refined Aβ42 ensemble still contains the fibril-like turn and β-strand as its major feature, but in the refined Aβ40 ensemble we see many fewer β-strands than in the molecular dynamics ensemble. Our revised picture of the two peptides shows that Aβ40 is less structured than Aβ42, with the most populated β-strand of Aβ40 forming near its N-terminus. Aβ42, with two additional residues at the C-terminus, forms more C-terminal hydrophobic interactions, often adopting a large loop that nucleates a fibril-like turn and β-strand near the middle of the peptide sequence. Thus, the Aβ42 C-terminus does not form a β-strand itself, but promotes β-structure at a different region of the sequence, while preventing the type of β-strands formed in the Aβ40 ensemble.

After fully characterizing the amyloid-β monomer ensemble, we were interested in studying an oligomer of amyloid-β, which is believed to be the toxic agent in Alzheimer's disease. In collaboration with the Schaffer group, we assessed the toxicity of an Aβ42 oligomer, known as the globulomer, on cultures of human cortical neurons. This oligomer, which can be prepared consistently and does not aggregate to form fibrils, was found to induce neuronal cell death, indicating that it could be a toxic complex of amyloid-β. This led us to an investigation of the Aβ42 globulomer structure, known to consist of β-sheets. One proposed model of the globulomer is based on NMR data from a small globulomer precursor. Another model of the globulomer derives from coarse grain simulations of amyloid-β prefibrils. We used molecular dynamics simulations to begin a comparison of these two models. Based on our preliminary simulations, the prefibrillar model seems to maintain a more stable β-sheet structure than the NMR-based model. However, so far the NMR-based model has only been simulated as a dimer unit, and may be more stable when more chains are added. We have also calculated NMR observables from each of the two models and we find that J-coupling and amide exchange experiments may be useful in determining which model more accurately represents the globulomer. Future NMR experiments as well as calculation of NOE data from the simulations will help to form a better picture of this toxic Alzheimer's oligomer. Like the amyloid-β monomer, the oligomer may occupy a range of conformational states that form a diverse ensemble, and therefore molecular dynamics simulations as well as NMR data are crucial to fully representing its structure.

To my grandfather, John Sigler Ball (August 2, 1914 – December 30, 2012), an inspiring scientist and human being.

Contents

Abbreviations Used

Aβ: amyloid-β

Aβ42: amyloid-β 1-42

Aβ40: amyloid-β 1-40

AD: Alzheimer's disease

APP: amyloid precursor protein

CHC: central hydrophobic cluster

IDP: intrinsically disordered protein

MD: molecular dynamics

REMD: replica exchange molecular dynamics

MRRE: multi-reservoir replica exchange

NMR: nuclear magnetic resonance

NOE: nuclear Overhauser effect

NOESY: nuclear Overhauser effect spectroscopy

TOCSY: total correlation spectroscopy

RDC: residual dipolar coupling

PRE: paramagnetic relaxation enhancement

CD: circular dichroism

CSI: chemical shift index

Rg: radius of gyration

RMSD: root mean squared difference

hPSC: human pluripotent stem cell

hESC: human embryonic stem cell

Acknowledgments

I would like to thank my adviser Prof. Teresa Head-Gordon for all of her help, advice, and support over the past six years as I have developed as a scientist. I would also like to thank Prof. David Wemmer, who has worked closely with me on the experimental NMR portion of the project, has been an invaluable resource, and also provided thoughtful edits to this thesis. Dr. Aaron Phillips, a former member of the Wemmer lab, was instrumental in starting this project and collecting the NMR data. I thank all current and former members of the Wemmer lab for helping me out as I struggled through the experimental aspects of the project. I also acknowledge my collaborators Prof. David Schaffer and Tandis Vazin for the work on human neuronal cell cultures.

Current and former members of the Head-Gordon lab have been key to my completion of this work and well-being during my time in graduate school. I would particularly like to thank Dr. Nicolas Fawzi for helping me with the NMR observable calculations and interpretation. Dr. Paul Nerenberg also provided advice and assistance with the simulation data analysis, and Francesca Vitilini helped with the amyloid-β ensemble analysis. I thank Dr. Alex Sodt for his wisdom and support, Shachi Katira for her friendship and camaraderie, and Asmit Bhowmick, Saurabh Belsare, and Liam O'Suilleabhain for helping build the lab back up again. I have also had the pleasure to work with and mentor several outstanding undergraduate researchers during my time at Berkeley. I would like to thank Richard Zhu and Anna Chen for their outstanding experimental efforts attempting to make a very difficult protein behave. I would like to acknowledge Timothy Balmorez, an impressive undergraduate student, and Sukanya Sasmal, a first-year graduate student, for their work on the amyloid-β oligomer simulations.

I also owe much of my success and happiness here at Berkeley to the Biophysics program. I would like to thank all of the students, faculty, and staff involved in the program for their herculean efforts that keep it functioning so well. In particular Eva Nogales, chair of the program, and Kate Chase, program administrator, have been important assets for me.

The wider scientific community has also been crucial to my scientific development, and I would like to thank the scientists that I have met and talked to at conferences and visits. Particularly my exposure to the Intrinsically Disordered Protein community through a biophysical society subgroup and two amazing Gordon Conferences has been critical to my work with amyloid-β and developing my interests for future research directions.

I would also like to acknowledge my professors and mentors in the physics department at Middlebury College who started me down this path with their wonderful teaching and encouragement.

My family has been the greatest support I could ever hope for, as a person and scientist. From the first dinner-time conversation about atoms, my parents have always encouraged me in my scientific pursuits, sharing in my excitement and reassuring me when I encountered frustration or disappointment. To my mom, Cynthia Lanners, thank you for showing me how to be a strong, intellectual woman. To my dad, Gene Ball, thank you for showing me that the most

interesting scientific phenomena are the ones we don't understand. Watch out, you're not the only PhD in the family anymore! Thanks to my brothers, Erick Ball and Ben Ball, for showing me I was doing something right by following in my path (two more PhDs to come!), but always making your choices and accomplishments your own, exploring scientific and technical problems beyond anything I've encountered. A big thanks also to Kelli Carnes, who is like a sister and has always been there for a hug and a good listen through it all.

Semihcan Doken, my partner, has been there for me through the six years of my PhD, always showing his constant support and forming my family here in Berkeley for the past two years. I thank him for the cups of tea late at night working, for the smiling face at the end of a hard day, and mostly for being there anytime to talk when I needed it. I thank him for encouraging my choices and goals and having faith that I can achieve them. I know I can count on this support during all my future endeavors.

Chapter 1

Introduction

Unlike many other diseases, Alzheimer's disease is not caused by an infectious external agent, but by a malfunction of the normal cellular processes that take place in the brain. Amyloid-β is the key protein involved in Alzheimer's Disease (AD), causing damage to neurons and debilitating symptoms in AD patients as part of a progressive neurodegenerative disorder (1, 2). In order to better understand this illness, it is imperative to thoroughly characterize the central player, amyloid-β (Aβ). This thesis investigates the underlying physics of Aβ and how physical properties may influence its behavior in the context of Alzheimer's disease. Below we provide an over view of the role of Aβ in AD, known characteristics of Aβ, and the experimental and computational techniques we employ to study Aβ. We conclude with a short summary of each chapter.

Role of Aβ in Alzheimer's Disease

Aβ is a peptide generated when Amyloid Precursor Protein (APP) is cleaved by two enzymes, β- and γ-secretase. APP is a transmembrane protein, and γ-secretase cleaves within the cell membrane, so Aβ is made up of what were the transmembrane and extracellular regions of APP. β-secretase can cleave APP at variable locations within the membrane, and therefore Aβ presents as peptides of various lengths, the most common of which are amyloid-β 1-42 (Aβ42) and amyloid-β 1-40 (Aβ40). The only difference between the Aβ42 and Aβ40 amino acid sequences is the two additional residues that Aβ42 has at the C-terminus, which are absent in Aβ40. Aβ42 is present at higher levels in the brains of AD patients than in healthy brains, although the reason for this is still unclear (1, 2).

The brains of AD patients are characterized by large plaques, composed of long, insoluble fibrils (3). These fibrils, in turn, are composed of aggregated Aβ peptides. Both Aβ40 and Aβ42 are present in these aggregates, although Aβ42 is more aggregation prone and more toxic than Aβ40 (4-10). Aβ peptides can also form smaller soluble oligomeric structures, either as precursors to the insoluble fibrils or as an alternate pathway (11). These soluble oligomers are now believed to be more toxic than the fibrils, however the mechanism of toxicity is unknown (12, 13). Characterization of the individual Aβ peptides can provide physical information that is important to understanding both oligomer and fibril formation, as well as explaining differences between Aβ40 and Aβ42.

Intrinsic Disorder of Aβ

Monomeric Aβ is an unusual protein because it does not form a stable folded structure as the typical functional protein does. Instead, both Aβ40 and Aβ42 have been characterized as intrinsically disordered proteins (IDPs) (14, 15), which fluctuate between different conformational states, frequently sampling different dihedral angles along the entire protein. IDPs are fundamentally flexible and cannot be described by a small number of folded conformations. While a natively folded protein is energetically trapped in a single state with high barriers to unfolding, an IDP can sample a "landscape" of energetic states, hopping between local minima. There is no deep energy minimum much lower on the IDP landscape, and therefore it tends to spend similar amounts of time in many weakly stabilized states (16). Experimentally IDPs can be identified using a variety of techniques that indicate lack of stable secondary and tertiary structure, flexibility, and exposure to solvent (17, 18).

Aβ has no evidence of stable secondary or tertiary structure in the monomeric state (14, 15, 19, 20). This makes it particularly difficult to characterize the Aβ peptide. The Aβ fibril does have stable tertiary structure, forming a repetitive intermolecular β-sheet that extends in the direction of fibril growth, with each individual chain running perpendicular to the fibril axis (21). The Aβ oligomer is less well characterized than the fibril, but seems also to have stable β-sheet structure, although it may still have transient, disordered tertiary contacts (11, 22). It is unknown whether any of this fibrilar or oligomeric structure is present in the Aβ monomer.

Molecular Dynamics

Insights into biological systems that are difficult to probe experimentally can be provided by computational simulation using molecular dynamics (MD). This technique employs physical theories and energy functions to calculate the expected velocities on each particle in a system, then incrementally moves the entire system forward through time, producing an incredibly detailed movie of the system's evolution. We can perform these simulations at atomic resolution, yielding a detailed structure of a protein molecule at many points in time. The energy functions used in these simulations require approximations, but are designed to accurately reproduce experimental data (23-26). These energy functions describe the energy landscape of the system, and MD allows all states on this landscape to be sampled as the system evolves, generating information about what structures are possible, the relative populations of each structure, and the timescales of molecular motion.

One of the limitations of MD simulations is the accessible timescale. While a complicated system may only access certain states on the timescale of milliseconds, most atomistic MD simulations have only just begun to access this timescale (27, 28). Short simulations can result in sampling that is restricted to a small area of the energy landscape, so that only configurations corresponding to one energy minimum are observed although experimentally many more states are populated by the system. Advanced techniques can be used to improve sampling. One such technique is replica exchange molecular dynamics (REMD), which uses many parallel simulations at multiple temperatures to gain access to kinetically restricted areas of the energy landscape (26, 29-31). The high temperature system traverses energy barriers more easily than those at lower temperatures, and a Monte Carlo acceptance

criterion is used to pull structures from the higher temperature and correctly populate the lowest desired temperature with states corresponding to a Boltzmann distribution. Using such a technique, MD simulations may generate a set of states corresponding to all likely configurations of a complicated system, representing each individual structure in atomic detail.

NMR Spectroscopy

The results of protein simulations should always be employed in conjunction with experiments to ensure accuracy and allow for improvement of the simulation theory. Nuclear magnetic resonance (NMR) spectroscopy can be used to collect information on protein structure and dynamic at the atomic level. One benefit of NMR compared to other techniques in structural biology is that it can be performed on a sample in native conditions. Protein samples are dissolved in buffer and placed in a very strong magnetic field where electromagnetic pulses can be used to probe the interactions between the atomic nuclei in the protein. These experiments can be performed with no modifications to the protein that would cause it to adopt configurations different from those present *in vivo*.

There are many different types of information about a protein that can be obtained from a single or multiple NMR experiments. Chemical shift data and through bond-coupling data, such as J-coupling constants, report on the secondary structure of the protein. Tertiary structure information is available from through-space coupling interactions like the Nuclear Overhauser Effect (NOE), as well as more complex experiments such as residual dipolar coupling. Because the relaxation rates of atomic nuclei are affected protein motion, dynamic information is also available from NMR. Combining all of these data provides an extensive characterization of the overall fold and secondary structure of a protein as well as which areas are more flexible and which are very rigid. These NMR observables can only be measured as an average over many individual protein molecules in the experimental sample, and since protein motion can be faster than the NMR experiment, time averaging also occurs.

Summary of Chapter 2

Many previous studies of Aβ have focused on small fragments rather than the full-length 40 or 42 residue peptide that is associated with Alzheimer's disease. These fragments are easier to simulate and work with experimentally while still exhibiting the aggregation behavior seen in the full-length peptide (19, 32-47). One example is Aβ21-30, which, like the longer Aβ42 peptide, is designated as an intrinsically disordered protein (26). In Chapter 2 (published in (48)) we use atomistic MD simulations and a variety of NMR experiments to characterize the Aβ42 disease peptide, comparing it to the Aβ21-30 fragment which was previously characterized using similar methods (26). We find that although both peptides are IDPs, their properties are very different.

Aβ21-30 has few possibilities for favorable inter-residue contacts, and therefore the majority of Aβ21-30 conformations are extended and unstructured, similar to a random coil ensemble. In contrast Aβ42 samples a diverse ensemble of partially structured conformations including α-helices, β-turns, and β-strands. The differences between the two peptides are especially apparent from the NOE data, which shows very few contacts between Aβ21-30 residues that are sequentially separated. Aβ42, however, generates many NOEs that are due to

long-range tertiary interactions only possible in a structured conformation. Although, Aβ42 resembles a random coil by some measures, it clearly exhibits significant structure and cannot be described so simply. Some of the structural features present in the Aβ42 ensemble may also be relevant for the formation of the toxic Aβ fibrils and oligomers. Thus it is crucial to study the more complex full-length Aβ42 protein in the context of Alzheimer's disease, rather than the simpler, unstructured Aβ21-30 fragment.

Summary of Chapter 3

The comparison of Aβ42 and Aβ21-30 is of interest from the perspective of IDP physics, however, differences between Aβ40 and Aβ42 are more relevant to Alzheimer's disease biology. Although the two peptides are almost identical in sequence, differing only be two C-terminal amino acids, Aβ42 is much more toxic and aggregation prone (4-10, 49). In Chapter 3 we extend the methods from Chapter 2 to compare the conformational ensembles of the Aβ40 and Aβ42 monomers. We find that the additional two residues on the Aβ42 C-terminus (Ile41 and Ala42) directly prevent structural features seen in the Aβ40 ensemble and promote structure that is unique to Aβ42.

NMR data including, especially NOE contacts, are interpreted using conformational ensembles generated from *de novo* MD simulations for both Aβ40 and Aβ42. These MD ensembles both contain many structural features and point to major differences in the β-strand populations for the two peptides. While Aβ42 often forms a β-strand with a turn that is also present in the Aβ fibril structure, Aβ40 forms a wide variety of other, non-fibril-like β-strands. These Aβ40 β-strands do not form in the Aβ42 ensemble because they are replaced by hydrophobic contacts with the C-terminus. The two Aβ42 C-terminal residues also promote the fibril-like turn when they form a C-terminal hydrophobic cluster. Thus, the purely MD generated Aβ conformational ensembles provide a plausible picture of why Aβ42 might form toxic AD fibril and oligomer structures more easily than Aβ40.

Summary of Chapter 4

Although *de novo* molecular dynamics presents an attractive method for constructing the complete conformational ensemble of an intrinsically disordered protein such as Aβ, several other techniques are also commonly used in the IDP field. These techniques, which we designate "knowledge-based approaches" incorporate experimental data directly in the creation of the IDP ensemble rather than including experimental data only as a validation of a computationally generated ensemble (50-63). One main knowledge-based method relies on generating a large pool of statistical coil IDP conformations, from which a smaller ensemble is selected based on agreement with experimental data (50, 52, 54, 58-63).

In Chapter 4 (published in (64)) we compare the performance of *de novo* MD and this knowledge-based method when constructing IDP conformational ensembles for Aβ40 and Aβ42. We also examine various types of NMR experimental data in terms of their ability to discern between vastly different IDP conformational ensembles. We find that the knowledge-based method depends crucially on the characteristics of the starting pool of structures. The

knowledge-based approach cannot successfully reproduce the ensemble of an IDP with significant structure such as Aβ if the starting structure pool resembles an extended random coil. A combined MD and knowledge-based approach, where *de novo* MD generated structures form a starting pool that is subsequently refined by NMR data, emerges as the best method of generating an experimentally accurate Aβ ensemble.

Summary of Chapter 5

After comparing several different possible ensembles that could represent the Aβ40 and Aβ42 IDPs in Chapter 4, we found that although the *de novo* MD generated ensembles performs well with respect to reproducing the experimental NMR data, the best Aβ40 and Aβ42 ensembles are created by further refining the MD ensembles against the NMR data itself. In Chapter 5 (published in (65)) these NMR-refined MD Aβ ensembles are examined from a structural and biological disease perspective. As in the *de novo* MD ensembles characterized in Chapter 3, we find that both Aβ40 and Aβ42 adopt many partially structured conformations within their ensembles and that there are key differences between the Aβ40 and Aβ42 ensembles.

It is experimentally difficult to distinguish structural differences between the Aβ40 and Aβ42 monomers from most types of NMR data. Homonuclear NOESY spectra of the two peptides show us that the long-range inter-residue contacts made be structures in the two ensembles differ. Heteronuclear NOE intensities also provide experimental evidence that Aβ40 has a more flexible backbone than Aβ42 at the C-terminus (66, 67). In Chapter 5 we investigate the specific structures that explain these observed differences. As in the *de novo* MD ensemble we see that Aβ42 forms a fibril-like turn, with β-strands on either side, structures that are absent from the Aβ40 ensemble. In fact, in the NMR-refined Aβ40 ensemble we observe few β-strand conformations. The role of the Aβ42 C-terminal residues, while still key, is also adjusted somewhat in the NMR-refined ensemble. We see that the Aβ42 C-terminus is more ordered than that of Aβ40 because it engages in hydrophobic side chain contacts at the C-terminus and in the middle of the protein. This long-range interaction induces the formation of the Aβ42 β-turn and β-strands. Again this prominent β-turn present in a subset of Aβ42 structures stands out as a possible precursor to toxic structures observed in Aβ fibrils and oligomers.

Summary of Chapter 6

Although our studies have focused primarily on the monomeric forms of the Aβ peptides associated with Alzheimer's disease, the toxic form of the peptide is believed to be a small soluble oligomer (12, 13). However, many different oligomers of Aβ have been observed experimentally, and it is difficult to isolate a single species with particular size and structure from the distribution of oligomers that form during Aβ aggregation. One proposed toxic oligomer that can be prepared consistently is the Aβ42 "globulomer" (68), however its effect on human neurons has not been specifically characterized. In Chapter 6 we collaborate with the Schaffer group (69, 70) to examine the toxicity of the globulomer (relative to the Aβ42 monomer) toward human cortical neurons. Our collaborators have derived cultures of these neuronal cells from human pluripotent stem cells, and employed them as an *in vitro* model for the study of AD. We see that globulomer binds to these neurons inducing cell death, and is

selective for glutamatergic over GABAeric neurons, consistent with AD toxicity patterns in patients. We also find that the globulomer exhibits culture age dependent cell binding. These results support the proposed role of the Aβ42 globulomer as a toxic agent contributing to AD pathology.

Summary of Chapter 7

Because the Aβ42 globulomer seems to be a toxic Aβ oligomer, knowledge of its molecular structure is important to understanding disease mechanisms and developing possible drug therapies that could disrupt these mechanisms. Some general structural information is known about the globulomer, for example it contains about 12 monomer chains and is has β-sheet secondary structure. One NMR study of the globulomer has been performed, however most the of the data collected reported on a smaller oligomer called the "preglobulomer" which is a precursor to the globulomer and believed to be similar in structure (71-75). In Chapter 7, we present our initial work evaluating two different models of the Aβ42 globulomer using all-atom molecular dynamics simulations. The first model is based on the preglobulomer NMR data, while the second model is a dodecamer generated from coarse grain simulations of the Aβ40 protofibril (69).

From simulations of the dimer repeating unit, we find that the preglobulomer model does not maintain stable β secondary structure at the C-terminus, where it forms a parallel β-strand, however this structure may be more stable when more chains are present in the full preglobulomer or globulomer structures. The dodecamer protofibril model does maintain β-sheet structure when it is converted from the coarse grain model and equilibrated in all-atom simulations. We have also begun to calculate NMR observables for both globulomer models that we can compare to existing and future NMR data. Estimated amide protection factors for the preglobulomer dimer model appear to be lower than those seen experimentally. Chemical shifts and J-coupling constants have been calculated for the protofibril model, but have not yet been quantitatively compared to experiments. Ongoing analysis of MD simulation data for both models will give a better picture of the possible Aβ42 globulomer structures, and future NMR experiments can determine which model is the most accurate. Combined with our characterization of the Aβ monomer ensembles, we will then have the basis for understanding the formation of a toxic species associated with Alzheimer's disease.

Chapter 2

Homogeneous and heterogeneous tertiary structure ensembles of Aβ 21-30 and Aβ42 peptides

The interplay of modern molecular simulation and high quality nuclear magnetic resonance (NMR) experiments has reached a fruitful stage for quantitative characterization of structural ensembles of disordered peptides. Amyloid-β 1-42 (Aβ42), the primary peptide associated with Alzheimer's disease, and fragments such as Aβ21-30, are both classified as intrinsically disordered peptides (IDPs). We use a variety of NMR observables to validate *de novo* molecular dynamics simulations in explicit water to characterize the tertiary structure ensemble of Aβ42 and Aβ21-30 from the perspective of their classification as IDPs. Unlike the Aβ21-30 fragment which conforms to expectations of an IDP that is primarily extended, we find that Aβ42 samples conformations reflecting all possible secondary structure categories and spans the range of IDP classifications from collapsed structured states to highly extended conformations, making it an IDP with a far more heterogeneous tertiary ensemble.[1]

Introduction

Intrinsically disordered proteins (IDPs), polypeptides that do not adopt a single or dominant tertiary conformation, are abundant in eukaryotic proteomes, comprising 25-30% of protein coding sequences found in the human genome (71, 73-75). Characterizing IDP structural diversity is important to understanding their role in functional signaling, regulation, and transcription, as well as their detrimental involvement in pathologies such as cancer and neurodegenerative diseases. Their characterization has presented new challenges in structural biology since traditional experimental techniques such as X-ray crystallography and electron microscopy cannot yield the same detailed structural information that has been paramount in our understanding of folded enzymes and well-formed protein complexes (73).

[1] Reproduced from (48) with permission.

Our current understanding of IDPs comes in part through contrasting their sequence and structural characteristics with those of folded proteins (73). For example, IDPs sequences are often significantly lower in hydrophobic residue content, have an over-representation of polar and/or charged amino acids, and have low sequence complexity compared to globular, folded proteins (74, 76). IDPs often have little secondary structure compared to the (non-denatured) unfolded ensembles of structured protein at native or near native conditions that retain α-helical and β-strand segments. The structural diversity of IDPs suggests that conformational sub-populations interconvert at a much faster rate than the folded to unfolded transition timescale of a structurally ordered protein. While folded proteins can be usefully categorized into classes based on well-ordered secondary or tertiary structure, IDPs are classified by their degree of disorder, i.e. whether they predominately are collapsed semi-ordered ensembles ("structured" molten globules or pre-molten globule states), collapsed disordered ensembles, or extended disordered ensembles (71, 77). These classification schemes have been useful for IDPs with known function.

Amyloid-β 1-42 (Aβ42) is the primary peptide associated with Alzheimer's Disease (AD), generated in the amyloidogenic pathway by proteolytic cleavage of the transmembrane Amyloid Precursor Protein (APP) by membrane-anchored β- and γ-secretases. While Aβ appears to interconvert between α-helical and random coil populations when part of the APP integral membrane protein (78), the cleaved Aβ42 peptide and fragments such as Aβ21-30 have been categorized as intrinsically disordered peptides (IDPs) (14, 15). Although both are classified as IDPs, it would be useful to know whether their structural ensembles are similar or dissimilar so that we might usefully interpret the physiological behavior of the full-length peptide or the physiological significance of the smaller Aβ21-30 fragment.

The interplay of modern molecular simulation and high quality nuclear magnetic resonance (NMR) experiments has reached a fruitful stage for quantitative characterization of structural ensembles of disordered peptides, which we review here. We use a variety of NMR observables to validate *de novo* molecular dynamics simulations in explicit water to contrast the *tertiary structure* ensemble of Aβ21-30 and Aβ42. We show that Aβ42 samples conformations reflecting all possible secondary structure categories and spans the range of IDP classifications from collapsed structured states to highly extended conformations, while Aβ21-30 has a relatively homogeneous extended structure ensemble, showing that the two amyloid-β peptides are very different classes of IDP.

Methods

NMR Experiments

The peptide preparation and NMR experiments on Aβ21-30 have been reported in a previous publication (26). Here we report on the experimental details of Aβ42. Recombinant Aβ42 peptide was purchased from a commercial source (rPeptide, Athens, GA) reported to be > 97% purity. The peptide was lyophilized from trifluoroacetic acid (TFA). The alkaline pretreatment of Aβ and preparation of LMW Aβ by filtration protocols outlined by Teplow (79) were used to prepare a monomeric solution of Aβ42. The lyophilized peptide was dissolved in 2 mM NaOH to produce a peptide concentration of 0.21 mg/ml with a pH of > 10.5. The peptide was then

sonicated for 1 min in a bath sonicator and lyophilized. It was then resuspended in 20 mM sodium phosphate buffer, pH 7.2, 0.01% (w/v) sodium azide. This protocol ensures that when the lyophilized peptide is dissolved in buffer it will not pass through its pI of ~5.31. Before collecting NMR data, the sample was filtered with a 0.22 μm filter to remove any fibril seeds and brought to a concentration of ~600 μM, pH 7.2.

NMR data were collected at 287 K at the NMR Facility at UC Berkeley on Bruker Avance II 800 and 900 MHz spectrometers, the latter equipped with a Bruker cryogenic probe. The data were processed using NMRPipe and peaks were assigned and analyzed using CARA (80). Chemical shifts were assigned using 2D ^{1}H-^{1}H Total Correlation Spectroscopy (TOCSY) (81, 82) and Nuclear Overhauser Effect Spectroscopy (NOESY) spectra (83-85). The TOCSY spectra were collected in a 90% H_2O (10% $^{2}H_2O$) buffer solution with a mixing time of 60 ms and in a 99% $^{2}H_2O$ buffer solution with a mixing time of 80 ms. The NOESY spectra were collected in a 90% H_2O (10% $^{2}H_2O$) buffer solution with mixing times of 100 ms and 200ms, and in a 99% $^{2}H_2O$ buffer solution with a mixing time of 200 ms. In the direct dimension, 1024-2048 points were collected while 256 complex points were collected in the indirect dimension for all spectra. The spectral width was 12 ppm in each dimension with 16-48 scans. 1D spectra were collected at time points before and after the experiments to measure the degree to which the monomeric peptide signal decreased due to aggregation. We also ensured that our NOESY data was resulting only from a monomeric form of the peptide by collecting a NOESY spectrum immediately after dissolving the peptide and another NOESY spectrum ~48 hours after dissolving the sample. We saw no difference between the two spectra except for a decrease in peak intensity due to aggregation, which eliminates the possibility that some NOEs could be a result of oligomers forming during peptide incubation. Peaks were identified at particular chemical shifts in the 200 ms mixing time NOESY spectra in both H_2O and $^{2}H_2O$, and those peaks that could be unambiguously assigned to two specific hydrogen atoms were used to calibrate the NOE peak intensities derived from simulation. The majority of peaks could not be clearly assigned to a unique pair of hydrogen atoms due to spectral overlap. These peaks were instead assigned a list of potential hydrogen pairs consistent with the observed peak frequencies. This list is made up of atoms within a 0.04 -0.08 ppm range around each observed peak.

De novo molecular dynamics simulations

The *de novo* molecular dynamics simulations of Aβ21-30 have been reported in a previous publication (26). Here we report on the simulation details of Aβ42. The peptide was modeled using the Amber ff99SB force field (23) and aqueous solvent represented by the TIP4P-Ew water model (25). We chose the ff99SB force field and TIP4P-Ew water model combination because it was used in the previous Aβ21-30 study (26), and subsequent studies support its clear superiority relative to other biomolecular simulation force fields (24, 26, 86-88).

In order to calculate equilibrium ensemble averages in the NVT ensemble, we used an Andersen thermostat, a leapfrog integrator with a 1.0 fs time step, and periodic boundary conditions. Particle Mesh Ewald was used for calculating long-range electrostatic forces, and a cutoff of 9.0 Å was used for the real space electrostatics and Lennard-Jones forces throughout the study. The initial structure for Aβ42 was built in an extended form using the *LEaP* module that is part of the AMBER package. Each structure was solvated in a water box such that there

were 10 Å of water surrounding the peptide on all sides (15,142 molecules of water, 580 nm^3 box volume), and three Na+ ions to neutralize the charge of the peptide. The structure was minimized and equilibrated with constant volume while raising the temperature to 300 K, then equilibrated for 2 ns under a constant pressure of 1 bar, maintained with a Berendsen barostat, to achieve the correct density. Then, a 2 ns 498 K simulation was run on the extended Aβ42 peptide to obtain a more collapsed starting structure for equilibrium ensemble simulations. Two different collapsed structures were chosen to start two independent replica exchange simulations. These systems were prepared by first removing the solvent, and then the previous solvation and equilibration steps were used to create a final box that contained 6,251 water molecules (193 nm^3 volume). The *sander* module of AMBER and the Multi-Reservoir Replica Exchange (MRRE) method presented by Ruscio *et al*. (14, 31) were used to achieve a Boltzmann weighted ensemble of 100,000 Aβ42 structures at the experimental temperature of 287 K, from 100 ns of equilibrated simulation. Details of the MRRE calculations for Aβ42 are reported in the Appendix.

In order to calculate time-correlation data on the 287 K reservoir, selected structures from this ensemble were used to run microcanonical (NVE) ensemble simulations using the pmemd module of AMBER10. 50 structures were chosen from each 287 K Aβ42 ensemble (after the 10 ns of equilibration), spaced 1 ns apart (ensuring complete decorrelation through swaps between replicas), and equilibrated at constant 287 K for 100 ps before running 20 ns of simulation for each initial structure. In total, the Aβ42 constant energy simulation time with starting conformations equilibrated at 287 K was 2 μs.

Calculation of NMR Observables

For both Aβ21-30 and Aβ42 we evaluate NMR chemical shifts and scalar-coupling constants (66, 67, 89), NOE peak intensities for not only backbone amide proton contacts (14) but the full set of hydrogen contacts that were measured by us, using the methods described in Fawzi *et al.* (26). For Aβ42 we also evaluate residual dipolar couplings (66, 67, 90).

Chemical shifts. We used the program SHIFTS (91) to calculate chemical shifts for H_α and amide hydrogen atoms and carbon chemical shifts. These values were compared directly to the chemical shift (in ppm) assigned to each hydrogen and carbon atom from the NMR spectra reported here, without any normalization. The C_α and C_β chemical shifts were also calculated and compared to experimental values from Hou *et al.* (14).

Scalar couplings. The *ptraj* module of AMBER was used to calculate the ϕ angles for each residue of Aβ21-30 or Aβ42 at every snapshot in the 287 K ensemble, and the J-coupling constant, $J(\phi)$ was evaluated for each snapshot using the Karplus equation (92)

$$J(\phi) = A\cos^2(\phi - 60) + B\cos(\phi - 60) + C \tag{2.1}$$

with coefficients A = 6.51, B = -1.76, and C = 1.60 corresponding to the original parameter set by Vuister and Bax (93). We note that the ϕ values used to calculate the J-coupling constants correspond to the dihedral angle defined by C^{i-1}, N^i, C_a^i, C^i, although the J-splitting seen experimentally is a result of coupling between two hydrogen atoms, related by the dihedral angle defined by HN^i, N^i, C_a^i, H_a^i. These two dihedral angles are geometrically related by a phase

factor of ~ 60° as per Eq. (2.1). This discrepancy was not properly accounted for in a previous publication, Fawzi *et al.* (26), and we show the corrected calculation in Results. To obtain an overall metric of how well the calculated coupling constants correspond to the experimental data, and compare this to other simulated systems, we calculate the figure of merit

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(\left\langle J_i \right\rangle_{sim} - J_{i,\exp t}\right)^2}{\sigma^2} \tag{2.2}$$

where N is the number of J-coupling constants measured, $\langle J_i \rangle_{sim}$ is the i^{th} calculated j-coupling constant averaged over all structures in the simulated ensemble, $J_{i,expt}$ is the i^{th} experimental J-coupling constant, and σ^2 is the uncertainty in $\langle J_i \rangle_{sim}$, which we expect to be dominated by the uncertainty in the Karplus parameters.

ROESY and NOESY Intensities. We use the method of Peter *et al*. (94) to calculate the spectral density function (which is related to the NOE or ROE peak intensity) from the short NVE simulations. Specifically, we use *ptraj* to analyze the NVE simulations. The *ptraj* output is the normalized correlation function for each pair of hydrogen atoms of the Aβ peptides (the normalization option is convenient as the values of the correlation function, output as decimal numbers with limited field width by *ptraj* are then all available to high precision, which is especially critical for distant ^{1}H-^{1}H pairs). These are calculated according to

$$C(\tau) = \left\langle \frac{1}{r^6(t)} \right\rangle^{-1} \left\langle \frac{P_2\left(\cos\chi_{t,t+\tau}\right)}{r^3(t)r^3(t+\tau)} \right\rangle \tag{2.3a}$$

where $r(t)$ is the vector between each hydrogen atom pair at time t, $\chi_{t,t+\tau}$ is the angle between the $r(t)$ and $r(t+\tau)$ vectors, and P_2 is the second order Legendre polynomial. We multiply these normalized correlation functions by $\langle r^{-6}(t) \rangle$, averaged over the individual trajectory, to obtain the expression

$$\left\langle \frac{P_2\left(\cos\chi_{t,t+\tau}\right)}{r^3(t)r^3(t+\tau)} \right\rangle \tag{2.3b}$$

for each of the independent 100 20-ns NVE simulations and average these values together, then re-normalize the average correlation function by dividing by $\langle r^{-6}(t) \rangle$, averaged over all NVE simulations (equivalent to dividing the entire correlation function by the value at t=0) . Each correlation function is then fit over a 5 ns range for τ to a multi-exponential form

$$C(\tau) \approx \sum_{i=1}^{N} a_i e^{-t/\tau_i} \tag{2.4}$$

with N = 1, 2, 3, or 4 and

$$\sum_{i=1}^{N} a_i = 1$$

using the MATLAB function *lsqcurvefit*, which fits curves in a least-squares sense (The MathWorks, Natick, MA). The value for N is determined using Mallow's C_p statistic (95),

$$\hat{R}(S) = \hat{R}_{tr}(S) + 2k\hat{\sigma}^2 \tag{2.5}$$

where S is the set of N covariates, R_{tr}, the training error, is the square difference between the correlation values and the fit exponential function, k is the number of regressors ($2N$ - 1), and σ^2 is the variance under the full model, which we estimate as 0.001 based on a selected number of spin vectors. The value of N that yields the lowest value for Mallow's C_p statistic is used as the fit, which is a better criterion for model selection than simply the training error, which is biased towards higher values of N.

The resulting time correlation function can be Fourier-transformed

$$J(\omega) = \int_{-\infty}^{\infty} C(\tau) e^{i\omega\tau} d\tau \tag{2.6}$$

to yield

$$J(\omega) = \frac{1}{5} \sum_{i=1}^{N} \left[a_i \frac{2\tau_i}{1 + \omega^2 \tau_i^2} \right] \tag{2.7}$$

following the convention for Fourier-transforms of exponentials. The correlation time constants, τ_i, have a practical upper bound imposed by the rotation of the entire peptide in solution. No vectors may remain correlated at timescales longer than the rotational correlation time of the entire molecule in the slowest dimension. In cases where the average correlation function for the vector between two hydrogen atoms is fit with a time constant greater than 15 ns (around 5% of cases), we replace the long time constant with a time constant of 15 ns, which is a reasonable value for the longest possible rotational time-scale, based on the Debye-Stokes theory for calculating the rotational correlation time of a spherical molecule,

$$\tau = \frac{4\pi a^3 \eta}{3kT} \tag{2.8}$$

where a is the molecule's radius, η is the viscosity of the solvent (in this case water), k is the Boltzmann constant, and T is the temperature (287 K). The average Aβ42 radius of gyration in our simulated ensemble is 13.0±4.5 Å. The most compact structures (radii of gyration of 8.5 Å) correspond to τ values of 0.8 ns using a spherical molecule, while the most extended (radii of gyration of 17.5 Å) correspond to τ values of 7.4 ns, less than half of our 15 ns limit on rotational correlation time. This limit on the correlation time has a physical basis, because global tumbling, independent of any local angle or distance fluctuations, causes a complete loss of correlation in

C(τ) as defined in Eq. (2.3a). We note that the traditional method for disordered peptides of using the same fast motional correction for all spin pairs was not sufficient to accurately predict the Aβ21-30 ROESY peaks, so we were obliged to calculate correlation times and spectral densities from NVE simulations for both peptides.

For Aβ21-30 we predict the ROESY spectra from our structural ensemble and dynamical trajectories by calculating the intensity

$$I(t_{\text{mix}}) = Xe^{-\Lambda t_{\text{mix}}} X^{-1} I(0) \tag{2.9}$$

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

$$\rho_{ii} = \sum_{j=1,\neq i}^{n} \frac{1}{10} K^2 \left[\frac{3}{2} J_{ij}(2\omega_0) + \frac{9}{4} J_{ij}(\omega_0) + \frac{5}{4} J_{ij}(0) \right] \tag{2.10a}$$

and off-diagonal elements

$$\sigma_{ij} = \frac{1}{10} K^2 \left[\frac{3}{2} J_{ij}(\omega_0) + J_{ij}(0) \right] \tag{2.10b}$$

where ρ is the direct dipolar relaxation rate and σ is the cross-relaxation rate for all proton pairs, and the constant factor *K* is defined as

$$K = \frac{\mu_0}{4\pi r_{eff}^{\ 3}} \hbar \gamma_a \gamma_b \tag{2.11a}$$

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. The effective *r* vector

$$r_{eff} = \left\langle \frac{1}{r^6(t)} \right\rangle^{-1/6} \tag{2.11b}$$

is the average of the r^{-6} values, which has then been converted back into units of distance. For Aβ42 we predict the NOESY spectra from our structural ensemble and dynamical trajectories by calculating the intensity from a different relaxation matrix **R**, composed of diagonal elements

$$\rho_{ii} = \sum_{j=1\neq i}^{n} \frac{1}{10} K^2 \left[3J_{ij}(2\omega_0) + \frac{3}{2} J_{ij}(\omega_0) + \frac{1}{2} J_{ij}(0) \right] \tag{2.12a}$$

and off-diagonal elements

$$\sigma_{ij} = \frac{1}{10} K^2 \left[3J_{ij}(\omega_0) - \frac{1}{2} J_{ij}(0) \right] \tag{2.12b}$$

We simulate all hydrogen atoms explicitly for each methyl group and hence calculate all pair correlation functions, including neighboring methylene and methyl group protons, and we ignore water proton coordinates as is the standard assumption in the NMR experiment.

We solve for the magnetization matrix at the mixing times that were used in both H_2O and ^2H_2O ROESY and NOESY experiments, removing all exchangeable hydrogen atoms to simulate ^2H_2O buffer conditions. For the H_2O experiment, we multiply all amine intensities by 0.9 to account for the presence of 10% ^2H_2O, and remove all basic amine and hydroxyl groups, which would not appear due to broadening caused by exchange with solvent. We then add all of the degenerate intensities from methyl groups and intra-residue spins that are indistinguishable in the NMR spectra.

Finally, the constant relating the t = 0 matrix, $\mathbf{M}(0)$ of unity to the experimental cross-peak intensity scale, is approximated by fitting a line of best fit (restrained to cross through the origin) to a plot of known experimental intensities (those that were unambiguously assigned) versus their corresponding calculated intensities. For this plot, the experimental intensity values are an average of the peak intensities measured on either side of the diagonal in order to compensate for base-line differences on either side of the diagonal. The slope of this line can then be used as a conversion factor for all calculated intensities. All reported intensities (experimental and simulated) are also normalized by an estimate for the smallest experimentally identifiable peak intensity, so that an intensity of 1 indicates the smallest intensity that should be visible experimentally. After NOE peaks were calculated from the simulated ensemble the intensities were compared with the experimental intensity and expressed as multiples of smallest experimentally identifiable peak. This generated a distribution of simulated NOE peak intensities corresponding to every possible atomic contact. For these relative intensities, only peaks with intensities above 1 are expected to be experimentally visible above noise.

Due to the complexity of the noise distribution in the Aβ42 spectra relative to the Aβ21-30 spectra, uncertainties in Aβ42 peak intensity due to noise were estimated from simulation (again expressed relative to the smallest identifiable peak). We used a factor of 0.44 for the H_2O spectrum and 0.50 for the ^2H_2O spectrum, and this noise was assumed to have a normal distribution (although there are other non-random noise features in the NMR data). Based on simulated NOE intensities for the H_2O spectrum, we were able to estimate the number of NOE peaks below 1 that should statistically be made visible due to noise increasing their intensities above 1, and the number of NOE peaks above 1 which noise should decrease below 1, rendering them unobservable. These statistical estimates were made by binning the simulated peaks by intensity. For each bin below 1, a normal distribution with standard deviation $\sigma = 0.44$ was used to determine the probability that random experimental noise would increase the lowest intensity in the bin to a value above 1. This probability is given by

$$P_i = 1 - \Phi\left(\frac{(1 - W_i)}{\sigma} \right) \tag{2.13}$$

where Φ is the cumulative distribution function of the normal distribution with a standard deviation of one, centered at zero, and W_i is the intensity of the weakest NOE peak in bin i. The number of peaks that are increased above 1 due to noise is then given by the sum over all bins below 1,

$$N_{fn} = \sum_{i}^{N_b} P_i n_i \tag{2.14}$$

where n_i is the number of peaks in bin i. This number N_{fn} is a reasonable estimate for the number of false negatives that should statistically be observed if all true NOE peaks are simulated with the correct intensities. The analogous procedure was performed for the bins above 1 to determine the number of intensities that should statistically drop below 1 due to noise, i.e. the number of false positives expected if all simulated peaks are exactly correct. We used a bin size of σ/2 = 0.22. The distribution of peak intensities was obtained from the simulated H_2O NOE because experimental NOE intensities below 1 are not visible by definition, however the intensity distribution should be quite similar to experiment because simulated NOE intensities were scaled to fit experiment and normalized by the lower bound for experimentally visible peak intensity. No peaks in high-noise regions of the experimental spectra (noise from experimental artifacts such as t1 noise) were used in this procedure because these regions do not have the same normally distributed noise with σ = 0.44.

Residual Dipolar Couplings. We use the PALES (96) program to calculate residual dipolar coupling (RDC) values by residue for each structure in our simulated Aβ42 ensemble. The program computes the RDC by using steric properties of the molecule to generate an alignment orientation. Then, the angle between the internuclear vector and the external magnetic field is used to calculate the RDC for each snapshot. We calculate the RDC for the backbone amide bond vector for each Aβ42 residue. The average value over the ensemble for each residue is then compared to the Aβ42 experimental data from Yan *et al.* (67). The simulation values are multiplied by a constant to put them on the same scale as the experimental data, which is determined by experimental peptide concentration.

Ensemble Structural Analysis

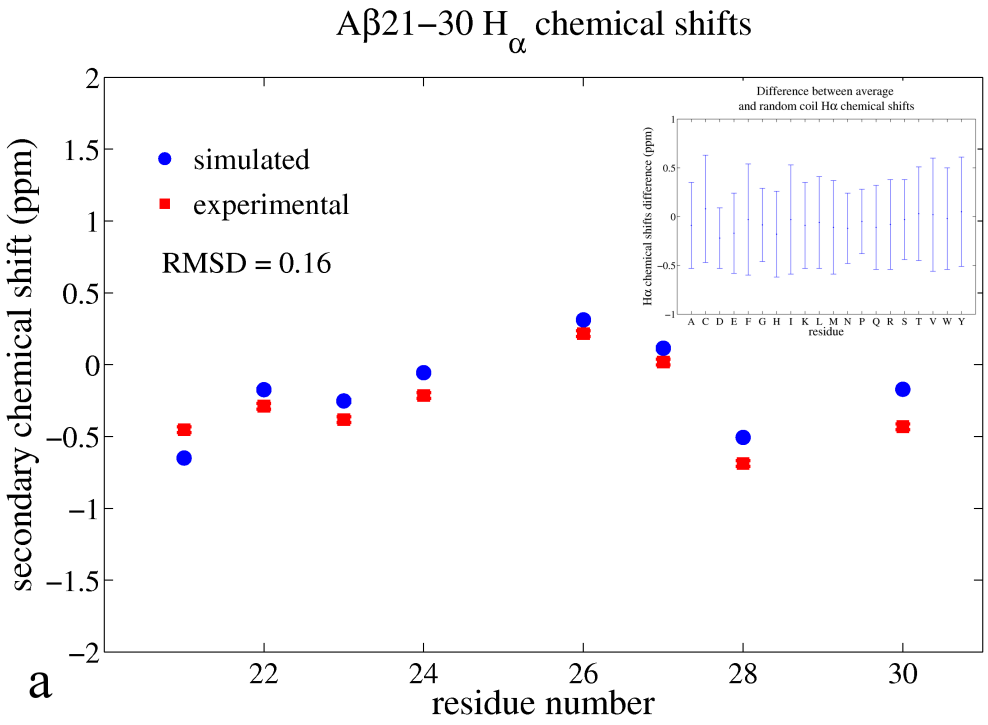
Structural analysis of the *de novo* molecular dynamics simulations of Aβ21-30 and the Aβ42 287 K ensembles was performed using *ptraj*, perl scripts and MATLAB (The MathWorks, Natick, MA) scripts. *Ptraj* was used to identify regions of secondary structure in the peptide backbone using the DSSP criteria (97). We also used *ptraj* to calculate radii of gyration and to identify the most commonly formed hydrogen bonds and salt-bridges. We used a cutoff of 3.5 Å between heavy atoms for identifying hydrogen bonds and a 4.0 Å cutoff for salt-bridges with a 60° angle cutoff for both. In-house scripts were used to identify particular secondary structure features and hydrogen bonds that are present simultaneously in the Aβ42 ensemble.

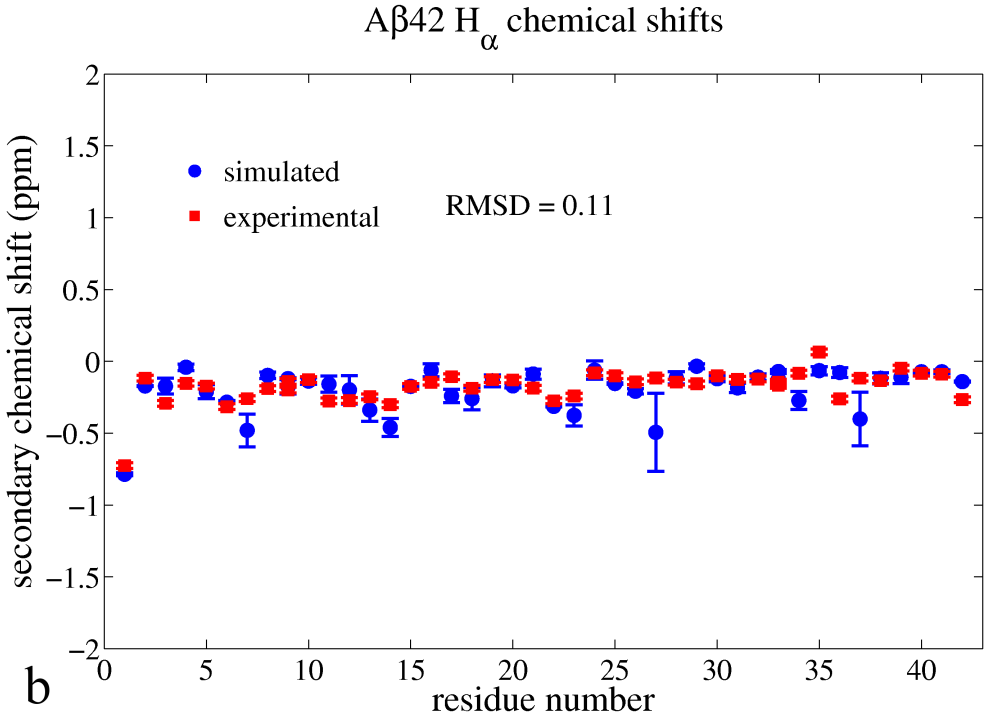
Results

Chemical Shifts

We compare the Aβ21-30 and Aβ42 experimental and simulated chemical shifts for H_α and H_N (Figure 2.1)[2], as well as C_α and C_β (Figure 2.2), for each residue; the experimental carbon chemical shift data for Aβ42 was taken from Hou *et al.* (14), otherwise the experimental data was generated in the Wemmer lab. Residue specific random-coil chemical shifts are subtracted from both experimental and simulated chemical shifts (98). The agreement is very good between measured and calculated C_α, C_β, H_α, and most H_{SC} chemical shifts, while the calculated amide hydrogen chemical shift values are less good, especially so for the Aβ21-30 peptide. This is consistent with previous results showing that SHIFTS and other chemical shift calculators have difficulty predicting amide hydrogen values due to the high sensitivity of chemical shifts to hydrogen bonding geometry (91, 99). In summary, the chemical shift comparisons serve as one indication that our simulations have sampled the correct conformational landscape measured in the experimental ensemble.

[2] This and all other figures in this chapter are reproduced from (48) with permission.


Aβ21−30 Hα chemical shifts
simulated
experimental
RMSD = 0.16
secondary chemical shift (ppm)
residue number
Difference between average
and random coil Hα chemical shifts
Hα chemical shifts difference (ppm)
residue
a


Aβ42 Hα chemical shifts
simulated
experimental
RMSD = 0.11
secondary chemical shift (ppm)
residue number
b

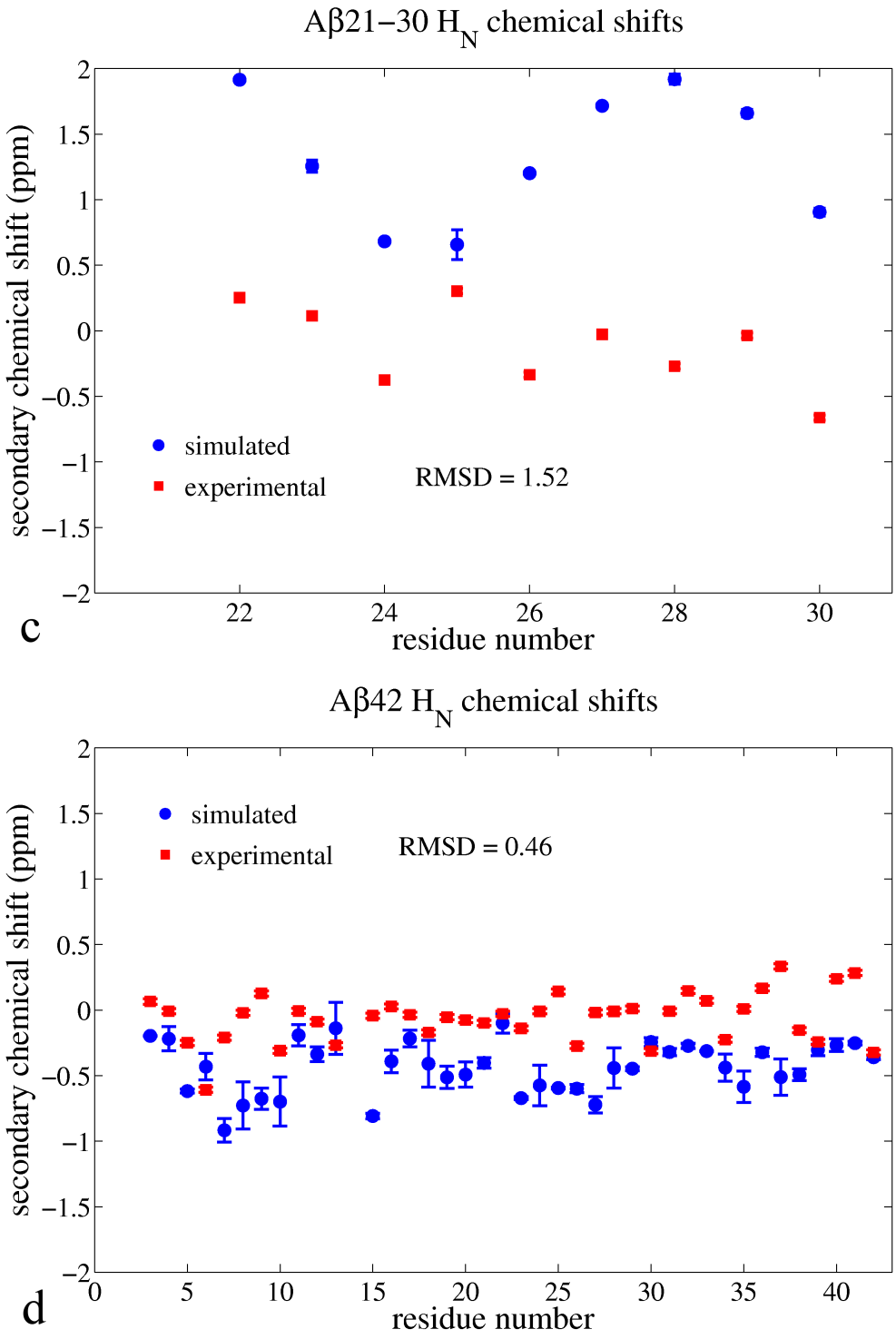
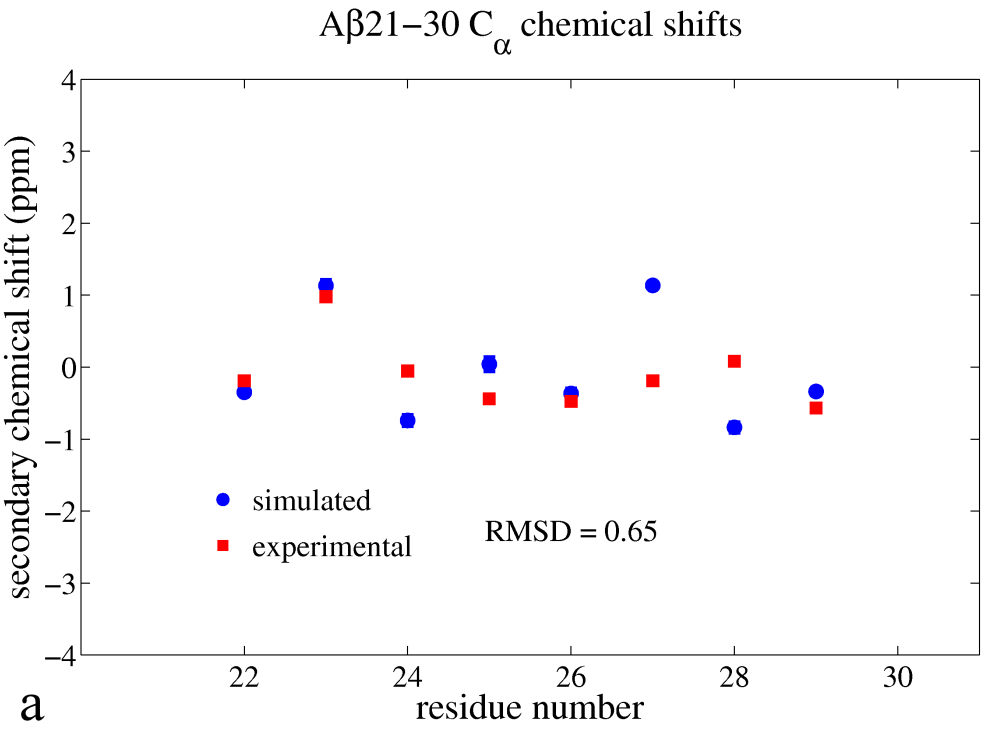
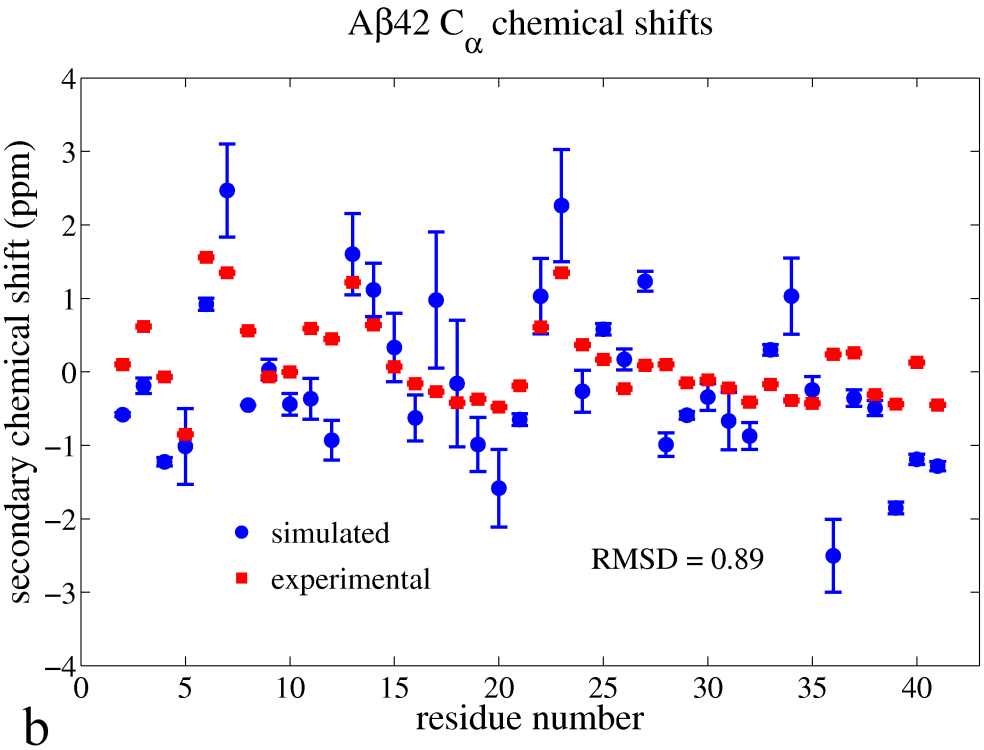


Figure 2.1. *Experimental and calculated proton secondary chemical shifts by residue.* (a) Aβ21-30 H_α chemical shifts, (b) Aβ42 H_α chemical shifts, (c) Aβ21-30 H_N chemical shifts, and (d) Aβ42 H_N chemical shifts. Red squares represent experimental data, while blue circle represent the data calculated from simulation. Random coil residue specific values are subtracted from both experimental and simulation values. The experimental data for Aβ21-30 data is from Fawzi *et al*. **(26)**. Insert (a): difference between average chemical shift for each residue and random coil H_α chemical shift from Merutka *et al*. **(98)**.


Aβ21−30 Cα chemical shifts
secondary chemical shift (ppm)
residue number
simulated
experimental
RMSD = 0.65
a


Aβ42 Cα chemical shifts
secondary chemical shift (ppm)
residue number
simulated
experimental
RMSD = 0.89
b

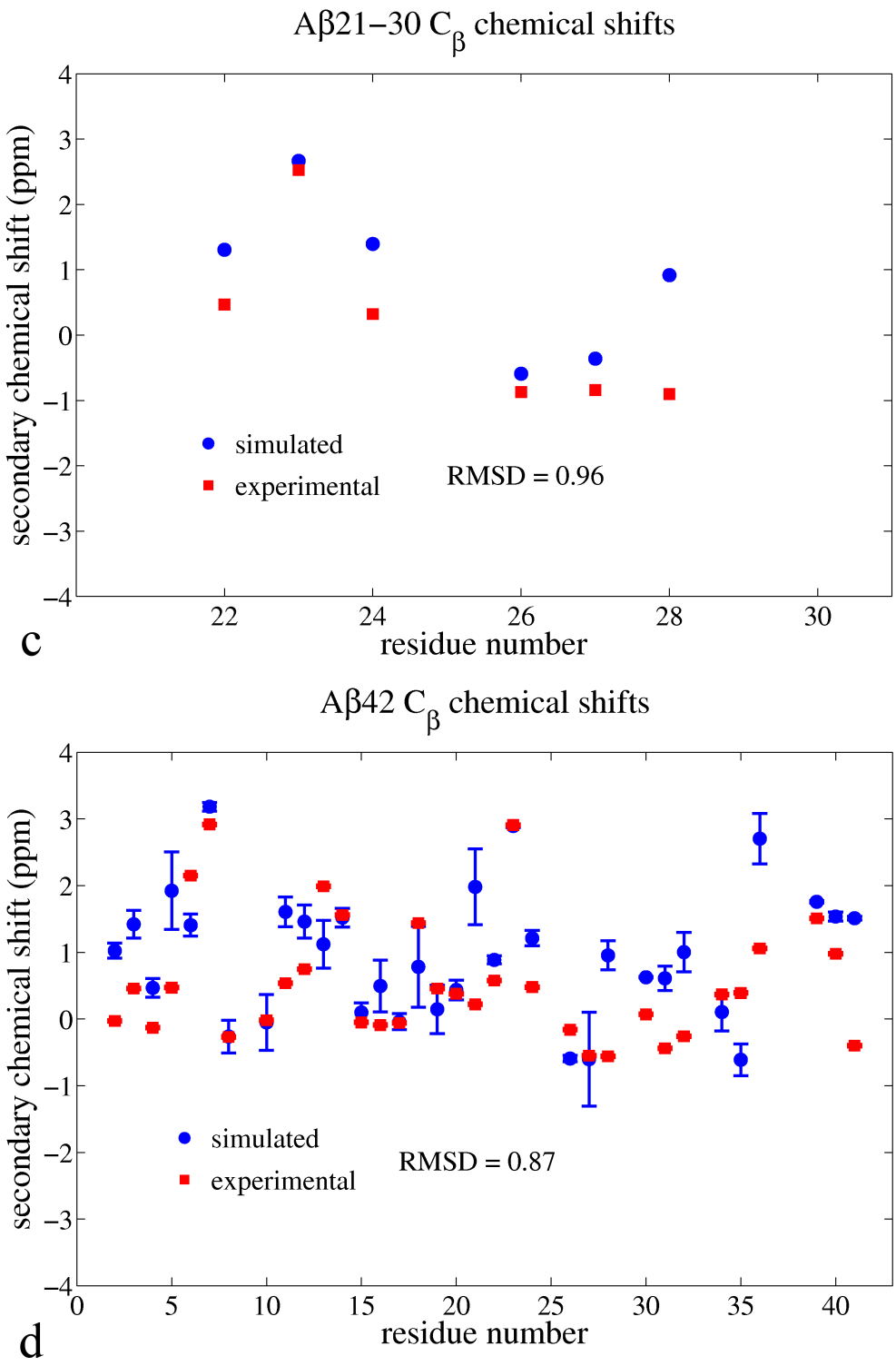


Figure 2.2. *Experimental and calculated carbon secondary chemical shifts by residue.* (a) Aβ21-30 C_α chemical shifts, (b) Aβ42 C_α chemical shifts, (c) Aβ21-30 C_β chemical shifts, and (d) Aβ42 C_β chemical shifts. Red squares represent experimental data, while blue circle represent the data calculated from simulation. Random coil residue specific values are subtracted from both experimental and simulation values. The experimental data for Aβ42 are taken from Hou *et al*. (14) while the Aβ21-30 data is from Fawzi *et al*. (26).

In general the chemical shifts by residue assume values near the average shift for that residue type from the BMRB database (100): 98% of the hydrogen chemical shift rms differences are within one standard deviation, and these trends do not vary greatly across peptide sequence, indicating that the ensemble does not show a strong preference for α-helical or β-sheet conformations. In fact, the rms difference between experimental chemical shifts and average chemical shifts is on the same order as the difference between the experimental and simulated chemical shifts. Furthermore, we note that average chemical shifts by residue do not differ greatly from random coil chemical shifts (98) (Figure 2.1a insert). This distinction is important when comparing the structural ensembles of Aβ21-30 and Aβ42. We show later that while the tertiary ensemble of the Aβ21-30 fragment peptide is largely a random coil, Aβ42 samples both extended and collapsed states containing significant amounts of β-turn, α-helical, or β-strand secondary structure, and therefore its average chemical shift value should not be interpreted as a random coil ensemble. Therefore the chemical shift data does not discriminate between the two IDP peptide ensembles.

J-Couplings

We next consider the comparison of experimental and our simulated $^{3}J_{H^{N}H^{\alpha}}$ scalar coupling constants for Aβ21-30 (Figure 2.3a) and Aβ42 (Figure 2.3b), using the established Karplus parameters from Vuister and Bax (93). We initially compare our simulations against the experimental data reported in our previous study on Aβ21-30 (26) and that of Yan and co-workers for Aβ42 (67). Overall, the average J-coupling constants do not reflect any well-defined secondary structure for either IDP.

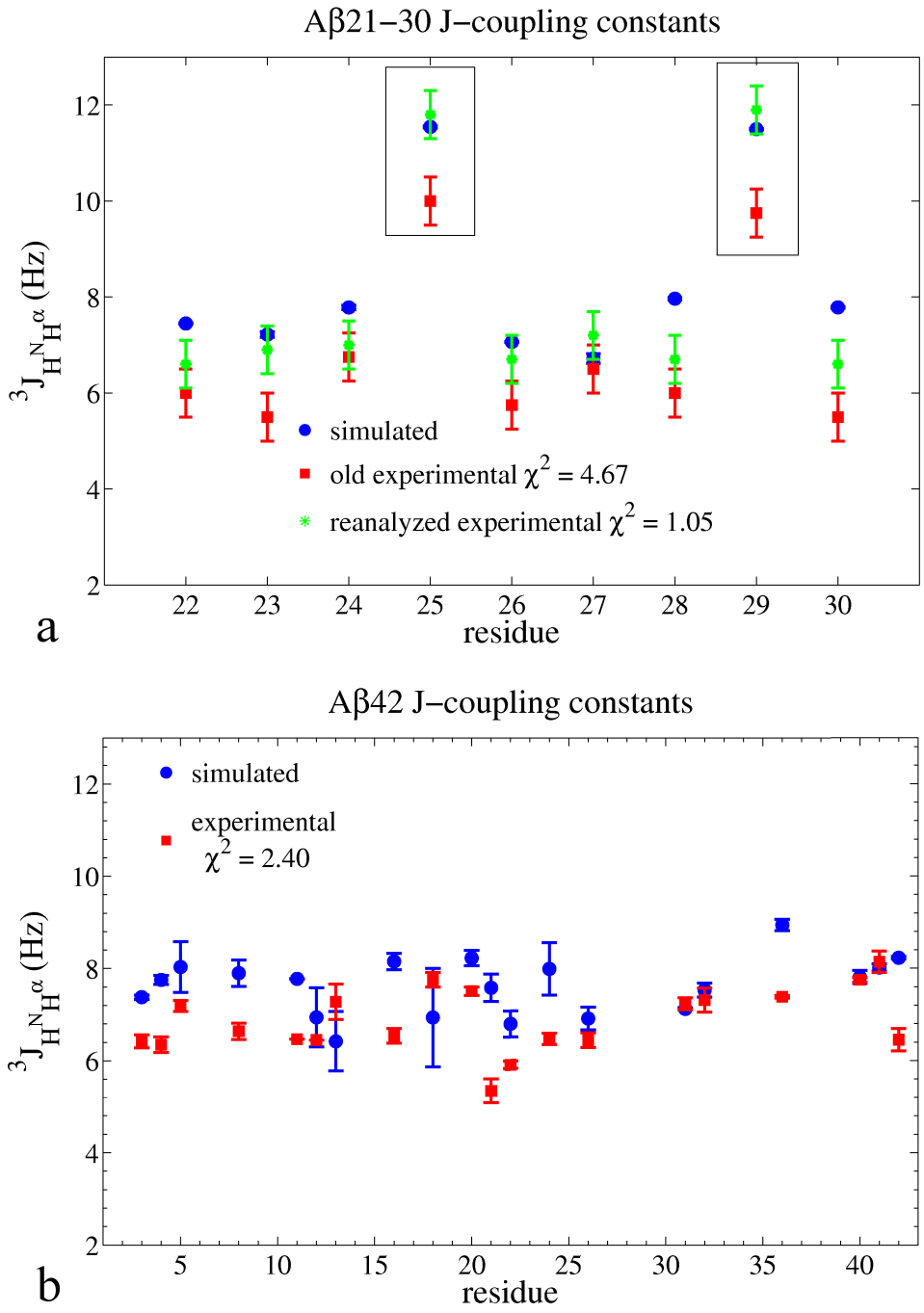


Figure 2.3. *J-coupling constants for backbone amides for (a) Aβ21-30 and (b) Aβ42.* The red symbols are the Aβ42 experimental data from Yan *et al*. (67) and the Aβ21-30 data from Fawzi *et al*. (26). The green stars are the reprocessed experimental data for Aβ21-30. The blue circles are the data calculated from the simulation ensemble using the Vuister and Bax Karplus parameters (93). Glycine residues are boxed to emphasize that they can only be measured as the sum of two coupling constants and therefore have higher values than the other residues. Simulation uncertainty bars represent rms difference between two independent simulations and the average. The Yan *et al*. data (67) has not accounted for a relaxation correction that makes J-couplings determined from a $H_N H_a$ 3D experiment consistently lower than those from COSY splittings (14) by a small amount (maybe as high as ~10% (93)); if this relaxation effect is accounted for, then the qualitative agreement between experiment and simulation is excellent.

We note that our original calculated $^3J_{H^N H^\alpha}$ constants for Aβ21-30 used an incorrect definition of the ϕ angle (26), and Figure 2.3a displays the simulation data using the correct definition of the ϕ angle in Eq. (2.1). We see that the calculated J-coupling constants (blue symbols) do not agree as well with the experimental data (red symbols) for either Aβ42 or Aβ21-30, thereby implying that our ensemble of structures does not sample the same dihedral angle distribution as that of the peptide in the NMR sample, and thereby yields average J-coupling values that differ significantly from experiment.

However, we have determined that for Aβ21-30 most of the disagreement is likely due to experimental error of the original studies reported in (26). We went back to the experimental data and reprocessed it to remove inappropriate truncation of, and application of a cosine window function to, the time-domain data in the original processing, yielding significantly larger values of $^3J_{H^N H^\alpha}$ than the original splittings we reported. In order to conclusively rule out dangers of interpreting apparent splittings in spectra with either large peak widths or also truncation artifacts, we also performed lineshape analysis. The (new) directly measured splittings were within +/- 0.2 Hz of the $^3J_{H^N H^\alpha}$ obtained by least squared non linear fitting of Eq. (6.5) in (101) to a 1D slice through the peak maxima. These $^3J_{H^N H^\alpha}$ values are also significantly higher on average than the values we previously reported. These new experimental values are shown in Figures 2.3a (green symbols) which are shifted upward by ~0.5-1.0 Hz relative to the original data. We also see much improved agreement between simulation and experiment, providing further support of the accuracy of our simulated ensemble for Aβ21-30. We would like to point out that the high coupling constants for the glycine residues (Figure 2.3a, boxed) are sums of the coupling constants for both glycine H_α atoms, which are measured as one splitting experimentally.

For Aβ42 there are experimental differences in the $^3J_{H^N H^\alpha}$ scalar coupling constants reported in two independent NMR studies (14, 67). The earlier 2004 NMR study report scalar coupling values of 7.0-8.0 Hz for most residues, with some residues exhibiting values as high as 9.0 Hz (26). The later NMR study (67) determined an average scalar coupling values of ~6.9 Hz across all residues, with a highest value of 8.14 Hz. However, the later NMR study that derived J-coupling values from a 3D $H_N H_a$ experiment did not take into account relaxation effects (102); if these were accounted for then the scalar couplings would be shifted upward by ~10% (93) relative to that reported in Figure 2.3b. Roughly speaking, accounting for the relaxation effect would provide excellent qualitative agreement of the Yan *et al*. data (67) with the earlier NMR study, and with our simulation of $^3J_{H^N H^\alpha}$ scalar couplings for Aβ42 that determines an average scalar coupling values of ~7.6 Hz across all residues. We attribute any remainder of J-coupling discrepancies to insufficient parameterization of the intrinsic backbone dihedral angle preferences that over-sample the extended β-region of the Ramachandran map (φ = -120°, ψ = 145°) relative to the slightly more collapsed polyproline II (PP II) conformation (φ = -75°, ψ = 145°). In recent work by Nerenberg and Head-Gordon, using the same force field combination of ff99SB and TIP4P-Ew, we found that agreement of the $^3J_{H^N H^\alpha}$ coupling constant with experiment is complete for a glycine tri-peptide (χ^2 = 0.00(37)) (88). However, in the same study, we found that agreement between experiment and simulation diminishes steeply for alanine and valine tri-peptides (χ^2 = 1.44 and 1.01 respectively), and the discrepancy is amplified as peptide length

increases, a known shortcoming of current force-fields when used to simulate unfolded proteins (24, 88).

Residual Dipolar Couplings

Table 2.1 compares the experimental RDC values (67) of Aβ42 to those calculated from our simulated ensemble using PALES (96). We compare all available RDCs for 33 residues reported in (67), unlike previous work (90) that only compares their simulated RDC data against 19-22 of the 33 experimental RDC values. The RMS difference between simulation and experiment is 2.12 Hz, however, this error is dominated by Lys16 which has a very large experimental RDC value compared to the other residues. Excluding Lys16, our RMSD is lowered to 1.73 Hz, which is on the order of the experimental and simulation error bars (Figure 2.4). In our opinion, it is difficult to interpret the structural content of the experimental RDCs for an IDP (66, 67, 90), and therefore we primarily present the RDCs to provide further experimental validation of the simulated ensemble.

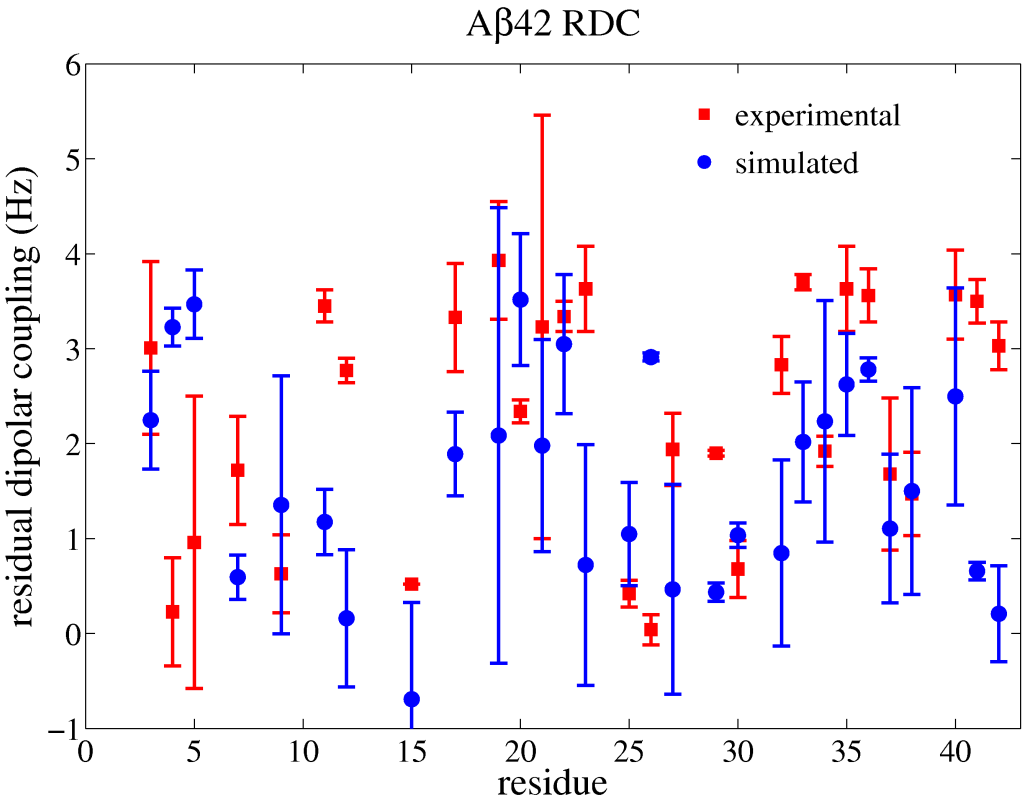


Figure 2.4. *Experimental vs. calculated residual dipolar couplings for Aβ42.* The red symbols are experimental data for Ab42 from Yan *et al*. (67). The blue circles are the data calculated from the simulation ensemble.

Residue	Experimental RDC	Experimental error	RDC from Simulation	Simulation error
Glu 3	3.01	0.91	2.25	0.52
Phe 4	0.23	0.57	3.23	0.2
Arg 5	0.96	1.54	3.47	0.36
Asp 7	1.72	0.57	0.59	0.23
Gly 9	0.63	0.41	1.36	1.36
Glu 11	3.45	0.17	1.18	0.34
Val 12	2.77	0.13	0.16	0.72
Gln 15	0.52	0	-0.69	1.02
Lys 16	7.59	2.6	0.73	1
Leu 17	3.33	0.57	1.89	0.44
Phe 19	3.93	0.62	2.09	2.4
Phe 20	2.34	0.12	3.52	0.69
Ala 21	3.23	2.23	1.98	1.12
Glu 22	3.34	0.16	3.05	0.73
Asp 23	3.63	0.45	0.72	1.27
Gly 25	0.42	0.14	1.05	0.54
Ser 26	0.04	0.16	2.91	0.04
Asn 27	1.94	0.38	0.47	1.1
Gly 29	1.9	0.03	0.44	0.1
Ala 30	0.68	0.3	1.04	0.13
Ile 32	2.83	0.3	0.85	0.98
Gly 33	3.7	0.08	2.02	0.63
Leu 34	1.92	0.16	2.23	1.27
Met 35	3.63	0.45	2.62	0.54
Val 36	3.56	0.28	2.78	0.12
Gly 37	1.68	0.8	1.11	0.78
Gly 38	1.47	0.44	1.5	1.09
Val 40	3.57	0.47	2.5	1.14
Ile 41	3.5	0.23	0.66	0.09
Ala 42	3.03	0.25	0.21	0.51

Table 2.1. Aβ42 RDC values from experiment (67) and the simulated ensemble.

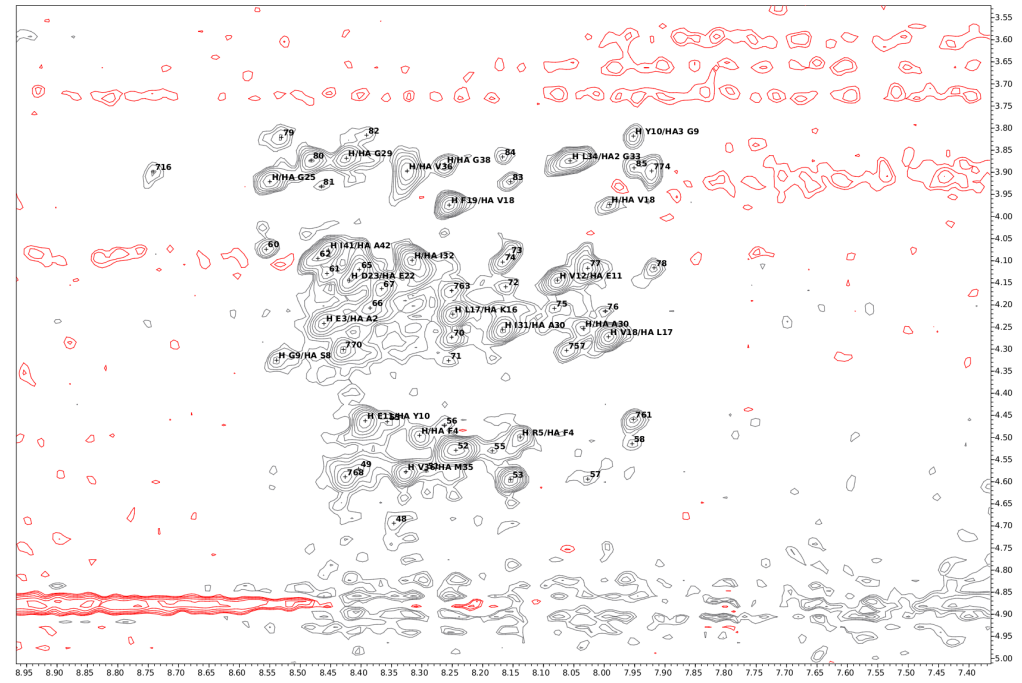
ROESY and NOESY Peaks

In our original high field NMR study of Aβ21-30 (26), the 2D ROESY experiments yielded a set of 155 assigned ROESY crosspeaks, of which 28 were weak medium range ROE interactions. The medium range ROEs comprised several i,i+2 and i,i+3 interactions and two extremely weak i,i+4 interactions; no longer range ROE crosspeaks were observed and no strong patterns of α-helical or β-strand contacts were evident. In the case of Aβ21-30, all experimental peaks were assignable. The role of simulation for Aβ21-30 was to ascertain whether the medium range interactions were occurring simultaneously, as would be typical in structure calculations on folded proteins which assume that weak NOE or ROE cross peaks correspond to large (~4.5-6.0Å) upper distance restraints on a single well-defined structural population. In fact we found

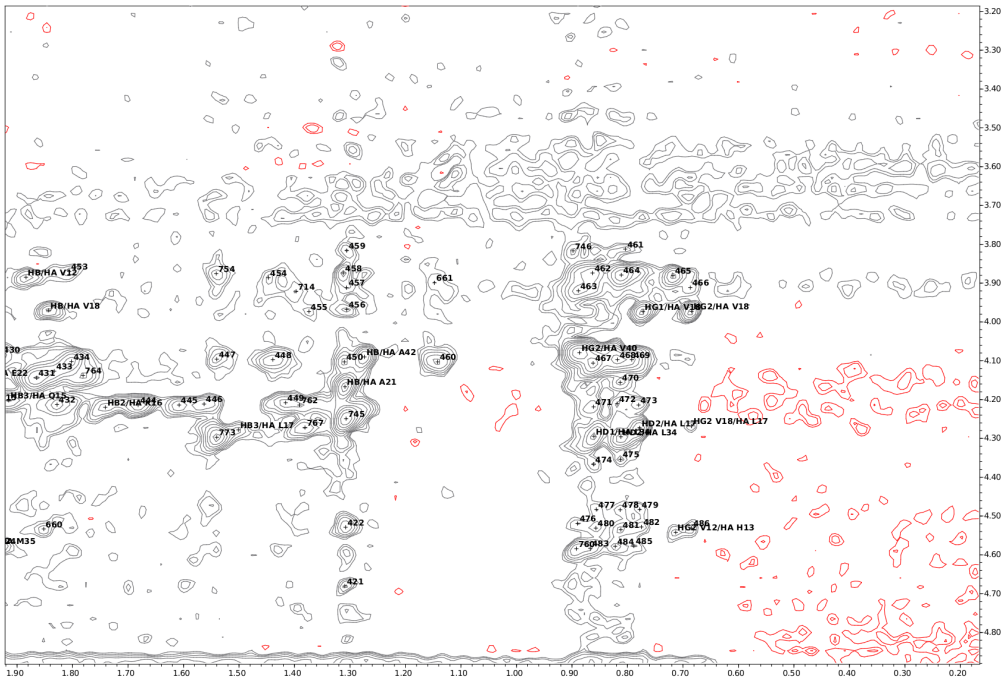
that the peptide ensemble involves significant disorder and hence it is inappropriate to use the standard structure determination methods.

We determined very good agreement between the simulated and experimentally observed ROE cross peaks, indicating that our ff99SB/TIP4P-Ew simulations provide a well-validated estimate of the ensemble of structures interrogated by the experiments, allowing them to be used to describe the full structural ensemble diversity. We found that the conformational ensemble of Aβ21-30 was dominated by an unstructured population that lacked any consensus secondary structure or hydrogen-bonded interactions. The remaining minority population involved ~14% population of β-turn structure centered at Val24 and Gly25, which in turn populated contacts between Asp23 and Ser26. The simulations also indicate that the Asp23 to Lys28 salt bridge, important to the fibril structure (103), was formed in ~7% of the ensemble. Nonetheless, while Aβ21-30 exhibits some residual structure, it is largely an extended random coil peptide. Experimental and simulation data suggest that the structured populations may increase upon lengthening of the Aβ peptide to larger fragments (32, 36, 104, 105), and that is what we consider next for Aβ42.

Figure 2.5 shows a region of the experimental $^{1}H_2O$ NOESY spectrum of Aβ42 that emphasizes the difficulties in obtaining unambiguous tertiary structure assignments for longer IDPs with spectra that are congested (giving multiple possible peaks assignments, Figure 2.5a) and have appreciable variation in noise (Figure 2.5b). We define an experimental NOE peak as "assignable" based on the definition that there is one dominant short-ranged contact that must contribute the majority toward the peak intensity, and that its intensity is above experimental noise uncertainties. Therefore, due to degeneracy of chemical shifts, only 196 of the 705 NOESY peaks seen in the H_2O and $^{2}H_2O$ NMR spectra could be assigned from the experimental data alone. Of these assignable peaks, 11 were found to be non-sequential and non-intra-residue, and these peaks had significantly lower experimental intensities than many other assigned peaks. Therefore 509 experimental peaks, 165 of which are I to i+2 or greater, have multiple possible assignments for Aβ42, far different than that seen for Aβ21-30.



a



b

Figure 2.5. *Plot of the experimental NOESY spectra.* (a) a region of the spectra with large chemical shift degeneracy (47 H chemical shifts along the x dimension and 52 H chemical shifts along the y dimension), and (b) a region of the spectra with high T1 noise and large background signal arising from the water peak.

In order to assign more of the NOEs, we performed *de novo* calculations of the NOE intensities using our MD simulation and described in the Methods section, in which we use the experimentally assigned peaks to place the simulation intensities on the absolute scale of the NMR experiment (Figure 2.13). We also determine a scaling factor based on the smallest peak that can be identified clearly above the noise in the experimental spectra - and express experimental and simulated intensities as a multiple of this threshold. Only peaks simulated with intensity above 1 are considered visible above noise. With this scaling factor determined from the $^{1}H_2O$ and $^{2}H_2O$ experimental data, 176 of the 196 assigned experimental peaks were predicted as visible from the simulation data. None of the 11 long range assigned peaks was predicted by simulation to be above 1, but 10 of them were predicted above 0.1. A factor of 10 in intensity corresponds to a factor of $10^{1/6} = 1.5$ difference in distance, thus the predicted distances are just outside the range of giving a detectable peak. We note that 10 of these peaks involved i to i+2 and i to i+3 contacts, with only 1 peak that involved an i to i+4 contact.

The remaining 509 unassigned experimental NOEs were compared to the simulation data by summing the calculated NOE intensities for all proton pairs that had the same chemical shift of the experimentally observed peak. Of these 509 unassigned peaks, the simulation predicted 355 peaks, of which 54 had no sequential or intra-residue contact. Only 223 of these 355 contacts were dominated by a single atomic contact included, while 132 of these simulated peaks had two or more contacts contributing with similar intensities. Therefore there were 174 false negatives (not predicted to be visible), 122 of which could not result from an intra-residue or sequential contact. In addition to the 531 simulated peaks that were consistent with experimentally observed peaks, there were 383 false positive peaks (calculated from simulation that were not seen in experiment). These data are summarized in Table 2.2.

Peak Types	All NOE Peaks	Long range peaks (i to i+2 or greater)
Experimentally observed peaks	**705**	**176**
Experimentally assigned peaks	196	11
Experimentally unassigned peaks	509	165
Simulated peaks	914	188
Simulated peaks agree experimentally assigned peaks	176	0
Simulated peaks agree experimentally unassigned peaks	355	54
False negatives	174	122
False negatives found in high noise regions	-62	-47
False negatives explained by atomic contacts on same residue pairs	-66	-29
False negatives inconsistent with experiment	**46**	**46**
False positives	383	134
False positives found in high noise regions	-82	-19
False positives explained by atomic contacts on same residue pairs	-223	-81
False positives made up of numerous weak contacts	-16	-6
False positives inconsistent with experiment	**62**	**28**

Table 2.2. The summary of experimental and calculated NOEs statistics, and analysis of the number of false positives and false negatives.

In total, the simulations yielded 174 false negatives and 383 false positives when compared to the NMR experiment. However, unlike our previous experiment on more concentrated samples of Aβ21-30, the noise across the NOESY spectra for Aβ42 was quite variable. We estimate that the average experimental uncertainty in the intensities due to noise is ± 0.44 of the threshold for the H_2O spectrum and ± 0.50 for the ^2H_2O spectrum. However, there are regions where noise move far outside these experimental ranges due to T1 noise, modulations from the water signal, and/or dense regions of overlapping large peaks that make peak intensities difficult to determine (see Figure 2.5). This required a more painstaking analysis of different regions of the spectra to determine our confidence in what are genuine false positives and false negatives in the NOESY assignments. For example, 82 false positive peaks were in regions of the spectra well above the baseline intensity uncertainties, which would prevent these 82 peaks from being observed experimentally. Of the false negatives, 62 of these peaks were observed in the high-noise regions of the experimental spectra, which could cause a weak peak to appear stronger.

Disregarding peaks that could not be seen due to high noise, there were 301 remaining false positives, 115 of which contain only residue i to i+2 or longer contacts, and 112 remaining

false negatives of which 75 are long-range. However, for many hydrogen-hydrogen pairs with NOEs observed but not predicted (false negatives), NOEs between other proton pairs within the same residue pair were not observed but were predicted (false positives). Thus although the simulations sometimes miss the correct inter-proton contacts, they do largely succeed in predicting correct residue contacts. Considering predictions between the proper residue pairs as correct removes 223 false positives and 66 false negatives (Table 2.2). Finally, the remaining 78 false positives were analyzed to see whether their intensities were dominated by a specific set of contacts: 16 of these peaks, although predicted to be above noise level, did not contain any dominant atomic contact of significant intensity, but instead were composed of many atomic contacts whose individual intensities were below half of the noise level. Considering that multiple crosspeaks would probably be spread over a wider range of shifts than used in constructing the simulated spectrum, these are not serious experimental disagreements.

The remaining 46 false negatives (out of 705 observed peaks) and 62 false positives (out of 914 simulated peaks) are weak as estimated by the experimental intensity uncertainty inherent in the H_2O (± 0.44) and ^2H_2O spectra (± 0.50), although we report them in the Appendix. If the noise is assumed to be normally distributed, based on our distribution of experimental intensities we would expect to see ~69 false negatives, compared to the 46 that we determined. Given the experimental variation in noise, our level of false negatives is not statistically significant. Similarly, if the noise is assumed to be normally distributed, based on our distribution of *simulated* peak intensities, we would expect to see ~29 false positive peaks, whereas we see 60. Even so, a vast majority of the false positives are short-range, and do not grossly influence the backbone structural ensemble. These data are also summarized in Table 2.2.

Peptide Structural Ensemble

Given the overall good agreement of the calculated NMR observables from *de novo* molecular dynamics with various NMR measurements, we proceed to a structural analysis of the simulated ensemble. Figure 2.6 shows the radius of gyration (R_g) distribution of Aβ42, which emphasizes that the peptide adopts conformations that vary extensively in the degree of compactness. It is interesting to compare this to a random coil polymer, in which the original Flory model (106) states that R_g scales with the number of residues, n

$$R_g = R_0 n^{\nu}, \tag{2.15}$$

where R_0 is an estimated persistence length and n depends on the quality of the solvent. In a good (denaturing) solvent $n \sim 0.598$, and assuming a persistence length of R_0=1.33 (107), a 42 amino acid peptide's average radius of gyration, $\langle R_g \rangle$ would be estimated to be 12.4±1.0 Å, consistent with what we find for the Aβ42 peptide in which $\langle R_g \rangle$ = 13.0±4.5 Å. However, the large standard deviation for $\langle R_g \rangle$ and the nature of its distribution (Figure 2.6) emphasizes that the Aβ42 peptide samples many conformations that are even more compact or more extended than expected from a good solvent model or estimates of the persistence lengths of denatured proteins. For example, if we assume a larger persistence length of R_0 = 2.1 (108), we get a value of $\langle R_g \rangle$ = 19.6 Å in a good solvent, or an $\langle R_g \rangle$ = 7.2 Å in a poor (water-like) solvent ($n \sim 0.33$), which better correlates with Aβ42's skewed R_g distribution that is peaked around 9-10 Å but with a long tail that samples conformations with an R_g as large as ~30 Å.

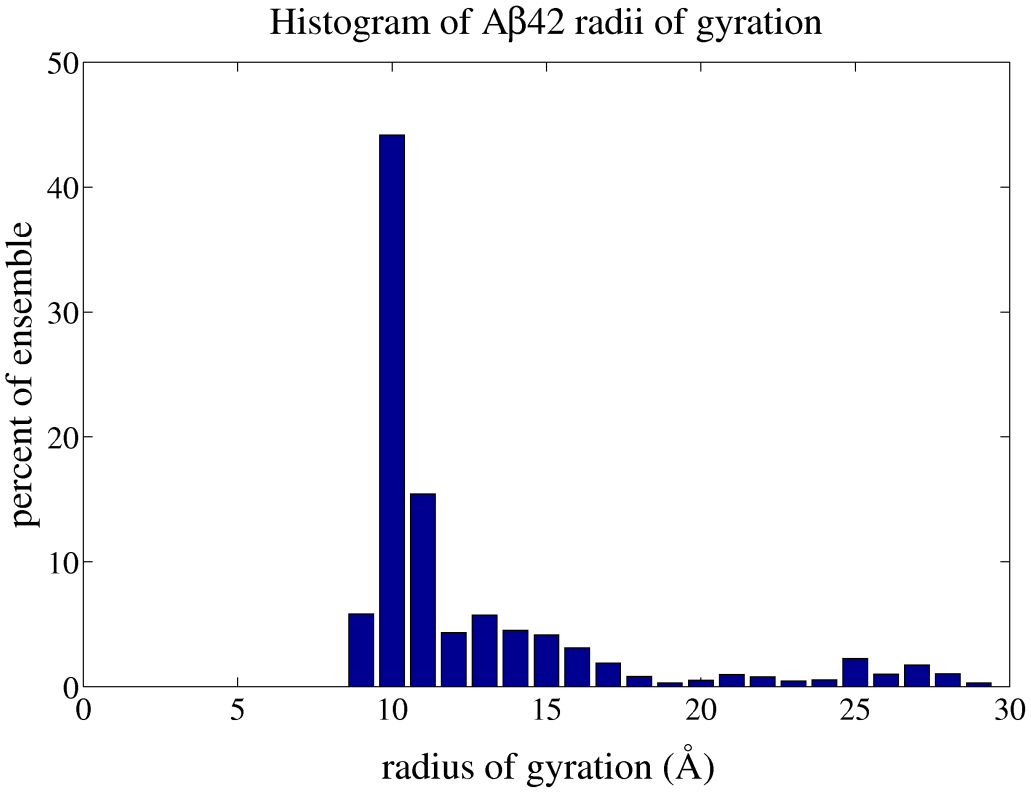


Figure 2.6. *Distribution of Aβ42 radius of gyration.* The peptide's radius of gyration ranges from ~9 Å, which is similar to that of globular proteins with the same number of residues, to ~30 Å, which is fully extended, with an average radius of gyration is 13±4.5 Å (Figure 2.4).

We also find that the more compact portion of the Aβ42 ensemble (~90% of the ensemble) is dominated by conformations with some type of secondary structure, hydrogen bonding, or ion pairing structural features. Figure 2.7 shows the probability of observing a given DSSP secondary structure category for each amino acid; DSSP defines secondary structure classes by distinct hydrogen bonding patterns and geometry (primarily dihedral angles), with α-helices and β-strands requiring the cooperative organization of repeats based on localized turns and bridges (97). We find that only 0.2% of the Aβ42 ensemble is completely lacking of any identifiable secondary structure! This is far different than our previous study of Aβ21-30 in which ~65% of the structural ensemble had no identifiable stabilizing secondary structure or consensus hydrogen bonds[26].

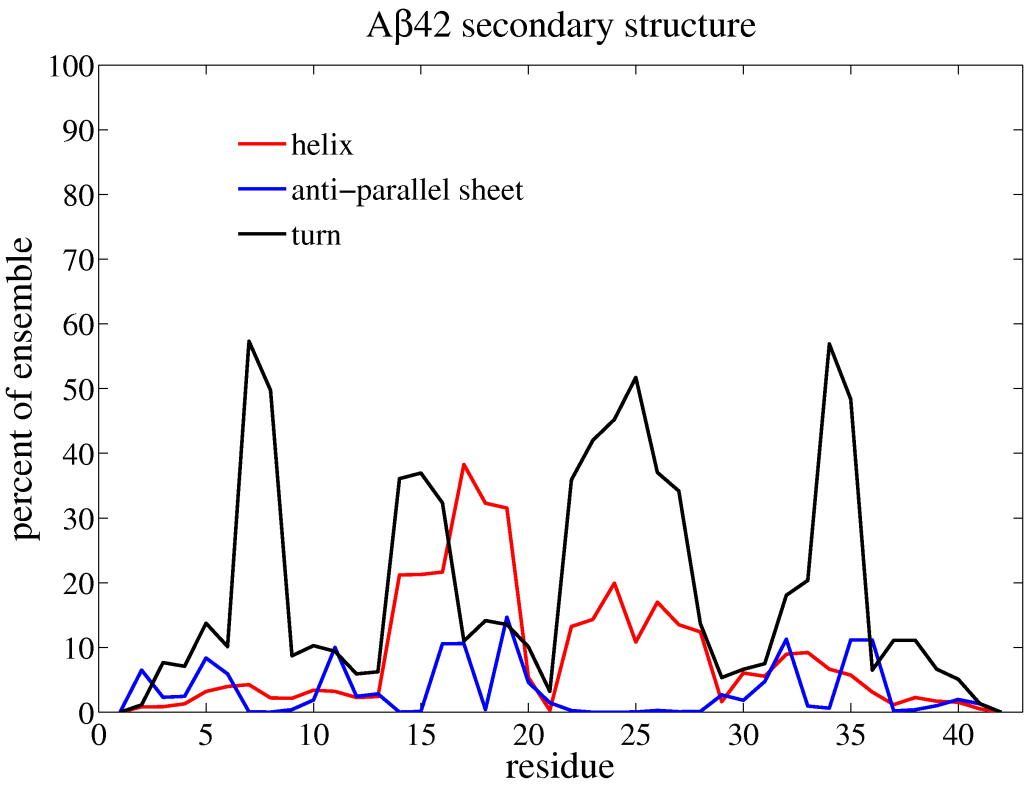


Figure 2.7. *Percentage of Aβ42 ensemble in different types of secondary structure by residue.* The red line represents α-helix, the blue line for anti-parallel strand, and the black line for β-turns.

In order to better characterize the Aβ42 peptide conformations that contain these turns, we focused on six pairs of residues that define the following turns: The first turn is defined at residues 7-8 (Turn7-8), the second turn at residues 14-15 (Turn14-15), the third turn at residues 18-19 (Turn18-19), the fourth turn at residues 24-25 (Turn24-25), the fifth turn at residues 26-27 (Turn26-27), and the final turn at residues 34-35 (Turn34-35). We also combine the α-helix structure and 3_{10}-helix classifications into one 'helix' classification in our secondary structure analysis because the same sequence region has the same tendency to form either helical structure. Residues 14 to 19 often adopt helical structure, as do residues 22 to 28, while residues 30 to 35 do so to a lesser extent. However, for residues 14-19 and 22-28, in 26% and 10% of the ensemble respectively, only part of the helix is formed while the other portion of the helical region forms one of the turns mentioned above. For this reason, we break these regions into 5 pairs of residues that often form helical structure: residues 14-15 (Helix14-15), residues 17-18 (Helix17-18), residues 23-24 (Helix23-24), residues 26-27 (Helix26-27), and residues 32-33 (Helix32-33). *These residues simply label the central regions of the helix, since DSSP requires that helices (as well and turns and β-strands) be supported by hydrogen bonds involving additional residues.* Finally, anti-parallel β-strand secondary structure is also observed in the ensemble. Anti-parallel β-strand structure is observed for residues 3-6 (Beta3-6), residues 10-13 (Beta10-13), residues 16-21 (Beta16-21) and finally residues 29-36 (Beta29-36). We do not display parallel β-strand structure because this structure was only seen at very low levels, and when the particular conformations were examined more closely, they were found to be 180° turns, rather than true parallel β-strands.

Table 2.3 summarizes the percentage of the time each of the above secondary categories appears in the Aβ42 conformational ensemble, the percentage of the time each occurs in isolation with no other secondary structure, and the frequency at which pairs of secondary structure elements form simultaneously. It is seen that Turn7-8 and Turn34-35 are the turns most commonly formed, and each is present in almost half of the simulation ensemble, and they often form simultaneously with each other (34% of the ensemble). For the minority of consensus secondary structure found in the conformational ensemble of Ab21-30, the dominant feature was ~14% of the ensemble forming a classic β-turn structure centered at Val24 and G25 bringing together Asp23 and S26 (26).

Secondary structure	**Secondary Structure Region Percentages**						
	Turn7-8	**Turn14-15/ Helix14-15**	**Turn18-19/ Helix17-18**	**Turn24-25/ Helix23-24**	**Turn26-27/ Helix26-27**	**Turn34-35**	**Helix32-33**
Observed	48.90	34.02 / 20.78	12.52 / 31.82	35.24 / 14.23	28.32 / 13.36	47.25	8.16
In isolation	3.11	1.24 / 0.59	0.19 / 0.00	3.22 / 1.04	8.41 / 2.68	2.36	0.89
Secondary structure	**Simultaneous Secondary Structure Pairing Percentages**						
	Turn7-8	**Turn14-15/ Helix14-15**	**Turn18-19/ Helix 17-18**	**Turn24-25/ Helix23-24**	**Turn26-27/ Helix26-27**	**Turn34-35**	**Helix32-33**
Turn7-8		24.28 / 16.91	10.03 / 26.81	24.87 / 7.68	14.48 / 9.14	33.75	2.54
Turn14-15			5.06 / 19.55	16.81 / 5.00	9.68 / 5.56	25.82	0.66
Turn18-19				8.28 / 2.08	5.14 / 2.96	10.52	0.09
Turn24-25					10.20 / 6.83	24.30	2.62
Turn26-27						14.16	0.73
Turn34-35							0.27
Helix14-15			6.74 / 11.85	12.78 / 3.56	5.96 / 4.76	16.05	1.82
Helix17-18				18.56 / 6.18	8.73 / 7.01	26.20	2.36
Helix23-24					2.92 / 1.49	7.28	4.09
Helix26-27						9.38	0.16

Table 2.3. *The categorization of secondary structure features most prevalent in the ensemble.* Each column is a different secondary structural feature (see text). In the first part of the table we report the observed secondary structure percentages, as well as the percentage of the time they are observed in isolation. The second half of the table reports percentages of the time that two secondary structural elements occur simultaneously.

The Aβ42 ensemble is dominated by secondary structure pairings with Turn7-8 and Turn34-35 (Figures 2.8a and 2.8b), impossible for Aβ21-30: Turn14-15, Helix17-18, and Turn24-25 are often present in conjunction with either of these turns in ~25% of the ensemble, Helix14-15 and Turn18-19 are most often paired with these turns in ~17% of the ensemble, and likewise Helix23-24 and Helix26-27 are most frequently present in conjunction with either Turn7-8 or Turn34-35 in ~9% of the ensemble. Table 2.2 also shows that Helix17-18 occurs with high frequency (32% of the ensemble) compared to the other helical segments, but it never occurs in isolation, while Helix32-33 is the least frequent of the secondary structure categories. Even though the helical conformational populations are diverse, it is clear that helical conformations are more populated in the N-terminus region of the Ab42 peptide (Figure 2.9a and 9b). We emphasize that while we have quantified the percentages of the ensemble that involve secondary structure pairings, most often these pairs are occurring with a third, fourth, or

sometimes a greater number of additional secondary structural elements, resulting in a combinatorial explosion in the structural diversity of metastable conformations.

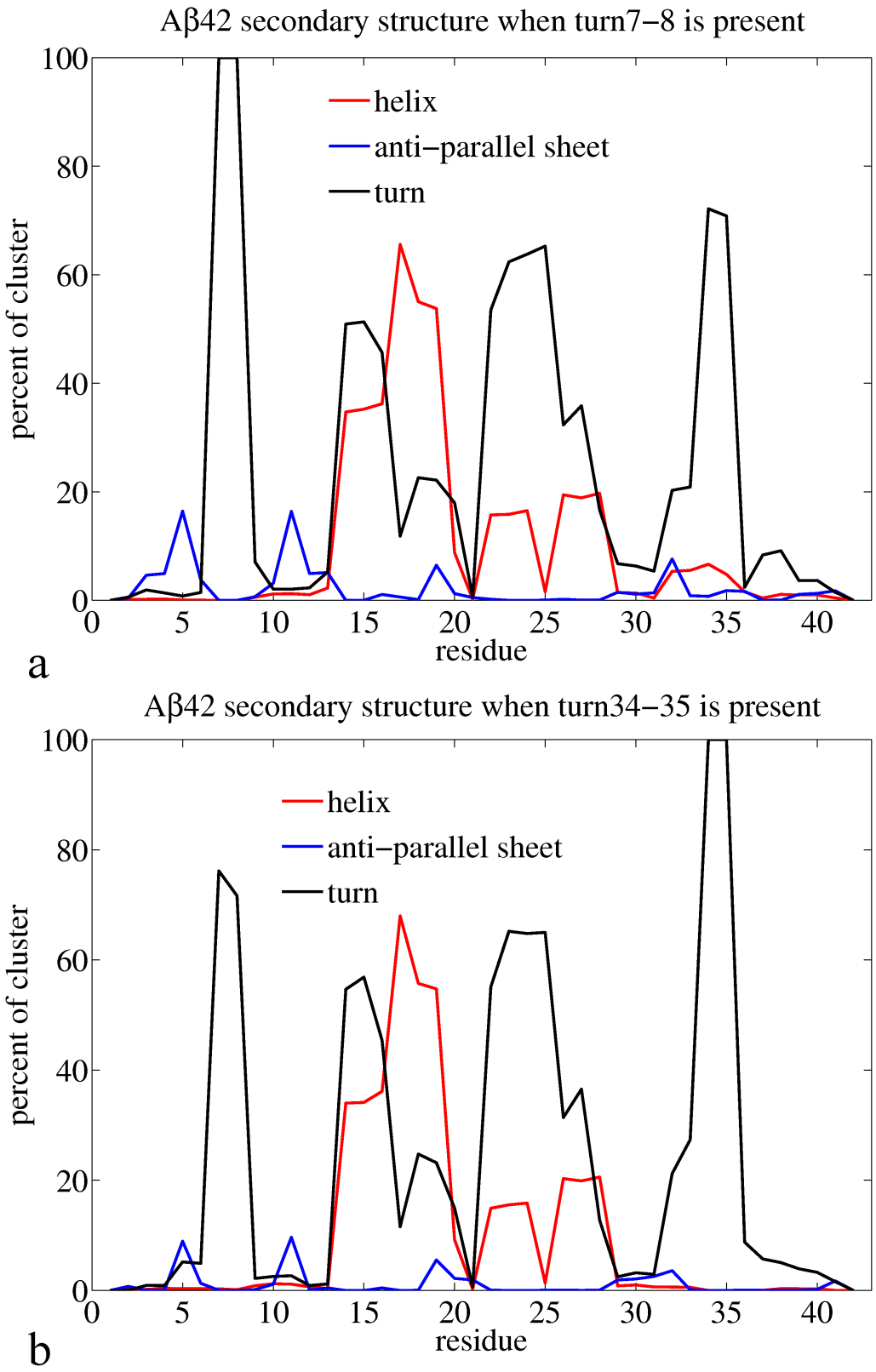


Figure 2.8. *Percent of Aβ42 ensemble in secondary structure by residue involving the most ubiquitous turns.* For fraction of ensemble where (a) Turn7-8 is always present and (b) Turn34-35 is always present.

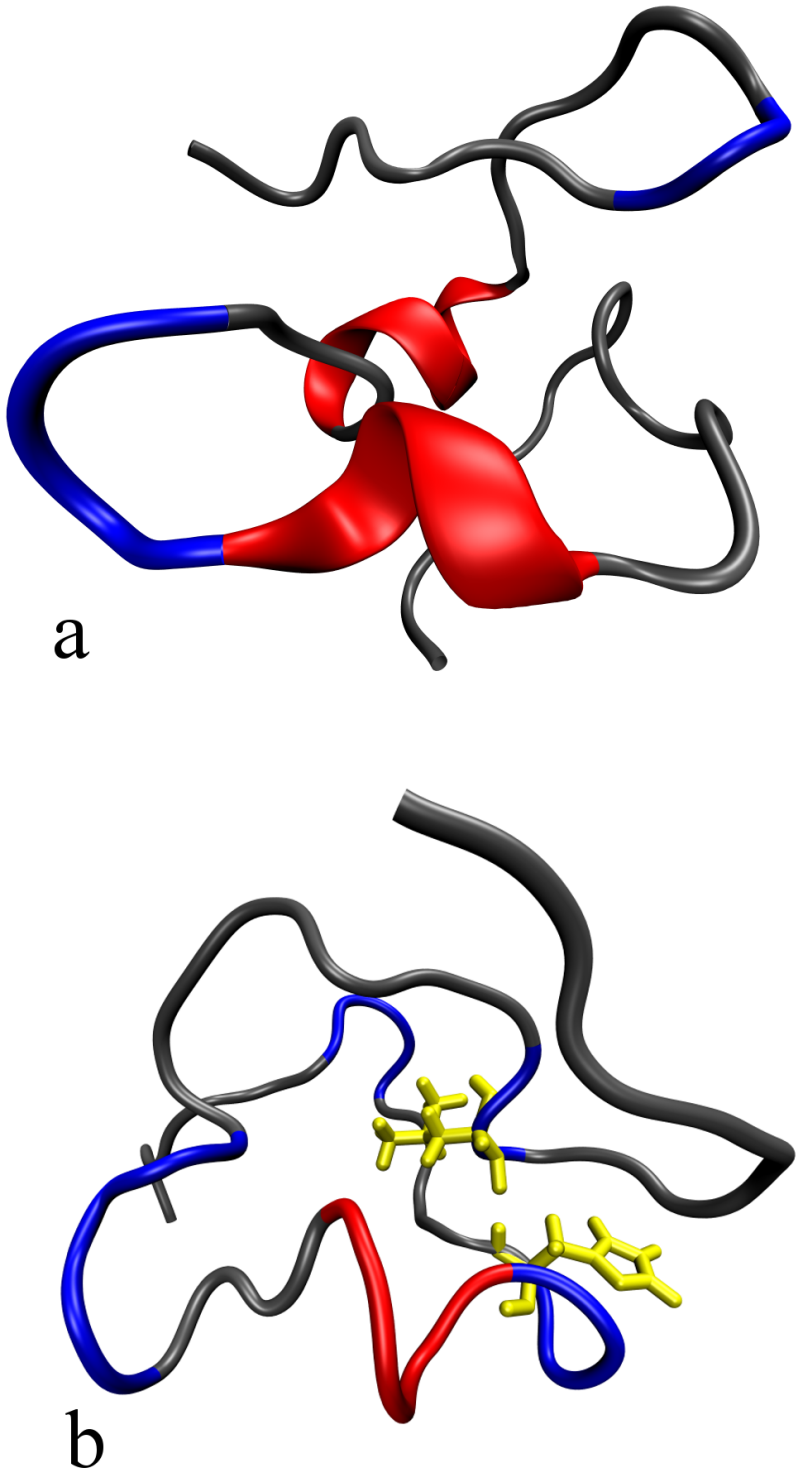


Figure 2.9. *Representative ribbon conformations of the Aβ42 populations.* Turns (blue) and Helix (red). (a) Turn7-8 and Turn24-25 and Helix14-15 and Helix26-27; each of these pairs occurs with high frequency with each other (Table 2.2) and (b) the hydrogen-bond between backbone residues 13-34, which occurs in ~40% of the ensemble, occurs with high frequency when many secondary structural elements occur simultaneously, such Turn7-8, Turn14-15, Turn24-25, and Turn26-27, and Helix17-18 shown here.

The most unusual secondary structure feature for Aβ42 involves Turn26-27. While it simultaneously pairs with either Turn7-8 or Turn34-35 in ~14% of the ensemble, 12% of the time the formation of Turn26-27 occurs simultaneously with β-strand structure involving Beta16-21 and Beta29-36, and this β-strand structure precludes the formation of many of the other secondary structures such as Turn14-15, Turn18-19, Turn24-25, Turn34-35, Helix14-15, Helix17-18, Helix23-24, Helix32-33, and of course Helix26-27, which are so common in the rest of the ensemble. In fact 8% of the time Turn26-27 occurs in isolation (compared to 0-3% for all other turns or helices, Table 2.3), i.e. with all other secondary structure categories completely absent, and when this happens the β-strand is almost always present (Figure 2.10a). This is unlike the case when the N-terminus anti-parallel β-strand (Beta3-6 and Beta10-13), nucleates around Turn7-8 (9% of the ensemble), because this localized structure always forms simultaneously with other secondary structure categories in the latter two thirds of the sequence (Figure 2.10b). Thus, while Turn26-27 is not the most dominant feature of secondary structure, present in only 28% of the ensemble, it is one of the most important features because it can have a much longer-range effect on the peptide's structure than all other secondary structure categories that are more structurally localized. There is still a significant percentage (9%) of the population where the Beta16-21 and Beta29-36 pairing forms without Turn26-27 present (Figure 2.10c), so that in total this β-strand occurs in 21% of the ensemble. In 5% of the ensemble, Beta16-21 forms a β-strand with some other part of the peptide, and in 4% of the ensemble Beta29-36 forms a β-strand with another region besides Beta16-21, although in both cases the pairing partner region is highly promiscuous.

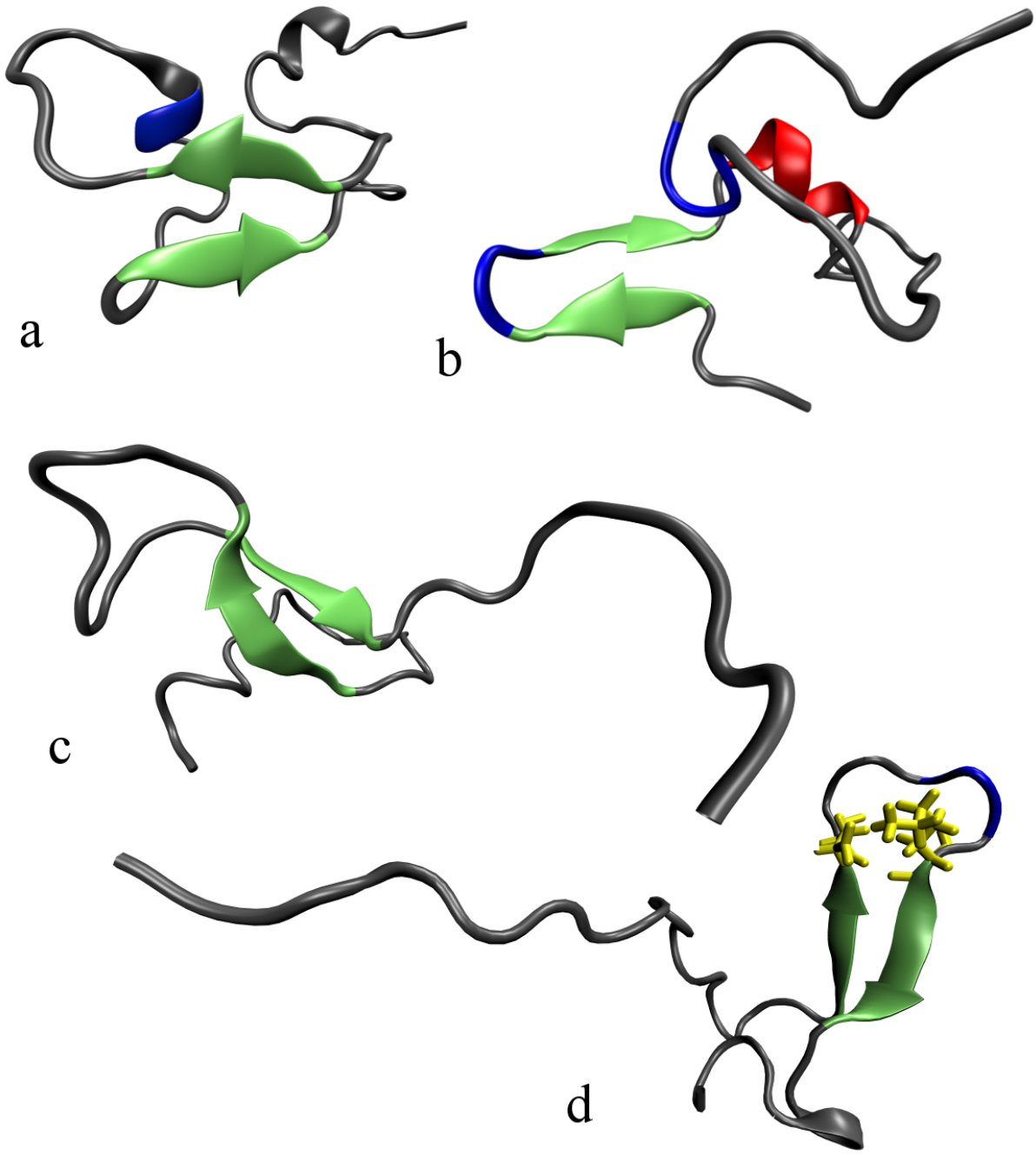


Figure 2.10. *Representative ribbon conformations of the Aβ42 populations.* Turns (blue), Helix (red), and β-strands (green). (a) Turn26-27 forms with β-strand (Beta16-21 and Beta29-36), and no other major secondary structure is present. (b) Turn7-8 nucleating β-strand Beta3-6 and Beta10-13, along with Helix14-15 and Helix17-18, and Turn34-35. (c) β-strand involving Beta16-21 and Beta29-36 in the absence of Turn26-27, or any other of the major secondary structure features. (d) Turn26-27 with β-strand Beta16-21 and Beta29-36, stabilized by a salt bridge between Glu22 and Lys28, in the absence of other secondary structure.

Other interesting structural features of the ensemble are the presence of particular hydrogen bonding partners, and Table 2.4 emphasizes that most of these are stabilizing the secondary structure categories above has required by the DSSP definitions. Stabilizing salt bridges are also found in the Aβ42 ensemble, in which Arg5 often participates in various salt-bridges: the largest occupancy (16% of the ensemble) is between the Arg5 and Glu3 side chains, and Arg5 also forms frequent salt-bridges with Asp1, Glu11, and Glu22. The Asp23-Lys28 salt-bridge, found in the Aβ42 fibril conformation as well as 7% of the Ab21-30 ensemble, and its competitor Glu22-Lys28, are found in 3% and 4% of the ensemble, respectively, and not surprisingly they co-form with helix and turns that they encapsulate: Turn24-25, Turn26-27, Helix23-24, Helix26-27. In only 1% of the population does a salt bridge stabilize the Beta16-19 and Beta29-36 β-strand (Figure 2.10d), which is thought necessary for stabilizing the Aβ42 monomer in the greater amyloid fibril assembly, although we only observe it for the competitor Glu22-Lys28 pairing (103, 109).

Residue - Residue hydrogen-bonds	**% *in Aβ42* Ensemble**	**Residue - Residue hydrogen-bonds**	**% *in Aβ42* Ensemble**
2 5	**6%**	**18 34**	**10%**
4 7	**11%**	**19 32**	**8%**
4 8	**7%**	**21 24**	**16%**
6 9	**50%**	**21 25**	**10%**
6 10	**13%**	**21 26**	**24%**
7 9	**6%**	**21 27**	**20%**
8 10	**6%**	**22 25**	**42%**
8 11	**5%**	**23 26**	**13%**
8 12	**5%**	**24 27**	**18%**
10 12	**6%**	**24 28**	**11%**
13 16	**50%**	**25 28**	**24%**
13 17	**39%**	**25 29**	**5%**
13 34	**41%**	**26 28**	**9%**
14 17	**7%**	**26 29**	**13%**
14 18	**5%**	**29 31**	**9%**
16 19	**40%**	**29 32**	**7%**
16 20	**12%**	**31 34**	**13%**
16 36	**10%**	**31 35**	**10%**
17 20	**30%**	**32 35**	**8%**
17 21	**9%**	**33 36**	**32%**

Table 2.4. *Hydrogen-bonds between residue pairs most frequently found (>5%) in Aβ42 ensemble.* Hydrogen bonds that define Turn7-8 (pink); Turn14-15/Helix14-15 (orange); Turn18-19/Helix17-18 (yellow); Turn24-25/Helix23-24 (green); Turn26-27/Helix26-27 (light blue); Turn34-35/Helix32-33 (purple); β-strands (brown). See Figure 2.10 where colors correspond to contact map.

Finally Figure 2.11 provides a contact map of strong experimental NOE intensities for Aβ42 that are dominated by a single contact. We note that these dominant NOEs correlate well both with the various secondary structures (including β-strands) and the presence of NOEs that "compact" structure (i.e., those which involve multiple secondary structure categories occurring simultaneously) that are described in Figures 2.6-2.10. Moreover, they are highly consistent with

the hydrogen bonds that occur with high frequency in Table 2.4 and lend strong support to the observed structural sub-populations.

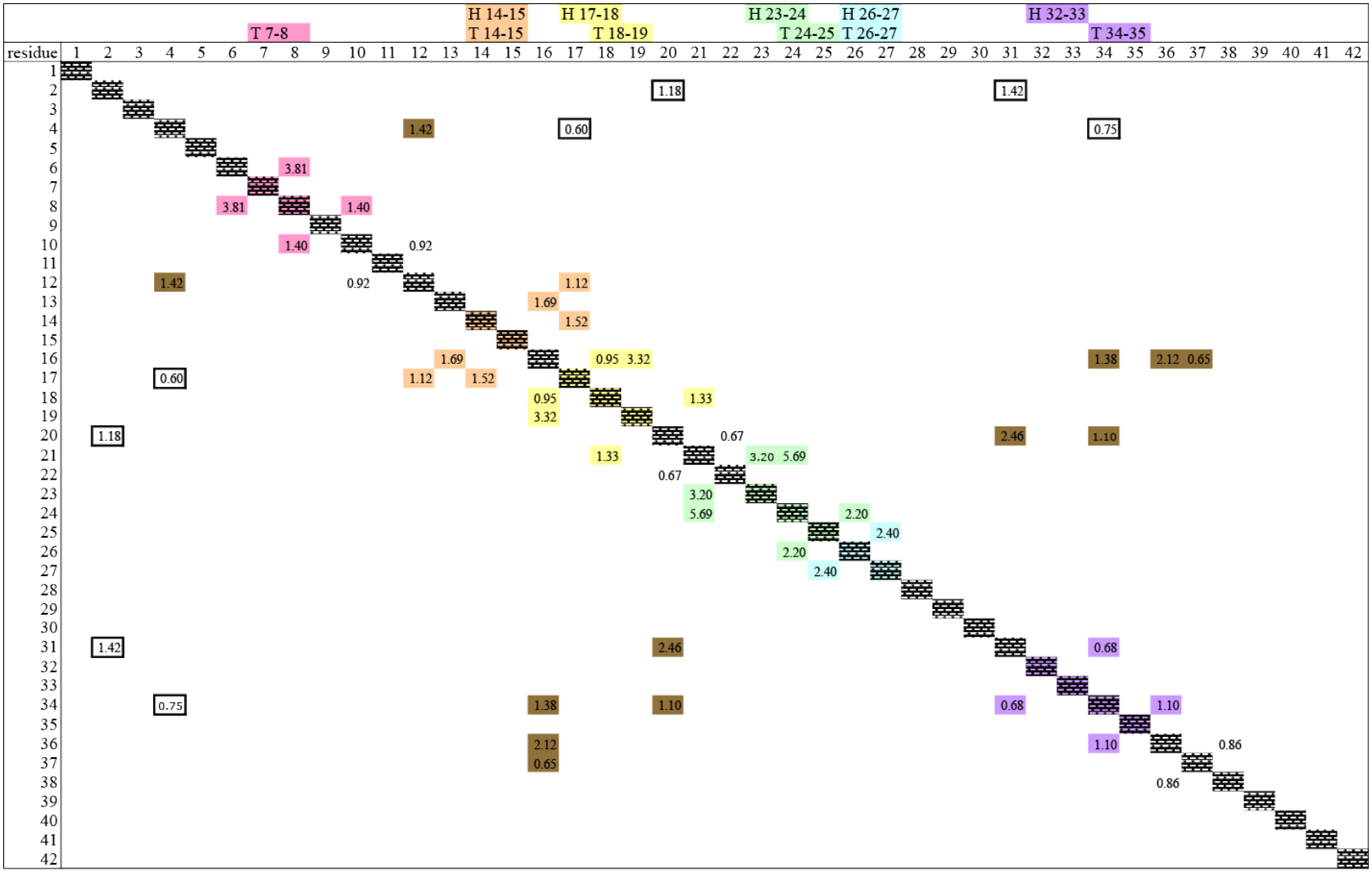


Figure 2.11. *Contact map of strong experimental NOE intensities for Aβ42 that are dominated by a single contact.* Strong experimental intensities that define Turn7-8 (pink); Turn14-15/Helix14-15 (orange); Turn18-19/Helix17-18 (yellow); Turn24-25/Helix23-24 (green); Turn26-27/Helix26-27 (light blue); Turn34-35/Helix32-33 (purple); β-strands are shown in brown; hydrogen bonds that "compact" structure, and which involve multiple secondary structure categories occurring simultaneously, are outlined in black. Intensity values are reported in each reported single contact pair.

Discussion and Conclusions

Our NMR/*de novo* MD study of the Aβ21-30 and Aβ42 conformational ensembles reveals a broad range of intrinsically disordered peptide structure. This range of structural complexity can be usefully described with techniques that combine the collection and analysis of both experimental NMR and simulation data using more modern force fields. Using a combination of NOESY cross-peaks, which reveal tertiary interactions via definitive intra-molecular contacts, and MD simulations that provide the structural context for such contacts in an equilibrium ensemble at conditions that match the experiment, we achieve a picture of two very different intrinsically disordered peptides. While the Aβ21-30 peptide conforms to a rather homogeneous ensemble consistent with an extended random coil, Aβ42 exhibits traits of an extremely heterogeneous ensemble of peptide conformations that contain a diversity of localized as well as long-range tertiary structure.

Of interest here is that functional but intrinsically disordered peptides and proteins are typically classified by having a dominant population corresponding to a particular "degree of disorder": collapsed semi-ordered ensembles (radius of gyration similar to that of a globular protein of same sequence length, typically containing well-formed secondary structure but little organized tertiary structure), collapsed disordered ensembles (radius of gyration typical of a globular protein and containing little well-formed secondary or tertiary structure), or an extended disordered ensemble (much larger radius of gyration than a globular protein and dominated by random coil). The Aβ42 peptide does not fall into any of these classifications exclusively, rather the NMR and MD show that its structural populations span the full range of classifications. For example, unlike molten globules (or ordered proteins), the Aβ42 hydrogen chemical shifts are not highly dispersed - 98% of the chemical shifts assume values within a standard deviation of the average shift for each residue type along the Aβ42 sequence. The J-coupling values also provide no evidence of secondary structure "blocks" at different points in the peptide sequence of Aβ42. Because Aβ42 samples so many distinct conformations, experimental observables that are based on simple ensemble averages like chemical shifts and spin-spin couplings yield values for these observables that are close to random-coil values. While this might be interpreted as a signature of an extended IDP, in fact the homogeneity of the chemical shifts and J-coupling are a result of averaging over many heterogeneous conformations involving compact, structured conformations, in addition to a significant population (~10%) of more extended conformers that span the R_g range from 9 Å to 30 Å. The NOESY experiment also confirms that these compact conformations must persist on long enough time-scales for the cross relaxation between protons to be detected, and thus these interactions must be stabilized by (likely shallow) free-energy minima. While the MD simulations show that many medium- to long-range NOEs in the Aβ42 NOESY spectra are reporting on semi-order in the collapsed conformations, we also observe numerous instances where a single peak is made up of many different contacts arising from different conformations with completely distinct tertiary structures and degree of compactness. By contrast, Aβ21-30 is more easily classifiable as an extended random coil.

The many strong NOESY cross-peaks observed experimentally for Aβ42, which we are able explain with specific attributes of the peptide populations from our simulated ensemble, allow us to corroborate or question specific structural features proposed in previous work. The

first NMR studies of the monomeric structural ensemble of Aβ42 (and Aβ40) by Hou *et al*. was directed toward changes that occurred upon oxidation of the Met35 side chain, a chemical change associated with impeded fibrillization of the peptide (14). Based on ^{1}H, ^{15}N, and ^{13}C NMR chemical shifts and backbone NOEs, that study found an absence of any well-defined secondary or tertiary structures, interpreting from the NMR data that the ensemble was well-classified as random, extended chains. Although our results show that Aβ42 is not really an extended random coil, that was a reasonable conclusion based on the more limited analysis of NMR data alone, since simple averaging over an ensemble of a highly diverse set of secondary and tertiary structures, a majority of which are collapsed, can generate average NMR observables consistent with extended random coil signatures. However, on the basis of NOE and chemical shift data, Hou *et al*. identified backbone C_α and C_β chemical shifts consistent with β-strand structure in two hydrophobic regions (Leu17-Ala21 and Ile31-Val36), as well as turn structures at two largely hydrophilic regions (Asp7-Glu11 and Phe20-Ser26). Those results are consistent with the present study in which we found significant β-strands (Beta16-21 and Beta29-36), as well as turns (Turn7-8, Turn24-25, and Turn26-27) largely in the same region of the sequence. What our study offers is a far richer picture as to how these secondary structural elements organize together to create a diverse set of collapse to extended conformations that contribute to the broad ensemble average.

A study by Sgouarkis and co-workers consisted of an MD simulation using the OPLS force field and the TIP3P water model, which they validated against measured scalar $^{3}J_{HNH\alpha}$ data (89). Although the agreement with the experimental J-coupling data was not particularly good, they nonetheless characterized the conformational ensemble of Aβ42 (and Aβ40) that pointed to a unique feature of the longer peptide forming a β-hairpin at Met35 and Val36, bringing together short anti-parallel β-strands at residues Ile31-Leu34 and G38-Ile41. This prediction of large population of β-strand structure in this region clearly contradicts the present study and the earlier Hou *et al.* NMR results since NOEs arising from these structure are not observed. More recently Sgourakis and co-workers performed a second MD study on Aβ42 (90), this time using a combination of ff99SB force field and the TIP4P-Ew water model that we successfully used on Aβ21-30 (26) and which we have used in this study. They analyzed 11,570 Aβ conformations using a clustering algorithm that yielded thousands of small clusters, with the six largest clusters having populations in the range 2-6% of the ensemble (90). This analysis does not provide much structural insight, and we found that classifications based on energetically stabilizing features such as hydrogen bonds and secondary structure to be more informative. Again they find central importance of the short β-strand at residues 38-40, which they argue may act as a conformational switch by forming alternative interactions with other β-strands along the sequence, in disagreement with this study and that of Hou and co-workers.

In summary, the full Aβ42 peptide and the shorter Aβ21-30 fragment we studied previously (26), represent two fundamentally different types of IDPs. Experimentally, while the average chemical shift and scalar coupling values suggest that they both are random coil polymers, it is only true for Aβ21-30, an outcome supported by a *de novo* molecular simulation ensemble that predicts its small number of 28 non-sequential NOESY cross-peaks (26). By contrast, the NMR observables for Ab42 represent an average over a rich and diverse set of tertiary structures, supported by a larger number of 179 non-sequential NOESY cross-peaks that are reasonably described by our MD ensemble. While the Aβ21-30 structural ensemble at 283K

is primarily extended, with ~65% of the structures containing no secondary structure or hydrogen-bonding pairing for any residue, the Aβ42 structural ensemble at 287 K has <1% of the population in which there is no secondary structure. In fact, key structures such as the β-strand involving Beta16-21 and Beta29-36 could even be crucial to oligomerization of fibril nucleation, although this is currently speculation. Aβ21-30 has the same primary sequence as a key middle segment of the Aβ42 peptide, however the additional interactions possible in the longer peptide shift the equilibrium to other stabilizing conformations dominated by a different set of residues, with the exception of ~5% of the population that contains the Asp23-Lys 28 salt bridge that is observed in both peptides. The full Aβ42 sequence may be necessary to interpret any physiological behavior of the peptide, because it is only the complete peptide that adopts a diverse but structured conformational ensemble.

APPENDIX

Multi-Reservoir Replica Exchange (MRRE) ensemble sampling method

The *sander* module of AMBER (110, 111) and the MRRE method presented by Ruscio *et al*. (30, 31) were used to achieve a Boltzmann weighted ensemble of Aβ42 structures at the experimental temperature of 287 K. Initially, parallel-tempering replica exchange (112) was used to converge a Boltzmann reservoir at 397 K using 13 replicas spaced from 447 K to 397 K (the spacing of replica temperatures at all stages of simulation was roughly exponential, ranging from every 5 K at the highest temperatures to every 2 K at the lowest). Exchanges were attempted every 1.0 ps. After the 397 K replica was sufficiently converged, the last 30 ns of simulation time was used to create a 397 K reservoir for the next set of replica exchange, taking 30,000 conformational snapshots spaced one picosecond apart.

In the next stage of MRRE, the 397K reservoir was used to exchange with the highest temperature replica of 18 replicas spaced between the 397 K reservoir and a low temperature replica of 335 K, using the Reservoir Replica Exchange method (30). After the 335 K replica converged, the final 30 ns of simulation were used to create a 335 K reservoir for the next set of replica exchange, taking 30,000 conformational snapshots spaced one picosecond apart. In the final stage of MRRE, the 335K reservoir exchanged with the highest temperature of 18 replicas that included the lowest temperature of 287 K, corresponding to the NMR experimental temperature. 50,000 structures from the final 50 ns from the converged 287 K replica were taken to construct the 287 K structural ensemble, and used to calculate conformational statistics. Aβ42 needed 40 ns to converge the 397 K reservoir, 60 ns to converge the 335 K reservoir, and 60 ns to converge the 287 K ensemble. The total simulation time combining every replica was 5.36 ms for Aβ42, resulting in 100 ns of equilibrated NVT simulation at the lowest temperature of 287 K.

Convergence of each intermediate reservoir, as well as the low temperature ensemble, was determined by comparing average intra-atomic distances for peptide hydrogen atoms between the two independent simulations using the *ptraj* module of the AMBER software package. Various statistics measuring agreement of these two ensembles were plotted over time (Figure 2.12), and generally, we focused on the short intra-atomic distances (<7.0 Å) that would have the greatest effect on experimental NOE measurements. When greater than 85% of the

short distances were converged within 10% of the total distance, the sampling was deemed sufficient.

Atom 1	Atom 2	Calculated H_2O NOE intensity	Experimental H_2O NOE intensity	Calculated ^2H_2O NOE intensity	Experimental ^2H_2O NOE intensity
Hε Tyr 10	Hγ2 Val 12	0.92	1.63	0.85	1.84
Hγ3 Glu 22	Hγ Val 24	0.58	3.29	0.52	
Hδ Tyr 10	Hγ2 Val 12	0.52	1.52	0.46	1.70
Hδ Phe 20	Hγ Val 24	0.47	1.98	0.43	3.44
Hε Phe 20	Hγ2 Glu 22	0.46	1.32	0.42	
Hδ1 Leu 17	Hε Phe 19	0.45	3.06	0.41	5.07
Hδ Phe 20	Hγ3 Glu 22	0.28	1.70	0.26	
Hβ Asp 7	Hδ Tyr 10	0.19	1.50	0.17	
Hγ2 Val 18	Hδ Phe 20	0.16	2.85	0.14	2.70
Hγ2 Val 18	Hε Phe 20	0.16	1.57	0.14	2.98
Hγ2 Val 18	Hζ Phe 20	0.06	1.52	0.06	1.63

Table 2.5. List of experimentally assigned NOE peaks that are not due to intra-residue or sequential contacts along with calculated NOE values from simulation for both H_2O and ^2H_2O.

Atom 1	Atom 2	Experimental NOE		Peaks in Spectrum 1		Peaks in Spectrum 2	
		H_2O	^2H_2O	H_2O	^2H_2O	H_2O	^2H_2O
Hγ Val 24	Hδ Phe 20	1.26	3.44	2	2	2	2
Hβ Asp 7	Hα Ala 21	3.36		1	0	2	2
Hβ2 Gln 15	Hγ2 Val 12	3.02		2	0	2	0
Hε Phe 20	Hγ2 Val 18	1.57	2.98	1	1	1	1
Hγ2 Val 18	Hδ Phe 20	1.50	2.70	2	2	2	2
Hδ2 Leu 17	Hδ Tyr 10	1.14	2.81	1	2	1	0
Hγ2 Val 12	Hβ Asp 7		2.62	0	1	0	0
H Val 24	Hα His 13	2.61		1	0	1	0
Hβ2 Asp 23	Hα Gly 25		2.43	0	1	0	0
Hδ2 His 13	Hη1 Arg 5	2.39		1	0	1	0
Hγ Val 24	Hδ Tyr 10	1.30	2.27	1	1	1	1
Hβ2 Glu 3	Hβ2 Asp 1	2.16		1	0	1	0
Hβ2 His 13	Hα Glu 11		2.07	0	1	0	1
Hε1 His 6	H Ala 30	2.03		1	0	1	0
Hβ2 Asp 1	Hδ2 His 13		2.02	0	1	0	0
Hδ2 His 6	Hα His 13		1.87	0	1	0	1
Hζ Phe 19	Hα Asp 23	1.72		1	0	0	0
H Asp 7	Hβ2 Arg 5	1.72		1	0	1	0
Hβ2 Glu 11	Hβ2 His 13	1.67		1	0	1	0
Hζ Phe 20	Hγ2 Val 18	1.52	1.63	1	2	2	2
Hδ22 Asn 27	H Ile 32	1.63		1	0	1	0
Hδ21 Asn 27	Hε22 Gln 15	1.57		1	0	1	0
Hε21 Gln 15	Hδ2 His 6	1.53		1	0	0	0
Hδ Tyr 10	Hβ Asp 7	1.50		1	0	0	0
Hδ2 His 13	Hβ3 Glu 11	1.49		1	0	1	0
Hδ2 His 13	Hα Glu 11	1.48		1	0	1	0
Hδ2 His 6	Hα Glu 3	1.47		1	0	1	1
Hδ2 Leu 17	Hε Tyr 10	0.60		2	2	0	1
Hδ2 His 14	Hβ Val 12	1.41		1	0	1	0
Hβ3 Asp 1	Hα Phe 4	1.36		1	0	0	0
H Ala 30	Hα Met 35	1.32		1	0	0	0
Hδ22 Asn 27	Hγ12 Ile 32	1.31		1	0	1	0
Hε21 Gln 15	Hγ2 Glu 22	1.29		1	0	1	0
Hδ Tyr 10	H His 13	1.29		1	0	1	0
Hγ2 Val 18	Hδ2 His 14	1.29		1	0	1	1
Hδ1 Leu 17	Hδ Tyr 10	1.26	0.66	1	1	1	1
Hγ3 Glu 3	H Ser 8	1.11		1	0	1	0
Hγ Leu 17	H Val 24	1.07		1	0	0	0
Hδ2 His 13	Hβ3 Tyr 10	1.04		1	0	1	0
Hβ Phe 4	H Tyr 10	1.03		1	0	1	0
H Leu 17	Hα Ser 8	1.03		1	0	0	0
H Glu 22	Hβ Ala 42	1.02		1	0	0	0
H His 13	Hβ2 Arg 5	1.00		1	0	1	0
Hβ3 His 13	Hε Tyr 10	0.89		1	0	2	1
Hβ3 Phe 19	Hδ1 Ile 41	0.88		1	0	2	1
Hε Lys 16	Hδ Phe 20		0.74	0	1	0	1

Table 2.6. List of false-negative NOE peaks (identified experimentally but not calculated from simulation) that cannot be explained by: the presence of high noise in that region of the spectrum or by a small simulation inaccuracy that identifies the correct residue interaction but misses the exact atomic contact. Listed are the experimental intensity for both the H_2O and ^2H_2O spectra, and the atomic contact which would produce this peak. If more than one atomic contact could produce this NOE peak, the contact is listed with the smallest difference in sequence between the two residues. The number of spectra in which peak is observed experimentally is also indicated.

MD H2O NOE intensity	MD D2O NOE intensity	atom 1	atom 2	MD contact intensity	atom 1	atom 2	MD contact intensity	atom 1	atom 2	MD contact intensity	atom 1	atom 2	MD contact intensity	atom 1	atom 2	MD contact intensity
		Hε1 HIS 13	Hγ GLN 15		Hε1 HIS 14	Hγ GLN 15										
	7.36			6.02			1.42									
		Hα VAL	Hβ2		Hβ2	Hα2		Hβ2 TYR	Hα GLY			Hα VAL			Hα GLY	
7.32	5.31	12	HIS 13	1.81	HIS 6	GLY 9	1.23	10	25	0.77	Hβ2 HIS 13	36	0.76	Hβ2 HIS 13	37	0.65
			Hα		Hβ3	Hα			Hα		Hα					
		Hε	GLY		TYR	GLY		Hε	GLY		GLY	Hε				
7.24	6.85	LYS 16	37	2.08	10	25	0.77	LYS 16	38	0.65	25	LYS 28	0.61			
		Hα VAL	Hα MET		Hα MET	Hα VAL										
6.21	5.43	12	35	3.80	35	36	0.74									
		H	H TYR		H HIS	H										
6.20	0.00	GLY 9	10	5.62	6	GLY 9	0.55									
		Hα GLU	Hα GLY		Hα GLY	Hα VAL		Hα GLU	Hα VAL							
6.03	4.28	22	25	2.62	38	39	1.27	11	12	1.10						
		Hβ2 TYR	Hβ MET		Hβ2	Hε MET		Hβ2 TYR	Hε MET							
5.91	5.58	10	35	2.34	HIS 6	35	0.68	10	35	0.67						
		Hδ2	Hα VAL		Hε22 GLN	Hα GLY		Hδ2	Hα GLY							
5.13		HIS 13	36	2.78	15	38	0.91	HIS 13	37	0.79						
		Hε1	Hβ2													
	5.11	HIS 6	SER 8	4.68												
		Hδ2	Hε MET		Hδ TYR	Hε MET		Hδ2	Hβ MET		Hδ TYR	Hβ MET		Hβ2 GLU	Hδ22 ASN	
4.95	3.74	HIS 6	35	1.05	10	35	1.00	HIS 6	35	0.99	10	35	0.92	22	27	0.88
		Hβ	Hβ MET		Hβ LEU	Hε MET		Hγ LEU	Hε MET		Hγ LEU	Hβ MET				
4.85	4.07	LEU 34	35	1.32	34	35	0.97	34	35	0.87	34	35	0.78			
		Hε1	Hβ3													
	4.61	HIS 6	SER 8	4.68												
		Hβ3 TYR	Hβ MET		Hβ3 TYR	Hε MET		Hβ3 GLN	Hε LYS							
4.52	3.84	10	35	2.34	10	35	0.67	15	16	0.57						
		Hα VAL	Hβ3		Hβ3	Hα VAL		Hβ3	Hα GLY							
4.11	2.74	12	HIS 13	1.81	HIS 13	36	0.76	HIS 13	37	0.65						
		Hα VAL	Hα		Hα	Hβ2		Hα	Hβ2							
3.99	2.33	12	HIS 13	1.26	PHE 20	SER 26	0.89	PHE 19	SER 26	0.72						

				Hβ2								
			Hε1	GLN								
		3.75	HIS 13	15	3.40							
				Hα				Hα				
			Hδ2	VAL		Hδ2		GLY				
		3.74	HIS 13	36	2.78	HIS 13		37	0.79			
				Hβ3								
			Hε1	GLN								
		3.66	HIS 13	15	3.40							
				Hε								
			Hβ2	TYR								
	3.51	3.29	SER 8	10	2.45							
				Hα								
			Hε1	GLY								
	3.41		HIS 13	37	3.07							
				Hε				Hβ				
			Hδ or ε	MET		Hδ or ε		MET				
	3.34	3.02	PHE 4	35	0.87	PHE 4		35	0.70			
				Hα								
			Hε1	GLY								
		3.25	HIS 13	37	3.07							
				Hα								
			Hε	GLY								
	2.77	2.13	PHE 20	29	0.96							
				H								
			Hα	GLY		Hα		H				
	2.69		ASP 23	25	1.56	ASP 7		GLY 9	1.10			
			Hβ	Hε		Hβ		Hβ2				
			ALA	MET		ALA		GLU				
	2.67	1.53	21	35	0.67	21		22	0.66			
				Hα								
			Hβ2	MET								
	2.66	2.55	HIS 13	35	2.22							
				Hα								
			Hζ	GLY		Hδ		Hβ2				
	2.63	2.33	PHE 20	29	0.57	PHE 20		SER 26	0.56			
				Hα								
			Hβ3	MET								
	2.58	2.62	HIS 13	35	2.22							
			Hβ2	Hα3								
	2.60	2.08	HIS 6	GLY 9	1.23							
			Hβ2	H								
	2.59		HIS 6	GLY 9	1.93							
			H	H								
			GLY	GLY								
	2.57		37	38	2.46							
				Hε				Hβ				
			Hζ	MET		Hζ		MET				
	2.49	2.14	PHE 4	35	0.78	PHE 4		35	0.67			

			Hε								
		Hβ3	TYR								
2.45	2.30	SER 8	10	2.45							
			Hα								
		Hδ	GLY								
2.42	1.92	LYS 16	25	1.22							
		Hβ	Hβ2		Hβ3	Hβ2		Hβ3	Hβ3		
2.41	2.00	ASP 7	SER 8	0.65	ASP 23	SER 26	0.61	ASP 23	SER 26	0.61	
		Hβ3	Hα2								
2.41	2.14	HIS 6	GLY 9	1.23							
		Hβ3	Hα3								
2.19	1.95	HIS 6	GLY 9	1.23							
		Hβ3	H								
2.16		HIS 6	GLY 9	1.93							
		Hγ	Hα								
		GLN	GLY								
2.12	1.60	15	37	1.04							
		Hα	H								
		GLU	GLY								
2.11		22	25	1.93							
		Hα	Hα								
		Hα	VAL		Hα	Hα					
1.93	1.68	ASP 23	24	0.74	ASP 1	ASP 7	0.67				
		Hα	Hβ3		Hα	Hβ3					
1.91		PHE 20	SER 26	0.89	PHE 19	SER 26	0.72				
		H									
		ALA	Hβ3		H PHE	Hβ3					
1.87		21	SER 26	1.12	20	SER 26	0.56				
		Hα	Hβ2		Hα	Hβ2					
		GLU	ASN		ALA	ASN					
1.77	1.73	22	27	1.03	21	27	0.61				
		Hα	Hβ2								
		VAL	ASN								
1.74	1.75	24	27	1.36							
		Hα	Hα								
		GLN	GLY								
	1.72	15	37	1.50							
			Hα		Hγ	Hα					
		Hβ	MET		LEU	MET					
1.72	1.36	LEU 34	35	0.93	34	35	0.55				
		Hα	Hα								
		TYR	GLY								
1.69	1.56	10	25	1.22							
		Hα	Hγ2								
		VAL	MET								
1.65	1.51	12	35	0.64							

		Hα VAL	Hγ3 MET					
1.65	1.51	12	35	0.64				
		Hβ3	Hδ TYR					
	1.64	SER 8	10	1.40				
		Hα GLY	Hα					
1.63	1.25	25	SER 26	1.58				
		Hα TYR	Hε MET					
1.55	1.47	10	35	0.61				
		Hε1	Hα GLN					
	1.46	HIS 13	15	1.21				
		Hβ	Hγ2 MET					
1.41	1.24	LEU 34	35	0.92				
		Hβ	Hγ3 MET					
1.41	1.24	LEU 34	35	0.92				
		Hγ2 VAL	Hε MET					
1.32	1.23	12	35	0.84				
		Hβ	Hβ3 ASN					
1.31	1.17	ALA 2	27	0.53				
1.31		H ASP 7	H GLY 9	1.14				
		Hδ2	Hγ GLN					
	1.12	HIS 14	15	1.00				
		Hδ2	H GLY					
1.10		HIS 13	37	1.06				
		Hε1	Hβ					
	1.05	HIS 6	ASP 7	1.21				

Table 2.7. List of false-positive NOE peaks (calculated from simulation but not identified experimentally) that cannot be explained by the presence of a high noise in that region of the spectrum or by a small simulation inaccuracy which identifies the correct residue interaction but misses the exact atomic contact. If the peak is made up of more than one atomic contact, the calculated intensity for each contact with an intensity greater than 0.5 for either H_2O or ^2H_2O is listed.

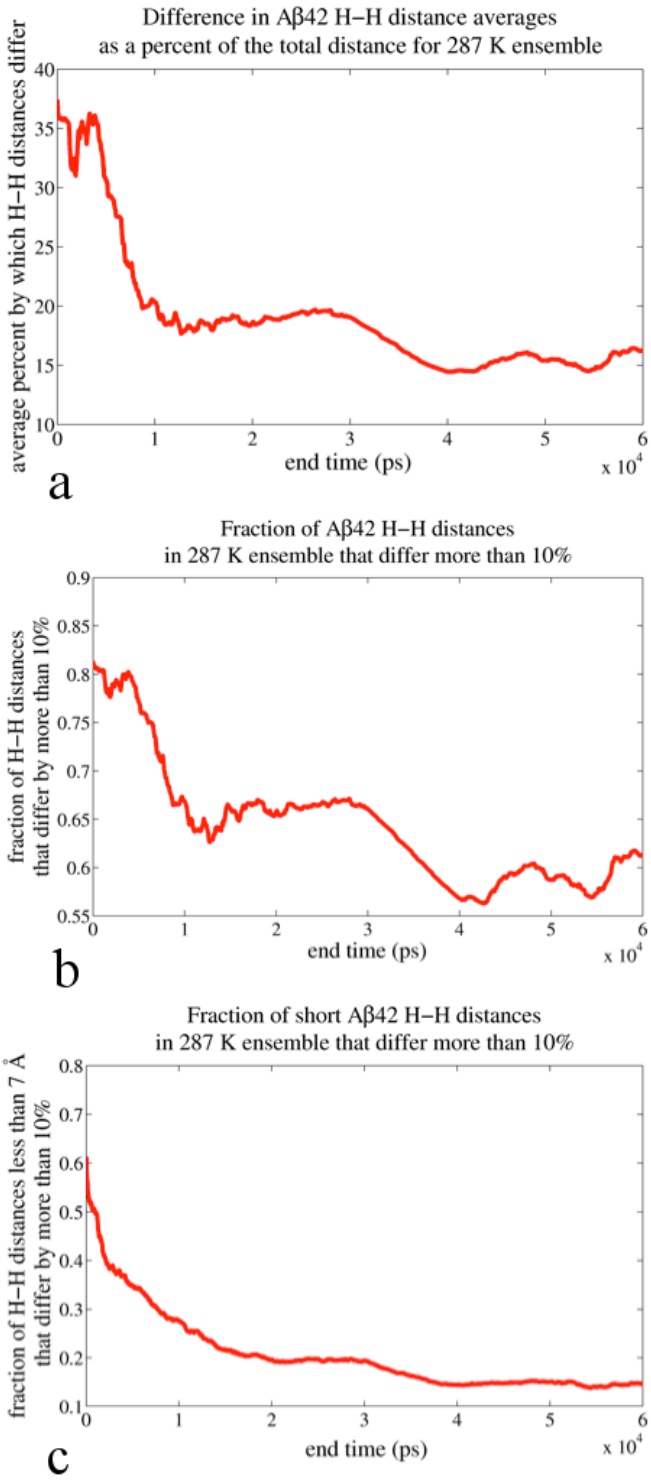


Figure 2.12. *Measure of convergence of 287 K Aβ42 simulated ensemble averaged over 50ns.* (a) The average percent of the mean by which Aβ42 inter-hydrogen distances from two simulations differ. (b) Fraction of all Aβ42 inter-hydrogen distances where the average distance for two simulations differs by more than 10% of the mean distance. (c) Fraction of short (less than 7 Å) inter-hydrogen distances where the average distance for two simulations differs by more than 10% of the mean distance.

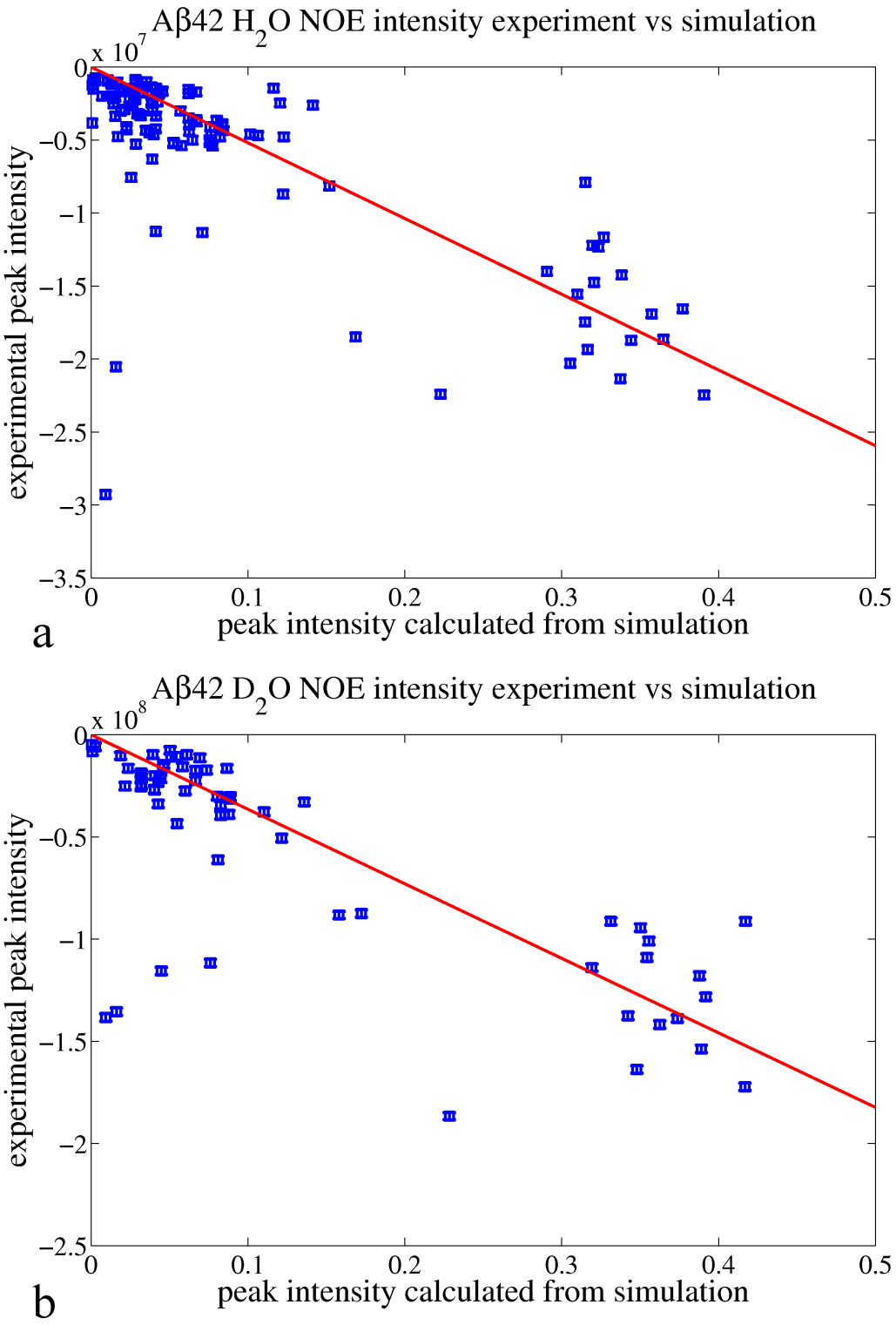


Figure 2.13. *Plot of assigned experimental NOE intensities to calibrate the simulated intensities* (a) H_2O NOESY spectrum and (b) ^2H_2O NOESY spectrum. The best-fit line is used to scale the simulation intensities to compare to experiment. The experimental intensities used are the average of the two peaks seen on either side of the diagonal, which differ somewhat due to base-line variation. Uncertainty bars correspond to the variation in peak intensity expected based on noise levels in areas of the spectrum that were not designated as 'high-noise regions'.

Chapter 3

Differences between the Aβ42 and Aβ40 monomer conformational ensembles generated from *de novo* molecular dynamics simulations

We have collected and simulated homonuclear ^{1}H NOESY spectra for the monomeric amyloid-β 1-40 (Aβ40) and amyloid-β 1-42 (Aβ42) peptides to show that Aβ42 forms a major β-strand between residues 16-21 and 29-36 in 30% of its structural ensemble, typically forming with a turn at residues 26-27 and with no other secondary structure present elsewhere along the chain, while Aβ40 favors a promiscuous set of β-strand conformations. We find that the two additional C-terminal residues of Aβ42, Ile41 and Ala42, form hydrophobic contacts with the same regions that form β-strands in the Aβ40 ensemble, directly disrupting these structures. Ile41 and Ala42 also form hydrophobic contacts in the C-terminus region, which promote β-strand structure that is compatible with known fibril forming regions of the Aβ sequence. This Aβ monomer structural information adds to the existing picture of how the two C-terminal residues of Aβ42 affect peptide aggregation rate, producing a plausible hypothesis for the Aβ42 fibrilization mechanism.

Introduction

The amyloid-β peptide is acknowledged to be a key molecular contributor to Alzheimer's Disease (AD), although the mechanism explaining disease symptoms and toxicity is not well understood (2). Because amyloid-β (Aβ) is generated through cleavage of the Amyloid Precursor Protein (APP), it is present in the brain as different length fragments of APP (113, 114). The most common of these are amyloid-β 1-40 (Aβ40) and amyloid-β 1-42 (Aβ42), which differ only at the C-terminus where Aβ42 has two additional hydrophobic residues (Ile41 and Ala42). Both peptides form insoluble fibrils and plaques in the extra-cellular space of the brain in AD patients (113, 114), but Aβ42 is more prevalent in these toxic aggregates (4-8, 49), and also

forms potentially toxic soluble oligomers that are not formed by Aβ40 (11, 115). Differences between Aβ40 and Aβ42 have also been observed *in vitro*. Aβ42 aggregates more quickly than Aβ40 (9, 10), and populates a larger distribution of oligomer sizes (116, 117). These striking behavioral dissimilarities indicate that a difference in sequence of only two residues has a substantial influence on structure and function for these disease peptides.

The Aβ40 and Aβ42 peptides have been characterized as intrinsically disordered in their monomeric state (14, 15), which makes structural investigations of their differences difficult. Intrinsically disordered proteins (IDPs) sample many different conformations that form an *ensemble* of structures, rather than folding to one stable structure. As shown in Chapter 2, IDP ensembles may consist of largely extended, random coil structures, as is the case for Aβ21-30 (26). In contrast the full length Aβ42 ensemble mostly consists of conformations containing secondary and tertiary structure (48). Although the energy landscape of an IDP does not contain a deep minimum corresponding to a stable folded structure, many shallow energy minima may exist, which are stabilized by structural features and inter-residue interactions (118). Because these minima are shallow, and the energetic barriers between them are small, the structural ensemble of an IDP can change more dramatically with small energetic variations due to differences in sequence. In the case of Aβ, a sequence difference of two residues causes a drastic difference in the behavior of Aβ40 and Aβ42, which may result from a similarly drastic difference in the monomer ensembles.

This study aims to assess these differences in the Aβ40 and Aβ42 structural ensembles using a combination of NMR spectroscopy and *de novo* molecular dynamics (MD) simulations. We have collected NOE data on both peptides that directly reports on tertiary structure contacts between residues. These data reveal differences in the long range contacts formed by Aβ40 and Aβ42 which are not accessible from other experimental observables such as circular dichroism (CD) spectra, NMR chemical shifts, or J-coupling constants (14, 67, 116, 119). The MD simulations add to this picture by allowing us to observe the individual conformations that populate the IDP ensembles. We use advanced simulation techniques to sample the Aβ40 and Aβ42 energy landscapes, creating ensembles of conformations without experimental biasing. By comparing observed NOEs with those calculated from the simulated ensembles, we can both validate and interpret the differences in structural sub-populations of the two peptides measured by NMR.

We find that the NMR and MD simulation data show that the Aβ42 ensemble contains a major β-strand between residues 16-21 and 29-36 in 30% of the monomer ensemble, typically forming with a turn at residues 26-27 and with no other secondary structure present elsewhere along the chain (48). Furthermore, Aβ40 exhibits far more promiscuous, but relatively unstable, β-strand structure compared to Aβ42, and does not populate the major β-strands and 26-27 β-turn that dominate the Aβ42 monomeric ensemble. The two additional C-terminal residues of Aβ42 are directly responsible for the differences in the populations of β-strands of the two amyloid peptides. Ile41 and Ala42 make hydrophobic contacts in regions of the sequence that preclude the β-strands observed in the Aβ40 ensemble. Additionally, Ile41 and Ala42 often form a hydrophobic cluster in the Aβ42 C-terminal region that promotes a dominant β-strand

population seen in the fibril subunit structure, but which is infrequently observed in the Aβ40 ensemble.

Methods

NMR Experiments

The Aβ40 and Aβ42 monomers were purchased and prepared according to the protocol in (48). Briefly, recombinant Aβ40 and Aβ42 peptides were purchased from a commercial source (rPeptide, Athens, GA) reported to be > 97% purity. The peptide was lyophilized from trifluoroacetic acid (TFA). The alkaline pretreatment of Aβ and preparation of LMW Aβ by filtration protocols outlined by Teplow (79) were used to prepare a monomeric solution of Aβ. The lyophilized peptide was dissolved in 2 mM NaOH to produce a peptide concentration of 0.21 mg/ml with a pH of > 10.5. The peptide was then sonicated for 1 min in a bath sonicator and lyophilized. It was then resuspended in 20 mM sodium phosphate buffer, pH 7.2, 0.01% (w/v) sodium azide. This protocol ensures that when the lyophilized peptide is dissolved in buffer it will not pass through its pI of ~5.31. Before collecting NMR data, the sample was filtered with a 0.22 μm filter to remove aggregates and fibril seeds (79), and brought to a concentration of ~600 μM at pH 7.2.

TOCSY and NOESY ^{1}H-^{1}H homonuclear spectra were collected for both peptides at the NMR Facility at UC Berkeley on Bruker Avance II 800 and 900 MHz spectrometers, the latter equipped with a Bruker cryogenic probe. The data was processed as described in (48). A more detailed description is provided in the Appendix.

Molecular Dynamics Simulations

We computed 287 K and 311 K equilibrium ensembles of Aβ40 and Aβ42 peptide conformations using Multi-Reservoir Replica Exchange (MRRE) (23, 24, 31) and AMBER 11 (111). The peptides were modeled with the Amber ff99SB force field (23, 24) solvated with TIP4P-Ew water (25), which has been shown to be the current best force field for reproducing NMR observables (90) and ^{1}H-^{1}H NOE data (26, 48). Two independent MRRE simulations for each peptide generated final ensembles of 70,000 - 90,000 structures pulled from 0.1 μs of replica exchange simulation time at each temperature. We also performed 100 separate 20 ns microcanonical ensemble simulations for each peptide in order to calculate time-correlation data. Further details are presented in the Appendix.

Calculation of NMR Observables

We calculated chemical shifts, J-coupling constants, Residual Dipolar Couplings (RDCs), and ^{1}H-^{1}H NOEs from our 287 K Aβ40 ensemble with the same procedure used for Aβ42 (48). All details of how the procedure differed in the case of Aβ40 are available in the Appendix.

Heteronuclear NOEs. We also calculated ^{1}H-^{15}N NOEs for the Aβ40 and Aβ42 backbone N-H atoms from the 287 K ensembles and correlation times. We use the same method as for the ^{1}H-

^{1}H NOEs (described in the Appendix) to calculate the spectral density function for each pair of H-N atoms from the short NVE simulations, resulting in a function

$$J(\omega) = \sum_{i=1}^{N}\left[a_i \frac{2\tau_i}{1+\omega^2\tau_i^2}\right] \tag{3.1}$$

for the H-N backbone bond of each residue of Aβ40 and Aβ42 where τ_i are the correlation time constants. We then calculate the steady state NOE enhancement factor of the ^{15}N spin by the ^{1}H NOE from our structural ensemble and dynamical trajectories according to

$$\varepsilon_{NOE} = 1 + \frac{\gamma_H \sigma_{HN}}{\gamma_N R_z^{(N)}} \tag{3.2}$$

where γ_H and γ_N are the gyromagnetic ratios of ^{1}H and ^{15}N, respectively. The ^{1}H-^{15}N cross-relaxation rate constant is given by

$$\sigma_{HN} = \frac{1}{10}K^2\left[3J_{ij}(\omega_{0,H}+\omega_{0,N}) - \frac{1}{2}J_{ij}(\omega_{0,H}-\omega_{0,N})\right] \tag{3.3}$$

and the ^{15}N self-relaxation by

$$R_z^{(N)} = \frac{1}{10}K^2\left[\frac{3}{2}J_{ij}(\omega_{0,N}) + 3J_{ij}(\omega_{0,H}+\omega_{0,N}) + \frac{1}{2}J_{ij}(\omega_{0,H}-\omega_{0,N})\right] \tag{3.4}$$

where $\omega_{0,H}$ is the Larmor frequency of ^{1}H and $\omega_{0,N}$ is the Larmor frequency of ^{15}N, and the constant factor K is defined as

$$K = \frac{\mu_0}{4\pi r_{eff}^{\ 3}}\hbar\gamma_H\gamma_N \tag{3.5}$$

where μ_0 is the permeability of free space, and $\hbar$ is Planck's constant. The effective *r* vector

$$r_{eff} = \left\langle \frac{1}{r^6(t)} \right\rangle^{-1/6} \tag{3.6}$$

is the average of the r^{-6} values, which has then been converted back into units of distance.

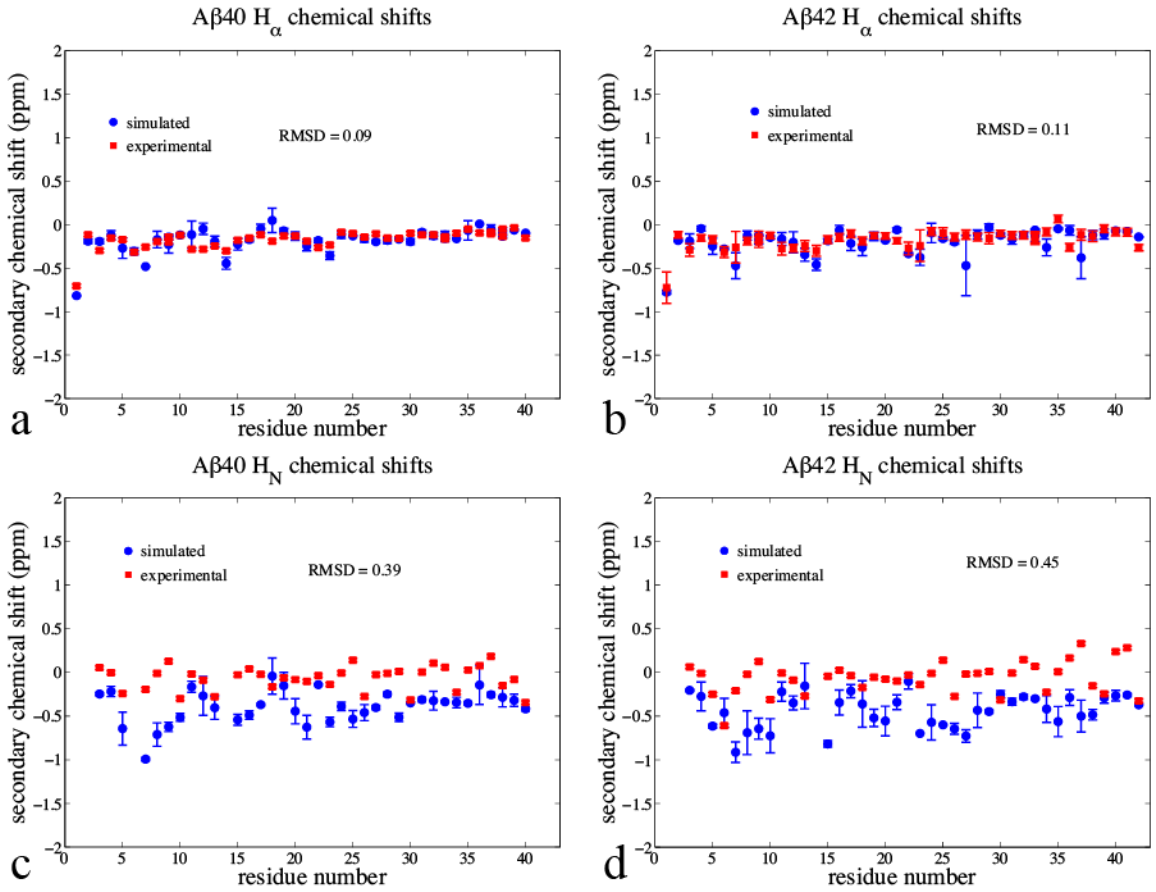
Ensemble Structural Analysis

Structural analysis of the *de novo* molecular dynamics simulations of Aβ40 and the Aβ42 287 K and 311 K ensembles was performed using *ptraj*, DSSP (97), perl scripts and MATLAB (The MathWorks, Natick, MA) scripts. The secondary structure designations we used to describe the Aβ40 and Aβ42 ensembles are all based on the DSSP criteria described by Kabsch and Sander

(97). We also used *ptraj* to identify the most commonly formed hydrogen bonds and salt-bridges and calculated distances between residue side chains. Side chain contacts were defined using a 7 Å cutoff between side chain centers of mass, and salt bridges were defined with 4 Å distance cutoff between heavy atoms and a 60° angle cutoff. In-house scripts were used to identify particular secondary structure features, side chain contacts, and hydrogen bonds that are present simultaneously in the Aβ42 and Aβ40 ensembles.

Results

Previous chemical shift and J-coupling data reported for the monomeric Aβ40 and Aβ42 peptides do not differ greatly from random coil values. However, in our simulated structural ensembles, we find that more than 99% of the Aβ40 and Aβ42 monomer conformers (snapshots during the trajectories) at 287 K and 311 K contain one or more elements of secondary structure somewhere along the peptide sequence. When averaged over the entire ensemble of each peptide, these yield average chemical shifts and J-couplings that do not give a clear indication of the secondary structure. Figures 3.1 and 3.2 display the agreement between experimentally measured chemical shifts and J-couplings (67), and those calculated from the Aβ40 and Aβ42 simulated ensembles. Because there is no large deviation from random coil values for either peptide, chemical shifts are similar for Aβ40 and Aβ42. The J-coupling values are shifted higher than random coil, but show no trends indicating different secondary structure for Aβ40 and Aβ42. The peptide residual dipolar couplings also show reasonable agreement between experiment and simulation values, but no clear differences between the two peptides (Figure 3.3).


Aβ40 Hα chemical shifts
simulated
experimental
RMSD = 0.09
secondary chemical shift (ppm)
residue number
a
Aβ42 Hα chemical shifts
simulated
experimental
RMSD = 0.11
secondary chemical shift (ppm)
residue number
b
Aβ40 HN chemical shifts
simulated
experimental
RMSD = 0.39
secondary chemical shift (ppm)
residue number
c
Aβ42 HN chemical shifts
simulated
experimental
RMSD = 0.45
secondary chemical shift (ppm)
residue number
d

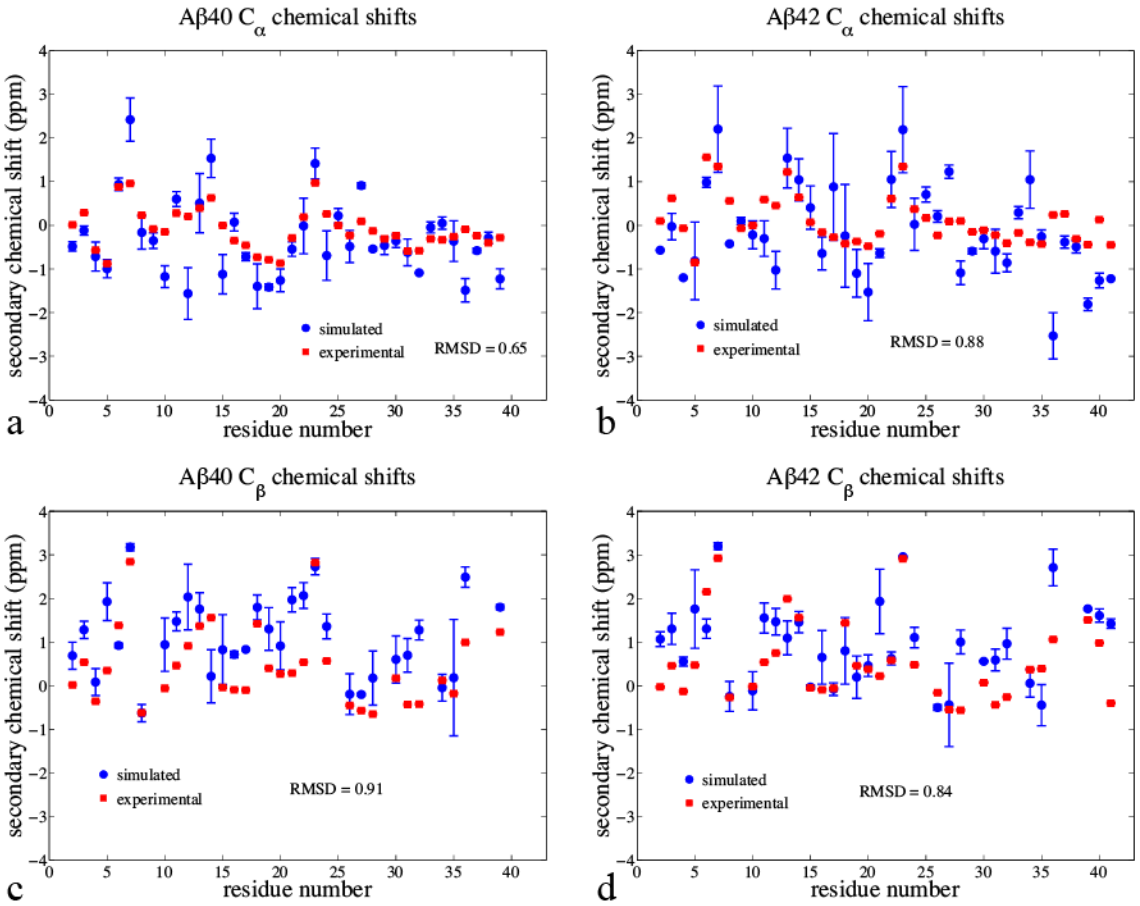


Figure 3.1. *Experimental and calculated proton and carbon secondary chemical shifts by residue.* (a) Aβ40 H_α chemical shifts, (b) Aβ42 H_α chemical shifts, (c) Aβ40 H_N chemical shifts, (d) Aβ42 H_N chemical shifts, (e) Aβ40 C_α chemical shifts, (f) Aβ42 C_α chemical shifts, (g) Aβ40 C_β chemical shifts, and (h) Aβ42 C_β chemical shifts. Red squares represent experimental data, while blue circles represent the data calculated from simulation. Random coil residue specific values are subtracted from both experimental and simulation values. The experimental carbon chemical shift data are taken from Hou *et al*. (14).

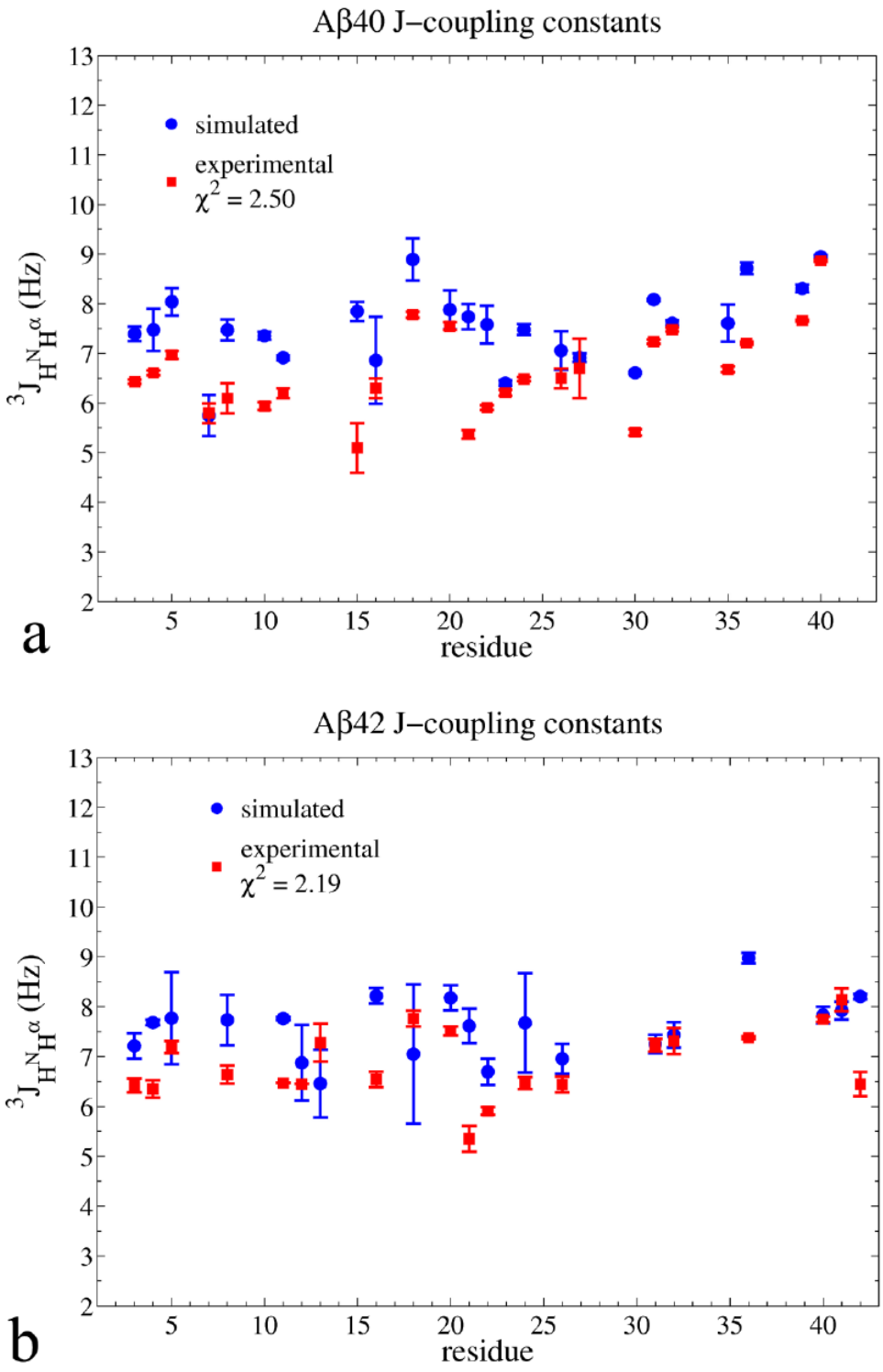


Figure 3.2. *J-coupling constants for backbone amides for (a) Aβ40 and (b) Aβ42.* The red squares are the experimental data from Yan *et al.* (67), blue circles are values calculated from the simulation ensemble. Simulation uncertainty bars represent rms difference between two independent simulations and the average. The experimental data has not accounted for a relaxation correction that makes J-couplings determined from a H_NH_α 3D experiment consistently lower than those from COSY splittings (14) by a small amount (as high as ~10% (93)); if this relaxation effect is accounted for, then the qualitative agreement between experiment and simulation is excellent.

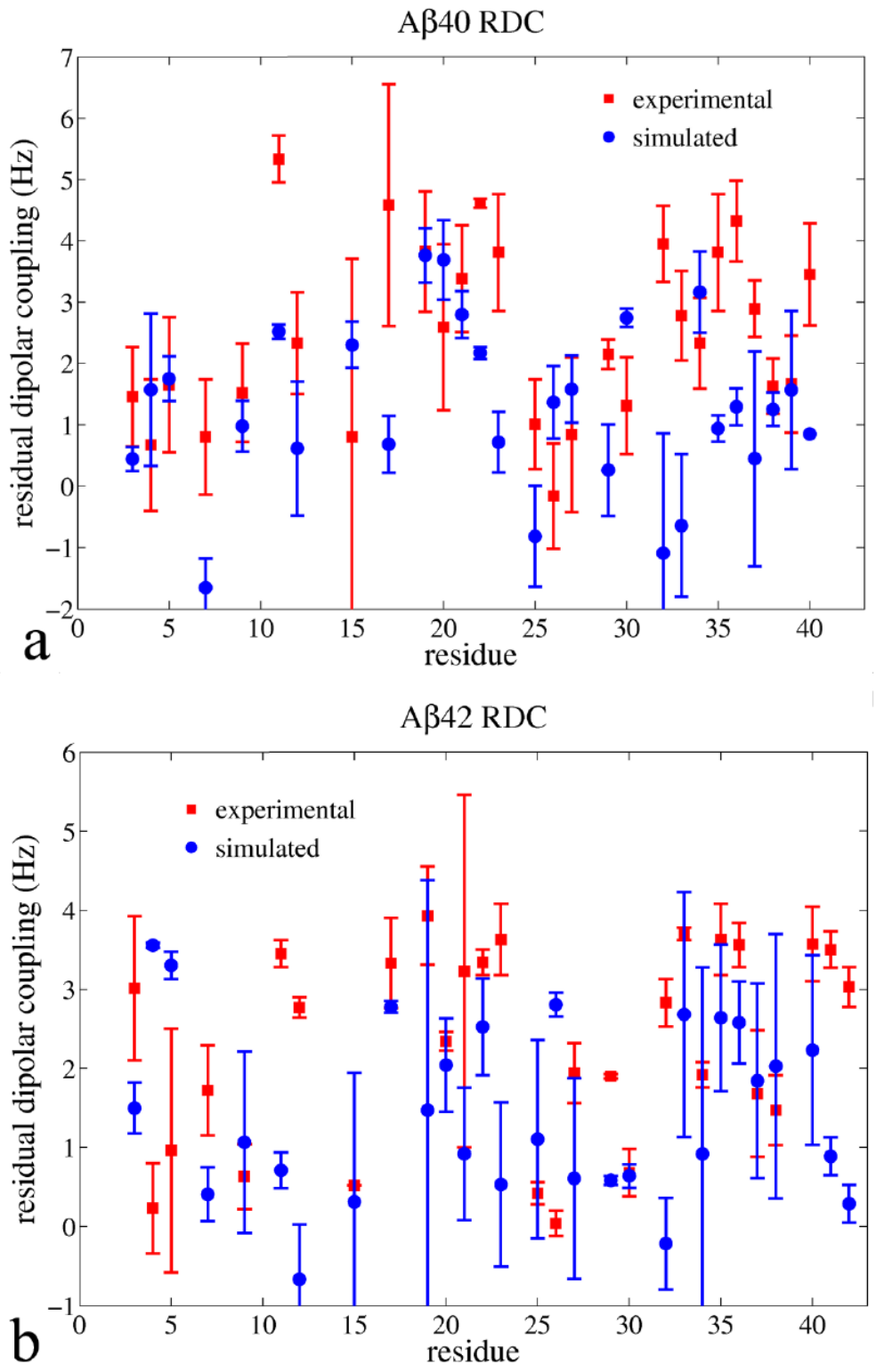


Figure 3.3. *Experimental vs. calculated residual dipolar couplings for Aβ40 (a) and Aβ42 (b).* The red squares are experimental data from Yan *et al*. (67). The blue circles are the data calculated from the simulation ensemble. The Aβ40 RMSD without Lys16 is 2.18 Hz, and the Aβ42 RMSD without Lys16 is 1.89 Hz.

The Aβ40 and Aβ42 simulated conformational ensembles provide more information about the structural differences between the two peptides. The propensities to form β-turn, anti-parallel β-strand, or helical structure are shown by residue for both Aβ40 and Aβ42 in Figure 3.4. We focus on these secondary structure classifications because they are the most common in the peptide ensembles (intra-molecular parallel β-strands are uncommon for such a short peptide). Although Aβ40 adopts some of the same structural features seen in the Aβ42 ensemble, such as a highly populated turn centered at residues 7-8 or a helix near Ser26 in ~15% of structures, the two peptides have substantially different secondary structure profiles overall. The primary difference is that Aβ40 forms turns and helices less often than Aβ42, while adopting a larger amount of various β-strand structures. When averaged over such a diverse structural ensemble, the NMR chemical shifts and J-couplings do not provide any discrimination between the relative secondary structure populations of the two peptides.

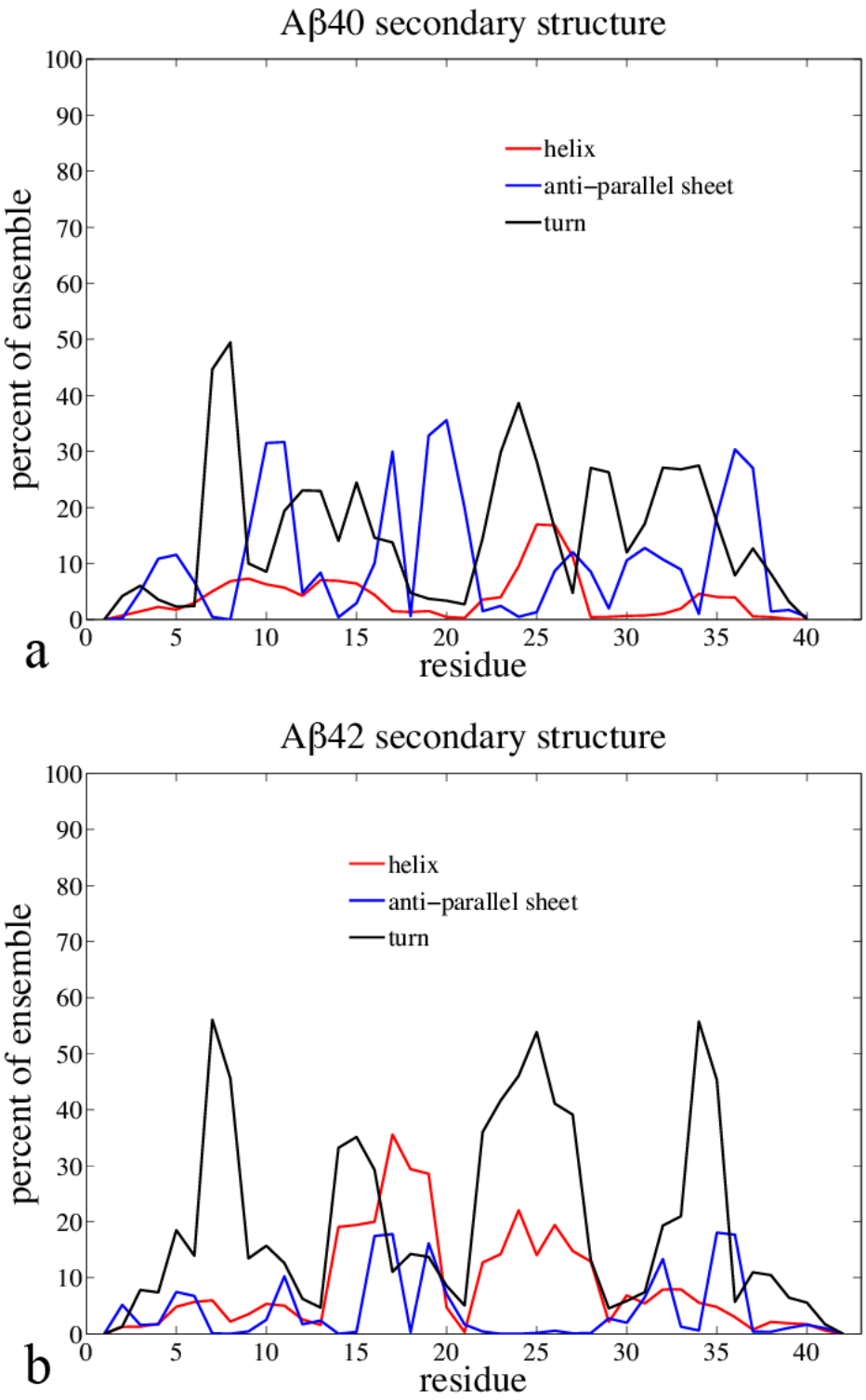


Figure 3.4. *Percentage of (a) Aβ40 simulated ensemble and (b) Aβ42 simulated ensemble in different types of secondary structure by residue.* The red line represents α-helix, the blue line for anti-parallel strand, and the black line for β-turns.

The high resolution NOE data collected at 900 MHz provide evidence for the differences in the β-strand content of the two peptides, with the information for all NOE data presented in Tables 3.1-3.8 in the Appendix. In Figure 3.5 we present a contact map of the strongest long-range NOEs for both peptides. The NOEs experimentally observed often have ambiguities in assignment (many resonances have the same chemical shift), and for these the simulated monomer ensembles are used to make assignments of the ambiguous cross peaks, as described in the Appendix. The contact map shows that the Aβ40 and Aβ42 monomers form some of the same medium-range NOE contacts, but the two peptides differ in the residues that form long-range contacts.

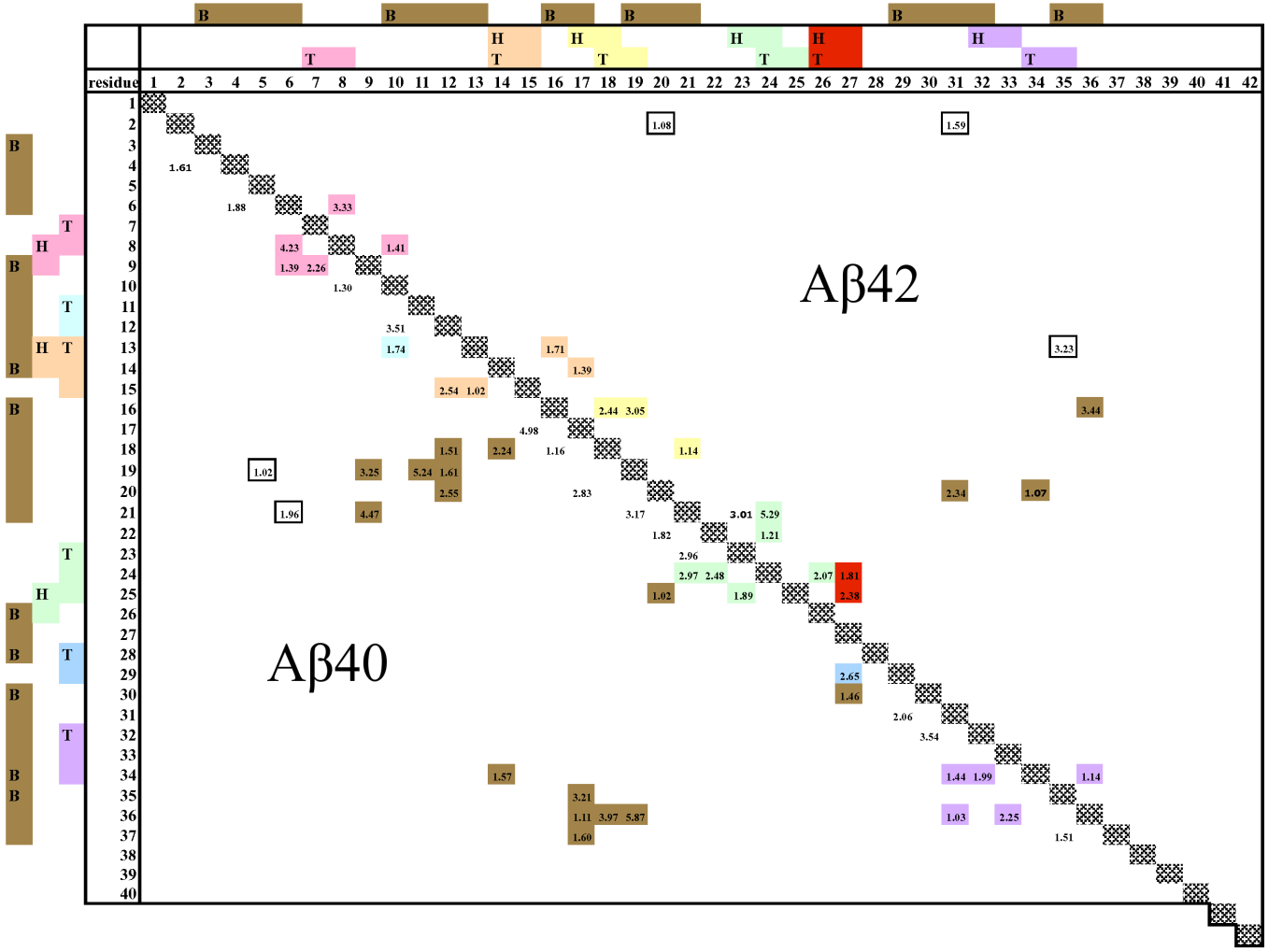


Figure 3.5. *Contact map of the strongest experimental NOE intensities for Aβ40 and Aβ42 that are dominated by a single contact.* We color the NOE contacts based on the structures in the simulated ensemble that contribute to the NOE intensity. NOEs that are due to turns or helices around residues 7-8 (pink); 11-12 (light teal); 13-15 (orange); 17-19 (yellow); 23-26 (green); 26-27 (red); 28-29 (blue); 32-35 (purple); NOEs due to β-strands are shown in brown; long-range contacts that "compact" structure, and which involve multiple secondary structure categories occurring simultaneously, are outlined in black. NOE intensity values calculated for the simulated ensemble are reported in each single contact pair.

In particular, we detected long-range NOEs for the Aβ42 monomer from Lys16 to Val36, Phe20 to Ile31, and Phe20 to Leu34, from the NMR data. The NOE between Phe20 H_{ε} and Leu34 $H_{\delta 1}$ is at a location in the NOESY spectrum where no short-range peaks occur, so we know from experiment alone that it must be due to a long-range contact, i to i+12 or greater. We see even more long-range NOEs resulting from the Aβ40 conformational ensemble (Figure 3.5). Aβ40 NOEs from Phe20 H_{ζ} to Gly25 H_{α}, Glu11 $H_{\beta 2}$ to Phe19 H_{δ}, Val12 $H_{\gamma 1}$ to Phe19 H_{α}, and Val12 $H_{\gamma 1}$ to Phe20 $H_{\beta 2}$ are not consistent with any assignment of residues closer in the sequence. Additionally, two long-range NOEs from Val18 $H_{\gamma 1}$ and $H_{\gamma 2}$ to Val36 H_{γ} cannot be assigned to a contact closer in sequence than i to i+6.

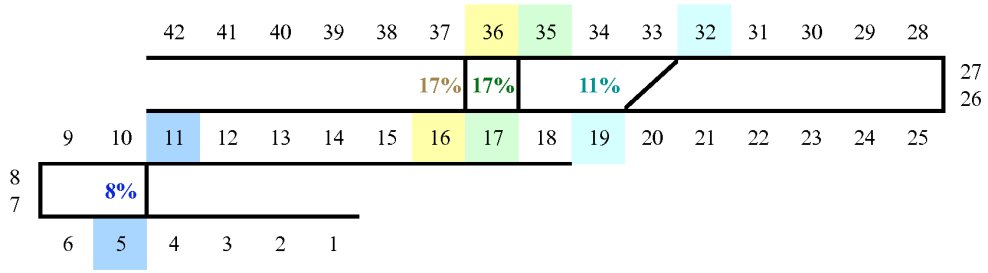
Our computationally derived Aβ40 and Aβ42 conformational ensembles (calculated with no NMR derived restraints) allow for more detailed structural interpretation of the long-range NOEs that we observe experimentally. These structural ensembles show that most of the long-range NOEs produced by each peptide are a result of hydrogen-bonded β-strand structure. However, different β-strands are formed in the Aβ40 ensemble than in the Aβ42 ensemble, and Figure 3.5 illustrates the most common β-strand pairs observed for each peptide, although we emphasize that in the ensemble these contacts do not all exist simultaneously. While the two common Aβ42 β-bridge partners can be represented in one diagram, Aβ40 forms several β-strands that bring very different parts of the Aβ sequence together, and we must use three separate diagrams to illustrate all of these contacts. We describe these β-strand structures more carefully below.

The simulated Aβ42 ensemble contains one β-strand between residues 3-6 and 10-13 in 8% of the ensemble, and another large β-strand between residues 16-21 and 29-36 (48). This large β-strand is the underlying structural feature responsible for the Aβ42 long-range experimental NOEs involving Lys16-Val36, Phe20-Ile31, and Phe20-Leu34. All or part of this 16-21 to 29-36 β-strand is present in 30% of the Aβ42 ensemble (see the Appendix), and is often stabilized by a hydrogen-bonded turn that is centered on residues 26-27, with no other major secondary structure present elsewhere in the chain (48). The 26-27 β-turn also stands out from other secondary structure in the Aβ42 ensemble because it forms without other turns or helices in 13% of the simulated ensemble, while each of the other five major turns exhibited by Aβ42 forms alone less than 3% of the time The unique 26-27 β-turn and intramolecular Aβ42 β-strand at residues 16-21 and 29-36 are compatible with the same 26-27 β-turn and the 16-21 and 29-36 β-strands that ultimately adopt the intermolecular arrangement of the stable mature fibril state. This is consistent with a number of MD simulations that highlight the importance of residues 23-28 for nucleating monomer folding (120), and detailed structural characterization of the amyloid-β fragments Aβ21-30 (26, 48, 121, 122) and Aβ10-35 (104, 105), as well as the importance of residues 16-22 that promote β-sheet structure as discussed in (123, 124).

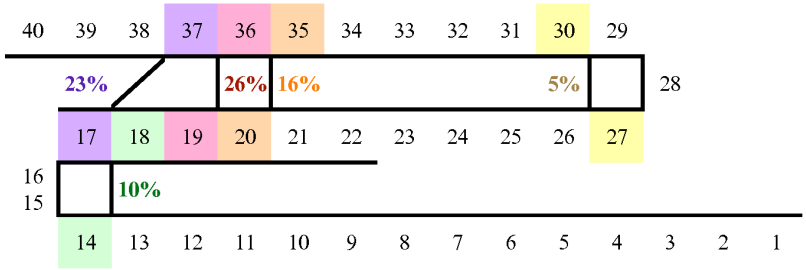
By contrast Aβ40 forms nine highly populated turns along its sequence, but no β-turn forms alone in more than 3% of the simulated ensemble. In fact, the 26-27 β-turn forms in only 4% of Aβ40 structures, compared to 32% of Aβ42 structures (see the Appendix). Instead, Aβ40 often forms a β-strand in that region of the peptide sequence: in 16% of the Aβ40 ensemble, at least one of the two residues, Ser26 or Asn27, is involved in a β-strand, while this happens in only 1% of the Aβ42 ensemble. In fact, Aβ40 contains regions with high β-strand occupancy all

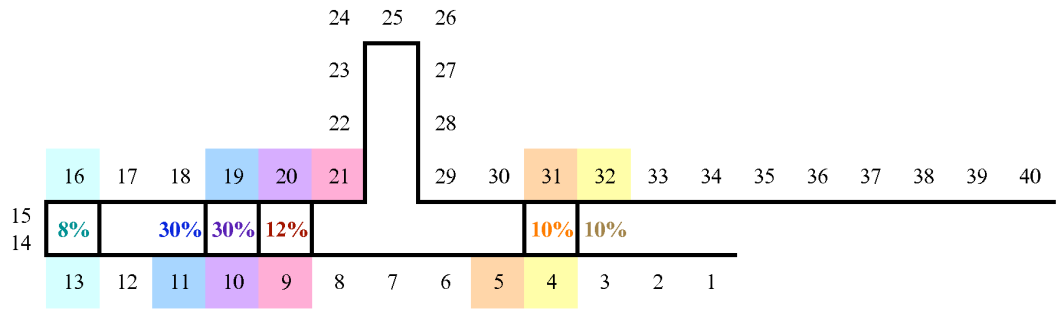
along its sequence, with different β-strand regions tending to mix-and-match alternative pairings to form a promiscuous set of β-strand sub-populations (Figure 3.6), unlike Aβ42, which is dominated by a single major β-strand.

Abeta42



Abeta40





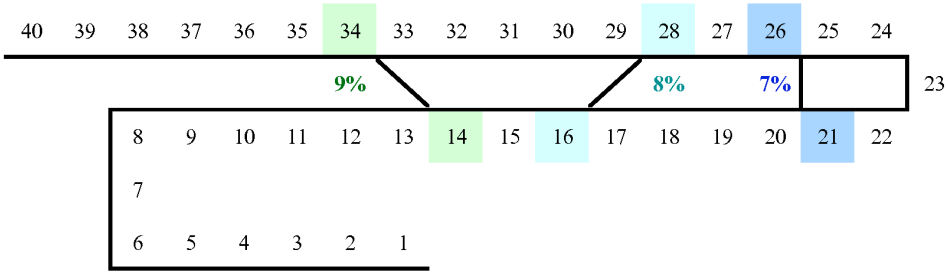


Figure 3.6. *Diagram of Aβ42 and Aβ40 β-strand partners in simulated ensembles.* This is a schematic showing which residues come together in a β-strand structure for each peptide. Residues of the same color are β-bridge partners, and the percent of the ensemble in which these two residues form a β-bridge is written between the two residues. The β-bridge contacts represented in each diagram may or may not form simultaneously in the peptide conformational ensemble.

The most populated β-strands in the Aβ40 simulated ensemble also provide structural explanation for the long-range Aβ40 NOEs observed experimentally and described above (Figure 3.5). A β-strand between residues 9-13 and 16-21 (39% of the ensemble) contributes to the Glu11-Phe19, Val12-Phe19, and Val12-Phe20 NOEs, while a β-strand between residues 17-20 and 35-37 (29% of the ensemble) is the underlying structure that causes the NOEs between Val18 and Val36. The β-strand pairing of 20-21 with 26-27 (7% of structures) contributes to the NOE from Phe20 to Gly25. Residues 3-6 also form a β-strand with 30-33 (10% of the ensemble), but this does not result in long-range NOEs that are as strong, in part because this strand is less populated.

Rather than the fibril-like 26-27 β-turn present in the Aβ42 simulated ensemble, Aβ40 exhibits a highly populated β-turn centered at residues 28 and 29 instead (26% of the ensemble), which is often flanked by the β-strand between residues 17-20 and 35-37, 22% of structures. However, the Aβ40 28-29 β-turn occurs with other turns or helices, even when the 17-20 and 35-37 β-strand is present, unlike the Aβ42 26-27 turn which often occurs alone. Finally, previous work by Yan and co-workers (125) examined side chain methyl groups, showing that Val18 is more ordered in Aβ40 compared to Aβ42. Our simulations provide an explanation for this experimental observation since we find Val18 participates in more NOE contacts and hydrogen bonds within the Aβ40 ensemble than in the Aβ42 ensemble (Figure 3.4 and Table 3.9). Representative Aβ40 and Aβ42 structures containing the primary β-strands of the two peptides at 287 K are presented in Figure 3.6.

We have described the major β-strand differences in the Aβ40 and Aβ42 ensembles that we observe using MD simulations, but a remaining question concerns the sequence difference that must be the underlying cause. The two C-terminal residues of Aβ42 (Ile42 and Ala42) have hydrophobic side chains, and most often interact hydrophobically with other residues, rather than through hydrogen bonding. Figure 3.7 shows the other residues that contact the Ile41 and Ala42 side chains most frequently in the Aβ42 structural ensemble. Many of these residues contacting Ile41 and Ala42 in the 287 K Aβ42 structural ensemble form alternate contacts comprising β-strands in the Aβ40 ensemble. This helps to explain why the β-strand structures that dominate the Aβ40 ensemble are not present in the 287 K Aβ42 ensemble, because in the Aβ42 ensemble they are substituted with C-terminal hydrophobic contacts. The most populated β-strands in the Aβ40 ensemble, between residues 9-13 and 16-21, residues 17-20 and 35-37, residues 3-6 and 30-33, and residues 20-21 and 26-27, together compose 52% of the Aβ40 ensemble. Ile41 and/or Ala42 contact residues Arg5, Tyr10, Lys16, Phe19, Lys28, Ile32, Gly 33, Leu34, Met35, and Val36 side chains in 45% of Aβ42 structures, thereby completely replacing the Aβ40 β-strand populations.

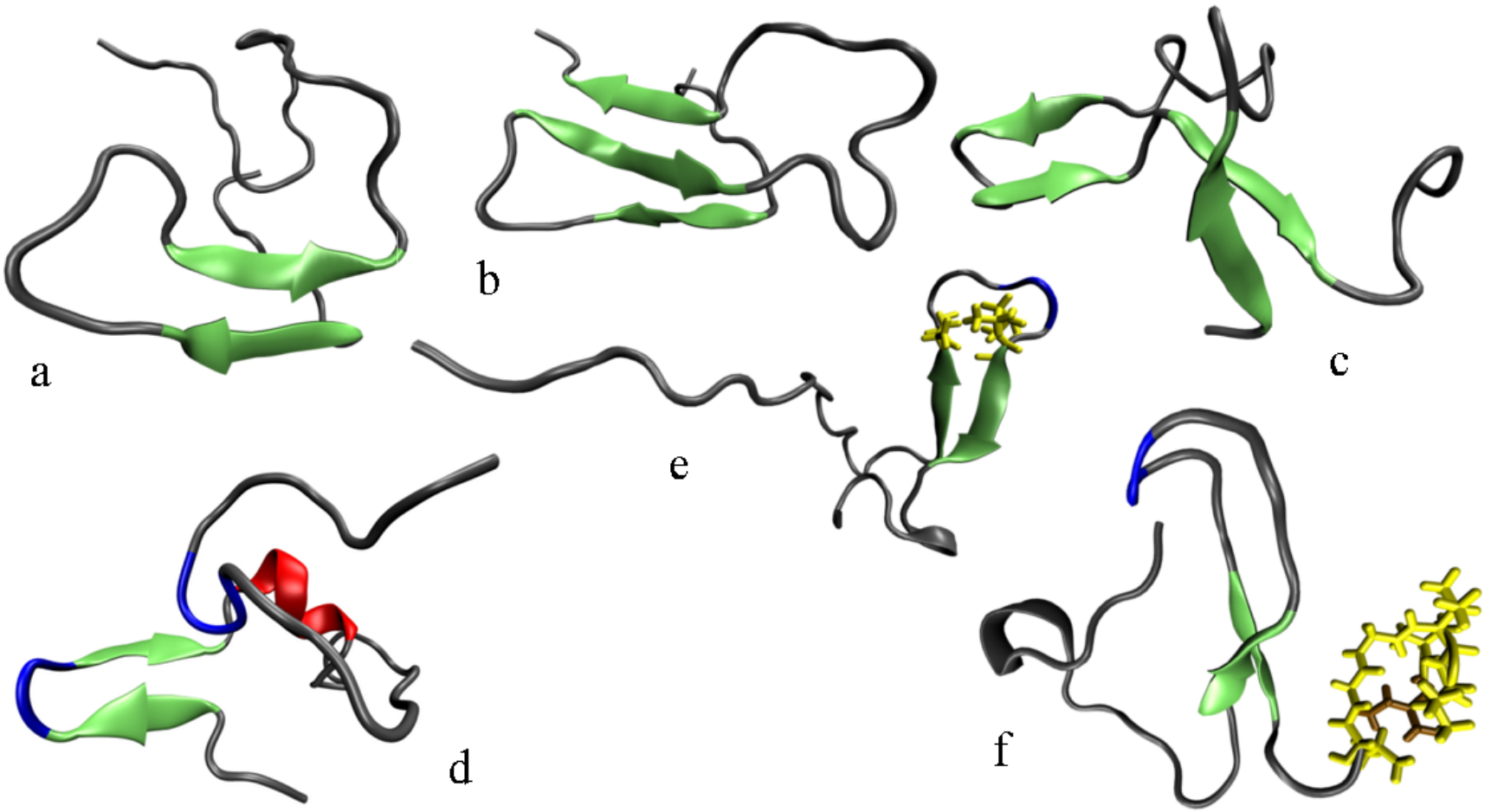


Figure 3.7. *Representative Aβ40 and Aβ42 conformations containing β-strands, selected from the simulated ensembles.* Turns (blue), Helix (red), β-strands (green). (a) Aβ40 forms a the 9-13 and 16-21 β-strand. (b) Aβ40 forms the 17-20 and 35-37 β-strand simultaneously with the strand at 9-13 and 16-21. (c) Aβ40 forms a β-strand at residues 3-6 and 30-33 along with a strand at 20-21 and 26-27. (d) Aβ42 forms a turn at residues 7-8 nucleating a β-strand at residues 3-6 and 10-13, along with a helix from residue 14 to 18 and a turn at 34-35. (e) Aβ42 forms the turn at residues 26-27 with the 16-21 and 29-36 β-strand, stabilized by a salt bridge between Glu22 and Lys28 (side chains in yellow), in the absence of other secondary structure. (f) Aβ42 forms the same turn at 26-27 and 16-21 and 29-36 β-strand with a C-terminal hydrophobic side chain cluster from residues 36-42 (yellow) centered around the side chain of Ile41 (brown). Figures 3.7d and 3.7e are reprinted with permission from (48), copyright 2011 American Chemical Society.

Experimental relaxation rates and ^{1}H-^{15}N NOE intensities reveal that the Aβ42 C-terminus is less flexible than that of Aβ40. Figure 3.9 shows that our calculated ^{1}H-^{15}N NOE values agree well with the experimental data from Yan and Wang (66). Unlike the ^{1}H-^{1}H NOEs, these data are unambiguous and no information about the simulated ensembles are used to assign them. In Figure 3.10 we show the comparison between the Aβ40 and Aβ42 calculated ^{1}H-^{15}N NOE intensities. As in the experiment we see higher values at the Aβ42 C-terminus, associated with more ordered structures of the Aβ42 C-terminal backbone. As represented in the Figure 3.8a, this increased order seems to be a result of hydrophobic interactions involving the Ile41 side chain. Furthermore, the high density of hydrophobic contacts at the C-terminus of the Aβ42 peptide displaces the β-turn at residues 28 and 29 seen in the Aβ40 ensemble with a turn that is closer to the N-terminus, stabilizing the β-turn at residues 26 and 27 that often occurs with the formation of the major 16-21 and 29-36 β-strand. The C-terminal hydrophobic clustering occurs with the 16-21 and 29-36 β-strand motif in 7% of the ensemble and with the 26-27 β-turn in 7% of the ensemble, while 6% of the ensemble contains the hydrophobic cluster with both the strand and turn, as shown in Figure 3.7f.

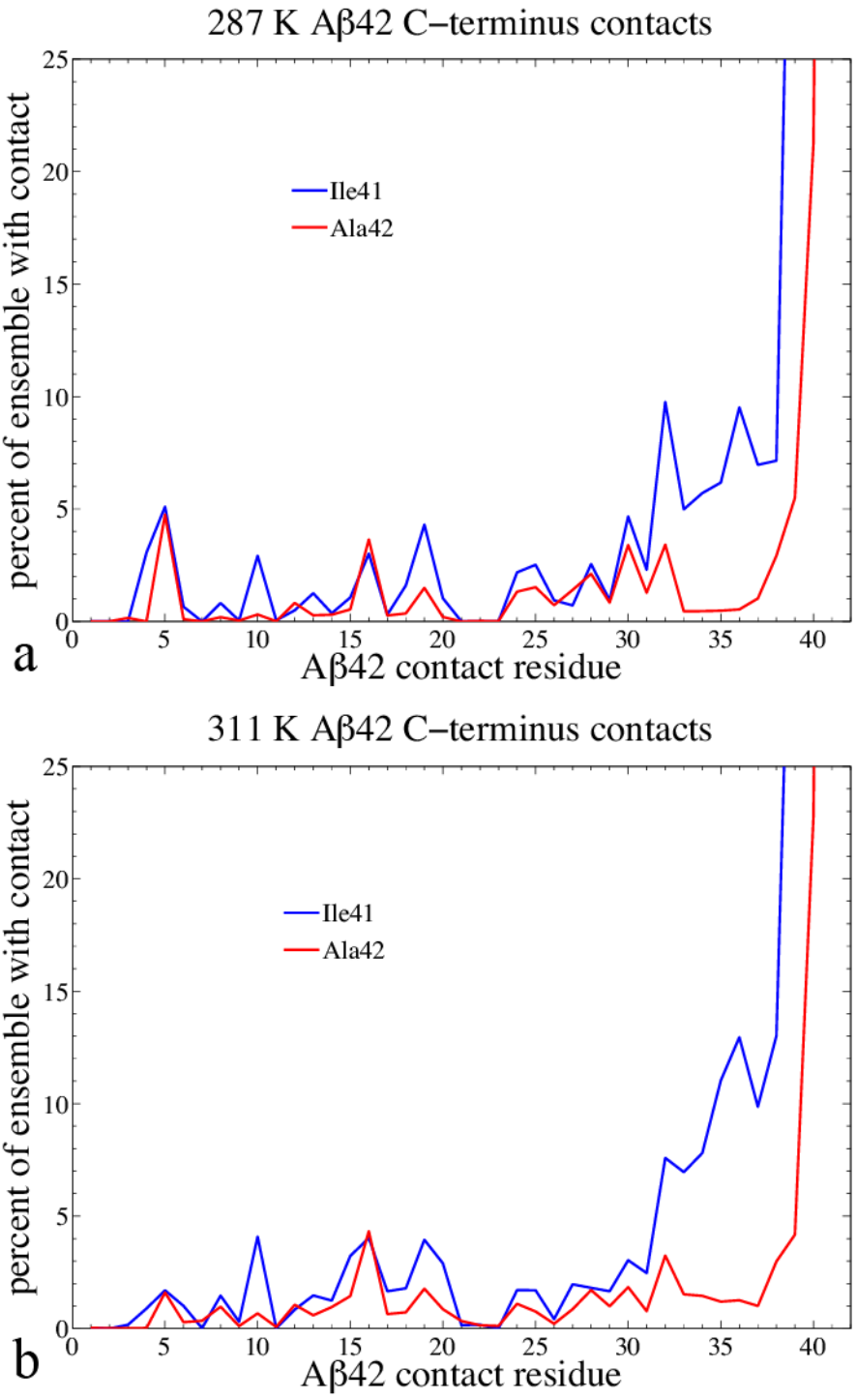


Figure 3.8. *Aβ42 Ile41 and Ala42 side chain contacts at 287 K (a) and 311 K (b).* This plot presents only contacts involving Ile41 in blue and only contacts involving Ala42 in red. The percent of the Aβ42 ensemble with the contact (Ile41 or Ala42) is shown on the X-axis, while the residue in contact with Ile41 or Ala42 is shown on the Y-axis. Contacts involving neither Ile41 nor Ala42 are not presented. Contacts are defined as a distance between side chain centers of mass of the two residues of less than 7 Å.

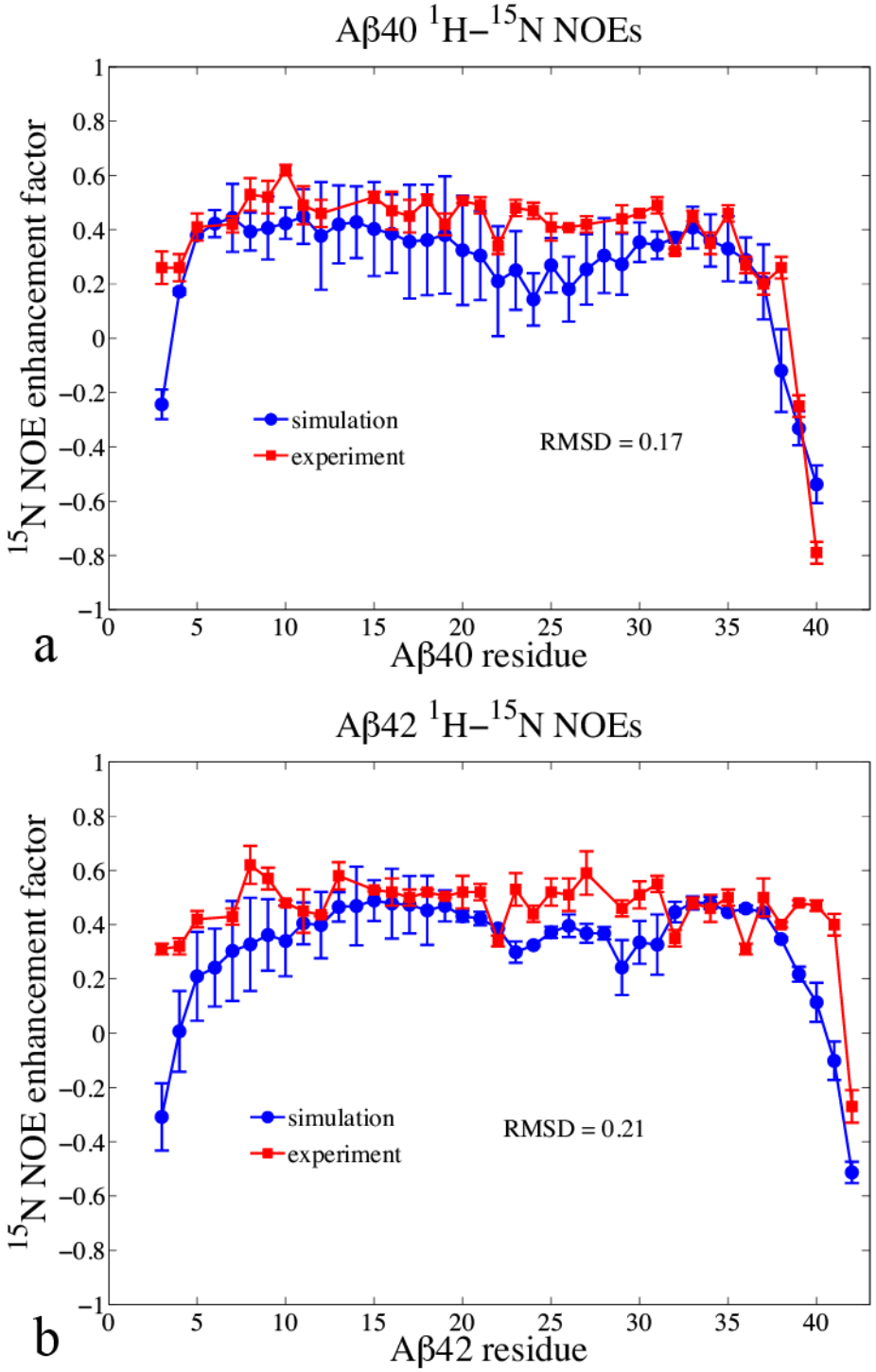


Figure 3.9. *Agreement with experiment of simulated Aβ40 (a) and Aβ42 (b) ^{1}H-^{15}N NOE.* The red squares are experimental data from Yan and Wang (66). The blue circles are the data calculated from the simulation ensemble.

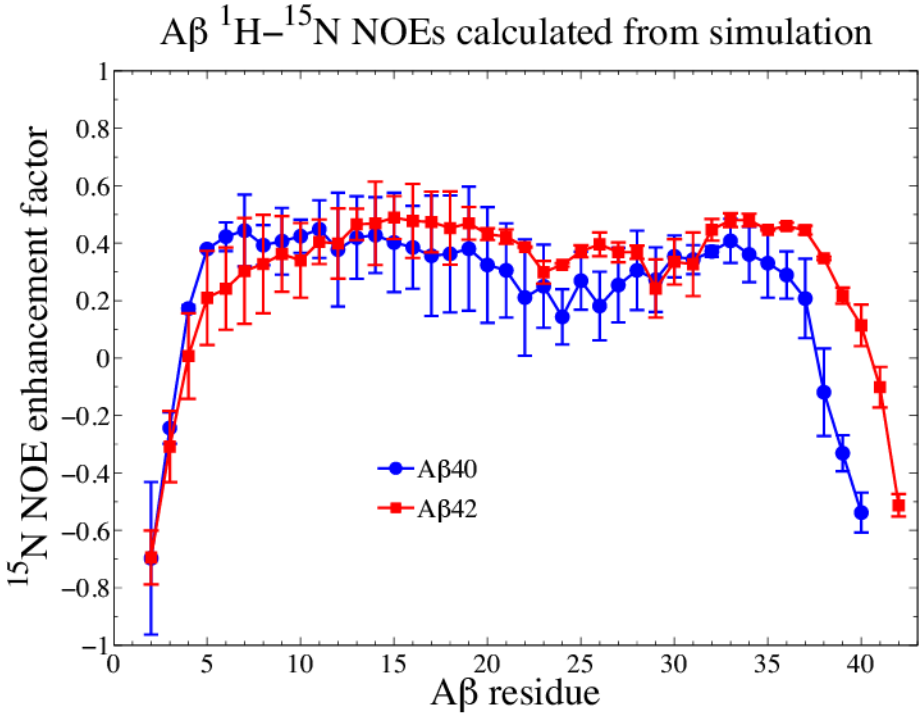


Figure 3.10. *Comparison of Aβ40 and Aβ42 steady-state ^{1}H-^{15}N resolved NOE enhancements calculated from MD simulations by residue.* The blue circles are the Aβ40 data and the red squares are the Aβ42 data.

We also converged and analyzed simulated Aβ40 and Aβ42 ensembles at 311 K, near physiological temperature. At the higher temperature, Aβ40's most populated β-strands at 287 K (Figure 3.7a and 3.7b) decrease, while the two β-strands present in the Aβ42 ensemble remain and even increase in population (Figure 3.7d and 3.7e). We attribute this stabilization of the Aβ42 β-strands with increasing temperature to an increase in hydrophobic clustering of the C-terminal residues as seen in Figure 3.8b. This temperature dependent MD ensemble data also helps to explain the results of circular dichroism experiments showing that the Aβ42 β-strand content is more stable than that of Aβ40 as the temperature of the sample is increased (119).

Discussion & Conclusions

Although Aβ40 and Aβ42 are intrinsically disordered peptides, we have shown through molecular dynamics simulations that their monomer ensembles are significantly structured, presenting a diverse set of secondary structures including α-helix, β-turns, and β-strands. These simulation results provide a level of detail not available from the CD and NMR experimental data, by allowing us to see the particular secondary structure features adopted by one peptide rather than an average over all of the peptides in an experimental sample. Our simulated ensembles also do a good job of reproducing the NMR data, especially for chemical shifts and NOEs, showing that the current MD force fields are able to accurately represent the energy landscape for an IDP. The quantitative comparison to experiments gives us enough confidence to interpret the differences we see between the Aβ40 and Aβ42 ensembles with respect to fibrilization and oligomerization.

Looking at the simulated ensembles of Aβ40 and Aβ42, we see that there are many differences between the structures occupied by the two peptides, even though their sequences are almost identical. The two additional residues on the C-terminus shift of Aβ42 prevent it from occupying the same combination of promiscuous β-strands that Aβ40 forms in its ensemble. Instead Aβ42 contains more hydrophobic interactions at the C-terminus, and a β-strand around a turn at residues 26-27, consistent with the Aβ fibril structure. Although we also see Aβ40 form a β-strand and turn at residues 28-29, this strand is not consistent with the aggregated fibril conformation of the peptide. Since Aβ42 has a cluster of hydrophobic interactions around its C-terminus, it cannot form the more C-terminal strand seen in Aβ40, and instead it forms a slightly shifted strand centered at the 26-27 turn. From the MD analysis we see a picture forming of how the Ile41 and Ala42 can cause such drastic differences in Aβ42 behavior when compared to Aβ40. Rather than residues 41 and 42 directly forming fibril-like β structure, their hydrophobic side chain contacts change the way that β-strands are formed at other locations of the peptide sequence, preventing β-strands that are not compatible with Aβ aggregated structure and promoting conformations that could more easily form fibril seeds. Such a large effect from a two residue addition is particular to the conformational ensembles of IDPs. The energy minima stabilizing the Aβ40 β-strand structures are shallow, and therefore they are easily disrupted and rearranged when Ile41 and Ala42 are added to create the Aβ42 ensemble. For a folded protein, adding two residues would be unlikely to eliminate the large energy minimum of the native state, as demonstrated by the common use of fusion tags for protein purification. Therefore, the intrinsically disordered nature of the Aβ peptides explains how they can have drastically different aggregation and disease behavior though they are very similar in sequence.

Because the Aβ40 and Aβ42 are IDPs, they also have very diverse structural ensembles. This makes complete MD sampling of the conformations they occupy difficult because the transitions between each of the local energy minima may be slower than the timescale accessed by MD simulations. We have addressed this problem by using the multi-reservoir replica exchange sampling technique (31, 81, 82), and looking for agreement between two independent simulations, however our simulated ensembles may still be incomplete with respect to their sampling of the full Aβ40 and Aβ42 energy landscapes. The level of disagreement we see between our NMR data and the simulated ensembles most likely a result of incomplete sampling, which results in improperly weighted MD ensembles. This raises the question of whether MD is the best computational technique for constructing an IDP ensemble to achieve proper Boltzmann weighting of the structures and the best agreement with experimental data. Because we observe such diverse structural features within the Aβ40 and Aβ42 ensembles, these disease peptides are ideal models on which to develop general methods for constructing accurate IDP ensembles.

APPENDIX

NMR experimental method

NMR data were collected at 287 K at the NMR Facility at UC Berkeley on Bruker Avance II 800 and 900 MHz spectrometers, the latter equipped with a Bruker cryogenic probe. The data were processed using NMRPipe and peaks were assigned and analyzed using CARA (80). Chemical shifts were assigned using 2D ^{1}H-^{1}H Total Correlation Spectroscopy (TOCSY) (81, 82) and

Nuclear Overhauser Effect Spectroscopy (NOESY) spectra (83-85). The TOCSY spectra were collected in a 90% H_2O (10% ^2H_2O) buffer solution with a mixing time of 60 ms and in a 99% ^2H_2O buffer solution with a mixing time of 80 ms. The NOESY spectra were collected in a 90% H_2O (10% ^2H_2O) buffer solution with mixing times of 100 ms and 200ms, and in a 99% ^2H_2O buffer solution with a mixing time of 200 ms. In the direct dimension, 2048 points were collected while 445-512 complex points were collected in the indirect dimension. The spectral width was 10 ppm in each dimension (9 ppm in the direct and 2 ppm in the indirect dimension for ^2H_2O) with 64 scans. We ensured that our NOESY data was resulting only from a monomeric form of the peptide by collecting a NOESY spectrum immediately after dissolving the peptide and another NOESY spectrum ~48 hours after dissolving the sample. We saw no difference between the two spectra except for a decrease in peak intensity due to aggregation, which eliminates the possibility that some NOEs could be a result of oligomers forming during peptide incubation. Peaks were identified at particular chemical shifts in the 200 ms mixing time NOESY spectra in both H_2O and ^2H_2O for Aβ40 and Aβ42, and those peaks that could be unambiguously assigned to two specific hydrogen atoms were used to calibrate the NOE peak intensities derived from simulation. The majority of peaks could not be clearly assigned to a unique pair of hydrogen atoms due to spectral overlap. These peaks were instead assigned a list of potential hydrogen pairs consistent with the observed peak frequencies. This list is made up of atoms within a 0.04 - 0.08 ppm range around each observed peak.

De novo molecular dynamics simulation method

The *de novo* molecular dynamics simulations of Aβ42 have been reported in a previous publication (48). The Aβ40 peptide was also modeled using the Amber ff99SB force field (23) and aqueous solvent represented by the TIP4P-Ew water model (25). We chose the ff99SB force field and TIP4P-Ew water model combination because it was used successfully in previous Aβ21-30 and Aβ42 studies to reproduce ^1H-^1H NOEs (26, 48).

In order to calculate equilibrium ensemble averages in the NVT ensemble, we used an Andersen thermostat, a leapfrog integrator with a 1.0 fs time step, and periodic boundary conditions. Particle Mesh Ewald was used for calculating long-range electrostatic forces, and a cutoff of 9.0 Å was used for the real space electrostatics and Lennard-Jones forces throughout the study. The initial structure for Aβ40 was built in an extended form using the *LEaP* module that is part of the AMBER package. The structure was solvated in a water box such that there were 10 Å of water surrounding the peptide on all sides (13,992 molecules of water, 540 nm^3 box volume), and three Na^+ ions to neutralize the charge of the peptide. The structure was minimized and equilibrated with constant volume while raising the temperature to 300 K, then equilibrated for 2 ns under a constant pressure of 1 bar, maintained with a Berendsen barostat, to achieve the correct density. Then, a 2 ns 498 K simulation was run on the extended Aβ40 peptide to obtain a more collapsed starting structure for equilibrium ensemble simulations. Two different collapsed structures were chosen to start two independent replica exchange simulations. These systems were prepared by first removing the solvent, and then the previous solvation and equilibration steps were used to create a final box that contained 6,136 water molecules (190 nm^3 volume).

The *sander* module of AMBER (111) and the Multi-Reservoir Replica Exchange (MRRE) method presented by Ruscio *et al*. (31, 110) were used to achieve a Boltzmann weighted ensemble of Aβ40 structures at the experimental temperature of 287 K and at the physiological temperature of 311 K. Initially, replica exchange was used to converge two independent Boltzmann reservoirs at 397 K using 13 replicas spaced from 447 K to 397 K. Exchanges were attempted every 1.0 ps. After the 397 K replica was sufficiently converged, the last 30 ns of simulation time was used to create a 397 K reservoir for the next set of replica exchange for each of the two independent trajectories, taking 30,000 conformational snapshots spaced one picosecond apart. In the next stage of MRRE, the 397 K reservoir was used to exchange with the highest temperature replica of 18 replicas spaced between the 397 K reservoir and a low temperature replica of 335 K. After the 335 K replica converged, the final 30 ns of simulation were used to create a 335 K reservoir for the next set of replica exchange. In the final stage of MRRE, the 335 K reservoir exchanged with the highest temperature of 18 replicas that included 311 K and 287 K. Approximately 50,000 structures from the final 50 ns of each of the independent converged 287 K and 311 K replica were used to construct the equilibrium ensembles. Aβ40 needed 40 ns to converge the 397 K reservoir, 70 ns to converge the 335 K reservoir, and 60 ns to converge the 287 K ensemble. The total simulation time combining every replica was 5.72 μs for Aβ40, resulting in 100 ns of equilibrated NVT simulation at 287 K and 311 K. A minority of these structures was extended outside of the simulation box, and the peptides were interacting with their periodic image. We removed these unphysical structures to yield 72,632 Aβ40 structures and 89,469 Aβ42 structures in their respective 287 K ensembles and 79,253 Aβ40 structures and 89,852 Aβ42 structures at 311 K.

Convergence of each intermediate reservoir, as well as the low temperature ensemble, was determined by comparing average intra-atomic distances for peptide hydrogen atoms between the two independent simulations using the *ptraj* module of the AMBER software package (111). Various statistics measuring agreement of these two ensembles were analyzed over time, and generally, we focused on the short intra-atomic distances (<7.0 Å) that would have the greatest effect on experimental NOE measurements. When greater than 90% of the short distances were converged within 10% of the total distance, the sampling was deemed sufficient.

In order to calculate time-correlation data on the 287 K reservoir, selected structures from this ensemble were used to run microcanonical (NVE) ensemble simulations using the pmemd module of AMBER10 (110). 50 structures were chosen from each 287 K Aβ40 ensemble (after the 10 ns of equilibration), spaced 1 ns apart (ensuring complete decorrelation through swaps between replicas), and equilibrated at constant 287 K for 100 ps before running 20 ns of simulation for each initial structure. In total, the Aβ40 constant energy simulation time with starting conformations equilibrated at 287 K was 2 μs.

Calculation of NMR observables methods

For both Aβ40 and Aβ42 we calculated NMR chemical shifts, scalar-coupling constants, residual dipolar couplings (RDCs), ^{1}H-^{1}H NOE peak intensities, and ^{1}H-^{15}N NOEs from our simulated ensemble data. To calculate hydrogen and carbon chemical shifts we used the program SHIFTS (91) as in Ball *et al.* (48). The scalar-coupling constants and the χ^2 figure of merit were also

calculated as described previously (48), using the original Vuister and Bax parameter set (93) for the Karplus equation (92). RDCs were calculated using PALES (96) and multiplied by a constant factor (48) for comparison to experimental data from *Yan et al.* (67).

Homonuclear NOE Intensities. Due to the complexity of the noise distribution in the Aβ40 and Aβ42 spectra, uncertainties in Aβ40 and Aβ42 peak intensities due to noise were estimated from simulation (again expressed relative to the smallest identifiable peak). We used a factor of 0.32 for the Aβ40 H_2O spectrum, 0.46 for the Aβ40 ^2H_2O spectrum, 0.44 for the Aβ42 H_2O spectrum, and 0.50 for the Aβ42 ^2H_2O spectrum, and this noise was assumed to have a normal distribution (although there are other non-random noise features in the NMR data). Using the simulated NOE intensities for the Aβ40 and Aβ42 H_2O spectra, we calculated the expected number of false-positive and false-negative NOEs for each peptide based on the noise level as in (48). For Aβ40, we would expect 36 false negatives and 26 false positives, and for Aβ42, we would expect 57 false negatives and 23 false positives.

Correction to Aβ42 secondary structure ensemble percentages

The 287 K Aβ42 ensemble previously reported (48) was missing 10% of the 100,000 simulated structures that were supposed to be part of the ensemble, and instead had duplicates of other structures in the Aβ42 ensemble. In this chapter we have corrected this mistake, presenting all analysis on the correct Aβ42 ensemble. We have also become aware that ~10% of the structures in the Aβ42 287 K ensemble were extended outside of the periodic box, and we removed these structures from the Aβ42 analysis in this chapter (see *De novo* molecular dynamics simulation method section). This correction to the Aβ42 ensemble resulted in mostly minor changes to the percentage of the Aβ42 ensemble containing certain structural features. We note that a turn at residues 26-27 was previously reported to appear in 28% of the ensemble, 8% of the time with no other turns or helices (48), but actually appears in 32% of the corrected ensemble, 13% of the time alone. We also previously reported that a β-strand between residues 16-21 and 29-36 forms in 21% of the ensemble (48), while it appears in 30% of the corrected ensemble. A β-strand at residues 3-6 and 10-13 was also previously reported in 9% of the ensemble (48), but actually appears in 8% of the corrected ensemble.

Peak Types	All NOE peaks	Long range peaks (i to i+2 or greater)
Experimentally observed peaks	**1088**	**267**
Experimentally assigned peaks	382	20
Experimentally unassigned peaks	706	247
Simulated peaks	1306	387
Simulated peaks agree experimentally assigned peaks	339	9
Simulated peaks agree experimentally unassigned peaks	562	146
False negatives	187	112
False negatives found in high noise regions	-86	-59
False negatives explained by atomic contacts on same residue pairs	-82	-34
False negatives inconsistent with experiment	**19**	**19**
False positives	405	235
False positives found in high noise regions	-78	-40
False positives explained by atomic contacts on same residue pairs	-283	-162
False positives made up of numerous weak contacts	-23	-19
False positives inconsistent with experiment	**21**	**14**

Table 3.1. Aβ40 summary of experimental and calculated NOEs statistics, and analysis of the number of false positives and false negatives.

Peak Types	All NOE Peaks	Long range peaks (i to i+2 or greater)
Experimentally observed peaks	**705**	**175**
Experimentally assigned peaks	196	11
Experimentally unassigned peaks	509	164
Simulated peaks	945	206
Simulated peaks agree experimentally assigned peaks	177	0
Simulated peaks agree experimentally unassigned peaks	360	55
False negatives	168	120
False negatives found in high noise regions	-59	-45
False negatives explained by atomic contacts on same residue pairs	-65	-31
False negatives inconsistent with experiment	**44**	**44**
False positives	408	151
False positives found in high noise regions	-86	-21
False positives explained by atomic contacts on same residue pairs	-241	-91
False positives made up of numerous weak contacts	-19	-10
False positives inconsistent with experiment	**62**	**29**

Table 3.2. Aβ42 summary of experimental and calculated NOEs statistics, and analysis of the number of false positives and false negatives.

Atom 1	Atom 2	Calculated H_2O NOE intensity	Experimental H_2O NOE intensity	Calculated ^2H_2O NOE intensity	Experimental ^2H_2O NOE intensity
He TYR 10	Hg1 VAL 12	3.51	5.64	3.17	5.29
He TYR 10	Hg2 VAL 12	3.51	2.68	3.17	2.86
He PHE 19	Hb ALA 21	3.17	8.26	2.91	4.49
Hd TYR 10	Hg1 VAL 12	2.97	5.86	2.67	5.34
Hd TYR 10	Hg2 VAL 12	2.97	2.99	2.67	2.45
Ha ASN 27	H GLY 29	2.65	1.97	0.00	0.00
Ha ASP 23	H GLY 25	1.89	2.08	0.00	0.00
Hb ASP 7	H GLY 9	1.82	1.80	0.00	0.00
Hb ALA 2	H PHE 4	1.61	3.01	0.00	0.00
Ha SER 8	H TYR 10	0.78	2.25	0.00	0.00
H GLU 22	Hg VAL 24	0.74	2.68	0.00	0.00
Hb3 SER 8	H TYR 10	0.67	10.32	0.00	0.00
Hg2 VAL 18	He PHE 20	0.62	5.84	0.56	4.51
Ha2 GLY 37	H VAL 39	0.52	11.59	0.00	0.00
Hg2 VAL 18	Hd PHE 20	0.47	4.09	0.42	4.15
Hd LEU 17	He PHE 19	0.45	8.16	0.40	5.23
Hd LEU 17	Hz PHE 19	0.29	8.25	0.26	6.35
Ha LEU 17	Hz PHE 19	0.27	2.55	0.26	
Hg2 VAL 18	Hb3 PHE 20	0.13	3.51	0.10	1.59
Hg2 VAL 12	Hd2 HIS 14	0.12	5.49	0.10	2.90

Table 3.3. List of Aβ40 experimentally assigned NOE peaks that are not due to intra-residue or sequential contacts along with calculated NOE values from simulation for both H_2O and ^2H_2O.

Atom 1	Atom 2	Calculated H_2O NOE intensity	Experimental H_2O NOE intensity	Calculated ^2H_2O NOE intensity	Experimental ^2H_2O NOE intensity
He TYR 10	Hg2 VAL 12	0.93	1.63	0.86	1.84
Hg3 GLU 22	Hg VAL 24	0.68	3.29	0.60	
Hd1 LEU 17	He or d PHE 19	0.61	3.06	0.56	5.07
Hd TYR 10	Hg2 VAL 12	0.52	1.52	0.45	1.70
Hd PHE 20	Hg VAL 24	0.47	1.98	0.43	3.44
He PHE 20	Hg2 GLU 22	0.35	1.32	0.32	
Hb ASP 7	Hd TYR 10	0.26	1.50	0.23	
Hg2 VAL 18	He PHE 20	0.18	1.57	0.17	2.98
Hg2 VAL 18	Hd PHE 20	0.17	2.85	0.15	2.70
Hd PHE 20	Hg3 GLU 22	0.17	1.70	0.15	
Hg2 VAL 18	Hz PHE 20	0.07	1.52	0.07	1.63

Table 3.4. List of Aβ42 experimentally assigned NOE peaks that are not due to intra-residue or sequential contacts along with calculated NOE values from simulation for both H_2O and ^2H_2O.

Atom 1	Atom 2	Experimental NOE Intensity		Number of Peaks in Spectrum	
		H_2O	^2H_2O	H_2O	^2H_2O
Ha2 GLY 37	H VAL 39	5.57		2	0
Hb2 GLU 22	Hg3 MET 35	5.17		1	0
Hg2 VAL 18	Hb3 PHE 20	2.03	1.59	2	2
Hb2 PHE 20	Hb3 LYS 16		1.98	0	1
Hb2 PHE 20	Hb2 LEU 34		1.98	0	1
Hb2 PHE 20	Ha GLU 11		1.54	0	1
Hb ASP 7	H TYR 10	1.46		1	0
He1 HIS 6	Ha GLU 3	1.44		1	0
Hg2 VAL 18	Hb3 TYR 10		1.34	0	1
H VAL 40	Ha2 GLY 38	1.33		1	0
H GLU 11	Ha SER 8	1.26		1	0
Hd21 ASN 27	Ha GLU 22	1.23		1	0
Hb ASP 7	He TYR 10	1.23		2	0
H TYR 10	Ha ARG 5	1.22		1	0
He21 GLN 15	Ha GLU 11	1.18		1	0
H ASP 23	Ha SER 26	1.16		1	0
H GLY 9	Ha GLU 11	0.94		1	0
H VAL 40	Ha3 GLY 38	0.91		1	0
H ALA 30	Hg3 LYS 28	0.70		1	0

Table 3.5. List of Aβ40 false-negative NOE peaks (identified experimentally but not calculated from simulation) that cannot be explained by: the presence of high noise in that region of the spectrum or by a small simulation inaccuracy that identifies the correct residue interaction but misses the exact atomic contact. Listed are the experimental intensities and number of peaks observed experimentally for both H_2O and ^2H_2O spectra, and the atomic contact which would produce this peak. If more than one atomic contact can produce this NOE peak, the contact is listed with the smallest difference in sequence between the two residues.

Atom 1	Atom 2	Experimental NOE		Number of Peaks	
		H_2O	^2H_2O	H_2O	^2H_2O
Hd PHE 20	Hg VAL 24	2.0	3.4	2	2
Hb ASP 7	Ha ALA 21	3.4		1	0
Hb2 GLN 15	Hg2 VAL 12	3.0		2	0
Hd2 LEU 17	Hd TYR 10	1.1	2.8	1	2
Hg2 VAL 12	Hb ASP 7		2.6	0	1
H VAL 24	Ha HIS 13	2.6		1	0
Hb2 ASP 23	Ha GLY 25		2.4	0	1
Hd2 HIS 13	Hh1 ARG 5	2.4		1	0
Hb2 GLU 3	Hb2 ASP 1	2.2		1	0
Hb2 HIS 13	Ha GLU 11		2.1	0	1
He1 HIS 6	H ALA 30	2.0		1	0
Hb2 ASP 1	Hd2 HIS 13		2.0	0	1
Hd2 HIS 6	Ha HIS 13		1.9	0	1
Hz PHE 19	Ha ASP 23	1.7		1	0
H ASP 7	Hb2 ARG 5	1.7		1	0
Hb2 GLU 11	Hb2 HIS 13	1.7		1	0
Hz PHE 20	Hg2 VAL 18	1.5	1.6	1	2
Hd22 ASN 27	H ILE 32	1.6		1	0
Hd21 ASN 27	He22 GLN 15	1.6		1	0
He21 GLN 15	Hd2 HIS 6	1.5		1	0
Hd TYR 10	Hb ASP 7	1.5		1	0
Hd2 HIS 13	Hb3 GLU 11	1.5		1	0
Hd2 HIS 13	Ha GLU 11	1.5		1	0
Hd2 HIS 6	Ha GLU 3	1.5		1	0
Hd2 LEU 17	He TYR 10	0.6		2	0
Hd2 HIS 14	Hb VAL 12	1.4		1	0
Hb3 ASP 1	Ha PHE 4	1.4		1	0
He PHE 20	Hg2 GLU 22	1.3		1	0
H ALA 30	Ha MET 35	1.3		1	0
Hd22 ASN 27	Hg12 ILE 32	1.3		1	0
He21 GLN 15	Hg2 GLU 22	1.3		1	0
Hd TYR 10	H HIS 13	1.3		1	0
Hg2 VAL 18	Hd2 HIS 14	1.3		1	0
Hd1 LEU 17	Hd TYR 10	1.3	0.7	1	1
Hg3 GLU 3	H SER 8	1.1		1	0
Hg LEU 17	H VAL 24	1.1		1	0
Hd2 HIS 13	Hb3 TYR 10	1.0		1	0
Hb PHE 4	H TYR 10	1.0		1	0
H LEU 17	Ha SER 8	1.0		1	0
H GLU 22	Hb ALA 42	1.0		1	0
H HIS 13	Hb2 ARG 5	1.0		1	0
Hb3 HIS 13	He TYR 10	0.9		1	0
Hb3 PHE 19	Hd1 ILE 41	0.9		1	0
He LYS 16	Hd PHE 20		0.7	0	1

Table 3.6. List of Aβ42 false-negative NOE peaks.

MD H_2O NOE Intensity	MD ^2H_2O NOE Intensity	Atom 1	Atom 2	MD Contact Intensity	Atom 1	Atom 2	MD Contact Intensity	Atom 1	Atom 2	MD Contact Intensity	Atom 1	Atom 2	MD Contact Intensity	Atom 1	Atom 2	MD Contact Intensity	Atom 1	Atom 2	MD Contact Intensity
		Hb3 HIS	Ha2 GLY		Hb2 HIS	Ha GLY		Hd ARG	Ha2		Hb2 PHE	Ha GLY		Hb3 HIS	Ha2 GLY		Hb3 HIS	Ha3 GLY	
8.10	6.50	6	9	1.40	14	29	1.15	5	GLY 9	0.88	20	25	0.80	6	33	0.72	6	33	0.72
		Ha2 GLY	Ha TYR		Ha HIS	Ha GLY													
6.83	4.99	9	10	1.37	14	29	1.10												
		Ha2 GLY	Hb3 TYR		Hb3 PHE	Ha GLY		Hb3 PHE	Ha3										
5.13	4.86	9	10	1.05	20	25	0.80	20	GLY 37	0.53									
		Ha VAL	Hb2 HIS		Hb2 HIS	Ha2 GLY		Hb2 HIS	Ha3		Hb2 PHE	Ha3 GLY							
4.03	4.72	12	13	1.97	6	33	0.72	6	GLY 33	0.72	20	37	0.53						
		He1 HIS	Hb3 SER																
	4.27	6	8	4.23															
		Ha PHE	Ha ASN																
3.68	4.07	20	27	3.88															
		Hb2 HIS	Ha2 GLY																
	2.50	6	9	1.40															
		Hd2 HIS	Ha GLY																
	2.14	14	29	0.59															
		Ha VAL	He1 HIS		He1 HIS	Ha GLY													
	2.13	12	14	0.67	14	29	0.60												
		H VAL	Ha3 GLY																
2.10		18	37	1.87															
		Hd PHE	Hb3 ASN																
1.96	2.08	20	27	1.40															
		Hd PHE	Hb2 ASN																
	1.95	20	27	1.40															
		He1 HIS	Hb2																
	1.93	13	GLN 15	1.36															
		He1 HIS	Hb3																
	1.83	13	GLN 15	1.36															
		Ha PHE	Hb3 ASN																
1.72	1.48	20	27	1.18															
		Hd2 HIS	Hb ASP																
	1.52	6	7	1.87															
		Ha ASP	Hb2 ASN																
1.39	1.26	23	27	1.31															
		Ha ASP	Hb3 ASN																
1.38	1.32	23	27	1.31															
		Hg1	Ha3 GLY																
1.19		VAL 18	37	1.07															
		Hζ PHE	Ha GLY																
	1.08	20	25	1.02															
		Hb2 PHE	Ha ASN																
1.06	1.07	20	27	0.90															

Table 3.7. List of Aβ40 false-positive NOE peaks (calculated from simulation but not identified experimentally) that cannot be explained by the presence of a high noise in that region of the spectrum or by a small simulation inaccuracy which identifies the correct residue interaction but misses the exact atomic contact. If the peak is made up of more than one atomic contact, the calculated intensity for each contact with an intensity greater than 0.5 for either H_2O or ^2H_2O is listed.

MD H_2O NOE Intensity	MD ^2H_2O NOE Intensity	Atom 1	Atom 2	MD Contact Intensity	Atom 1	Atom 2	MD Contact Intensity	Atom 1	Atom 2	MD Contact Intensity	Atom 1	Atom 2	MD Contact Intensity	Atom 1	Atom 2	MD Contact Intensity
7.86	5.72	Ha VAL 12	Hb2 HIS 13	1.82	Hb2 TYR 10	Ha GLY 25	1.32	Hb2 HIS 6	Ha2 GLY 9	1.04	Hb2 HIS 13	Ha VAL 36	0.71	Hb2 HIS 13	Ha GLY 37	0.63
7.85	6.91	He LYS 16	Ha GLY 37	1.91	Hb3 TYR 10	Ha GLY 25	1.32	He LYS 16	Ha GLY 38	0.59	Ha GLY 25	He LYS 28	0.54			
	7.15	He1 HIS 13	Hg GLN 15	5.80	He1 HIS 14	Hg GLN 15	1.43									
6.10	5.29	Ha VAL 12	Ha MET 35	3.56	Ha MET 35	Ha VAL 36	0.75									
6.07	4.04	Ha GLU 22	Ha GLY 25	2.52	Ha GLY 38	Ha VAL 39	1.27	Ha GLU 11	Ha VAL 12	1.08						
5.68	5.38	Hb2 TYR 10	Hb MET 35	2.14	Hb2 TYR 10	He MET 35	0.68	Hb2 HIS 6	He MET 35	0.66						
5.04		Hd2 HIS 13	Ha VAL 36	2.63	He22 GLN 15	Ha GLY 38	0.93	Hd2 HIS 13	Ha GLY 37	0.77						
5.02	4.23	Hb LEU 34	Hb MET 35	1.41	Hb LEU 34	He MET 35	0.89	Hg LEU 34	He MET 35	0.83	Hg LEU 34	Hb MET 35	0.82			
4.69	3.47	Hd2 HIS 6	He MET 35	1.00	Hd TYR 10	He MET 35	0.97	Hb2 GLU 22	Hd22 ASN 27	0.87	Hd2 HIS 6	Hb MET 35	0.87	Hd TYR 10	Hb MET 35	0.83
	4.62	He1 HIS 6	Hb2 SER 8	4.14												
4.32	3.69	Hb3 TYR 10	Hb MET 35	2.14	Hb3 TYR 10	He MET 35	0.68	Hb3 GLN 15	He LYS 16	0.53						
4.20	2.81	Ha VAL 12	Hb3 HIS 13	1.82	Hb3 HIS 13	Ha VAL 36	0.71	Hb3 HIS 13	Ha GLY 37	0.63						
	4.09	He1 HIS 6	Hb3 SER 8	4.14												
4.09	2.41	Ha VAL 12	Ha HIS 13	1.30	Ha PHE 20	Hb2 SER 26	0.86	Ha PHE 19	Hb2 SER 26	0.66						
	3.84	He1 HIS 13	Hb2 GLN 15	3.43												
3.77	1.10	Hb2 SER 8	Hd TYR 10	1.41	Hd TYR 10	Ha GLY 29	0.68	Ha2 GLY 9	Hd TYR 10	0.55						
	3.64	He1 HIS 13	Hb3 GLN 15	3.43												
	3.63	Hd2 HIS 13	Ha VAL 36	2.63	Hd2 HIS 13	Ha GLY 37	0.77									
3.60	3.38	Hb2 SER 8	He TYR 10	2.44	He TYR 10	Ha GLY 29	0.50									
3.34	3.01	Hd or e PHE 4	He MET 35	0.82	Hd or e PHE 4	Hb MET 35	0.59									
3.34		He1 HIS 13	Ha GLY 37	2.97												
3.32	2.87	Hd LYS 16	Ha GLY 25	2.22												
	3.16	He1 HIS 13	Ha GLY 37	2.97												

2.90	2.31	He PHE 20	Ha GLY 29	1.05								
2.81	2.63	Hζ PHE 20	Ha GLY 29	0.64	Hd PHE 20	Hb2 SER 26	0.64					
2.79	1.59	Hb ALA 21	He MET 35	0.72	Hb ALA 21	Hb2 GLU 22	0.69					
2.63		Ha ASP 23	H GLY 25	1.52	Ha ASP 7	H GLY 9	1.08					
2.62	2.43	Ha TYR 10	Ha GLY 25	2.04								
2.56	2.47	Hb2 HIS 13	Ha MET 35	2.10								
2.55		H GLY 37	H GLY 38	2.43								
	2.53	Ha GLN 15	Ha GLY 37	2.28								
2.51	2.34	Hb3 HIS 13	Ha MET 35	2.10								
2.44	2.30	Hb3 SER 8	He TYR 10	2.44								
2.43	1.48	Hg GLN 15	Ha GLY 37	1.11	Hg GLN 15	Ha GLY 38	0.91					
2.42	1.77	Hb2 HIS 6	Ha3 GLY 9	1.04								
2.32	2.06	Hζ PHE 4	He MET 35	0.75	Hζ PHE 4	Hb MET 35	0.52					
2.31		Hb2 HIS 6	H GLY 9	1.62								
2.25	1.99	Hb3 HIS 6	Ha2 GLY 9	1.04								
2.21	2.23	Ha VAL 24	Hb2 ASN 27	1.82								
2.04		Ha GLU 22	H GLY 25	1.85								
2.03		H ALA 21	Hb3 SER 26	1.07	H PHE 20	Hb3 SER 26	0.74					
2.02	1.23	Hb3 ASP 23	Hb2 SER 26	0.91	Hb ASP 7	Hb2 SER 8	0.62					
2.00	1.77	Hb3 HIS 6	Ha3 GLY 9	1.04								
1.94	1.88	He LYS 16	Ha GLY 38	0.59								
1.88		Ha PHE 20	Hb3 SER 26	0.86	Ha PHE 19	Hb3 SER 26	0.66					
1.86		Hb3 HIS 6	H GLY 9	1.62								
1.80	1.41	Hb LEU 34	Ha MET 35	0.94	Hg LEU 34	Ha MET 35	0.55					
1.78	1.62	Ha VAL 12	Hg2 MET 35	0.61								

1.78	1.62	Ha VAL 12	Hg3 MET 35	0.61				
1.63	1.55	Ha TYR 10	He MET 35	0.71				
1.60	1.57	Ha GLU 22	Hb2 ASN 27	0.89	Ha ALA 21	Hb2 ASN 27	0.59	
	1.60	Hb3 SER 8	Hd TYR 10	1.41				
1.56	1.24	Ha GLY 25	Ha SER 26	1.50				
	1.54	He1 HIS 13	Ha GLN 15	1.29				
1.39	1.22	Hb LEU 34	Hg2 MET 35	0.88				
1.39	1.22	Hb LEU 34	Hg3 MET 35	0.88				
1.37	1.27	Hg2 VAL 12	He MET 35	0.82				
1.26		H ASP 7	H GLY 9	1.07				
	1.05	Hg GLN 15	Ha GLY 38	0.91				
1.04		Hd2 HIS 13	H GLY 37	1.00				
	1.03	Hd2 HIS 14	Hg GLN 15	0.92				
	1.01	He1 HIS 6	Hb ASP 7	1.15				

Table 3.8. List of Aβ42 false-positive NOE peaks (calculated from simulation but not identified experimentally) that cannot be explained by the presence of a high noise in that region of the spectrum or by a small simulation inaccuracy which identifies the correct residue interaction but misses the exact atomic contact. If the peak is made up of more than one atomic contact, the calculated intensity for each contact with an intensity greater than 0.5 for either H_2O or ^2H_2O is listed.

Val18 Hydrogen bonds	Aβ40	Aβ42
Val12	12%	
His13	10%	
His14		5%
Gln15	9%	
Lys16	5%	
Ile32		7%
Leu34		14%
Val36	8%	
Total	**44%**	**26%**

Table 3.9. Val18 backbone hydrogen bonds present in more than 1% of the ensemble for Aβ40 and Aβ42. The brown boxes correspond to β-strands and the orange corresponds to a turn at residues 14 and 15.

Chapter 4

Determining the structural ensemble of intrinsically disordered disease peptides using computation and experiment

Intrinsically disordered proteins (IDPs) represent a new frontier in structural biology since the primary characteristic of IDPs is that structures need to be characterized as diverse ensembles of conformational sub-states. These proteins are often determined to be central to disease processes such as Alzheimer's and Parkinson's. In this Perspective we review two general but very different ways for combining NMR spectroscopy with theoretical methods to derive structural ensembles for the disease IDPs amyloid-β 1-40 and amyloid-β 1-42, which are associated with Alzheimer's disease. We compare the performance of *de novo* molecular dynamics and knowledge-based approaches for generating structural ensembles by assessing their ability to reproduce a range of NMR experimental observables. In addition to the comparison of computational methods, we also evaluate the relative value of different types of NMR data for refining or validating the IDP structural ensembles, culminating in an overview of future directions for improving the theoretical and computational infrastructure for both *de novo* molecular dynamics (MD) and knowledge-based methods. We suggest that a larger community effort in the structural biology of IDPs could be realized through the establishment of a high-throughput NMR-MD IDP structure 'solver', analogous to the advent of X-ray crystallography beamlines to solve 3D structures of folded proteins.[3]

1. Introduction

Recognition of the importance of intrinsically disordered proteins (IDPs) has gradually taken root in the field of structural biology over the past 14 years since the publication of the seminal

[3] Reproduced from (64) with permission.

review by Wright and Dyson (17). Intrinsic protein disorder can refer to either local, disordered regions of a folded protein, or a global protein disorder without any stable structure along the entire sequence. Locally disordered proteins have been commonly assessed in X-ray and NMR studies and have been well studied from a functional perspective. In this chapter we focus on the study of globally disordered proteins. These proteins adopt neither a single nor a small number of stable folded conformations and are very flexible, frequently sampling different dihedral angles along the backbone. The energy landscape of an IDP lacks a deep minimum, unlike a folded globular protein, and the IDP is likely to sample many local minima (118). Thus, an IDP structure can be described as an ensemble of conformational states, each at a different local minimum, weighted according to a Boltzmann distribution (126).

Numerous examples of functional IDPs have now been characterized and found to serve important roles in recognition, inhibition, molecular assembly, protein modification, and entropic chain activities (71, 73). Considerable progress has also been made in identifying protein sequences that are likely to be disordered, and using these prediction algorithms, IDPs are estimated to make up 25-30% of proteins encoded in the human genome (71, 73-75). Disease proteins represent an important subcategory of IDPs. Many amyloid proteins are unstructured in the soluble monomeric state, but form large insoluble fibrils upon aggregation (15, 127). These include proteins involved in neurodegenerative disorders such as Parkinson's disease, Huntington's disease, prion diseases, and Alzheimer's Disease (AD). While these proteins may have properties not associated with functional IDPs, techniques developed to study intrinsically disordered disease proteins are likely applicable to all globally disordered proteins. For more general information about IDPs, we direct the reader the following reviews (15, 75, 126, 128, 129).

The experimental identification of proteins with global intrinsic disorder can be performed using various spectroscopic techniques including Circular Dichroism (CD), NMR, infrared spectroscopy (IR), UV spectroscopy, and fluorescence spectroscopy (17, 18). CD and IR report on the amount of secondary structure, while lack of chemical shift dispersion in NMR spectra is a good indication of high flexibility. Hydrodynamic techniques such as SAXS, gel filtration, and dynamic light scattering can also aid in IDP identification as they report on the radius of the protein, which will be larger for an IDP or denatured protein than a folded protein of the same mass. Lack of a cooperative folding transition, solubility at high temperatures, and proteolytic sensitivity are also attributes of IDPs that are useful in forming a complete picture of a certain protein's level of disorder. A subset of these techniques is generally employed in order to form a consensus on a given protein's level of disorder.

Experimental methods for identifying IDPs are affective at showing whether or not a protein is folded, and up to this point have primarily been used to classify IDPs based on a signature degree of underlying order. The assumption is that all conformational sub-states can be categorized into one of the following: a collapsed semi-ordered ensemble (compact structures with some well-formed secondary structure), a collapsed disordered ensemble (compact conformations with little or no secondary structure), or an extended disordered ensemble (the least collapsed and most unstructured ensemble classification) (18, 73, 75, 77). More recently, however, increased importance has been placed on characterization of the conformational sub-states comprising these ensembles since they each may have distinct functional roles (18, 130-135) or lead to hypotheses about disease origin (65). Experimental approaches such as X-ray and

electron crystallography and microscopy have traditionally excelled at determining the structure of single folded proteins (136, 137) and large protein complexes (138). IDPs, however, are not amenable to these static structural determination methods (128).

IDPs represent a new frontier in structural biology in that the IDP structure must be characterized as a diverse ensemble of interconverting conformational sub-states, as opposed to a single dominant 3D structure (126). This necessitates an adjustment in the core methodology of protein structure determination for this class of protein. In order to achieve both better ensemble classification and detailed description of conformational sub-states, we must critically assess how we currently derive these complex structural ensembles from experiment and theoretical models.

NMR is the experimental tool of choice for characterizing the solution structure and dynamics of biological molecules since it characterizes the native distribution of conformations in an aqueous environment (50, 128, 139). Observables from these experiments include chemical shifts, which are characteristic of functional groups and their surrounding environment, and spin-spin couplings (J-couplings), which independently report on backbone dihedral angles. In addition, through-space dipole-dipole interactions give rise to the Nuclear Overhauser Effect (NOE) that reports on tertiary structure contacts, and more recently, residual dipolar couplings (RDCs) have been used to describe the relative orientation of spatially separated regions of a protein (140-142). Paramagnetic Relaxation Enhancements (PREs), which can produce longer distance restraints than NOEs have also been used in the context of IDPs (51-53, 143), however this measurement requires chemical modification of the protein with a nitroxide spin label or an amino-terminal copper binding motif, which may perturb the monomeric IDP conformations (129, 144); for the amlyoid-β PRE study by Fawzi and coworkers, an additional cysteine is added to the N-terminus of the peptide to which the spin label is attached (145).

IDPs typically convert between conformations faster than the ~ns-ms timescale of the NMR experiment, leading to an averaging of the NMR observables across structural sub-populations. This uniform average hinders the structural characterization of all the conformational sub-states, and can even obscure the overall ensemble classification, as we will see for the amyloid peptides. Building the connection between the averaged NMR observables and the complete IDP structural ensemble therefore depends critically on computational models (146). The goal of the computational model is to provide a properly weighted set of the diverse sub-populations of the IDP most consistent with the NMR observables and perhaps other experimental measures such as circular dichrosim (147), small angle X-ray scattering (54, 148), or PREs (52, 149). Thus, multiple types of NMR or other experimental observables are necessary for validation of the computational model (50, 148).

Currently there are two primary but very different computational approaches to building an IDP structural ensemble composed of properly weighted conformational sub-states, which can be loosely contrasted as first principle or *de novo* methods vs. knowledge-based approaches. The *de novo* approach implements molecular dynamics (MD) simulations based on the theoretical foundations of statistical mechanical sampling and model-derived potential energy surfaces. *De novo* MD generates a structural ensemble that is representative of given thermodynamic conditions, i.e. a Boltzmann weighted ensemble of conformational sub-populations, independent of experimental input. The complementary use of MD and NMR data to determine structure and

dynamics of folded and unfolded proteins has been a highly active area over the last two decades (150, 151), particularly for relaxation measurements that require a dynamical interpretation of the NMR data at the picosecond and nanosecond timescales (152). For the *de novo* MD method, multiple NMR or other experimental data are necessary to *validate* the MD ensemble through direct back-calculation of observables and comparison to experiment. Once validated, MD simulations provide a prediction of the complete IDP structural ensemble, allowing overall classification as well as the study of individual conformational sub-states, which can be analyzed with some confidence.

In contrast, we define knowledge-based approaches as those that use experimental NMR information directly to derive the structural ensemble. While MD is often used to generate atomistic predictions independent of NMR experimental input, as in our *de novo* method, a number of researchers have advanced the combination of applying knowledge from NMR to restrain the MD ensemble (51, 53, 55-57). Such methods are the foundation of NMR structure determination of folded proteins using experimentally derived conformational constraints based on chemical shifts, J-couplings, and NOE data embodied in software packages such as CANDID (153), CYANA (154), X-Plor-NIH (155, 156), and TALOS (157). MD simulations have been combined with RDC restraint data for folded proteins (56) that then allows for the analysis of other features of the ensemble, such as conformational fluctuations. NMR restrained MD has also been applied to IDPs such as α-synuclein, a disease protein indicted in Parkinson's disease, which incorporated distance restraints derived from PRE experiments in order to guide the MD so that the protein's radius of gyration distribution is in good agreement with the experimental value (55).

Other knowledge-based approaches for IDPs forgo MD simulations altogether and instead use an extensive set of statistical coil conformations (50, 58); this starting pool, which can be generated using a variety of heuristics, can be thought of as a basis set of structures. Subsequently the starting pool of structures is then culled for the subset of conformations that are in best agreement with experimental data to create the IDP ensemble. In the energy-minima mapping and weighting (EMW) method, Stultz and coworkers used end-to-end distance restraints to develop a pool of conformations with varying radii of gyration. They then selected, via Monte Carlo, a weighted ensemble of 15 structures to optimize the agreement with experimental ^{13}C and ^{15}N chemical shifts and J-couplings (59, 60). Blackledge and coworkers have developed the program *Flexible-Meccano* to create a pool of structures based on random coil backbone dihedral angles, on which they employ a genetic search algorithm in their ASTEROIDS software program to select structures that together best match experimental chemical shifts, PREs, or RDCs (50, 61, 62).

The ENSEMBLE method, developed by Forman-Kay and coworkers, typically defines the starting pool of IDP conformational states as an ensemble of extended or random coil states, with an option for biasing the secondary structure of the ensemble at certain places in the sequence that are known to be partially structured (52, 54, 58, 63). Structures are selected from this pool using a Monte Carlo selection algorithm with an energy-weighting scheme for each type of experimental input. The ENSEMBLE program includes modules for several different experimental data types including chemical shifts, RDCs, PREs, J-couplings, and contact distances derived from NOEs. Since ENSEMBLE is a user friendly and publically available software package (58), we review it as an example of a knowledge-based approach.

The primary objective of this chapter is to review the *de novo* and knowledge-based approaches for deriving IDP structural ensembles in context of the intrinsically disordered Alzheimer's disease peptides amyloid-β 1-40 (Aβ40) and amyloid-β 1-42 (Aβ42). For the knowledge-based approach we analyze the ENSEMBLE method and consider its performance against the null hypothesis that random coil or statistical ensembles can explain the experimental observations. This comparison exposes the relative utility of different types of NMR data for refining or validating the IDP computational ensemble. As we outline in this chapter, chemical shifts and RDCs do not differentiate between IDP conformational ensembles that are qualitatively different in both their degree of secondary structure and their tertiary contacts. In contrast, scalar couplings and NOEs provide much stronger discrimination between qualitatively different structural ensembles for IDPs such as Aβ40 and Aβ42.

We conclude that while *de novo* MD currently offers one of the best approaches for generating correct Boltzmann weighting of the sub-populations that make up an IDP structural ensemble, a combination of MD and knowledge-based methods yields the most experimentally coherent and refined set of results. The importance of including MD-derived conformations is especially evident for disease peptides that resist a classification based on a particular degree of order, such as amyloid-β. We describe future directions for improving the computational infrastructure of both *de novo* and knowledge-based methods to most reliably work together with available NMR data on IDPs. Finally we suggest that a larger community effort in the structural biology of IDPs could be realized by combining NMR, MD, and knowledge-based methods into a structure 'solver', analogous to the advent of X-ray crystallography structural beamlines that have successfully solved the 3D structures for thousands of folded proteins.

2. Amyloid-β as a Model IDP

Amyloid-β is an intrinsically disordered protein involved in Alzheimer's disease, which is cleaved from the transmembrane amyloid precursor protein in variable sequence-length forms. The most common of these are amyloid-β 1-40 (Aβ40) and amyloid-β 1-42 (Aβ42). These peptides are identical in sequence except at the C-terminus, where Aβ42 contains two additional residues, isoleucine and alanine. Both *in vivo* and *in vitro* these peptides aggregate to form long fibrils composed of β-sheet structure, however Aβ42 aggregates more quickly *in vitro* and is believed to be more toxic *in vivo*. In the soluble monomeric form, various experimental techniques detect no major secondary or tertiary structure content, consistent with their classification as statistical coil IDPs (14, 15, 19, 20). An important question at present is whether there are important structural differences in their conformational sub-states that might explain the greater disease virulence of Aβ42 compared to Aβ40. We refer the reader to the following reviews of amyloid-β from a disease perspective (1, 2, 122). Here they serve as an ideal case study in a comparison of *de novo* and knowledge-based methods for reliably characterizing IDP conformational ensembles for this important disease problem.

3. Generation of IDP Structural Ensembles

In order to classify a given IDP into one of the aforementioned three levels of disorder, we must first generate a set of alternative ensembles to compare to the available NMR data, eventually selecting the one that is best validated. Here we consider the creation of four qualitatively

different conformational ensembles. Firstly we consider the null hypothesis that Aβ can be well represented by a random coil (RC) ensemble. The second ensemble is also random, but incorporates bioinformatics-based knowledge about what secondary structure category is more likely for a given residue in the amino acid sequence. In this case, the random ensemble is biased to contain a statistical probability of predicted secondary structure on a per residue basis (Pred-SS). Thirdly, we employ a fully knowledge-based ENSEMBLE approach that generates structural ensembles that best conform to the NMR data based on selection from the Pred-SS ensemble (Pred-SS-ENS). Each of the above three ensembles is compared against the ensemble we have generated by *de novo* MD for each of the IDPs Aβ40 and Aβ42, which we designate as our fourth candidate ensemble method (MD). Based on our analysis of the first four ensembles, we also create an additional fifth ensemble (MD-ENS) that combines the knowledge-based and *de novo* MD approaches, by using ENSEMBLE to select structures from the *de novo* MD starting pool. Details of the ensemble generation protocols are given in the Appendix.

Table 4.1 shows the average radius of gyration (Rg) values for each ensemble. We see that the order from most extended to most compact proceeds as Pred-SS > Pred-SS-ENS > RC > MD ~ MD-ENS, and thus the alternative ensembles span a range of IDP classifications by the <Rg> measure. Figure 4.1[4] provides the propensities for the Pred-SS vs. *de novo* MD ensemble to form turns, anti-parallel β-strands, or helical structure by residue for Aβ42. We do not show the secondary structure profiles for the RC and Pred-SS-ENS ensembles since they are similar to or less structured than the Pred-SS ensemble (see the Appendix), while MD-ENS resembles the MD ensemble.

Aβ40 Peptide	**Average Property**					
Ensemble Type	Rg (Å)	χ^2 (Hα)	χ^2 (HN)	χ^2 (Cα)	χ^2 (Cb)	$^3J_{H^N H^\alpha}$
RC	16.9± 3.1	0.77	0.07	0.31	0.78	1.95 (7.16)
Pred-SS	19.3± 3.6	0.48	0.13	0.49	0.55	1.95 (7.13)
ENS-Pred-SS	17.4± 3.5	0.28	0.06	0.37	0.20	1.99 (7.46)
MD	14.7± 4.8	0.58	0.36	0.69	0.70	0.99 (1.82)
ENS-MD	15.0± 4.1	0.30	0.34	0.46	0.36	0.62 (0.72)
Aβ42 Peptide	**Average Property**					
Ensemble Type	Rg (Å)	χ^2 (Hα)	χ^2 (HN)	χ^2 (Cα)	χ^2 (Cb)	$^3J_{H^N H^\alpha}$
RC	17.4± 3.3	0.88	0.08	0.35	0.90	2.10 (8.29)
Pred-SS	19.9± 3.8	0.51	0.12	0.72	0.61	2.09 (8.20)
ENS-Pred-SS	18.4± 4.0	0.30	0.08	0.47	0.12	2.28 (9.73)
MD	13.1± 4.3	0.54	0.48	0.98	0.52	0.99 (1.83)
ENS-MD	13.1± 2.8	0.33	0.37	0.51	0.34	0.60 (0.67)

Table 4.1. Comparison between random coil (RC), predicted secondary structure (Pred-SS), de novo MD (MD), and ENSEMBLE optimized Pred-SS-ENS and MD-ENS ensembles. For the radius of gyration (Rg) values we report both the ensemble average and RMSD. For chemical shifts we report χ^2 that measures agreement between the computational ensembles and the experimentally measured chemical shifts; $\chi^2 < 1$ indicates no disagreement with experiment within SHIFTX calculator error. We also report the $^3J_{H^N H^\alpha}$ RMSD (χ^2).

[4] This and all other figures in this chapter are reproduced from (64) with permission.

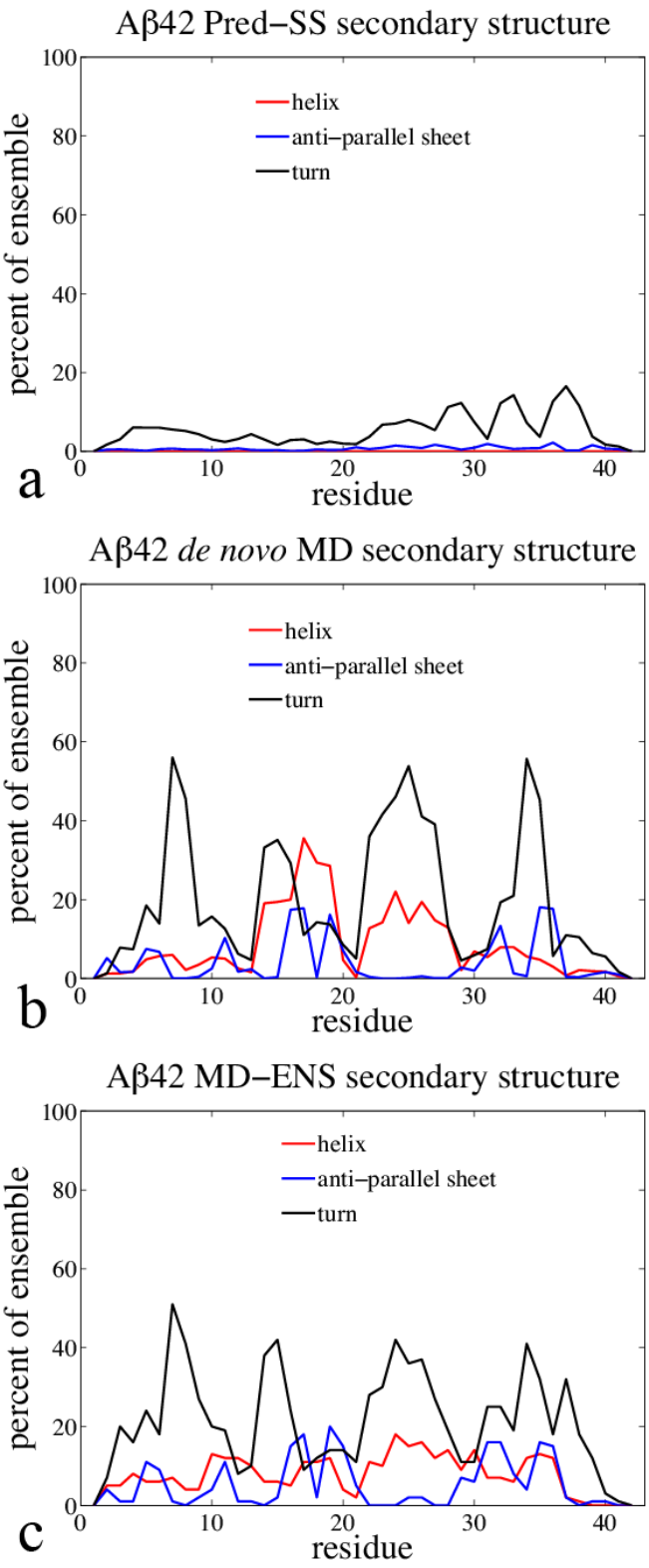


Figure 4.1. *Percentage of Aβ42 simulated ensemble in different types of secondary structure by residue for (a) the Pred-SS, (b)* de novo *MD, and (c) MD-ENS ensembles.* The red line represents helix, the blue line for anti-parallel strand, and the black line for β-turns.

We have previously shown that the Aβ40 and Aβ42 peptide samples some type of structured conformations in ~99% of its ensemble, including complex β-strand motifs (48, 65). From this, we conclude that the radius of gyration trends stem from the much larger propensity for the MD ensembles to form *cooperative* secondary structure and therefore to be more compact, as opposed to the random or knowledge-based ensembles that do not generate contiguous blocks of secondary structure, and hence are more extended on average. Although the secondary structure content of the MD-ENS ensemble resembles that of the MD ensemble, Figure 4.1 shows that there is some variation in the percentages with which certain residues adopt different types of secondary structure.

4. Experimental NMR Applied to IDPs

Any of the computational methods considered depend on both high quality and a wide range of experimental NMR data. The primary reason is that qualitatively different structural ensembles may reproduce a given NMR observable with the same fidelity, so that additional experimental observables are needed for validating different structural attributes of the same system (129, 158, 159). For analysis of Aβ40 and Aβ42 IDPs covered in this chapter, we have utilized chemical shift data from the Zagorski group (14) as well as J-coupling constants, RDCs, and heteronuclear ^{1}H-^{15}N NOEs for backbone amides from Wang and co-workers (66, 67, 125). Our group generated NOESY ^{1}H-^{1}H homonuclear spectra collected for both peptides Aβ40 and Aβ42 and reported elsewhere (48, 65). The data were processed as described in (48). Here we summarize the experimental characterization of Aβ40 and Aβ42 that is independent of any theoretical model of their IDP structural ensembles.

Chemical shifts. In folded proteins, the type of secondary structure at a particular residue affects the chemical environment of the hydrogen and carbon atoms in that residue, leading to signature chemical shift values for α-helices as well as β-sheets (160, 161). Chemical shifts have also been used to quantify secondary structure propensities in IDPs (62, 63, 162, 163). However, the Aβ40 and Aβ42 hydrogen and carbon chemical shifts are not highly dispersed, i.e. there are no obvious regions of the sequence displaying a clear trend toward cooperative α-helix or β-strand chemical shifts.

A more sound analysis is to examine chemical shift indices (CSI), i.e. deviations of chemical shift values from a random coil reference, that provide a better indication of whether the region is more likely to be in an α-helix or β-strand (160, 161) For example, a recently reported knowledge-based CSI "calculator" predicts similar β-sheet trends for Aβ40 and Aβ42, with highest percentages of β-sheets, ~15-20%, in the C-terminus when backbone nitrogen chemical shifts are included (163). If nitrogen chemical shifts are not included, for the reason that they have greater experimental error relative to the chemical shifts of other atom types, then the absolute probabilities of β-sheets were found to be very low, 1-5%, for both peptides in all regions of the sequence. This indicates that both Aβ IDP monomers contain some β-strands, but the β-strand population percentages are ambiguous if only chemical shift information is taken into account.

J-coupling constants. J-coupling constants are another measurement reflecting a protein's secondary structure because they report directly on the backbone φ dihedral angle for each

residue. The J-coupling values for Aβ40 and Aβ42 also provide no evidence of secondary structure "blocks" at different points in the peptide sequence, i.e. no persistence of J-coupling values that would be consistent with a dominant population of α-helical or β-strand secondary structure as in a folded protein. However, the J-couplings for both peptides are shifted upward from random coil to yield values mostly between 6.0-8.5 Hz. This is consistent with an extended random coil ensemble or the presence of β-strand structure (14), although population percentages are unknown.

Homonuclear NOE. Unlike chemical shifts and J-couplings, homonuclear NOEs are not as highly averaged over all conformations in the IDP ensemble, but rather report on persistent tertiary structure contacts between pairs of atoms in the protein. An NOE cross-peak between two atoms is a result of through-space coupling between the nuclei of those atoms and indicates that the atoms must come into close contact in a significant percentage of structures in the IDP ensemble.

The ^{1}H-^{1}H homonuclear NOE cross-peaks for Aβ40 and Aβ42 provide evidence for compact tertiary conformations held together by a number of long-range contacts that must persist on long enough time-scales for the cross-relaxation between protons to be detected, and thus these interactions must be stabilized by (likely shallow) free-energy minima. Table 4.2 provides a summary of the ^{1}H-^{1}H homonuclear NOE cross-peaks for Aβ40 and Aβ42 that can be uniquely assigned from experiment alone (typically from an atom on residue i to one on residue i+2 or closer). Although the NMR spectra is so congested that we can't uniquely assign all the cross-peaks, we can provide a lower bound on *n* for contacts from residue *i* to residue *i*+*n* that cannot be assigned from just experimental information alone. For both Aβ40 and Aβ42 about 20% of the cross-peaks are long-ranged (defined to be residue i to residue i+2 or greater) contacts, however Table 4.2 makes clear that the two peptides differ substantially in their ^{1}H-^{1}H homonuclear cross-peaks, and therefore must sample different tertiary structures. However we can't uniquely formulate a structural model of sub-populations based on the experimental ^{1}H-^{1}H homonuclear NOE data in isolation as we have shown previously for Aβ40, Aβ42, and the fragment Aβ21-30 peptides (26, 48, 65).

	Aβ40	**Aβ42**
Total NOE peaks	1108	707
Intra-residue and sequential	828 (362)	537 (185)
i to i+ 2 (assigned)	142 (20)	68(9)
i to i+ 3 (assigned)	58	36 (1)
i to i+ 4 (assigned)	40	20 (1)
i to i+ 5	11	13
i to i+ 6	7	4
i to i+ 7	10	9
i to i+ 8	2	3
i to i+9	2	6
i to i+10		1
i to i+11	2	
i to i+12	1	3
i to i+13	2	
i to i+14	2	3
i to i+20	1	2
i to i+21		1
i to i+24		1

Table 4.2. Summary of experimental NOEs determined for Aβ40 and Aβ42 independent of any computational model. Only ~25% of peaks for each peptide are assignable from experiment alone. For the remaining ~75% peaks that cannot be assigned, peak intensities may be composed of a single pair contact of which several pair contacts are possible assignments, or may be composed of multiple, fractional contact pairs. In either case we provide the experimental lower bound, *n*, for *i* to *i*+*n* contacts for the unassigned peaks. Parentheses define the number of peaks uniquely assigned to one proton pair from experimental information alone for each category.

Heteronuclear NOE. The ^{1}H-^{15}N heteronuclear NOE is based on coupling between two atoms that are covalently bonded together, so it does not depend on the protein structure, but only on the dynamics of the backbone for each residue. If the backbone is more structured, the amide bond motion will have a longer timescale and the ^{1}H-^{15}N NOE signal will be stronger. This is another measurement where the difference between the Aβ40 and Aβ42 peptides is evident. The C-terminus of the Aβ42 backbone has higher ^{1}H-^{15}N NOE intensities and slower relaxation rates than the C-terminus of Aβ40, indicating that Aβ42 is more structurally ordered in this region (66, 164).

Residual Dipolar Couplings (RDCs). Unlike the other NMR measurements described above, RDC experiments cannot be conducted in a normal aqueous environment. The RDC measurement requires that the sample be aligned along a universal axis and therefore must be conducted in an anisotropic medium such as 10% polyacrylamide gels (67). The alignment of the amide bond for each residue of the protein backbone can then be determined with respect to the protein's global alignment by measuring the RDC. RDCs have been useful in structure prediction for folded proteins, and have also been applied to unfolded proteins and IDPs (165, 166). However, the RDC may be difficult to interpret in the case of a diverse ensemble of protein conformations. The main concern is the timescale of interconversion between the different

conformational states of the IDP. Interpretation of RDCs for IDPs generally assumes that the proteins are able to realign in the medium on a faster timescale than that of the exchange between conformational states, which may not be the case. Although the RDCs for amyloid-β are non-uniform along the peptide sequence, and show different trends between Aβ40 and Aβ42, they are difficult to interpret structurally for these reasons.

Disorder classification from NMR. Together, this combination of NMR data supports the fact that there are important structural differences between the IDP structural ensembles of Aβ40 and Aβ42. However, the NMR evidence is mixed with regard to the overall ensemble characterization of disorder for the amyloid-β peptides. While the chemical shift and J-coupling data seem to indicate a highly disordered peptide with a small amount of β-strand or extended structure, the homonuclear and heteronuclear NOE data indicate a more compact and structured ensemble. The highly averaged experimental information makes it difficult to define and properly weight all conformational sub-populations of the two peptides, and hence a computational model is necessary to interpret the NMR data. However the range of NMR observables can provide an important validation suite to judge the quality of different computationally generated structural ensembles.

5. Evaluation NMR Observables from Structural Ensembles

To validate each of the five alternative ensembles against the experimental NMR data, we need a method of calculating the chemical shifts, J-coupling constants, RDCs, and ^{1}H-^{1}H NOEs as averages over the entire computationally-generated structural ensemble for comparison with experimental values. Below, we review the procedures for evaluating NMR observables (48).

Chemical shifts. General purpose chemical shift calculators such as SHIFTX (99) and SHIFTS (91) describe the isotropic shielding of the applied magnetic field for the given atom, a quantity that depends sensitively on the local electronic structure environment (91, 99, 167). Even for folded proteins with a dominant native conformer, each atom type can exist in many different local environments, and for disordered peptides and proteins the ensemble average reflects an even more diverse set of chemical environments. This makes an accurate calculation of chemical shifts quite a challenge for IDPs. For this review we used SHIFTX rather than SHIFTS to calculate chemical shifts for consistency with the ENSEMBLE program, but the results generated by the two programs are consistent when applied to amyloid-β (26, 48). To determine chemical shifts of the entire IDP ensemble, we took an average of the SHIFTX values for each individual structure, consistent with the experimental average.

J-coupling constants. To calculate the scalar coupling constants, $^{3}J_{HNH^{\alpha}}$, we used the Karplus equation (92)

$$J(\phi) = A\cos^2(\phi - 60) + B\cos(\phi - 60) + C \tag{4.1}$$

where ϕ indicates the protein backbone dihedral angle, with coefficients A = 6.51, B = −1.76, and C = 1.60 corresponding to the parameter set by Vuister and Bax (93). We note that the original experimental J-coupling data from Wang and co-workers (67) has been corrected for a missing relaxation that makes scalar couplings determined from the H_NH_{α} 3D experiment

consistently lower than those from COSY splittings by a small amount (93) (from ~1-5%). The J-coupling values were also an average over all structures in the ensemble, consistent with the experimental average.

Residual Dipolar Couplings. We used the standard method in the field for calculation of RDCs, the PALES (96) program, for each structure for the five different Aβ40 and Aβ42 ensembles. The program computes the RDC by using steric properties of the molecule to generate a global alignment orientation. Then, the angle between the backbone amide bond vectors and the external magnetic field is used to calculate the RDC for each conformation, and the RDCs are averaged over all conformations of a given ensemble. To compare with the ENSEMBLE program, we also calculated RDCs using a local alignment program developed in the Forman-Kay lab, where 15 residue segments along the protein are aligned separately over the ensemble of structures (141). The local RDCs (L-RDCs) are also averaged over all conformations of a given ensemble. This local alignment has lower computational cost and has been shown to give similar results to PALES, hence L-RDCs, rather than RDCs generated from a global alignment algorithm, are optimized in the standard implementation of the ENSEMBLE approach. We also note that the PALES alignment and RDC calculation were developed for folded proteins, and their application to IDPs assumes individual IDP conformations behave similarly to folded proteins during the RDC experiment, which may not be the case for amyloid-β.

Homonuclear NOEs. We also evaluate the ^{1}H-^{1}H NOESY spectra as in reference (48) by calculating the intensity of the NOE cross-peaks

$$I\left(t_{\text{mix}}\right) = Xe^{-\Lambda t_{\text{mix}}} X^{-1} I\left(0\right) \tag{4.2}$$

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

$$\rho_{ii} = \sum\nolimits_{j=1,\neq i}^{n} \frac{1}{10} K^2 \left[3J_{ij}\left(2\omega_0\right) + \frac{3}{2} J_{ij}\left(\omega_0\right) + \frac{1}{2} J_{ij}\left(0\right) \right] \tag{4.3a}$$

and off-diagonal elements

$$\sigma_{ij} = \frac{1}{10} K^2 \left[3J_{ij}\left(2\omega_0\right) - \frac{1}{2} J_{ij}\left(0\right) \right] \tag{4.3b}$$

that are comprised of appropriate combinations of the spectral density functions

$$J\left(\omega\right) = \sum_{i=1}^{N} \left[a_i \frac{2\tau_i}{1+\omega^2\tau_i^2} \right] \tag{4.4}$$

evaluated at the relevant Larmor frequencies, ω, and where *K* is given by

$$K = \frac{\mu_0}{4\pi r^3_{\text{eff}}} \hbar \gamma_{\text{H}}^{\ 2} \tag{4.5}$$

.

γ_H is the gyromagnetic ratio of ^{1}H, μ_0 is the permeability of free space, and $\hbar$ is Planck's constant. r_{eff} is the distance between the hydrogen atoms raised to -6 power, averaged over all structures in the ensemble and then raised to the -1/6 power to convert back to units of distance. We account for all hydrogen atoms explicitly (including all methyl or methylene groups) and hence calculate r_{eff} and correlation functions for every pair of hydrogen atoms. The spectral density function for each atom pair is calculated as the Fourier transform of the correlation function for the pair vector (48). We ignore water proton coordinates, as is the standard assumption in the NMR experiment.

Heteronuclear NOEs. Finally we calculate ^{1}H-^{15}N NOEs by evaluating the steady state NOE enhancement factor of the ^{15}N spin by the ^{1}H NOE according to

$$\varepsilon_{NOE} = 1 + \frac{\gamma_H \sigma_{HN}}{\gamma_N R_z^{(N)}} \tag{4.6}$$

where γ_H and γ_N are the gyromagnetic ratios of ^{1}H and ^{15}N, respectively. The ^{1}H-^{15}N cross-relaxation rate constant is given by

$$\sigma_{HN} = \frac{1}{10} K^2 \left[3J_{HN}\left(\omega_{0,H} + \omega_{0,N}\right) - \frac{1}{2} J_{HN}\left(\omega_{0,H} - \omega_{0,N}\right) \right] \tag{4.7}$$

and the ^{15}N self-relaxation by

$$R_z^{(N)} = \frac{1}{10} K^2 \left[\frac{3}{2} J_{HN}\left(\omega_{0,N}\right) + 3J_{HN}\left(\omega_{0,H} + \omega_{0,N}\right) + \frac{1}{2} J_{HN}\left(\omega_{0,H} - \omega_{0,N}\right) \right]. \tag{4.8}$$

In this case $J_{HN}(\omega)$ is the spectral density function for the ^{1}H-^{15}N covalently bonded pair. Note that the NOE calculations require correlation time information about the vector between each pair of atoms given by τ in Eq. (4.4).

This dynamic information is naturally supplied by the *de novo* MD method, which allows direct measurement of the autocorrelation of the inter-atomic vector over the time of the simulation. However, it is not available for the RC, Pred-SS and knowledge-based ensembles. This is an inherent limitation of ensembles generated from a static perspective only, which we discuss further below. Since time information is not available for the static ensembles, we can only evaluate the NOEs for these ensembles under the assumption of one average correlation time applied to all pairs of protons (we use an average correlation time derived from our MD simulations). Of course the *de novo* MD method can account for the timescales explicitly and more importantly for the fact that different pairs of hydrogen atoms do decay on different timescales.

6. Validation of Different Structural Ensembles with NMR

Chemical shifts. Most knowledge-based approaches for NMR structure determination of IDPs depend heavily on site-specific chemical shift values, δ. In the case of Aβ, and possibly other

IDPs that form a diverse set of secondary structure, chemical shifts do not provide strong discrimination among different structural ensembles or structural interpretations. This is a very different situation from folded proteins, where chemicals shifts are used in detailed structural refinement. Here we review why chemical shifts are not particularly useful for defining or validating the Aβ structural ensemble.

First we note that the calculated chemical shifts have an uncertainty that is independent of the quality or type of structural ensemble, and results from approximations of the SHIFTX or SHIFTS calculators themselves. Other research groups have reported the uncertainty, σ^2 (ppm), for these calculators, with the value depending on the atom type and its bonding chemistry (91, 99). Therefore the best way compare the agreement of various IDP ensembles with chemical shift data is to calculate chemical shift χ^2 values

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(\left\langle \delta_{i,\,\mathrm{calc}} \right\rangle - \delta_{i,\,\mathrm{exp}}\right)}{\sigma_i^2}\,. \tag{4.9}$$

We calculate χ^2 by taking the difference between the experimental chemical shift and the shift calculated from each of the structural ensembles and normalizing it by the calculator uncertainty. Reported uncertainties (root mean squared difference, RMSD, from experiment) for the SHIFTX calculator (99) are $\sigma^2 = 0.23$ ppm for H_α, $\sigma^2 = 0.49$ ppm for H_N, $\sigma^2 = 0.98$ ppm for C_α, and $\sigma^2 = 1.10$ ppm for C_β. Any dominant error due to the underlying structural ensemble would then correspond to values of $\chi^2 > 1$.

Table 4.3 displays the χ^2 agreement between experimentally measured proton (48) and carbon (14) chemical shifts with those generated from each structural ensemble for both Aβ40 and Aβ42. Experimental chemical shift data reported for the monomeric Aβ40 and Aβ42 peptides do not differ greatly from random coil values, and therefore the RC ensemble falls within χ^2 uncertainty. Since the Pred-SS ensemble shows no *cooperative* secondary structure (Figure 4.1a), it remains largely equivalent to the RC ensemble as deduced by chemical shifts. The *de novo* MD structural ensemble is also in good agreement with the chemical shift data, however ~99% of the MD generated Aβ conformations contain one or more elements of cooperative secondary structure somewhere along the peptide sequence (Figure 4.1b). The reason that the MD ensemble is also in good agreement with the experimental chemical shifts is that averaging over a large ensemble of cooperatively formed secondary structure and tertiary contacts yields average chemical shifts that are consistent with random coil values. For example, averaging the chemical shifts of all folded proteins in the PDB results in averages very similar to random coil values (48, 100).

Aβ40 Peptide	**Average Property**			
Ensemble Type	RDC - PALES (Hz)	RDC-Local (Hz)	H_2O NOEs	D_2O NOEs
RC	1.48	1.55	4.33 (0.57)	5.32 (0.53)
Pred-SS	1.54	1.35	3.85 (0.58)	4.43 (0.54)
ENS-Pred-SS	1.47	0.26	3.74 (0.57)	4.79 (0.48)
MD	2.22	1.88	1.15 (0.74)	3.22 (0.55)
ENS-MD	1.69	0.18	1.22 (0.70)	3.66 (0.51)
Aβ42 Peptide	**Average Property**			
Ensemble Type	RDC - PALES (Hz)	RDC-Local (Hz)	H_2O NOEs	D_2O NOEs
RC	1.33	1.42	2.27 (0.33)	1.35 (0.61)
Pred-SS	1.37	1.27	1.74 (0.35)	1.00 (0.66)
ENS-Pred-SS	1.07	0.39	2.56 (0.34)	1.00 (0.68)
MD	2.25	2.14	1.25 (0.67)	0.58 (0.80)
ENS-MD	2.13	0.33	1.51 (0.62)	0.73 (0.76)

Table 4.3. Comparison between random coil (RC), predicted secondary structure (Pred-SS), *de novo* MD (MD), and ENSEMBLE optimized Pred-SS-ENS, and MD-ENS ensembles. We report RMSDs for the RDC calculator PALES and L-RDCs evaluated with ENSEMBLE using local alignments. The NOEs are back-calculated from the structural ensembles as described in Section 4. We evaluate the RMSD normalized by the largest NOE intensity, RMSDN and (correlation coefficient, r) with the H_2O and D_2O experiments.

ENSEMBLE optimization of the Pred-SS and MD starting pools improves the χ^2 values, but all are within the calculator uncertainty. Not surprisingly, if the knowledge-based ENSEMBLE approach were biased by chemical shift data alone, they would show little deviation from their starting "soup", and the structural interpretation would be highly dependent on the starting ensemble. For this reason we conclude that NMR chemical shifts do not provide any qualitative discrimination between the alternative ensembles, at least for the Aβ40 and Aβ42 IDPs.

J-coupling constants. In contrast to chemical shifts, J-couplings provide an important way to discriminate between random or extended IDPs and those that are more collapsed with cooperative secondary structure and tertiary structure contacts. Figure 4.2 illustrates this by plotting the agreement between experimentally measured $^3J(H_N,H_\alpha)$ (67, 90), and those calculated from the RC, Pred-SS, Pred-SS-ENS, *de novo* MD, and MD-ENS ensembles for Aβ40 and Aβ42. It is evident that the MD ensembles are far superior to the RC, Pred-SS, and Pred-SS-ENS ensembles when validated against scalar couplings, yielding an RMSD across all residues of 0.60-0.99 Hz, whereas the other ensembles yield much larger RMSDs of ~1.95-2.28 Hz (Table 4.3). Therefore J-couplings provide an experimental measure for discriminating among qualitatively different structural ensembles for the amyloid peptides.

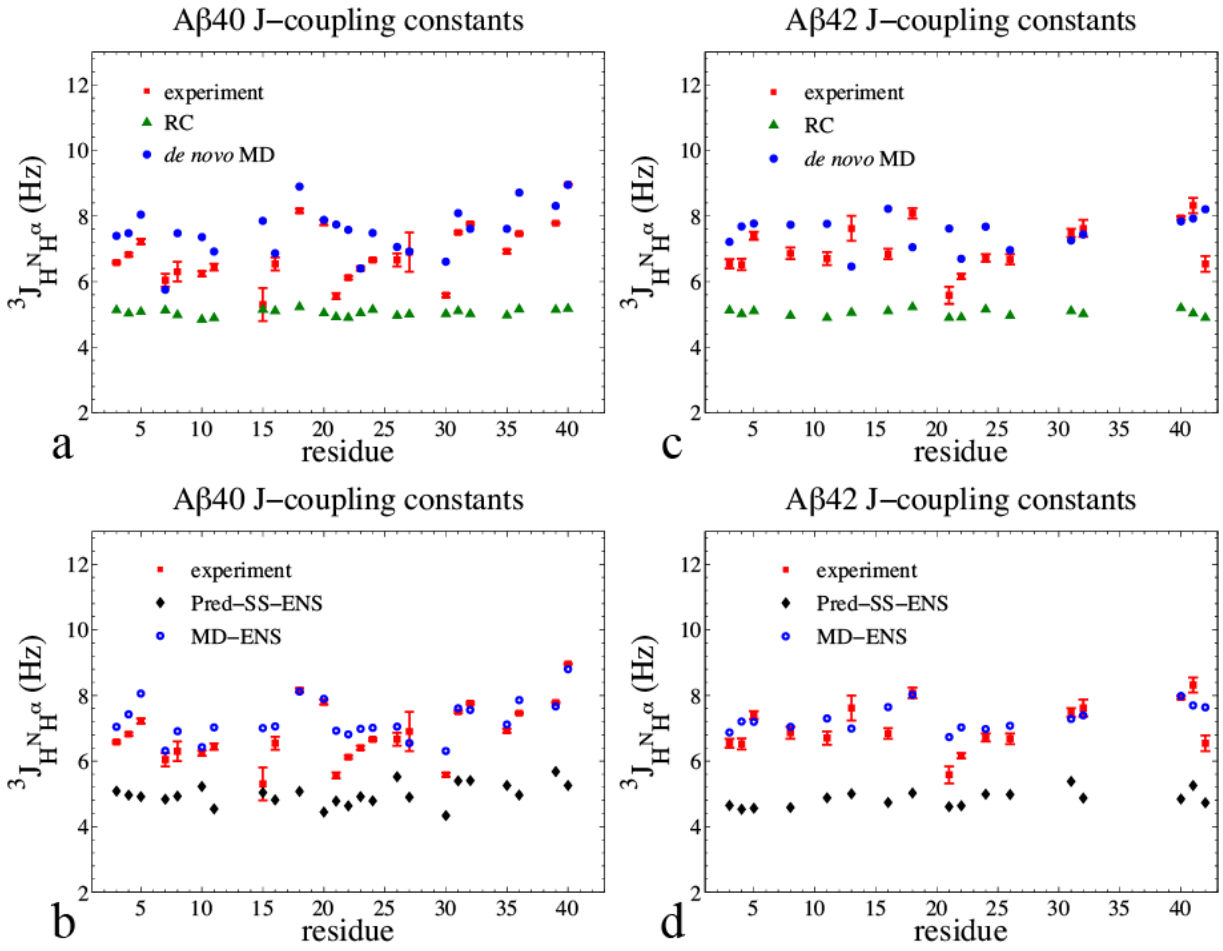


Figure 4.2. *J-coupling constants for backbone amides for Aβ40 and Aβ42.* (a) Aβ40 experimental J-coupling constants (red squares) compared to RC (green triangles) and *de novo* MD (solid blue circles). (b) Aβ40 experimental J-coupling constants (red squares) compared to Pred-SS-ENS (black diamonds) and MD-ENS (blue circles). (c) Aβ42 experimental, RC, and *de novo* MD J-coupling constants. (d) Aβ42 experimental, Pred-SS-ENS, and MD-ENS J-coupling constants. The experimental data from Yan *et al.* (67) have been corrected to account for T_{1sel} relaxation and bring J-couplings determined from a $H_N H_\alpha$ 3D experiment to be consistent with those from COSY splittings (14).

We attribute the poor performance of the RC, Pred-SS, and Pred-SS-ENS ensembles to a structural distribution that too uniformly samples the ϕ angles. This can be seen from the Karplus Eq. (4.1) using the Vuister and Bax parameters, in which a uniform ϕ average from -180° to 180° would reduce to J=A/2+C or ~4.85 Hz, consistent with the across the board value of ~5 Hz calculated for the RC, Pred-SS, and Pred-SS-ENS ensembles (Figure 4.2). An option to introduce new structures created by TraDES to the starting pool during the selection process is available in the ENSEMBLE package, however this would be unlikely to improve J-couplings because the additional structures would still not have cooperative secondary structure. By contrast the *de novo* MD ensemble yields values in a range of ~6.5-8.5 Hz that is consistent with experiment, and with a pool of conformations containing cooperative secondary structure. The MD-ENS is able to refine the J-coupling values further, lowering the RMSD from experiment (Figure 4.2, Table 4.3). Figure 4.1c emphasizes that the refined MD-ENS for Aβ42 contains largely the same distribution of cooperative secondary structure, with certain populations shifting

by 5-15%, thereby providing an independent validation of the *de novo* MD ensemble. Thus, the absolute residue by residue $^{3}J(H_N,H_\alpha)$ values, which are highly sensitive to backbone conformational preferences for the ϕ angle, provide important proof that cooperative secondary structure generated by *de novo* MD is a better structural prediction than the RC, Pred-SS, and Pred-SS-ENS ensembles.

Residual dipolar couplings. Table 4.3 provides the assessment of the five alternative ensembles for Aβ40 and Aβ42 using RDC values evaluated residue by residue using the PALES program (96) and L-RDCs based on local alignments (141). While the RC and Pred-SS ensembles yield lower RMSD values, 1.3 - 1.5 Hz, they are marginally better than the *de novo* MD RMSD of 2.2 Hz. This is in part due to the fact that experimental RDC uncertainties for IDPs are larger (~0.9 Hz for Aβ40 and ~0.5 Hz for Aβ42) than the uncertainty observed for folded proteins of ~0.1 Hz (168). In addition, there are large uncertainties in the accuracy of RDC calculators using programs such as PALES (96). In fact, the reported RMSD of the PALES calculator for folded proteins is ~2.0 Hz, on the same order as the RMSD for the *de novo* MD ensemble. While the ENSEMBLE method does significantly lower the RMSD for L-RDCs for the Pred-SS-ENS and MD-ENS ensembles, the corresponding RMSD based on the global alignment using PALES is marginally better than the Pred-SS and *de novo* MD starting pools. Hence for this particular application on disordered amyloid peptides, RDCs are not a particularly good experimental metric for differentiating among the different ensembles, and substantial disagreement between RDCs based on local and global alignments are observed.

Homonuclear NOEs. Finally, we consider the performance of the different ensemble methods for reproducing ^{1}H-^{1}H homonuclear NOE cross-peaks. We presented the NOE data collection for the Aβ42 peptide in which ~700 cross-peaks are observed in the NOE spectra, but only ~200 can be uniquely assigned from experimental information alone (Table 4.2). The remaining cross-peaks do not have a clear independent assignment (and in fact require a computational model to interpret them). Therefore in this review we only compare the different methods against the NOE cross-peaks that can be assigned by experiment alone. We note that quantitatively reproducing NOE intensities is a very high bar since peak volumes are extremely sensitive to r^{-6} distance averaging. Geometric imperfections in the conformational ensemble where contact differences differ by a factor of $2^{1/6}$ (difference between 1 Å and 1.12 Å) will double the corresponding intensity value, thereby driving up the RMSD error for all ensembles. Large absolute NOE intensities especially tend to dominate the RMSD error; we mitigate this effect by normalizing the RMSD (RMSDN) by the experimental intensity for each NOE.

Table 4.3 shows that the predicted set of ^{1}H-^{1}H NOEs from *de novo* MD is better than any other ensemble, with RMSDNs that are significantly lower by 2-3 fold and with much higher correlation coefficients. The NOE validation clearly indicates that the *de novo* MD ensemble with its cooperative secondary structure is a better representation of Aβ40 and Aβ42 than are the RC, Pred-SS, or Pred-SS-ENS ensembles, which have no cooperative secondary structure. As discussed in Section 5, the statistically generated and knowledge-based ensembles agree relatively poorly with the NOE observables since no correlation times are known, and hence the MD-ENS ensemble is in somewhat worse agreement with the experimental NOEs than the *de novo* MD ensemble (Table 4.3). The NOE validation emphasizes that an IDP's diverse set of conformations gives rise to a heterogeneous set of correlation times that must be described in order to validate against experimental NOEs.

Heteronuclear NOEs. We further emphasize that the calculation of heteronuclear NOEs, being a purely dynamical measurement, is only possible with the *de novo* MD method. Figure 4.3 shows a comparison of the experimental ^{1}H-^{15}N NOE intensities, measured by Yan and Wang (66), and those derived from our MD simulation for Aβ42 and Aβ40, showing overall excellent agreement. Unlike the ^{1}H-^{1}H NOEs, these assignments are unambiguous. We find that, as in the experiment, there is an increase in ^{1}H-^{15}N NOE intensities calculated from simulation for residues 35-40 for Aβ42 compared to Aβ40, indicating that the longer peptide experiences slower dynamics at the C-terminus. This difference in experimental ^{1}H-^{15}N NOEs for Aβ42 and Aβ40 has previously been interpreted as evidence that Aβ42 has greater structural rigidity in the C-terminus compared to Aβ40 (66, 164).

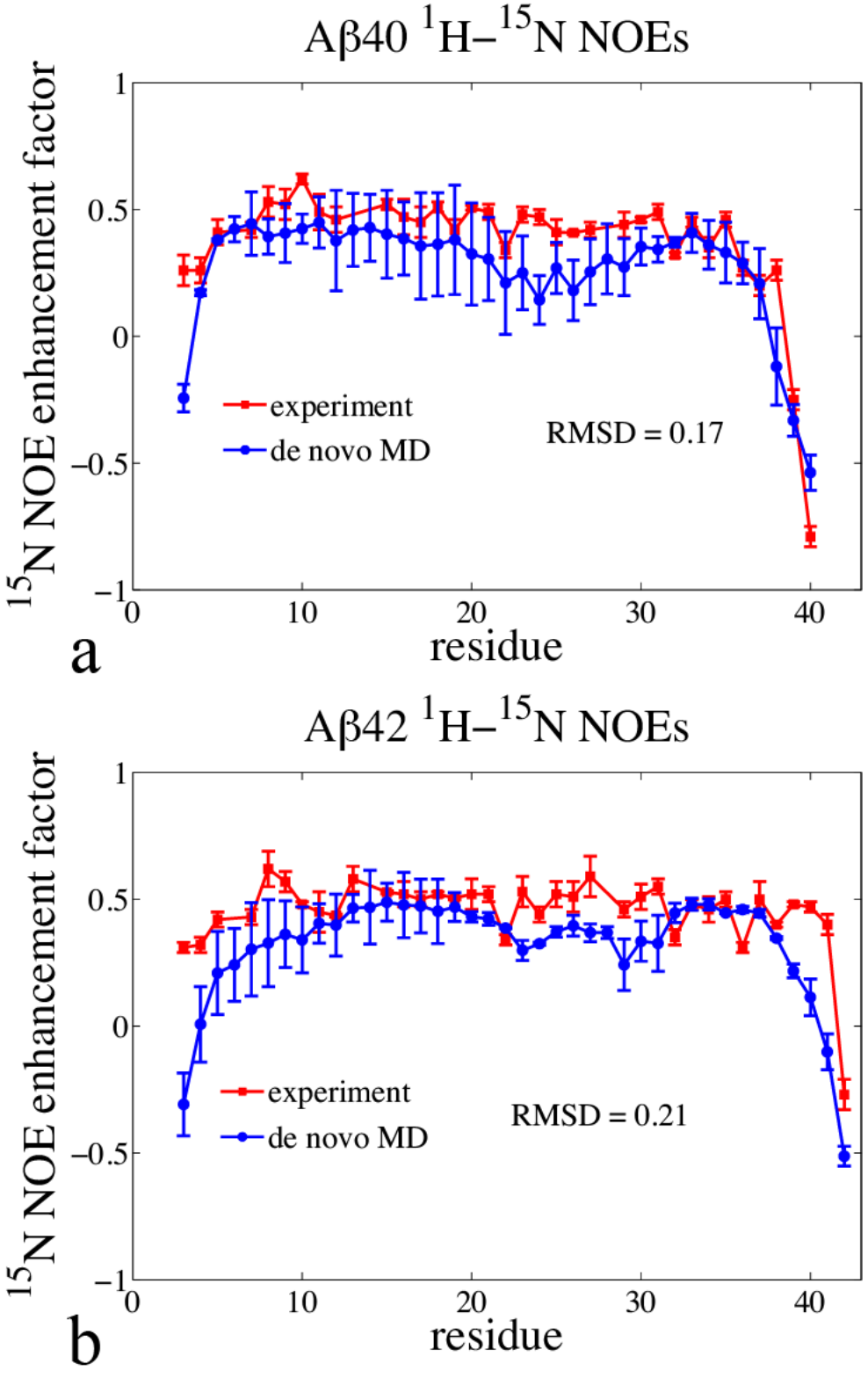


Figure 4.3. *Agreement with experiment of simulated (a) Aβ40 and (b) Aβ42 ^{1}H-^{15}N NOE.* The red squares are experimental data from Yan and Wang (66). The blue circles are the data calculated from the *de novo* MD ensemble.

7. The Structural Differences Between IDPs Aβ42 and Aβ40

Based on the MD and MD-ENS structural ensembles, which are best validated against the NMR data, we find that both peptides have a diverse set of secondary structure elements including turns, helices, and anti-parallel and parallel β-strands. However, the most significant difference in the structural ensembles of the two IDPs is the type of β-hairpins and β-strands they populate. The analysis of the sub-populations of structure show that Aβ42 forms a major population of an anti-parallel β-hairpin involving residues 16-21 with residues 29-36. These particular β-strands are the same β-strands adopted by the peptide in the insoluble amyloid fibril state. By contrast, Aβ40 forms an alternative anti-parallel β-hairpin between residues 9-13 and 16-21 that is incompatible with the amyloid fibril β-strands. The fact that Aβ42 fibrillizes more readily than Aβ40 thus could arise from the differences observed in β-strand sub-populations between the two IDPs in their monomeric forms (65). A more detailed analysis of the amyloid IDPs can be found in a companion publication (65).

8. Summary and Future Directions

We have shown that the MD and MD-ENS structural ensembles for the IDPs Aβ40 and Aβ42 yield substantially better agreement with a range of NMR data than the random coil or statistical ensembles that are typically used with knowledge based approaches, even after refinement with all available NMR data. The MD ensembles are qualitatively different than random coil or statistical ensembles in that the sub-populations are richly structured, contain a diverse set of secondary structures including α-helix, β-turns, and β-strands, and span the full range of compact to fully extended conformations. Furthermore, while MD generated ensembles are Boltzmann weighted, the knowledge-based approaches give equal statistical weight to all conformations and thus are likely inconsistent with statistical mechanical weightings that are inherent to the NMR experiment.

We have shown that some types of NMR data may not be helpful for discriminating among qualitatively different structural ensemble of IDPs. In particular, averages over a diverse set of cooperative secondary structure conformations yield experimental values of chemical shifts that are superficially consistent with values expected from a random coil ensemble. Furthermore, if the chemical shifts are not highly dispersed along the sequence of a particular IDP, such as is found for the amyloid-β peptides, then the chemical shifts have limited value as experimental input or as a validation measure. J-couplings do provide some discrimination between randomly generated conformations and a diverse population of cooperative secondary structure. In fact J-couplings were a decisive factor in determining that the RC and Pred-SS ensembles are not well-validated models of amyloid-β structure.

Unlike others who have used RDC data to help interpret IDP or unfolded protein structural ensembles, we found RDCs to be only marginally useful for Aβ40 and Aβ42. This may be due to limitations of RDC calculators such as PALES (96), which were originally developed and successfully applied to folded proteins, but which are reported to have large uncertainties in their predicted RDC values. Furthermore, the agreement between global alignment algorithms such as PALES (96) diverge significantly from RDCs evaluated from localized alignments (141) for Aβ40 and Aβ42, indicating that in cases like this the ENSEMBLE package should be

employed using the PALES calculator to fit RDCs, which is possible though not standard (169). More research may be necessary to apply programs like PALES to disordered proteins, which likely do not align in an anisotropic medium in the same way as folded proteins, in part due to the timescale of interconversion of the conformational sub-states. For example, conventional methods for calculating RDCs cannot be applied to the motion of multi-domain biomolecules (170), and the local conformational sampling and long-range structure need to be simultaneously accounted for because they both affect the experimental RDC data (50). However, progress is being made in using RDCs to provide meaningful structural information for other IDPs (50, 56, 61, 140, 170, 171). We speculate that success is greatest when all sub-populations of the IDP ensemble are homogeneously classifiable (as extended disordered for example), so that assumptions about the global alignment properties of the IDP can be made.

We have demonstrated that homonuclear ^{1}H-^{1}H NOE intensities and heteronuclear ^{15}N-^{1}H NOEs are by themselves discriminating with regard to the tertiary contacts and backbone dynamics, respectively, that define the important structural differences between the two Aβ peptides. Nonetheless, a correct picture of the IDP ensemble based on the experimental NOE data would not be possible without a computational model providing details of individual structures. In turn, although the homonuclear NOEs are averaged over all sub-populations, they are still vital for deducing whether a given ensemble contains sub-populations of structure with the right tertiary contacts to give rise to the observed cross-peaks in the spectra. Because these cross-peak intensities rely directly on the decay timescales of correlated proton distances, the NOEs for IDPs are reporting on a heterogeneous population of timescales. One of the primary limitations of the statically generated ensembles is that they are not associated with any information about motional timescales that can be used to calculate NOE observables. Relaxation times can be used with the ENSEMBLE method, although they are incorporated as structural rather than dynamic constraints (54, 58, 148). This is a genuine strength of the *de novo* MD methods, especially for ^{15}N-^{1}H relaxation measurements, which are incompatible with validation of the static ensembles.

We believe that the primary limitation of knowledge-based methods applied to the difficult amyloid-β case is three-fold. We note that while there will be quantitative differences between ENSEMBLE and other knowledge-based approaches such as ASTEROIDS, qualitatively the problems will be similar. First, there is no requirement for generating a complete and representative starting "basis set" of conformations to select the final ensemble from; i.e. these methods cannot use the NMR data effectively to select for compact structures with elements of cooperative secondary structure if the initial pool of structures is largely composed of extended random coil structures. Both ENSEMBLE and ASTEROIDS have relied on statistical coil ensembles as the starting pool of structures, and while some "on-the-fly" addition of new structures is possible with these methods, they do not yet support formation of complicated β-strand motifs (50, 58). Metrics of ensemble heterogeneity, such as those developed by the Onuchic and Stultz research groups, will continue to be useful as we explore the range of IDPs that cannot be easily classified based on their level of disorder (118, 172, 173). Second, for certain classes of IDPs such as amyloid-β, optimization of structures to reproduce chemical shifts does not discriminate among qualitatively different structural ensembles. Third, the optimization phase of the knowledge-based approaches relies on approximations to NMR observables, which may diverge from a global property, as for L-RDCs, or from the dynamical

origins of NOE intensities. At the same time, the *de novo* MD method is not quantitatively perfect, and therefore the MD ensemble provides an excellent start state for subsequent refinement by knowledge-based approaches. An unambiguous future direction for the structural biology of IDPs is the combined use of knowledge-based approaches and MD that supplies Boltzmann weighted conformational sub-states as well as heterogeneous timescales of motion.

Currently, the application of truly *de novo* MD techniques to generate IDP conformational ensemble pushes these methods to the edge of their capabilities. Because there are many degrees of freedom to sample in a protein conformational landscape, MD simulation methods rely on computationally cheap, but potentially inaccurate, theoretical models. In particular, a number of research groups are currently exploring whether the use of empirical fixed-charge force fields and advanced sampling techniques such as replica exchange are sufficient to provide an accurate and complete ensemble of structures (26, 31, 86-88, 174-177).

The predictive quality of biomolecular simulations depends on the accuracy of the empirical potential energy functions, which were first developed more than two decades ago for folded proteins (178-180). While much progress has been made on adapting the protein force fields to better describe energetics and structure, they continue to be co-parameterized with rather antiquated aqueous solvent models such as TIP3P or SPC that are deficient in modeling bulk water liquid. Because IDPs are generally more solvent-exposed than their folded globular counterparts (181, 182), we have argued that these default water models limit the predictive capability of *de novo* MD, especially when applied to IDPs. We have demonstrated this deficiency on the Aβ21-30 fragment (which is well-classified as an extended random coil (48)), showing that the TIP3P model poorly predicts the NOE data (26). Use of the TIP4P-Ew model – a superior theoretical model for bulk water – greatly improves the solvation description of Aβ21-30 so that *de novo* MD gives excellent agreement with the experimental NOE cross-peaks and relaxation times (26). This has motivated the use of TIP4P-Ew in other NMR simulation applications (48, 90). Overall, a number of affiliated research groups have great interest in improving the theoretical models for molecular simulation, including full reparameterization of the AMBER protein force fields with better descriptions of water solvent (88, 174), better modeling of short-ranged anisotropic interactions (which we believe will especially improve agreement with NOE data) through advanced multipole electrostatic models (183), and eventual movement from fixed charge to polarizable force fields (184-191).

Advanced sampling techniques allow faster sampling of conformational ensembles by artificially reducing free energy barriers between IDP conformational states. A Monte Carlo or reweighting step on the resulting ensemble states ensures a Boltzmann distribution. These techniques offer significant reductions in both simulation time and compute power needed for extensive sampling when compared to a single long MD trajectory (138, 192-194). Our group and others have used parallel tempering replica exchange (RE) to increase sampling (26, 90, 195). However, the MC or reweighting steps often require a significant amount of time and parallel computing resources so that further improvements are still necessary. For example, we have developed a modification of the RE technique, which we term multi-reservoir replica exchange (MRRE), which reduces computational cost and total simulation time compared to standard replica exchange, reaching the same degree of ensemble convergence (31, 48). A very different approach integrates out irrelevant or equilibrated degrees of freedom to devise a coarse-grained model that is easier to sample. For example, Terakawa and Takada used RE to characterize

fragments of an disordered p-53 N-terminal domain sequence with an all-atom model (177). These ensembles were used to construct a coarse grained potential using an iterative Boltzmann inversion procedure. Coarse grained simulations then sampled the ensemble of the full-length IDP, which was converted back to an all-atom representation and used to calculate RDC and SAXS data.

Advances in harnessing extensive computational resources to perform MD simulations have also improved ability to simulate IDPs, and increased automation in allocating processors for and monitoring these large simulations could further reduce the time and resources needed to produce complete conformational ensembles. Many of us in the field of computational chemistry have benefitted from the rapid evolution of multicore processors and GPUs, along with emerging GPU MD accelerators (196, 197). Impressive work by Pande and coworkers has increased their computational capacity through the distributed computing platform Folding (198). This platform takes advantage of the idle computing power of thousands of personal computers by enlisting them to run short MD simulations. Combining many of these short simulations has resulted in the accumulation of large total simulation times (> 200 μs) for sampling the ensembles of intrinsically disordered Aβ peptides (199). D. E. Shaw and coworkers have built a special-purpose MD simulator, Anton, staying well ahead of the Moore's Law curve of expected improvements in standard processor performance, so that today it performs MD calculations that beat current more conventional state of the art hardware by 2 orders of magnitude (200). Using this advanced computing power they are able to simulate the unfolded state of a ~85 residue protein for one continuous 200 μs simulation, achieving extensive sampling of the structural ensemble (176). Long-range contacts seen in this MD simulation show good correspondence with PRE NMR data, and there is also reasonable agreement between calculated and experimental relaxation rates and radius of gyration estimates.

All together, the productive interplay between NMR experiments, *de novo* MD simulations, and knowledge-based approaches, along with supporting models, algorithms, and computer hardware, gives us an ability to accurately identify structures present in IDP ensembles and use that knowledge to gain previously inaccessible functional insights (48, 51, 53, 55, 59, 60, 90, 177, 201-211). To further improve techniques for studying disordered proteins, we as a community could establish a high throughput computational infrastructure to predict IDP structural ensembles using a combination of MD and NMR as appropriate experimental restraints as per knowledge-based approaches that do not violate Boltzmann weighting. This would be similar to the establishment of X-ray crystallography beamlines for the rapid solution of folded protein structures that was launched during the structural genomics era. The ultimate goal in both cases is to use structural information to drive the formation of hypotheses about protein function. Based on the success of using structural information for functional characterization of folded proteins and complexes, we hope and expect that structural knowledge of IDP ensembles can provide similar insight into IDP function and enable development of molecular hypotheses for disease IDPs.

APPENDIX

Details of generating the different structural ensembles

RC ensemble. To generate a RC or Pred-SS structural ensemble we used the TraDES software package (212, 213). TraDES generates a random coil (RC) ensemble of structures by building each conformation one residue at a time and picking the dihedral angles according to their probability from the Ramachandran plot. This procedure also avoids steric clashes, but does not attempt to distinguish between energetically favorable and unfavorable conformations.

Pred-SS ensemble. There is also an option to bias the TraDES ensemble to preferentially sample α-helix or β-strand regions along the sequence according to a bioinformatics prediction of secondary structure, which we label the Pred-SS ensemble. Many bioinformatics tools allow prediction of a protein's secondary structure based on its sequence alone by comparing the sequence to known structures in the PDB. We employ the Psi-Pred V3.0 (214, 215) server, for this purpose, feeding it the Aβ40 and Aβ42 sequences and receiving a prediction of either extended, helical, or coil structure for each residue, as shown in Table 4.4. For both peptides, PsiPred predicts blocks of extended, β-structure with high confidence, consistent with the structure that Aβ adopts in the aggregated fibril state (21). This does not guarantee that the resulting conformations will contain true cooperative secondary structure (such as β-strand), since the secondary structure state of each residue is picked independently of other residues by TraDES. The RC and Pred-SS ensembles each contain 100,000 structures as recommended in (58).

Pred-SS-ENS ensemble. As an example of a knowledge-based approach we consider the ENSEMBLE software package, which selects from a large starting pool (basis set) of structures, typically generated by TraDES, a subset of 100 conformations that best conform to various NMR experimental data supplied to it (see Section 5 on calculation of experimental observables from structures). The Pred-SS-ENS, was selected by the ENSEMBLE program from a starting 'soup' consisting of the Pred-SS ensemble structures. We supplied H_α, H_N, C_α, and C_β chemical shifts, J-coupling constants, RDCs and NOEs for both Aβ40 and Aβ42. We used default values of the ENSEMBLE program input parameters and the default output of a 100-structure ensemble. We ran each ENSEMBLE optimization for 48 hours on a Cray XE6 at National Energy Research Scientific Computing Center (NERSC), during which ~500 rounds of ENSEMBLE optimization steps were completed. We found that the resulting Pred-SS-ENS ensemble satisfied all of the experimental criteria besides RDCs according to the ENSEMBLE software. Based on our subsequent analysis, the default convergence criteria for RDCs appear to be too strict. We therefore selected those ensembles with the best fit to the NMR data overall, which we found to be sufficient for quantitative analysis.

MD ensemble. We created the fourth ensemble with *de novo* molecular dynamics simulations of Aβ40 and Aβ42 using the Amber ff99SB force field (23) and aqueous solvent represented by the TIP4P-Ew water model (25), which we chose because previous studies support its clear superiority relative to other biomolecular simulation force fields (24, 26, 48, 90). We simulated each amyloid-β peptide in a cubic box containing 6,251 water molecules for Aβ42 and 6,136 water molecules for Aβ40, with three Na+ ions to neutralize the charge of the peptide. The

sander module of AMBER (111) was used in conjunction with Multi-Reservoir Replica Exchange (MRRE) method (31) to generate ~2 μs of MD trajectories, from which we created a Boltzmann weighted ensemble of 72,632 - 89,469 structures each for Aβ40 and Aβ42 respectively at the experimental temperature of 287 K.

MD-ENS ensemble. After reaching the conclusion that no true cooperative secondary structure can be generated from the purely knowledge-based method that selects from a pool of random structures, we decided to construct a fifth ensemble that uses a knowledge-based approach combined with MD. For this ensemble (MD-ENS), we employed the same ENSEMBLE procedure described above for the Pred-SS-ENS ensemble except that we used the *de novo* MD ensemble as the starting pool of structures from which the experimentally optimized 100-structure ensemble was selected. NOE data was not included in our final MD ENSEMBLE refinement because a more structured ensemble could not satisfy the large number of distance restraints. We again had the same problem satisfying RDC convergence criteria, but otherwise the MD-ENS ensemble was optimized well to the NMR data after ~500 ENSEMBLE rounds.

Details of using the ENSEMBLE software

For knowledge-based approach we used the ENSEMBLE software package (58) to select a final ensemble of 100 structures from a starting pool of structures. The ENSEMBLE method formulates energy functions that score structures favorably when they agree with an experimental observable and unfavorably when they do not. To generate the Pred-SS-ENS and MD-ENS ensembles we performed two sets of ENSEMBLE optimization. The first used the Pred-SS ensemble as the starting 'soup' of structures while the second used our *de novo* MD ensemble. In our first attempt to select a final ensemble that agreed with experiment we supplied multiple atom type chemical shifts, J-coupling constants, RDCs, and NOE contacts that we could assign directly from the experimental data for both Aβ40 and Aβ42. For the NOEs, we did not have specific distance restraints for the contacts, so we set the distance to a maximum of 7.0 Å for each contact. We used default values of the ENSEMBLE program for the experimental observable target energies. As recommended (58), we set chemical shifts and NOE distances to converge first, before converging J-coupling constants or RDCs. We ran each ENSEMBLE optimization for 48 hours on a Cray XE6 at NERSC, during which ~500 rounds of ENSEMBLE optimization steps were completed.

After this first ENSEMBLE attempt we saw that while the both chemical shifts and NOE distance restraints converged for the Pred-SS-ENS ensemble, in the MD-ENS calculation the NOE distance restraints did not converge. For the MD-ENS optimization the NOE distance restraint energies were extremely high (~800 in the best ensemble, compared with the target energy of 49). This was probably due to the approximation of the NOE cross-peaks as simple distance restrains, the large number of NOE distance restraints being optimized (177 for Aβ42 and 340 for Aβ40), and the fact that the MD starting pool was more diverse than the Pred-SS pool. We then attempted a second MD-ENS ENSEMBLE run without the NOE distance restraints, and this time the chemical shifts did converge.

The second stage of the ENSEMBLE runs optimized against J-coupling constants and RDCs. This round was able to successfully converged the J-coupling constants to the default energy tolerance, but not the RDCs. Although the RDC energies had improved (from ~80 to 51

on energy units), their convergence had stagnated so as to not meet the ENSEMBLE convergence tolerance of >0.2 Hz, keeping a constant energy of ~51, even with additional computation time of ~5000 ENSEMBLE optimization rounds. In actuality, the RDCs were sufficiently converged to permit the analysis in sections 5 and 6, suggesting that the default RDC tolerance was too tight. We therefore selected the final ensembles of 100 structures with the lowest energy according to the ENSEMBLE weighting function, in order to give the overall best performance against the chemical shift, J-coupling and RDC data. We then used these final ensembles as the Pred-SS-ENS and MD-ENS ensembles in our subsequent analyses.

	Aβ40				Aβ42			
Residue	structure prediction	prediction confidence	TraDES % extended	TraDES % coil	structure prediction	prediction confidence	TraDES % extended	TraDES % coil
1	C	9	0	100	C	9	0	100
2	C	4	0	100	C	4	0	100
3	C	2	0	100	C	2	0	100
4	C	1	0	100	C	1	0	100
5	C	1	0	100	C	1	0	100
6	C	1	0	100	C	0	0	100
7	C	4	0	100	C	4	0	100
8	C	6	0	100	C	6	0	100
9	C	5	0	100	C	5	0	100
10	E	0	0	100	E	0	0	100
11	E	5	50	50	E	5	50	50
12	E	7	70	30	E	7	70	30
13	E	7	70	30	E	8	80	20
14	E	6	60	40	E	6	60	40
15	E	4	40	60	E	5	50	50
16	E	8	80	20	E	8	80	20
17	E	9	90	10	E	9	90	10
18	E	8	80	20	E	8	80	20
19	E	8	80	20	E	8	80	20
20	E	7	70	30	E	7	70	30
21	E	5	50	50	E	5	50	50
22	E	0	0	100	E	0	0	100
23	C	1	0	100	C	1	0	100
24	C	2	0	100	C	2	0	100
25	C	7	0	100	C	7	0	100
26	C	8	0	100	C	8	0	100
27	C	8	0	100	C	8	0	100
28	C	4	0	100	C	4	0	100
29	C	2	0	100	C	2	0	100
30	E	1	10	90	E	1	10	90
31	E	8	80	20	E	9	90	10
32	E	9	90	10	E	9	90	10
33	E	9	90	10	E	9	90	10
34	E	9	90	10	E	9	90	10
35	E	7	70	30	E	7	70	30
36	E	5	50	50	E	6	60	40
37	E	0	0	100	E	0	0	100
38	E	0	0	100	E	0	0	100
39	E	2	20	80	E	8	80	20
40	C	9	0	100	E	9	90	10
41					E	4	40	60
42					C	9	0	100

Table 4.4. Aβ40 and Aβ42 predicted secondary structure. For Aβ40 and Aβ42 the predicted secondary structure and confidence in that prediction are presented (214, 215). This corresponds to a TraDES ensemble generated with extended dihedral angles for those residues predicted to have extended structure. The percent of the TraDES ensemble that has that residue extended corresponds to the confidence of the prediction, and the rest of the TraDES ensemble is generated with random coil structure.

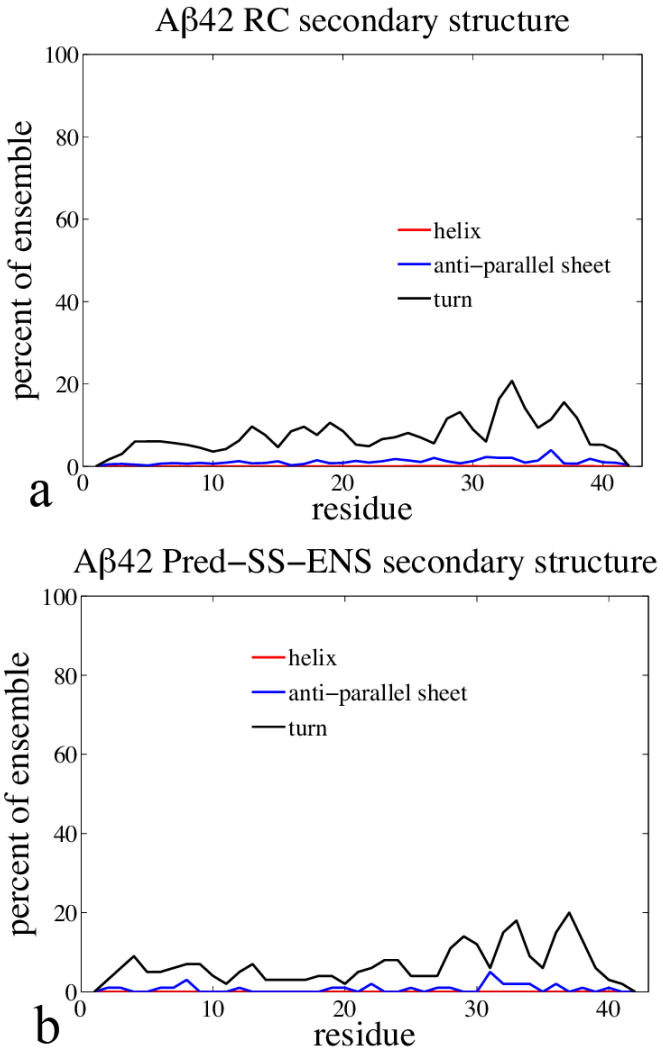


Figure 4.4. *Percentage of Aβ42 simulated ensemble in different types of secondary structure by residue for (a) the RC, and (b) Pred-SS-ENS ensembles.* The red line represents helix, the blue line for anti-parallel strand, and the black line for β-turns.

Chapter 5

Differences in β-strand populations of monomeric Aβ40 and Aβ42 based on molecular dynamics simulations refined using NMR data

Using homonuclear ^{1}H NOESY spectra, with chemical shifts, $^{3}J_{H^{N}H^{\alpha}}$ scalar couplings, residual dipolar couplings, and ^{1}H-^{15}N NOEs, we have optimized and validated the conformational ensembles of the amyloid-β 1-40 (Aβ40) and amyloid-β 1-42 (Aβ42) peptides generated by molecular dynamics simulations. We find that both peptides have a diverse set of secondary structure elements including turns, helices, and anti-parallel and parallel β-strands. The most significant difference in the structural ensembles of the two peptides is the type of β-hairpins and β-strands they populate. We find that Aβ42 forms a major anti-parallel β-hairpin involving the central hydrophobic cluster (CHC) residues (16-21) with residues 29-36, compatible with known amyloid fibril forming regions, while Aβ40 forms an alternative but less populated anti-parallel β-hairpin between the CHC and residues 9-13, that sometimes forms a β-sheet by association with residues 35-37. Furthermore, we show that the two additional C-terminal residues of Aβ42, in particular Ile41, directly control the differences in the β-strand content found between the Aβ40 and Aβ42 structural ensembles. Integrating the experimental and theoretical evidence accumulated over the last decade, it is now possible to present monomeric structural ensembles of Aβ40 and Aβ42 consistent with available information that produce a plausible molecular basis for why Aβ42 exhibits greater fibrillization rates than Aβ40.[5]

Introduction

Alzheimer's Disease (AD) is characterized by insoluble fibrils and plaques in the extra-cellular space within the brain that are largely composed of the two cleaved products of the amyloid

[5] Reproduced from (65) with permission.

precursor protein (216), amyloid-β 1-40 (Aβ40) and amyloid-β 1-42 (Aβ42) (113, 114). Although these two peptides differ only in Aβ42 having two additional hydrophobic residues at its C-terminus, Ile41 and Ala42, Aβ42 has been shown to be more significant in disease development. Aβ42 is more prevalent in the insoluble aggregates and causes more extensive damage to neuronal cell cultures than Aβ40 (4-8, 49, 126), and Aβ42 aggregates and fibrillizes much more quickly *in vitro* than Aβ40 (9, 10, 71, 73), demonstrating that the addition of these two C-terminal residues has a significant effect on the physiological and biophysical behavior of the two peptides.

Monomeric forms of Aβ40 and Aβ42 have been classified as intrinsically disordered peptides (IDPs), meaning that they populate a diverse set of conformational states as opposed to a single dominant folded conformation (14, 15, 64, 71, 73-75). However, when part of the ordered fibril state, both peptides adopt highly similar morphologies, with β-strands running orthogonal to the fibril axis, which organize further into intermolecular β-sheets that can extend to microns in length (21, 40, 103, 109, 217, 218). Since Aβ40 and Aβ42 adopt similar structures when part of the fibril, differences in the monomeric conformational ensembles could provide a starting point for understanding the greater predisposition of the Aβ42 peptide for faster fibrillization, aberrant oligomerization, or disease outcomes compared to Aβ40. In particular, does the addition of the two hydrophobic residues Ile41 and Ala42 produce any changes in the monomeric conformational ensemble for Aβ42 with respect to Aβ40?

We have collected ^{1}H NOESY spectra for the Aβ40 and Aβ42 monomers that in fact show differences in their structural ensembles, which are not evident from previous circular dichroism (CD) spectra or NMR chemical shift and J-coupling experiments (14, 67, 116, 119). While the NOESY data yield important differences in residue contacts observed for Aβ40 vs. Aβ42, these NMR experiments can only provide an ensemble-averaged picture of the tertiary contacts that occur, and not in what combinations they are present in specific, significantly populated conformers. As we have shown previously in our comparison study of Aβ21-30 (26) and Aβ42 (48), and more recently in a review of different computational approaches for generating IDP conformational ensembles (18, 64, 130-135), *de novo* molecular dynamics (MD) simulations provide one of the best approaches for most reliably characterizing the structural ensembles sampled by Aβ42 and Aβ40 as monomers.

We find that the MD simulation data, further refined with the ENSEMBLE method (52, 54, 63, 141)) and validated against a range of experimental NMR data including ^{1}H NOEs, show that both peptides have a diverse set of secondary structure elements including turns, helices, and parallel and anti-parallel b-strands in ~99% of the ensemble conformers. However, the most significant difference in the structural ensembles of the two peptides is the type of β structures they populate. We find that Aβ42 forms a major anti-parallel β-hairpin involving the central hydrophobic cluster (CHC) residues 16-21 and residues 29-36, typically forming with a turn at residues 26-27 which is rarely present in the Aβ40 ensemble (48). This dominant sub-population is consistent with the β-strands and β-turn that form an intermolecular β-sheet steric zipper (43, 219) in models of the Aβ40 and Aβ42 fibril structures based on solid state NMR (103). Instead, Aβ40 forms an alternative but less populated anti-parallel β-hairpin between the CHC and

residues 9-13, that sometimes forms a β-sheet when the CHC associates with a third β-strand comprising residues 35-37.

We find that the two additional C-terminal hydrophobic residues of Aβ42 sharply increases the hydrophobic clustering between residues 39-40 and 31-36 for Aβ42 compared to Aβ40, and when Ile41 is included the number of structures with hydrophobic contacts with 31-36, it increases to a decisive majority of the Aβ42 ensemble. This hydrophobic clustering is directly responsible for the differences in the populations of secondary structure, and β-strand content in particular, of the two amyloid peptides. When our new experimental NOEs and simulated ensemble results are placed in the context of experimental and theoretical evidence accumulated over the last decade (20, 66, 116, 119, 125, 164, 220-224), we believe that a good consensus has been reached on the monomeric structural ensembles of amyloid-β and the differences that exist between Aβ40 and Aβ42. Based on this consensus, the underlying structural differences between the two Aβ monomeric ensembles are in themselves sufficient to provide clear and testable hypotheses for why the nucleation step for fibrillization may be more difficult for Aβ40 compared to Aβ42.

Methods

NMR Experiments

The Aβ40 and Aβ42 monomers were purchased and prepared according to the protocol in (48). A more detailed description is provided in a previous publication (50, 64, 128, 139) and extensively in the Appendix. In this work we also corrected the experimental $^{3}J_{H^{N}H^{\alpha}}$ values by Wang *et al.* using the method described by Vuister and Bax (93), also described in the Appendix. This correction accounts for the effect of selective T1 relaxation so that the resulting J-coupling constants are comparable to those measured more accurately from COSY splittings.

Molecular Dynamics Simulations

We computed equilibrium ensembles of Aβ40 and Aβ42 peptide conformations at 287 K and 311 K using Multi-Reservoir Replica Exchange (MRRE) (31) and AMBER 11 (111, 129, 144). The peptides were modeled with the Amber ff99SB force field (23, 24) and solvated with TIP4P-Ew water (25), which is currently the best force field combination for reproducing NMR observables of flexible peptides (90) and ^{1}H-^{1}H NOE data (26, 48). Two independent MRRE simulations for each peptide generated final ensembles of 70,000 - 90,000 structures pulled from 0.1 μs of replica exchange simulation time at each temperature. We also performed 100 separate 20 ns microcanonical ensemble simulations for each peptide in order to calculate time-correlation data. Further details can be found in the Appendix to Chapter 3.

ENSEMBLE Refinements

We also consider the ENSEMBLE software package (52, 54, 63, 141), which selects from a large starting pool or "basis set" of structures a subset of conformations that best conform to various NMR experimental data supplied to it. We performed an ENSEMBLE optimization using the *de novo* MD ensemble as the starting 'soup' of structures, and we supplied chemical

shifts, J-coupling constants, and RDCs for both Aβ40 and Aβ42. We used default values of the ENSEMBLE program for the experimental observable target energies and selected ensembles of 100 structures from the soup to best match the NMR data, combining the 20 best ensembles for a total of 2,000 structures in the final MD-ENS ensembles. Further details are given in the Appendix.

Calculation of NMR Observables

We back-calculated chemical shifts, J-coupling constants, Residual Dipolar Couplings (RDCs) based on local (141) and global alignments (96), and ^{1}H-^{1}H NOEs from our 287 K Aβ40 and Aβ42 refined ensembles with the same procedure used for Aβ42 (48) and for IDPs (51, 53, 55-57, 64). All details of how the procedure differed in the case of Aβ40 are available in the Appendix. Finally, we calculated ^{1}H-^{15}N NOEs for the Aβ40 and Aβ42 backbone N-H atoms from the 287 K refined ensembles and MD correlation times. We use the same method as for the ^{1}H-^{1}H NOEs (described in the Appendix) to calculate the spectral density function for each pair of H-N atoms from the short NVE simulations. We then calculate the steady state NOE enhancement factor of the ^{15}N spin by the ^{1}H NOE from our structural ensemble and dynamical trajectories as described elsewhere (64, 153), and which we recapitulate in the Appendix.

Results

Summary of Experimental NMR Data for Aβ40 and Aβ42

We find that the hydrogen and carbon chemical shifts for both Aβ40 and Aβ42 do not differ significantly from random coil values (Figure 5.7)[6], and based on analysis of chemical shifts using webserver http://www-vendruscolo.ch.cam.ac.uk/d2D/ (160, 161), both peptides have significant β-strand content if backbone nitrogen chemical shifts are considered (163). However if nitrogen chemical shifts (which have larger experimental uncertainty compared to hydrogen and carbon chemical shifts) are not included in the webserver calculation, then the absolute probabilities of β-strand structure were found to be very low for both peptides.

J-coupling values also provide no strong evidence of structural differences between the Aβ peptides (Figure 5.8), i.e. there are no secondary structure "blocks" at different points in the peptide sequence that would be consistent with a dominant population of α-helical or β-strand secondary structure (as in a folded protein) that is different between the two peptides. However, the scalar couplings for both peptides are shifted upward from random coil to yield values mostly between 6.0-8.5 Hz, consistent with an extended random coil ensemble or the presence of β-strand structure, and there are certainly quantitative differences in the scalar couplings between the peptides. While these highly averaged data may imply that the two peptides do not have substantially different structural ensembles, ^{1}H-^{15}N, and ^{1}H NOE data do provide more information about important structural differences between the conformational ensembles of IDPs Aβ40 and Aβ42.

[6] This and all other figures in this chapter are reproduced from (65) with permission.

Experimental RDCs (Figure 5.9 and 5.10) are also difficult to interpret structurally because the timescale of interconversion between IDP conformational states may be on the same order as the timescale of the protein realignment in the anisotropic medium. However, RDCs for Aβ40 and Aβ42 vary along the peptide sequences and show differences between the two peptides. Thus, these RDC data contain information about Aβ40 and Aβ42 structural differences, which can help to determine the correct ensembles.

Our high field ^{1}H NOE data identifies 707 crosspeaks for Aβ42 and 1108 crosspeaks for the Aβ40 peptide, but only 382 and 196 of these crosspeaks, respectively, can be uniquely assigned from experimental information alone. This is due to the fact that the NOESY spectra are crowded, different ^{1}H atoms have the same chemical shift, and many NOE peaks have multiple possible assignments (Figure 5.11). In the case of a strong NOE where only one of the possible assignments is a short-range interaction, we can confidently assign the peak. Most of the assigned cross peaks are intra-residue or sequential peaks, and 147 of them are a result of the same ^{1}H-^{1}H contacts occurring in the Aβ40 and Aβ42 ensembles. However Table 5.1 shows that 235 of the crosspeaks are unique to Aβ40 and 49 are uniquely present for Aβ42. Therefore, the NOEs which can be assigned from experimental information alone already indicate that the structural ensembles are different between the two peptides.

	Aβ40	**Aβ42**	**Both peptides**
Total Assigned NOE peaks	382	196	147
Intra-residue and sequential	362	185	142
i to i+ 2	20	9	5
i to i+ 3		1	
i to i+ 4		1	

Table 5.1. Summary of experimentally assignable NOEs determined for Aβ40 and Aβ42. Only ~25% of peaks for each peptide are assignable from experiment alone, and the third column indicates the assigned peaks that are present in both the Aβ40 and Aβ42 spectra.

We cannot assign the longer-range NOE peaks uniquely to one ^{1}H-^{1}H contact because all possible assignments are long-range and therefore it is unclear which is correct. However, we can see that the Aβ40 and Aβ42 spectra are different (Figure 5.11) and many of the longer-range NOEs present in the Aβ40 ensemble are not present for Aβ42 and vice versa. For example, if we look at NOE peaks that because of the chemical shifts (which restrict residue types involved) cannot be a result of any contact between residues closer than i to i+5, we see that 19 out of 40 of the Aβ40 NOEs are not present in the Aβ42 spectrum and 28 out of 46 of the Aβ42 NOEs are not present in the Aβ40 spectrum (Table 5.2). This indicates that there are several long-range NOEs for each peptide that are unique to its structural ensemble, and therefore the two amyloid peptide ensembles have distinct structurally features. Further details on the ^{1}H NOE's are presented in the Appendix in Tables 5.4 and 5.5. For further interpretation a computational model is needed that is validated by the available NMR data and yet expands upon the molecular structure information that would explain the NOE differences found between the two peptides.

	Aβ40	**Aβ42**
Total NOE peaks	1108	707
Intra-residue and sequential	828	537
i to i+ 2	142	68
i to i+ 3	58	36
i to i+ 4	40	20
i to i+ 5	11 (2)	13 (8)
i to i+ 6	7 (3)	4 (1)
i to i+ 7	10 (8)	9 (6)
i to i+ 8	2 (2)	3 (0)
i to i+9	2 (0)	6 (1)
i to i+10		1 (0)
i to i+11	2 (2)	
i to i+12	1 (0)	3 (0)
i to i+13	2 (2)	
i to i+14	2 (1)	3 (1)
i to i+20	1 (1)	2 (1)
i to i+21		1 (0)
i to i+24		1 (0)

Table 5.2. Summary of remaining experimental NOEs determined for Aβ40 and Aβ42. For the remaining ~75% peaks that cannot be assigned, peak intensities may be composed of a single pair contact of which several pair contacts are possible assignments, or may be composed of multiple, fractional contact pairs. In either case we provide the *experimental* lower bound, *n*, for *i* to *i*+*n* contacts for the unassigned peaks for each peptide. The value in parentheses indicates the number of these peaks that are present in the other peptide spectra as well. This value is given only for contacts from i to i+5 or greater. Some data reproduced from (64, 157).

Experimental Validation of Theoretically Derived Structural Ensembles for Aβ40 and Aβ42

In a recent review we considered the effectiveness of *de novo* MD for generating IDP structural ensembles for Aβ40 and Aβ42 (56, 64), as compared to random coil or statistically biased secondary structure ensembles, or selecting a subset of structures from the random or statistical ensembles that give a best fit to experimental NMR data, such as developed in the ENSEMBLE (52, 54, 63, 141) and ASTEROIDS (50, 61, 62) approaches. The performance of a given computational method was judged by the ability of a given model to reproduce chemical shifts, J couplings, and RDCs based on local (L-RDC) and global alignments, and experimentally assignable ^{1}H-^{1}H NOEs, as averages over the entirety of their conformational ensembles.

We showed that ensembles of structures based on random coil or statistical conformational distributions perform poorly, and there were no subset of structures from these ensembles that could substantially optimize their agreement with the NMR data for Aβ40 and Aβ42. Instead, ensembles incorporating structural members from the *de novo* MD calculations that contain significant amounts of *cooperative* secondary structure content gave much better

agreement with all NMR data. Table 5.3 summarizes the quality of the *de novo* MD ensemble compared to an additional step of refining the *de novo* MD ensemble using knowledge from NMR experiments conducted on the Aβ40 and Aβ42 peptides using the ENSEMBLE software package (MD-ENS). We note that the MD and MD-ENS ensembles reproduce the chemical shift data equally well, although the chi-squared statistic is lower for the MD-ENS calculation. It is evident from Table 5.3 that while the MD-ENS structural ensemble is better optimized against scalar couplings and L-RDCs by construction, RDCs based on global alignments are improved (Figure 5.9) but not to the same extent as L-RDCs (Figure 5.10). NOE's based on experimental assignment are predicted equally well by MD-ENS when the correlation times from the *de novo* MD simulations are used.

Average Property	**Aβ40 MD**	**Aβ40 MD-ENS**	**Aβ42 MD**	**Aβ42 MD-ENS**
H_α χ^2	0.58	0.30	0.54	0.33
H_N χ^2	0.36	0.34	0.48	0.37
C_α χ^2	0.69	0.46	0.98	0.51
C_β χ^2	0.70	0.36	0.52	0.34
$^3J_{H^N H^\alpha}$	0.99 (1.82)	0.62 (0.72)	0.99 (1.83)	0.54 (0.56)
RDC (Hz)	2.22	1.69	2.25	2.13
L-RDC (Hz)	1.88	0.18	2.14	0.33
H_2O NOEs (assigned)	1.15 (0.74)	1.12 (0.74)	1.25 (0.67)	1.21 (0.68)
D_2O NOEs (assigned)	3.22 (0.55)	3.19 (0.55)	0.58 (0.80)	0.57 (0.80)
^1H-^{15}N NOEs	0.17		0.21	

Table 5.3. Comparison between *de novo* MD (MD) and experimentally optimized MD (MD-ENS) structural ensembles for Aβ40 and Aβ42. For chemical shifts we report χ^2, which measures agreement between the computational ensembles and the experimentally measured chemical shifts within SHIFTX (99) calculator error. We also provide the RMSD between experiment and calculated ensembles for $^3J_{H^N H^\alpha}$ and (χ^2). RMSDs for RDCs are based on either global alignments using PALES (96) or on local alignments (L-RDCs) evaluated with ENSEMBLE (141). For NOEs we consider the simulated agreement with experiment for NOEs that can be assigned from the spectrum alone. The RMSD is normalized by the NOE intensity value for each peak to yield the N-RMSD and (correlation coefficient, r) for the H_2O and D_2O experiments. In this work we have used the decay timescales of specific proton pairs from the *de novo* MD simulation to inform the calculation of the MD-ENS NOEs. Finally we provide the RMSD between experiment and *de novo* MD for the heteronuclear NOE's; these are a purely dynamical phenomena and hence can only be derived from the MD simulation. Some data reproduced from (58, 64).

We found that the Aβ40 *de novo* MD ensemble is more extensively optimized using the ENSEMBLE method compared to the Aβ42 peptide. For Aβ40 nearly half the residues across the sequence exhibit a decrease in the percentage of the ensemble where they are involved in β-strand structure. By contrast for Aβ42 there are fewer changes in the qualitative features of the ensemble and the ENSEMBLE optimization amplifies the primary β-hairpin that is discussed in more detail below. Because the optimization of J-coupling and RDC data results in changes in each type of secondary structure at the Aβ42 C-terminus, it is difficult to draw a direct connection between the change in the observable value and difference in the structures present. However Table 5.8 gives more detailed changes observed in the calculated MD vs. MD-ENS for each peptide.

Overall, the quantitative agreement between experiment and back-calculations with structures from MD-ENS for chemical shifts (Figure 5.7), scalar couplings (Figure 5.8), and L-RDCs evaluated with ENSEMBLE (141), and the ^{1}H-^{15}N and assignable NOEs are very good (Table 5.3), although the Aβ40 $^{2}H_2O$ spectrum agreement is less good than the other NOEs. Furthermore, given the *de novo* or MD-ENS ensemble of ^{1}H-^{1}H contacts, and using the corresponding timescales given by *de novo* MD simulations to calculate NOEs with the MD-ENS, we can also predict the assignments of the unknown experimental cross peaks (Tables 5.4 and 5.5). We also provide the experimentally assignable cross-peaks not due to intra-residue or sequential contacts and agreement with MD-ENS (Tables 5.6 and 5.7). We therefore choose to analyze the MD-ENS structural populations for the Aβ40 and Aβ42 peptides given its consistent high quality performance against all available experimental NMR data.

Structural Ensemble of Aβ40 and Aβ42

Given the strong validation against a range of experimental NMR data, we now use the MD-ENS structural ensembles to determine what differences there are between the Aβ40 and Aβ42 IDPs. The MD-ENS structural ensembles of Aβ40 and Aβ42 show stark differences between the two peptides. Figure 5.1 shows the propensities of each peptide to form β-turns, helical structure, or intramolecular β-bridges, β-hairpins or β-strands by residue, as averages over their conformational ensembles. As we found for Aβ42 (48), Aβ40 is a highly heterogeneous tertiary ensemble, which samples conformations reflecting all possible secondary structure categories and is composed of a range of collapsed structured states to highly extended conformations. Although Aβ40 samples some conformations very similar to ones seen in the Aβ42 ensemble, such as a highly populated turn centered at residues 7-8 or a helix near Ser26, the two peptides have substantially different secondary structure profiles overall. The simulated structural ensembles show that most of the long-range NOEs produced by each peptide are a result of hydrogen-bonded β-structure, however, different β-strand associations are formed in the Aβ40 ensemble than in the Aβ42 ensemble.

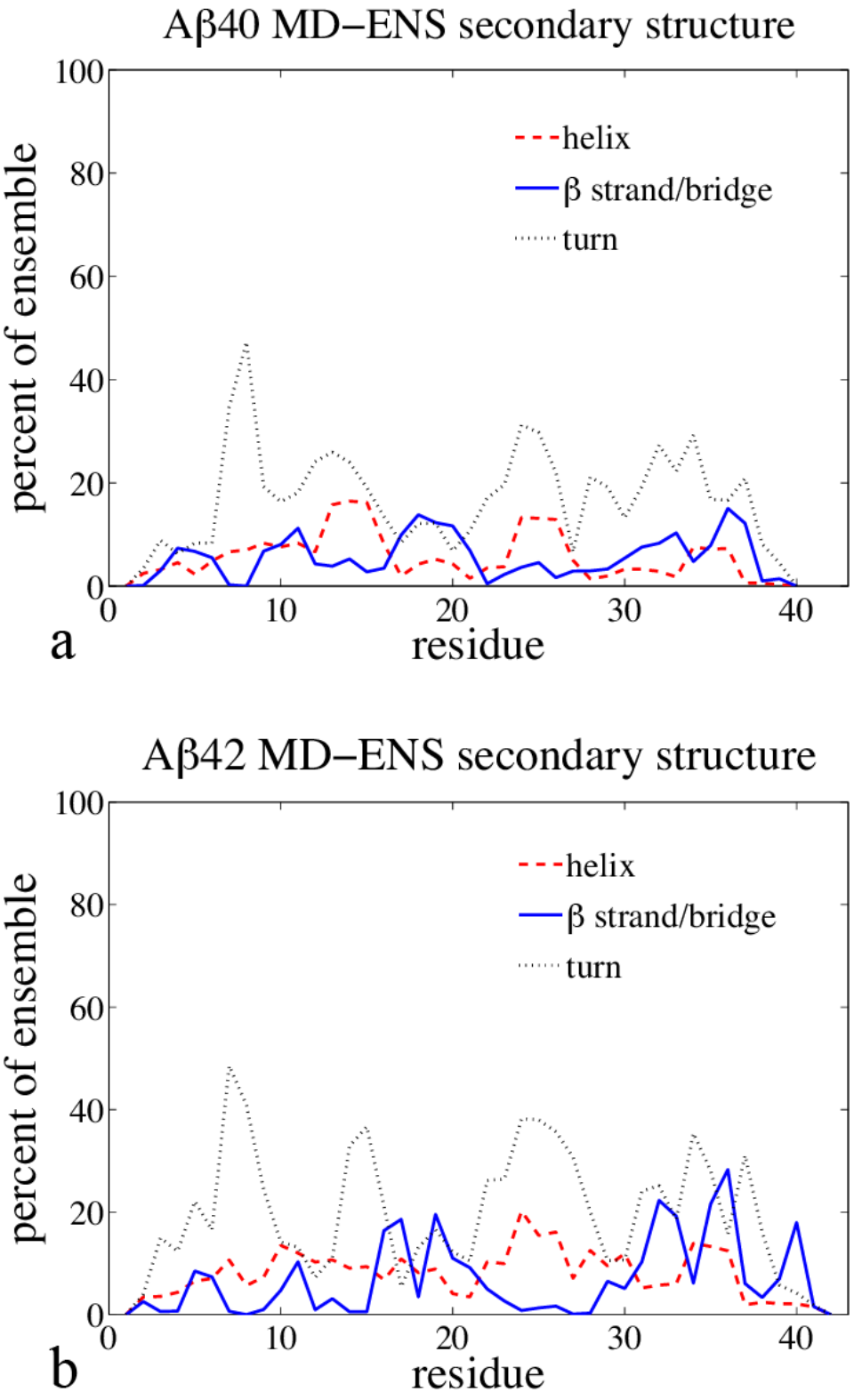


Figure 5.1. *Percentage of (a) Aβ40 and (b) Aβ42 simulated ensemble in different types of secondary structure by residue.* The dashed red line represents α-helix, the solid blue line for β-bridges or β-strands, and the dotted black line for β-turns.

Figure 5.2 is a contact map from the MD-ENS simulated ensemble for each peptide (the corresponding *de novo* MD simulated contact map is given in Figure 5.12). The long-range contacts are clearly different in the Aβ40 and Aβ42 ensembles, and many of these long-range contacts are due to β-strand formation. Clearly visible in the Aβ42 contact map are two β-hairpin sub-populations between the CHC residues 16-21 and 29-36 in ~34% of the ensemble (Figure 5.3a and 5.3b); one is defined by β-strand pairing of residues 16-17 with 35-36 (~16%) and the other by β-strand pairing of residues 17-21 with 29-34 (~18%). Furthermore, residues 26-27 form a β-turn in ~22% of the population, half of which also occurs with the dominant anti-parallel β-hairpin, consistent with the same 26-27 β-turn and the 16-21 and 29-36 β-strands that ultimately adopt the intermolecular arrangement of the stable mature fibril state. This feature is also consistent with a number of MD simulations that highlight the importance of residues 23-28 for nucleating monomer folding (120) and supported by detailed structural characterization of the amyloid-β fragments Aβ21-30 (26, 48, 121, 122) and Aβ10-35 (104, 105), as well as the importance of residues 16-22 that promote β-sheet structure as discussed in (123, 124).

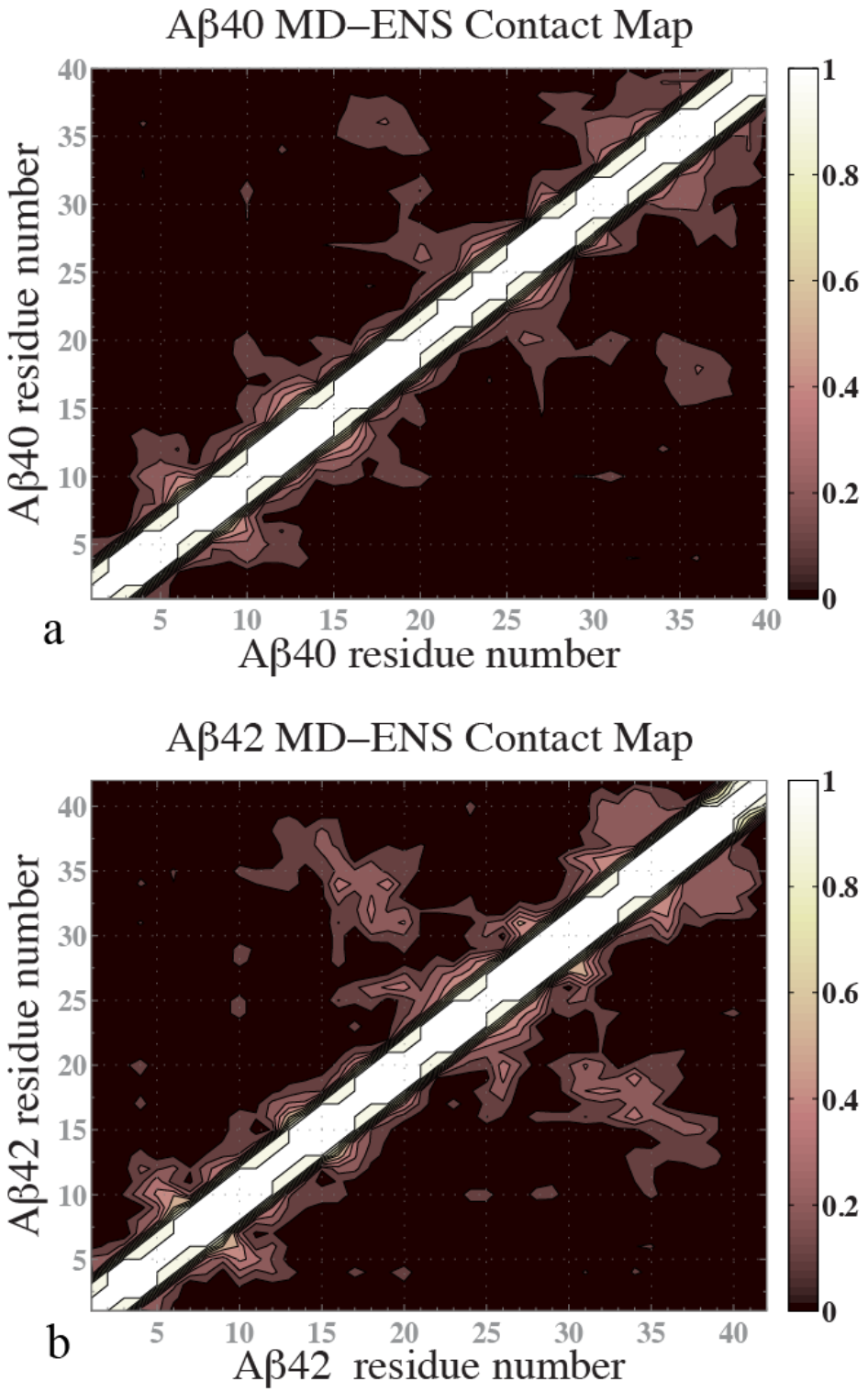


Figure 5.2. *Contact map of the simulated ensembles of (a) Aβ40 and (b) Aβ42.* This contact map gives the frequency of interaction between each pair of residues in the peptide MD-ENS simulated ensembles. White indicates contacts present in 100% of the ensemble and black indicates contacts never seen in the ensemble. We define two residues to be in contact if any of their heavy atoms are within 5 Å of each other.

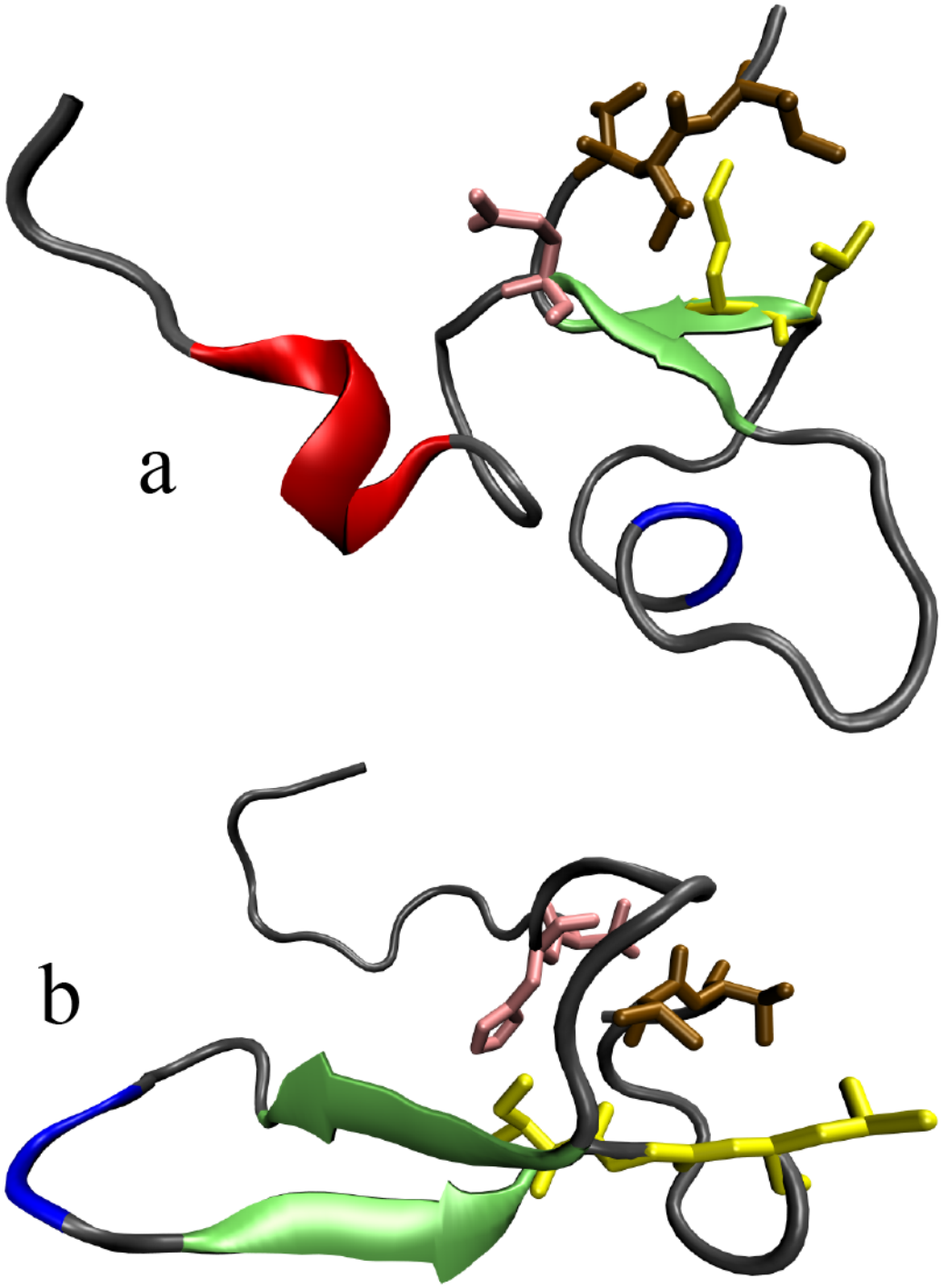


Figure 5.3. *The dominant 16-21 and 29-36 β-hairpin population for Aβ42.* (a) Aβ42 forms a β-turn at residues 26-27 with the 16-17 and 35-36 β-hairpin. The hydrophobic side chains of residues 39-41 (brown) also fold back to contact the side chains of residues 34-35 (yellow). The side chain of Gln15 (pink) caps the end of the β-hairpin by contacting residues 37-38. (b) Aβ42 forms the 26-27 β-turn and 17-21 and 29-33 β-hairpin with a C-terminal hydrophobic side chain interaction between 39-40 (brown) and 32-34 (yellow). Residues 12-13 (pink) also interact with the C-terminus around residue 38. Turns (blue), Helix (red), β-strands (green).

In contrast, Aβ40 forms an alternative, less populated anti-parallel β-hairpin between the CHC and residues 9-13 (Figure 5.4a) in ~10% of its ensemble, that sometimes includes CHC association with a third β-strand comprising residues 35-37 to define a β-sheet (Figure 5.4b). We note that Val18, at the center of the CHC, is in the middle of this Aβ40 β-hairpin and β-sheet. In fact, previous work by Yan and co-workers (125) examined side chain methyl groups, showing that Val18 is more ordered in Aβ40 compared to Aβ42. Our simulations provide an explanation for this experimental observation since we find that Val18 participates in more β-bridge or strand contacts within the Aβ40 ensemble (~14%) than in the Aβ42 ensemble (~3%). In the Aβ42 ensemble Val18 is found near the ends of the two β-strands involving the CHC and is less ordered as a result, due to fraying.

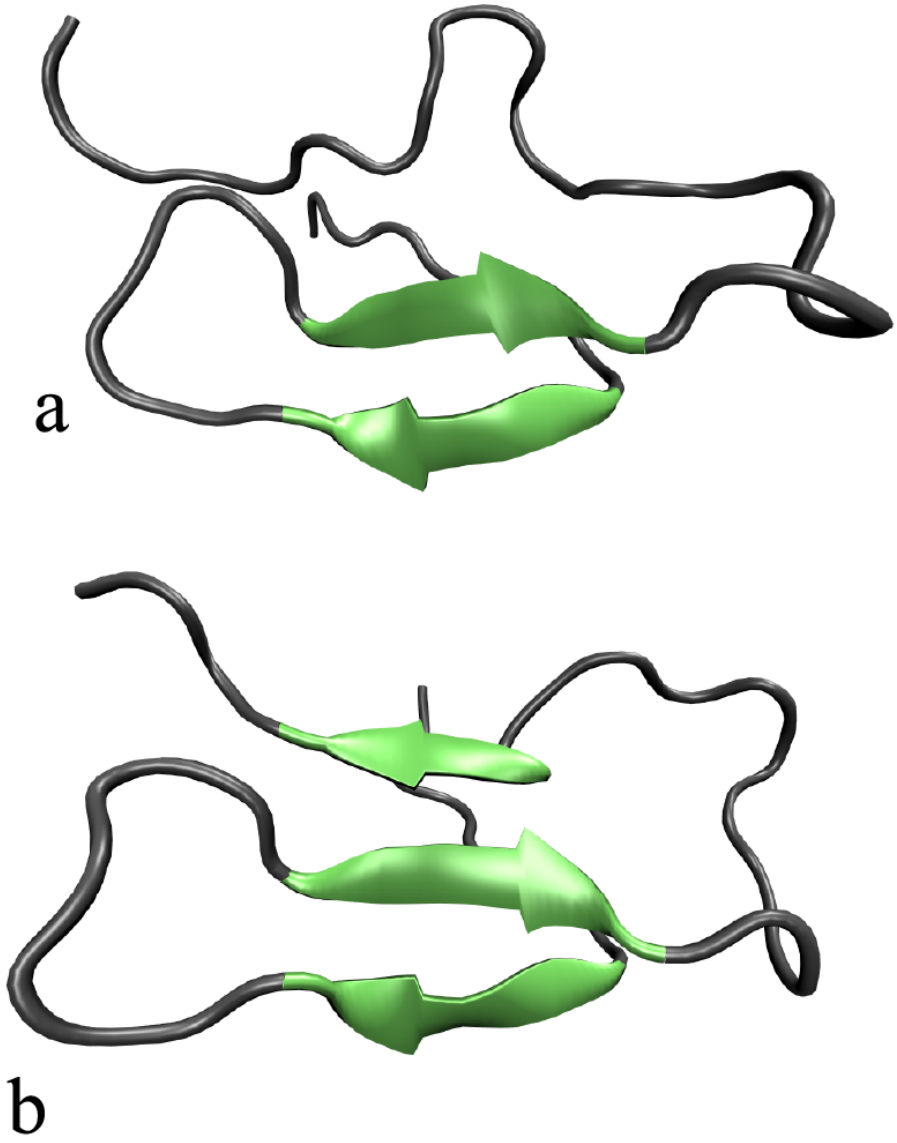


Figure 5.4. *The dominant β-hairpin and β-sheet population for Aβ40.* (a) Aβ40 forms the 9-13 and 16-21 β-hairpin. (b) The hairpin interaction between 17-20 and 35-37 occurs simultaneously to form a β-sheet with 3 strands. β-strands (green).

However, each peptide exhibits small additional sub-populations of anti-parallel and parallel β-strand associations, although most are defined by two hydrogen bonds only (i.e. a β-bridge). Figure 5.5a-5.5c provides three additional populations of β-strand structure for the Aβ42 ensemble that are worth mentioning. The first is an increased amount of anti-parallel β-strand association between residues 3-6 and 10-13 which comprises ~10% of the Aβ42 ensemble, although we emphasize that 7% of the Aβ42 conformers are only stabilized by a single β-bridge. The second is a parallel β-strand association between residues 21-23 and 36-38 in ~8% of the ensemble, half the time exhibiting only one β-bridge. Finally there is an anti-parallel β-hairpin formed by residues 34-36 and 39-40 in 6% of the ensemble that is negligibly populated in the Aβ40 ensemble. This last β-hairpin is consistent with that found in previous MD studies on Aβ42 (90, 195, 225), but it is not significantly populated and actually is subsumed into a larger sub-population involving hydrophobic clustering in the C-terminus that is a direct result of Ile41 and Ala42.

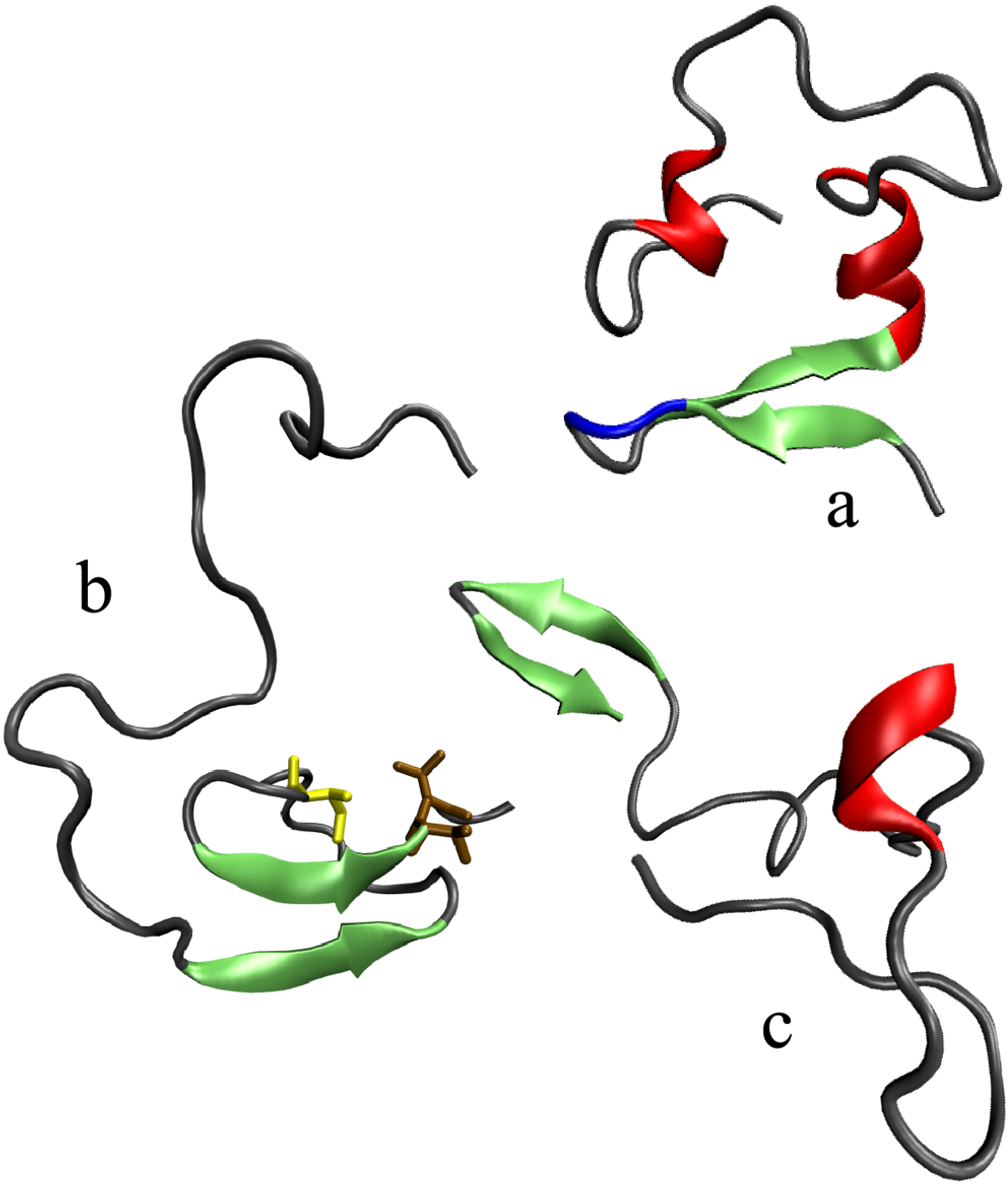


Figure 5.5. *Small sub-populations of Aβ42 conformations containing β-strands.* (a) Aβ42 forms a β-turn at residues 7-8 nucleating β-strand pairing of residues 3-6 and 10-13, along with a helix from residue 14-17 and at 32-35. (b) Aβ42 forms a parallel β-strand association between residues 21-23 and 36-38 while the 39-40 side chains (brown) contact Ile32 (yellow). (c) A C-terminal β-hairpin formed by residues 34-36 and 39-40. Turns (blue), Helix (red), β-strands (green).

The Role of Ile41 and Ala42

Based on characterization of our simulated ensembles, we have found that the very different populations of β-strand structure for Aβ40 and Aβ42 are consequences of the two additional hydrophobic residues in Aβ42, Ile41 and Ala42, which can form inter-residue contacts not available to Aβ40. These two residues do not form hydrogen bonds in a significant portion of the ensemble, but they are able to form hydrophobic interactions. We observe increased hydrophobic clustering between residues 39-41 and 31-36 for Aβ42, and this C-terminus clustering occurs frequently with contacts between residues 37-38 and residues 12-16. These interactions are visible in the contact map (Figure 5.2b). The new intramolecular contacts in the Aβ42 ensemble isolate the CHC from the N-terminus and the C-terminus to preclude β-hairpin formation with either the 9-13 or 35-37 β-strands observed in the Aβ40 ensemble. Instead the CHC of Aβ42 is most frequently encased in a loop defined by residues 15-38 or 16-36 that promote the 26-27 β-turn and/or β-hairpin that are compatible with known amyloid fibril forming regions. Furthermore, the more isolated N-terminus of the Aβ42 ensemble forms some type of β-bridge or β-strand association between residues 3-6 and 10-13 in ~10% of the ensemble, while Aβ40 forms it in only 3% of its ensemble. Parallel β-strand association between residues 21-23 and 36-38 also always occurs with the hydrophobic contacts between 39-41 and 31-36, which is why it never occurs in the Aβ40 ensemble.

Our simulated ensembles are also consistent with the slower relaxation rates and increased ^{1}H-^{15}N NOE intensities seen experimentally that indicate that the Aβ42 backbone is more ordered at the C-terminus than Aβ40 (66, 164). However, it is not known experimentally if order in the C-terminus arises from a populated helix or β-strand or from hydrophobic clustering often observed in disordered or unfolded states of proteins (143, 226). In Figure 5.6 we provide a comparison of the simulated ^{1}H-^{15}N NOE intensities for Aβ42 and Aβ40, which are in excellent agreement (64, 160, 161) with the experimentally measured values by Yan and Wang (125) (Table 5.3). We find there is an increase in ^{1}H-^{15}N NOE intensities calculated from simulation for residues 35-40 for Aβ42 compared to Aβ40 (the same seen experimentally), indicating that the longer peptide is more ordered at the C-terminus. We attribute this to the many hydrophobic interactions involving the Val40 side chain with residues 31-36 that make up 45% of the Aβ42 ensemble compared to 13% of the Aβ40 ensemble, and when Ile41 is included the hydrophobic clustering increases to close to 60%. Example Aβ42 structures in which the dominant β-hairpin and 26-27 β-turn form along with a C-terminal hydrophobic contact between 39-41 and 31-36 are shown in Figure 5.3a and 5.3b.

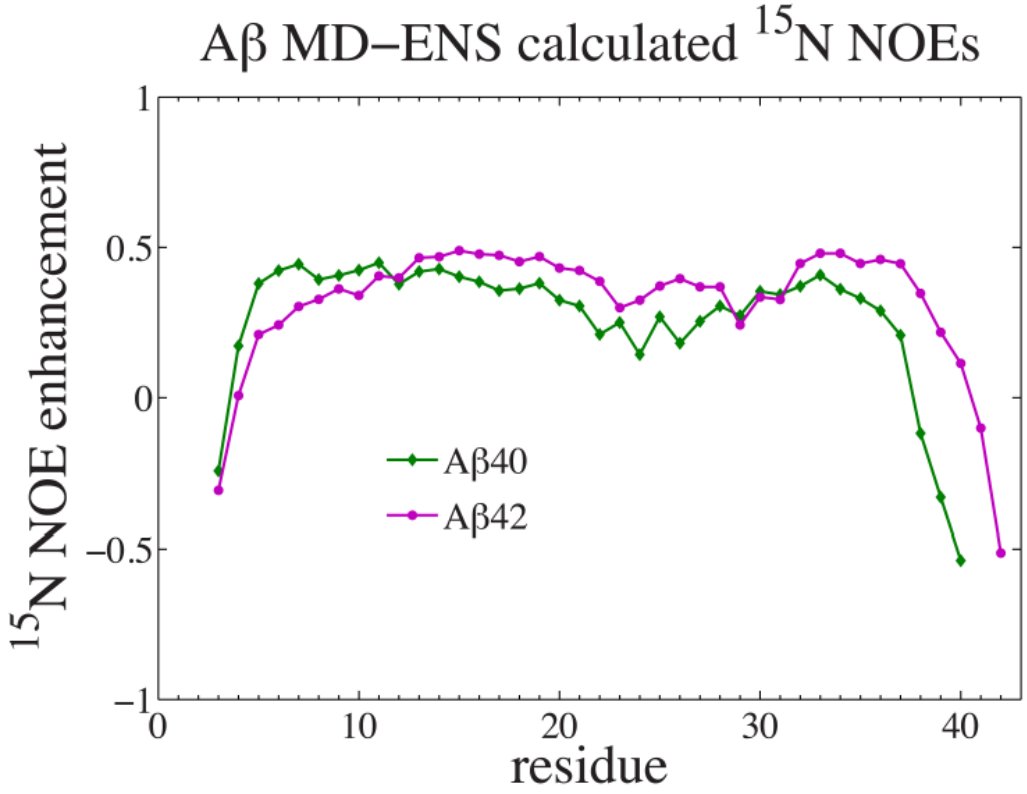


Figure 5.6. *Comparison of simulated Aβ40 and Aβ42 ^{1}H-^{15}N NOEs.* The MD results show that the C-terminus is more ordered for Aβ42 when compared to Aβ40.

When we analyze the *de novo* MD Aβ40 and Aβ42 ensembles derived at 311 K, near physiological temperature, we find that both peptides exhibit a decrease in population of the major turns and helices at the increased temperature, while Aβ40's most populated β-strands at 287 K melt out to yield significantly reduced percentages at the higher temperature. By contrast, the β-strands present in the Aβ42 ensemble are more stable and persist as the temperature increases, strengthened by the increase in hydrophobic clustering of the C-terminal residues which is expected to become more prominent with increasing temperature. This is consistent with the experimental finding from circular dichroism that the Aβ42 β-strand content is more stable than that of Aβ40 as the temperature of the sample is increased (119).

Discussion

For the past decade, Alzheimer's researchers have been interested in understanding why Aβ42 is much more aggregation prone than Aβ40, despite their nearly identical sequences. Since fibrillization of Aβ has been shown to follow a nucleation-dependent polymerization mechanism (26, 48, 65, 227), the kinetic data imply that the nucleation barrier is smaller for Aβ42 than for Aβ40. We suggest that the underlying structural differences between the two Aβ monomeric ensembles identify three possible factors for why the nucleation step is more difficult for Aβ40 compared to Aβ42.

The first is that in order to oligomerize or aggregate into a fibril-forming conformation containing intermolecular β-sheets at CHC residues 16-21 and 29-36, the Aβ peptide must overcome the free energy cost of breaking up any competing β-strand alignments. We have found that the CHC region of Aβ40 forms a different set of β-strand pairings than Aβ42. The fact that in the *de novo* MD ensemble the alternative β-strands for Aβ40 are less populated at higher temperatures means that the rate of fibrillization would increase with temperature, consistent with what is seen in fibrillization experiments (228, 229). The second is the presence of the two additional C-terminal residues Ile41 and Ala42 of Aβ42 that prevents the longer peptide from forming the intramolecular β-sheet seen in the Aβ40 ensemble. Finally, the increased hydrophobic clustering at the Aβ42 C-terminus isolates the CHC within a loop comprising residues 15-38, placing it in register with the 29-36 β-strand to form an intramolecular β-hairpin. These same β-strands are also aligned in the intermolecular β-sheets exhibited in the insoluble fibrillar states.

Thus, while the data do not make a direct connection between the Aβ40 and Aβ42 monomer conformational ensemble data and oligomer and fibril energetics or formation kinetics, our new data allows us to comment on other proposed oligomerization pathways based on assumptions about the monomeric starting point. Several previous NMR studies have observed that the Aβ42 C-terminal residues are less flexible than those of Aβ40, leading some groups to surmise that the C-terminus is preordered in a β-sheet conformation similar to that occupied by fibrils and oligomers, and this contributes to Aβ42's increased aggregation propensity (20, 66, 119, 125, 164). Our data, however, indicate that this reduced motility of the Aβ42 C-terminus is a result of an extensive network of hydrophobic contacts in ~60% of the ensemble rather than β-hairpin hydrogen bonds involving residues 34-36 and 39-40, which occur in only 6% of the ensemble.

Mutational studies have argued that extended conformations at residues 41 and 42 of Aβ42 and a turn at residues 38-39 are important for aggregation (220, 221), leading Irie *et al.* to conclude that Aβ42 forms an intramolecular, anti-parallel β-hairpin between residues 40-42 and 35-37 with turns at residues 33-34 and 38-39 (223). Again, our data contradict this picture in that we observe primarily hydrophobic contacts between these regions in the monomer ensemble, although we do find a small population of β-hairpin in this region of the sequence. Other mutation studies show that Aβ42's aggregation propensity is related to the hydrophobicity of Ile41 and Ala42 (116, 220, 221), supporting our picture of a C-terminal hydrophobic cluster. We also observe that the C-terminal hydrophobic cluster contacts the central hydrophobic cluster at residues 16-21, which often accompanies the formation of the residue 26-27 β-turn, in a very similar conformation to that seen by Maji and coworkers via photo-induced cross-linking (224).

Conclusions

Previous all-atom simulation studies (89, 90, 120, 123, 195, 230-236) and experimental CD and NMR spectroscopy (14, 66, 67, 116) have sought to characterize the differences between the monomer ensembles of the Aβ40 and Aβ42 peptides. The CD studies indicated that both peptides should be primarily classified as random coil (116, 119), consistent with the same classification that was derived from chemical shift and J-coupling measurements (14, 66, 67).

We showed previously that a random coil ensemble for Aβ40 and Aβ42 does not agree well with the available experimental NMR data (64, 67). Even when we considered the assumption of an ensemble that uses direct secondary structure prediction algorithms for Aβ40 and Aβ42 (but with no cooperative secondary structure), the resulting ensembles did not agree with the experimental data. Further optimization using the ENSEMBLE method to refine the random or statistical ensembles of conformational sub-populations, an accepted procedure for generating IDP structural ensembles (54, 63, 141, 162), showed no improvement.

Here we have taken a different approach and used the *de novo* MD results to provide a different "basis set" for selection of conformational states using the ENSEMBLE method. In this case the monomer ensembles of Aβ40 and Aβ42 have heterogeneous structure, presenting a diverse set of α-helix, β-turns, and β-strands. Based on the optimized MD-ENS structural ensemble of Aβ40 and Aβ42, which show very good agreement with the available NMR data, back-calculations of chemical shifts were also found to be consistent with random coils or ensembles with statistical or predictive assignments of secondary structure (64, 96).

We have demonstrated that homonuclear ^{1}H-^{1}H NOE intensities and ^{15}N-^{1}H heteronuclear NOEs are more discriminating with regard to the tertiary contacts that define the important structural differences between the two Aβ peptides. J-coupling constants and RDCs provide additional quantitative information about the differences between Aβ40 and Aβ42 ensembles when combined with simulation data. Our study is further distinguished by the productive interplay of molecular simulation to first simulate the NOE observables and thus validate the theory, which in turn can be used to further refine and interpret the NMR data for Aβ42 and Aβ40. It is important to emphasize that developing a structural model of the Aβ monomer ensembles based on the experimental NOE data, which are averaged by rapid exchange among conformers, would not be possible without the MD simulations providing details of individual structures.

Our data reveal how the addition of residues 41 and 42 drastically changes the conformational landscape of the Aβ42 peptide by increasing hydrophobic interactions within the C-terminus that exclude the formation of intramolecular β-hairpins formed frequently in the Aβ40 ensemble. The major β-hairpin populated in the Aβ42 ensemble is a consequence of the increased hydrophobic interactions, resulting in increased propensity for a β-turn at residues 26-27 and increasing the proximity of β-strands involving CHC residues 16-21 and 29-36, compatible with a stable pre-fibrillar oligomeric species known as the globulomer (68, 115) and various polymorphs of the fibril structure (224).

The results presented here, along with experimental and theoretical evidence accumulated over the last decade, now provide a fairly consistent picture of the monomeric ensembles of amyloid-β and the differences between Aβ40 and Aβ42. The combination of studies unifies our understanding that the hydrophobicity of residues 41 and 42 is crucial to the behavioral difference between Aβ40 and Aβ42 (116, 220, 221), and that the Aβ42 C-terminus folds in on itself (220, 222, 223), reducing its flexibility compared to the Aβ40 C-terminus (20, 66, 119, 125, 164). Our data contradict only the hypothesis that the Aβ42 monomer C-terminal structure is significantly populated by a β-hairpin involving residues 34-36 and 39-40. Instead, the Aβ42 C-terminus forms primarily hydrophobic contacts, a classic feature of the disordered or unfolded

state (143, 226), which indirectly promotes β-hairpin structure that is compatible with known fibril forming regions of the Aβ sequence.

Finally, the results here emphasize that the disease associated amyloid-β peptides, although clearly classified as IDPs, do not necessarily conform to the standard computational model assumptions or experimental expectations that have been so useful in characterization of functional IDPs (74) or IDPs with simpler helical structure motifs (162). In particular, successful use of NMR optimization approaches such as ENSEMBLE required the diverse cooperative secondary structure populations derived from *de novo* MD to achieve good agreement with the NMR data, rather than the commonly assumed random coil or statistical ensembles incorporating secondary structure as the possible conformational populations.

APPENDIX

NMR experimental method

Recombinant Aβ40 and Aβ42 peptides were purchased from rPeptide (Athens, GA) reported to be > 97% purity. The peptide was lyophilized from trifluoroacetic acid (TFA). The alkaline pretreatment of Aβ and preparation of LMW Aβ by filtration protocols outlined by Teplow (79) were used to prepare a monomeric solution of Aβ. The lyophilized peptide was dissolved in 2 mM NaOH to produce a peptide concentration of 0.21 mg/ml with a pH of > 10.5. The peptide was then sonicated for 1 min in a bath sonicator and lyophilized. It was then resuspended in 20 mM sodium phosphate buffer, pH 7.2, 0.01% (w/v) sodium azide. This protocol ensures that when the lyophilized peptide is dissolved in buffer it will not pass through its pI of ~5.31. Before collecting NMR data, the sample was filtered with a 0.22 μm filter to remove aggregates and fibril seeds (79), and brought to a concentration of ~600 μM at pH 7.2. TOCSY and NOESY ^{1}H-^{1}H homonuclear spectra were collected for both peptides at the NMR Facility at UC Berkeley on Bruker Avance II 800 and 900 MHz spectrometers, the latter equipped with a Bruker cryogenic probe. The data was processed as described in (48).

The data were processed using NMRPipe and peaks were assigned and analyzed using CARA (80). Chemical shifts were assigned using 2D ^{1}H-^{1}H Total Correlation Spectroscopy (TOCSY) (81, 82, 141) and Nuclear Overhauser Effect Spectroscopy (NOESY) spectra (83-85, 168). The TOCSY spectra were collected in a 90% H_2O (10% $^{2}H_2O$) buffer solution with a mixing time of 60 ms and in a 99% $^{2}H_2O$ buffer solution with a mixing time of 80 ms. The NOESY spectra were collected in a 90% H_2O (10% $^{2}H_2O$) buffer solution with mixing times of 100 ms and 200ms, and in a 99% $^{2}H_2O$ buffer solution with a mixing time of 200 ms. In the direct dimension, 2048 points were collected while 445-512 complex points were collected in the indirect dimension. The spectral width was 10 ppm in each dimension (9 ppm in the direct and 2 ppm in the indirect dimension for $^{2}H_2O$) with 64 scans.

Given that Aβ40, and Aβ42 in particular, are aggregation prone, our group and many others have used the alkaline pretreatment and filtration protocols outlined by Teplow and co-workers (79) in order to prepare a monomeric solution of Aβ. Other studies (66, 237-242), which have observed large oligomeric Aβ populations in equilibrium with the monomer, have used different peptide preparation protocols that either do not remove fibril seeds or do not remove

them as effectively as the Teplow method, and therefore have likely greater amounts of aggregates. Therefore, we review previously obtained NMR data generated from preparation protocols that are the same or similar to the Teplow method.

Hou *et al.* used translational diffusion and sedimentation equilibration experiments to show that the Aβ peptide samples prepared by the Teplow method are mostly monomeric (14) at concentrations between 0.15 mM and 0.80 mM, suggesting that our experimental conditions of 0.60 mM contain primarily the Aβ monomer. Nonetheless, during the course of our NMR experiments we observed a loss in intensity of resonances, particularly for Aβ42, reflecting formation of some aggregated peptide. However, the NOESY spectra taken shortly after preparing a sample (when little aggregate is present) are the same as at later times (when more aggregate is present) in that no new signals are observed. Furthermore, the NMR chemical shifts and line widths do not vary over time or as a function of concentration, consistent with rapid relaxation for the aggregate due to its high molecular weight (14, 66).

Although signals from the aggregated species are not observed, in principle peptides could exchange between aggregated and monomeric states, and NOEs reflecting the aggregated state could be detected in free peptide. A careful study of relaxation parameters for Aβ solutions prepared as ours showed that the interactions of free peptide with aggregates are weak (so little peptide is bound to the aggregate), that the kinetics of exchange are relatively slow, and that the relaxation in the bound state is very rapid (145, 243). These features show that exchange transferred NOEs from aggregated states do not contribute to the observed NOEs, which must then arise from the monomeric state. In summary, all available evidence makes a convincing case that the NMR data that we and others have collected on properly prepared Aβ40 and Aβ42 species are fully representative of the monomeric form.

Peaks were identified at particular chemical shifts in the 200 ms mixing time NOESY spectra in both H_2O and D_2O for Aβ40 and Aβ42, and those peaks that could be unambiguously assigned to two specific hydrogen atoms were used to calibrate the NOE peak intensities derived from simulation. The majority of peaks could not be clearly assigned to a unique pair of hydrogen atoms due to spectral overlap; these peaks were instead assigned a list of potential hydrogen pairs consistent with observed peak frequencies, made up of atoms within a 0.04-0.08 ppm range around each observed peak.

ENSEMBLE Refinements

We first attempted to converge chemical shifts before converging J-coupling constants or RDCs, as recommended in Krzeminski *et al.* (58). We ran each ENSEMBLE optimization for 48 hours on a Cray XE6 at NERSC, during which ~500 rounds of ENSEMBLE optimization steps were completed. While the chemical shifts and J-coupling constants were able to meet their target values in that amount of time, the RDCs were unable to converge to their target energy even after ~5000 ENSEMBLE runs, although when we compared the RDC values to the experimental data the agreement was good, leading us to conclude that the target was set too stringently. We therefore selected those ensembles that did the best overall on reproducing the NMR data, picking the ensemble with the lowest energy according to the ENSEMBLE weighting function for quantitative analysis. We then combined the 20 lowest energy ensembles of 100 structures to create a final ensemble of 2000 structures.

Ensemble Structural Analysis

Structural analysis of the *de novo* molecular dynamics simulations of Aβ40 and the Aβ42 287 K and 311 K ensembles was performed using *ptraj*, DSSP (97), perl scripts and MATLAB (The MathWorks, Natick, MA) scripts. The secondary structure designations we used to describe the Aβ40 and Aβ42 ensembles are all based on the DSSP criteria described by Kabsch and Sander (97). We also used *ptraj* to identify the most commonly formed hydrogen bonds and salt-bridges and calculated distances between residue side chains. Side chain contacts were defined using a 7 Å cut-off between side chain centers of mass, and salt bridges were defined with 4 Å distance cut-off between heavy atoms and a 60° angle cut-off. In-house scripts were used to identify particular secondary structure features, side chain contacts, and hydrogen bonds that are present simultaneously in the Aβ42 ensemble.

Calculation of NMR observables methods

For both Aβ40 and Aβ42 we calculated NMR chemical shifts, scalar-coupling constants, residual dipolar couplings (RDCs), ^{1}H-^{1}H NOE peak intensities, and ^{1}H-^{15}N NOEs from our simulated ensemble data.

Chemical Shifts. Figure 5.7 shows the agreement between experimentally measured hydrogen and carbon chemical shifts and that calculated as averages over the MD-ENS structures using the program SHIFTX (99).

Homonuclear NOE Intensities. We use the method of Peter *et al*. (94) to calculate the spectral density function (which is related to the NOE peak intensity) from the short NVE simulations. Specifically, we use *ptraj* to analyze the NVE simulations. The *ptraj* output is the normalized correlation function for each pair of hydrogen atoms of the Aβ peptides. These are calculated according to

$$C(\tau) = \left\langle \frac{1}{r^6(t)} \right\rangle^{-1} \left\langle \frac{P_2\left(\cos\chi_{t,t+\tau}\right)}{r^3(t)\,r^3(t+\tau)} \right\rangle \tag{5.1a}$$

where $r(t)$ is the vector between each hydrogen atom pair at time t, $\chi_{t,t+\tau}$ is the angle between the $r(t)$ and $r(t+\tau)$ vectors, and P_2 is the second order Legendre polynomial. We multiply these normalized correlation functions by $\langle r^{-6}(t)\rangle$, averaged over the individual trajectory, to obtain the expression

$$\left\langle \frac{P_2\left(\cos\chi_{t,t+\tau}\right)}{r^3(t)\,r^3(t+\tau)} \right\rangle \tag{5.1b}$$

for each of the independent 100 20-ns NVE simulations and average these values together, then re-normalize the average correlation function by dividing by $\langle r^{-6}(t)\rangle$, averaged over all NVE simulations (equivalent to dividing the entire correlation function by the value at t=0) . Each correlation function is then fit over a 5 ns range for τ to a multi-exponential form

$$C(\tau) \approx \sum_{i=1}^{N} a_i e^{-t/\tau_i} \tag{5.2}$$

with $N = 1, 2, 3,$ or 4 and

$$\sum_{i=1}^{N} a_i = 1$$

using the MATLAB function *lsqcurvefit*, which fits curves in a least-squares sense (The MathWorks, Natick, MA). The value for N is determined using Mallow's C_p statistic (95, 170),

$$\hat{R}(S) = \hat{R}_{tr}(S) + 2k\hat{\sigma}^2 \tag{5.3}$$

where S is the set of N covariates, R_{tr}, the training error, is the square difference between the correlation values and the fit exponential function, k is the number of regressors ($2N$ - 1), and σ^2 is the variance under the full model, which we estimate as 0.001 based on a selected number of spin vectors. The value of N that yields the lowest value for Mallow's C_p statistic is used as the fit, which is a better criterion for model selection than simply the training error, which is biased towards higher values of N.

The resulting time correlation function can be Fourier-transformed

$$J(\omega) = \int_{-\infty}^{\infty} C(\tau) e^{i\omega\tau}\, d\tau \tag{5.4}$$

to yield

$$J(\omega) = \sum_{i=1}^{N} \left[a_i \frac{2\tau_i}{1 + \omega^2 \tau_i^2} \right] \tag{5.5}$$

following the convention for Fourier-transforms of exponentials. The correlation time constants, τ_i, have a practical upper bound imposed by the rotation of the entire peptide in solution. No vectors may remain correlated at timescales longer than the rotational correlation time of the entire molecule in the slowest dimension. In cases where the average correlation function for the vector between two hydrogen atoms is fit with a time constant greater than 15 ns (around 5% of cases), we replace the long time constant with a time constant of 15 ns, which is a reasonable value for the longest possible rotational time-scale as described in (48).

We predict the Aβ40 and Aβ42 NOESY spectra from our structural ensemble and dynamical trajectories by calculating the intensity from a different relaxation matrix **R**, composed of diagonal elements

$$\rho_{ii} = \sum_{j=1\neq i}^{n} \frac{1}{10} K^2 \left[3J_{ij}(2\omega_0) + \frac{3}{2} J_{ij}(\omega_0) + \frac{1}{2} J_{ij}(0) \right] \tag{5.6a}$$

and off-diagonal elements

$$\sigma_{ij} = \frac{1}{10} K^2 \left[3J_{ij}(2\omega_0) - \frac{1}{2} J_{ij}(0) \right]. \tag{5.6b}$$

We simulate all hydrogen atoms explicitly for each methyl group and hence calculate all pair correlation functions, including neighboring methylene and methyl group protons, and we ignore water proton coordinates, as is the standard assumption in the NMR experiment.

We solve for the magnetization matrix at the mixing times that were used in both H_2O and ^2H_2O NOESY experiments, removing all exchangeable hydrogen atoms to simulate ^2H_2O buffer conditions. For the H_2O experiment, we multiply all amine intensities by 0.9 to account for the presence of 10% ^2H_2O, and remove all basic amine and hydroxyl groups, which would not appear due to broadening caused by exchange with solvent. We then add all of the degenerate intensities from methyl groups and intra-residue spins that are indistinguishable in the NMR spectra.

Finally, the constant relating the $t = 0$ matrix, $\mathbf{M}(0)$ of unity to the experimental cross-peak intensity scale, is approximated by fitting a line of best fit (restrained to cross through the origin) to a plot of known experimental intensities (those that were unambiguously assigned) versus their corresponding calculated intensities. For this plot, the experimental intensity values are an average of the peak intensities measured on either side of the diagonal in order to compensate for base-line differences on either side of the diagonal. The slope of this line can then be used as a conversion factor for all calculated intensities. All reported intensities (experimental and simulated) are also normalized by an estimate for the smallest experimentally identifiable peak intensity, so that an intensity of 1 indicates the smallest intensity that should be visible experimentally. After NOE peaks were calculated from the simulated ensemble the intensities were compared with the experimental intensity and expressed as multiples of smallest experimentally identifiable peak. This generated a distribution of simulated NOE peak intensities corresponding to every possible atomic contact. For these relative intensities, only peaks with intensities above 1 are expected to be experimentally visible above noise.

Heteronuclear NOEs. We use the same method as above to calculate the spectral density function for each pair of H-N atoms from the short NVE simulations for the H-N backbone bond of each residue of Ab40 and Ab42. We then calculated the steady state NOE enhancement factor of the ^{15}N spin by the ^1H NOE from our structural ensemble and dynamical trajectories according to

$$\varepsilon_{\mathrm{NOE}} = 1 + \frac{\gamma_{\mathrm{H}} \sigma_{\mathrm{HN}}}{\gamma_{\mathrm{N}} R_z^{(\mathrm{N})}} \tag{5.7}$$

where γ_H and γ_N are the gyromagnetic ratios of ^1H and ^{15}N, respectively. The ^1H-^{15}N cross-relaxation rate constant is given by

$$\sigma_{\mathrm{HN}} = \frac{1}{10} K^2 \left[3J_{ij}\left(\omega_{0,\mathrm{H}} + \omega_{0,\mathrm{N}}\right) - \frac{1}{2} J_{ij}\left(\omega_{0,\mathrm{H}} - \omega_{0,\mathrm{N}}\right) \right] \tag{5.8}$$

and the ^{15}N self-relaxation by

$$R_z^{(\mathrm{N})} = \frac{1}{10} K^2 \left[\frac{3}{2} J_{ij}\left(\omega_{0,\mathrm{N}}\right) + 3J_{ij}\left(\omega_{0,\mathrm{H}} + \omega_{0,\mathrm{N}}\right) + \frac{1}{2} J_{ij}\left(\omega_{0,\mathrm{H}} - \omega_{0,\mathrm{N}}\right) \right] \tag{5.9}$$

where $\omega_{0,\mathrm{H}}$ is the Larmor frequency of ^{1}H and $\omega_{0,\mathrm{N}}$ is the Larmor frequency of ^{15}N, and the constant factor K is defined as

$$K = \frac{\mu_0}{4\pi r_{\mathrm{eff}}^3} \hbar \gamma_{\mathrm{H}} \gamma_{\mathrm{N}} \tag{5.10}$$

where μ_0 is the permeability of free space, and $\hbar$ is Planck's constant. The effective r vector

$$r_{\mathrm{eff}} = \left\langle \frac{1}{r^6(t)} \right\rangle^{-1/6} \tag{5.11}$$

is the average of the r^{-6} values, which has then been converted back into units of distance.

J-Coupling Correction. In order to account for the effect of selective T_1 relaxation on the J-coupling constants measured using quantitative J-correlation and obtain experimental J-coupling constants comparable to those measured from COSY splittings, we used the method described by Vuister and Bax (93) involving Eq. 4 in their paper,

$$\begin{aligned} \mathrm{d}\left\langle \mathbf{I}_y^{\mathrm{N}} \right\rangle / \mathrm{d}t &= \pi J_{\mathrm{HH}} \left\langle 2\mathbf{I}_x^{\mathrm{N}} \mathbf{I}_z^{\alpha} \right\rangle - \left\langle \mathbf{I}_y^{\mathrm{N}} \right\rangle / T_{2\mathrm{HN}} \\ \mathrm{d}\left\langle 2\mathbf{I}_x^{\mathrm{N}} \mathbf{I}_z^{\alpha} \right\rangle / \mathrm{d}t &= -\pi J_{\mathrm{HH}} \left\langle \mathbf{I}_y^{\mathrm{N}} \right\rangle - \left\langle 2\mathbf{I}_x^{\mathrm{N}} \mathbf{I}_z^{\alpha} \right\rangle / T_{2\mathrm{HN}} - \left\langle 2\mathbf{I}_x^{\mathrm{N}} \mathbf{I}_z^{\alpha} \right\rangle / T_{1\mathrm{sel}} , \end{aligned} \tag{5.12}$$

describes the time dependence of the in-phase and antiphase amide proton magnetization during the dephasing period, 2ζ,. We solve this set of coupled differential equations to obtain

$$\left\langle 2\mathbf{I}_x^{\mathrm{N}} \mathbf{I}_z^{\alpha} \right\rangle \Big/ \left\langle \mathbf{I}_y^{\mathrm{N}} \right\rangle = \frac{-2\pi J_{\mathrm{HH}} \tanh\left(C_4 t/2\right)}{1/T_{1\mathrm{sel}} \tanh\left(C_4 t/2\right) + C_4} , \tag{5.13a}$$

where

$$C_4 = \left[\left(1/T_{1\mathrm{sel}}\right)^2 - 4\pi^2 J_{\mathrm{HH}}^2 \right] , \tag{5.13b}$$

which is dependent on the selective T_1 relaxation time constant, T_{1sel}. If T_{1sel} is assumed to be infinite, meaning that magnetization components giving rise to diagonal and cross-peaks relax at the same rate, then Eq. 5.13a reduces to

$$\left\langle 2\mathbf{I}_x^{\mathrm{N}}\mathbf{I}_z^{\alpha}\right\rangle \Big/ \left\langle \mathbf{I}_y^{\mathrm{N}}\right\rangle = -\tan\left(\pi J_{\mathrm{HH}} t\right). \tag{5.14}$$

This was the assumption made by Wang *et al.* when originally calculating the J-coupling constants, J_{app}, so in order to find the true magnetization ratio, we must plug in J_{app} to Eq. 5.14, along with the dephasing time $2\zeta = 26.1$ ms for t,

$$\left\langle 2\mathbf{I}_x^{\mathrm{N}}\mathbf{I}_z^{\alpha}\right\rangle \Big/ \left\langle \mathbf{I}_y^{\mathrm{N}}\right\rangle = -\tan\left(\pi J_{\mathrm{app}} 2\zeta\right). \tag{5.15}$$

Then, we can use this ratio to calculate the true J_{HH}, setting the full form of the equation (Eq. 5.14) equal to Eq. 5.15,

$$-\tan\left(\pi J_{\mathrm{app}} 2\zeta\right) = \frac{-2\pi J_{\mathrm{HH}} \tanh\left(C_4\, 2\zeta/2\right)}{1/T_{1\mathrm{sel}} \tanh\left(C_4\, 2\zeta/2\right) + C_4}, \tag{5.16}$$

and solving for J_{HH}. To calculate the T_{1sel} time necessary for Eq. 5.16, we sum the cross relaxation on the amide hydrogen of each residue due to every other hydrogen atom in the peptide, according to

$$R_{1i} = \sum_{j \neq i} \sigma_{ij}\,, \tag{5.17}$$

where

$$\sigma_{ij} = \frac{1}{10} K^2 \left[3J_{ij}\left(2\omega_0\right) - \frac{1}{2} J_{ij}\left(0\right)\right], \tag{5.18}$$

for the hydrogen atom pair, calculated from the correlation times of the H-H vector in our MD simulations. r_{eff} is also calculated based on there H-H distances in our simulated ensemble. T_{1sel} for each residue is then simply the inverse of R_1.

Figure 5.8 compares the experimentally measured scalar-coupling constants (with the correction described above) and values calculated as averages over the MD-ENS conformations as described previously (48), using the original Vuister and Bax parameter set (93) for the Karplus equation (92).

Residual dipolar couplings. Figure 5.9 compares the experimentally measured RDCs and those calculated as averages over the MD-ENS of structures using PALES (96) and multiplied by a constant factor (48) for comparison to experimental data from *Yan et al.* (67).

Peak Types	All NOE peaks	Long range peaks (i to i+2 or greater)
Experimentally observed peaks	**1108**	**280**
Experimentally assigned peaks	382	20
Experimentally unassigned peaks	726	260
Simulated peaks	1660	543
Simulated peaks agree experimentally assigned	356	11
Simulated peaks agree experimentally unassigned	652	206
False negatives	100	64
False negatives found in high noise regions	-42	-31
False negatives: atomic contacts on same residue pairs	-51	-26
False negatives inconsistent with experiment	**7**	**7**
False positives	652	326
False positives found in high noise regions	-206	-101
False positives: atomic contacts on same residue pairs	-364	-171
False positives made up of numerous weak	-10	-7
False positives inconsistent with experiment	**72**	**47**

Table 5.4. Summary of statistics of Aβ40 experimental NOEs and calculated NOEs from MD-ENS, and analysis of the number of false positives and false negatives.

Peak Types	All NOE Peaks	Long range peaks (i to i+2 or greater)
Experimentally observed peaks	**707**	**170**
Experimentally assigned peaks	196	11
Experimentally unassigned peaks	511	159
Simulated peaks	1007	173
Simulated peaks agree experimentally assigned peaks	178	1
Simulated peaks agree experimentally unassigned peaks	361	46
False negatives	168	123
False negatives found in high noise regions	-69	-54
False negatives explained by atomic contacts on same residue pairs	-67	-37
False negatives inconsistent with experiment	**32**	**32**
False positives	468	126
False positives found in high noise regions	-141	-30
False positives explained by atomic contacts on same residue pairs	-249	-60
False positives made up of numerous weak contacts	-9	-5
False positives inconsistent with experiment	**69**	**31**

Table 5.5. Summary of statistics of Aβ40 experimental NOEs and calculated NOEs from MD-ENS, and analysis of the number of false positives (peaks that were predicted from simulation but not seen experimentally) and false negatives (peaks seen experimentally but not predicted to be visible above the noise).

Atom 1	Atom 2	Calculated H_2O NOE intensity	Experimental H_2O NOE intensity	Calculated ^2H_2O NOE intensity	Experimental ^2H_2O NOE intensity
Hα ASN 27	H GLY 29	6.03	1.97	0.00	
Hε TYR 10	Hγ1 VAL 12	4.80	5.64	5.09	5.29
Hε TYR 10	Hγ2 VAL 12	4.80	2.68	5.09	2.86
Hδ TYR 10	Hγ1 VAL 12	4.71	5.86	4.92	5.34
Hδ TYR 10	Hγ2 VAL 12	4.71	2.99	4.92	2.45
Hβ ASP 7	H GLY 9	4.31	1.80	0.00	
Hε PHE 19	Hβ ALA 21	3.93	8.26	4.19	4.49
Hα ASP 23	H GLY 25	3.85	2.08	0.00	
Hβ ALA 2	H PHE 4	3.31	3.01	0.00	
Hα SER 8	H TYR 10	1.95	2.25	0.00	
Hα2 GLY 37	H VAL 39	1.33	11.59	0.00	
Hβ3 SER 8	H TYR 10	1.29	10.32	0.00	
Hδ1 LEU 17	Hε PHE 19	1.28	8.16	1.34	5.23
Hγ2 VAL 18	Hε PHE 20	1.07	5.84	1.13	4.51
Hγ2 VAL 18	Hδ PHE 20	0.96	4.09	0.99	4.15
H GLU 22	Hγ VAL 24	0.80	2.68	0.00	
Hδ1 LEU 17	Hξ PHE 19	0.80	8.25	0.84	6.35
Hγ2 VAL 12	Hδ2 HIS 14	0.43	5.49	0.43	2.90
Hγ2 VAL 18	Hβ3 PHE 20	0.36	3.51	0.34	1.59
Hα LEU 17	Hξ PHE 19	0.22	2.55	0.25	

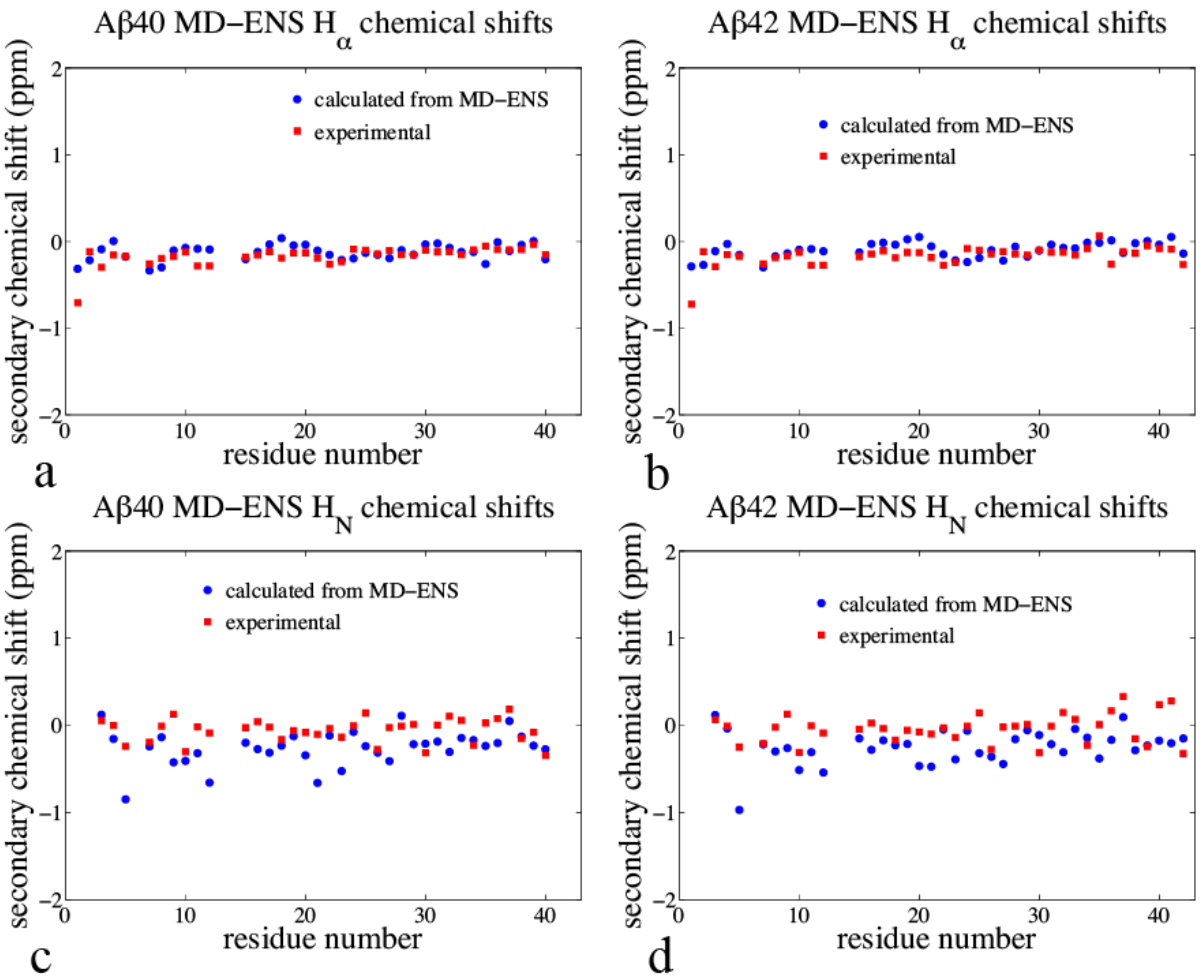
Table 5.6. List of Aβ40 experimentally assigned NOE peaks that are not due to intra-residue or sequential contacts along with calculated NOE values from MD-ENS ensemble for both H_2O and ^2H_2O.

Atom 1	Atom 2	Calculated H_2O NOE intensity	Experimental H_2O NOE intensity	Calculated ^2H_2O NOE intensity	Experimental ^2H_2O NOE intensity
Hε TYR 10	Hγ2 VAL 12	1.22	1.63	0.98	1.84
Hδ TYR 10	Hγ2 VAL 12	1.00	1.52	0.79	1.70
Hγ3 GLU 22	Hγ VAL 24	0.94	3.29	0.72	
Hβ ASP 7	Hδ TYR 10	0.84	1.50	0.69	
Hδ PHE 20	Hγ VAL 24	0.82	1.98	0.66	3.44
Hδ1 LEU 17	Hδ or ε PHE 19	0.75	3.06	0.60	5.07
Hε PHE 20	Hγ2 GLU 22	0.73	1.32	0.59	
Hγ2 VAL 18	Hδ PHE 20	0.39	2.85	0.29	2.70
Hγ2 VAL 18	Hε PHE 20	0.36	1.57	0.29	2.98
Hδ PHE 20	Hγ3 GLU 22	0.28	1.70	0.21	
Hγ2 VAL 18	Hξ PHE 20	0.23	1.52	0.18	1.63

Table 5.7. List of Aβ42 experimentally assigned NOE peaks that are not due to intra-residue or sequential contacts along with calculated NOE values from MD-ENS ensemble for both H_2O and ^2H_2O.

	Aβ40			Aβ42	
Residue	**β-strand MD**	**β-strand MD-ENS**	**Change between MD and MD-ENS**	**β-strand MD**	**β-strand MD-ENS**
10	32%	8%	**J-coupling decrease**		
11	32%	11%	**L-RDC increase**		
17	32%	10%	**L-RDC increase**		
19	33%	12%	**L-RDC decrease**		
20	36%	12%	**L-RDC decrease**		
21	20%	7%	**J-Coupling decrease**		
26	9%	2%	**L-RDC increase**		
27	12%	3%	**L-RDC increase**		
32			**J-coupling decrease**	17%	22%
33			**L-RDC increase**	4%	19%
36	36%	15%	**J-Coupling decrease**		
37	28%	12%	**L-RDC increase**		
40			**J-coupling decrease**	5%	18%
				α-helix MD	**α-helix MD-ENS**
15			**L-RDC increase**	19%	9%
16			**L-RDC increase; J-coupling decrease**	20%	7%
17			**L-RDC increase**	36%	11%
18			**J-Coupling increase**	29%	8%
19			**L-RDC increase**	29%	9%

Table 5.8. List of largest changes in calculated NMR observables between *de novo* MD and MD-ENS for Aβ40 and Aβ42.


Aβ40 MD–ENS Hα chemical shifts
calculated from MD–ENS
experimental
secondary chemical shift (ppm)
residue number
a
Aβ42 MD–ENS Hα chemical shifts
calculated from MD–ENS
experimental
secondary chemical shift (ppm)
residue number
b
Aβ40 MD–ENS HN chemical shifts
calculated from MD–ENS
experimental
secondary chemical shift (ppm)
residue number
c
Aβ42 MD–ENS HN chemical shifts
calculated from MD–ENS
experimental
secondary chemical shift (ppm)
residue number
d

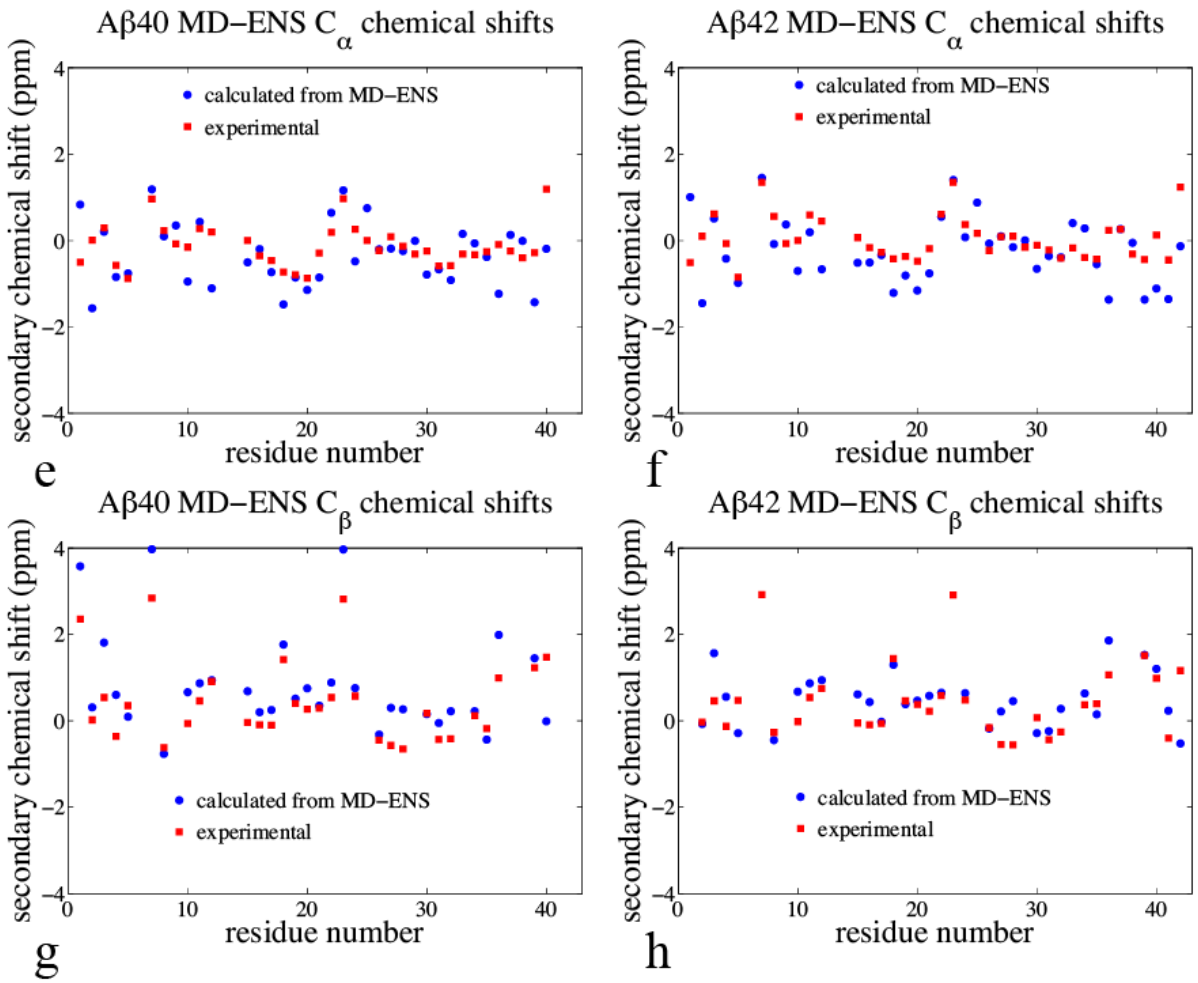


Figure 5.7. *Experimental and calculated proton and carbon secondary chemical shifts by residue.* (a) Aβ40 H_α chemical shifts, (b) Aβ42 H_α chemical shifts, (c) Aβ40 H_N chemical shifts, (d) Aβ42 H_N chemical shifts, (e) Aβ40 C_α chemical shifts, (f) Aβ42 C_α chemical shifts, (g) Aβ40 C_β chemical shifts, and (h) Aβ42 C_β chemical shifts. Red squares represent experimental data (the size of the marker represents the uncertainty), while blue circles represent the data calculated from MD-ENS simulation. Random coil residue specific values taken from are subtracted from both experimental and simulation values. The experimental carbon chemical shift data are taken from Hou *et al*. (14).

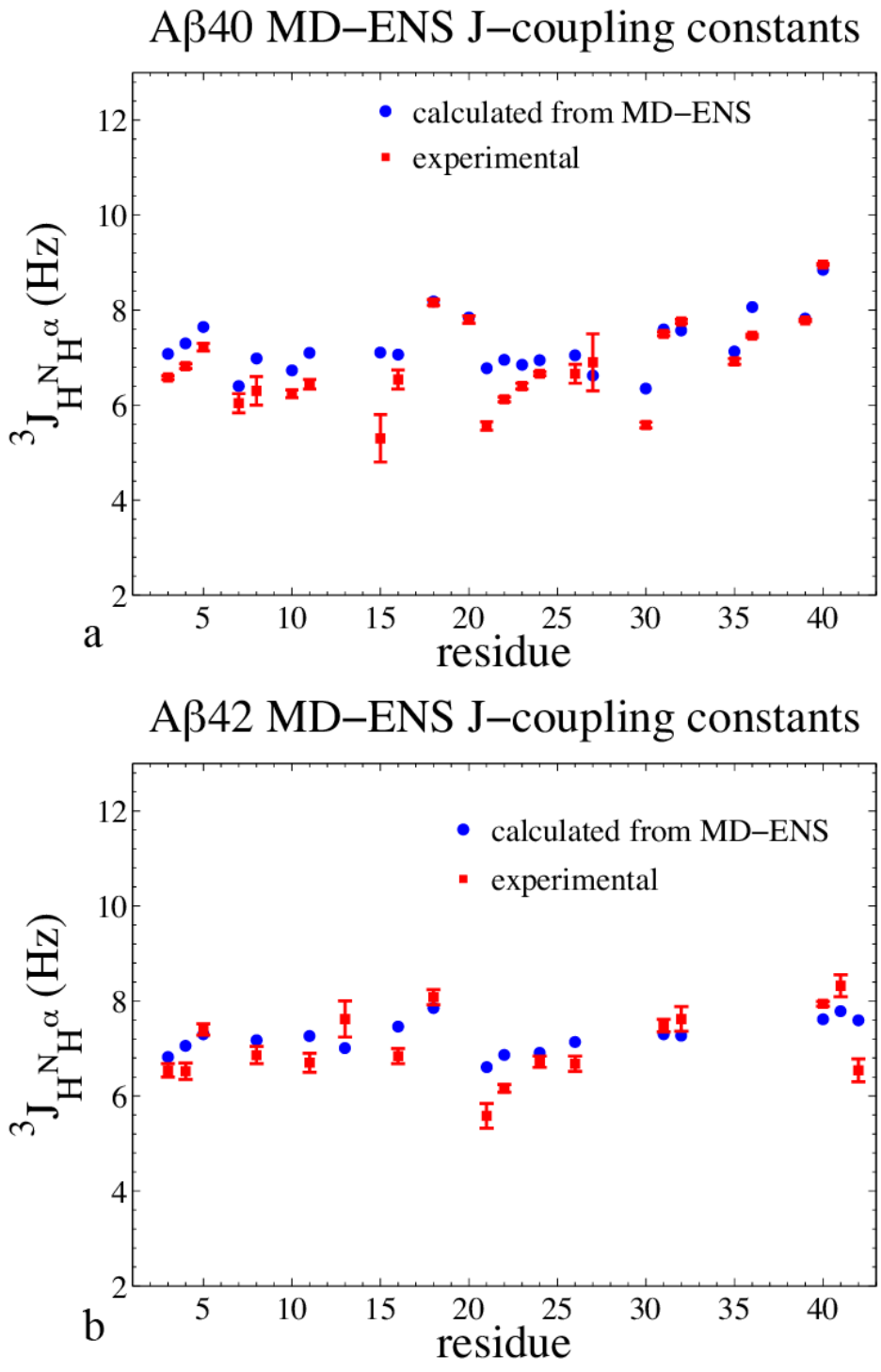


Figure 5.8. *J-coupling constants for backbone amides for (a) Aβ40 and (b) Aβ42.* The red symbols are the experimental data from Yan *et al.* (67). Simulation uncertainty bars represent rms difference between two independent simulations and the average. The experimental data has not accounted for a relaxation correction that makes J-couplings determined from a $H_N H_\alpha$ 3D experiment consistently lower than those from COSY splittings (14) by a small amount (as high as ~10% (93)); if this relaxation effect is accounted for, then the qualitative agreement between experiment and MD-ENS simulation is excellent.

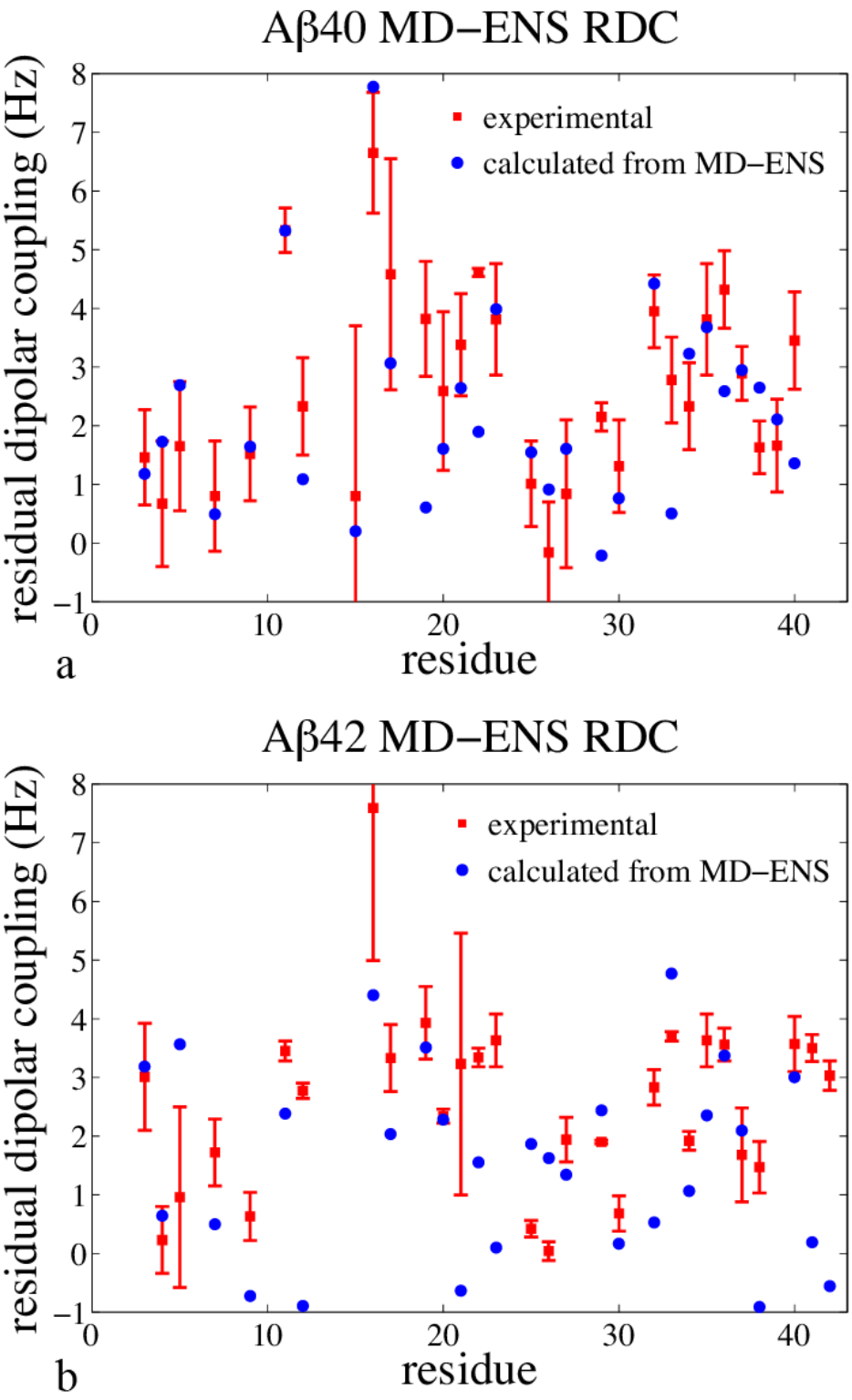


Figure 5.9. *Experimental vs. calculated residual dipolar couplings based on global alignments calculated from PALES for Aβ40 (a) and Aβ42 (b).* The red symbols are experimental data from Yan *et al*. (67). The blue circles are the data calculated from the simulation ensemble.

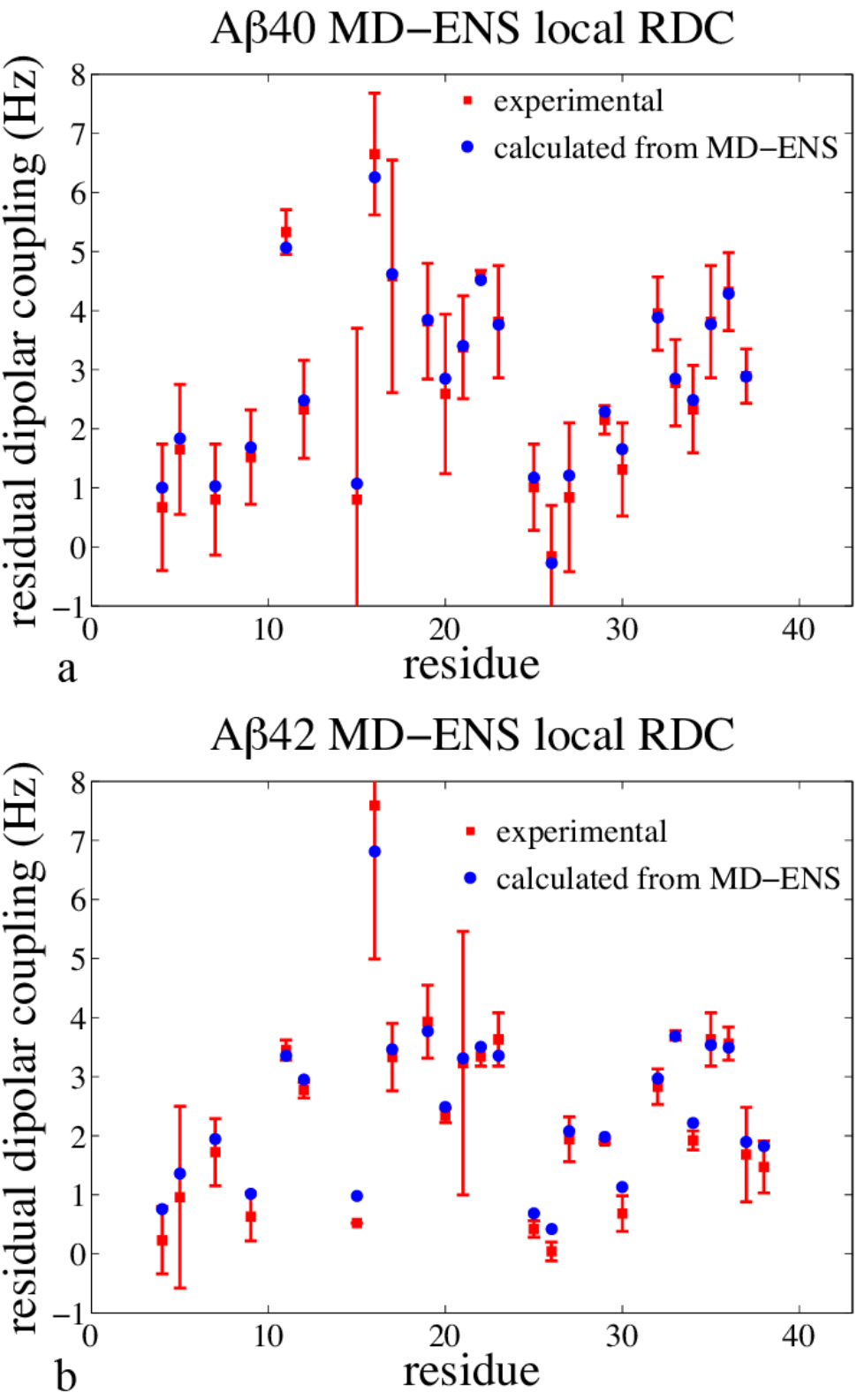


Figure 5.10. *Experimental vs. calculated residual dipolar couplings based on local alignments calculated from ENSEMBLE* (141) *for Aβ40 (a) and Aβ42 (b).* The red symbols are experimental data from Yan *et al*. (26). The blue circles are the data calculated from MD-ENS.

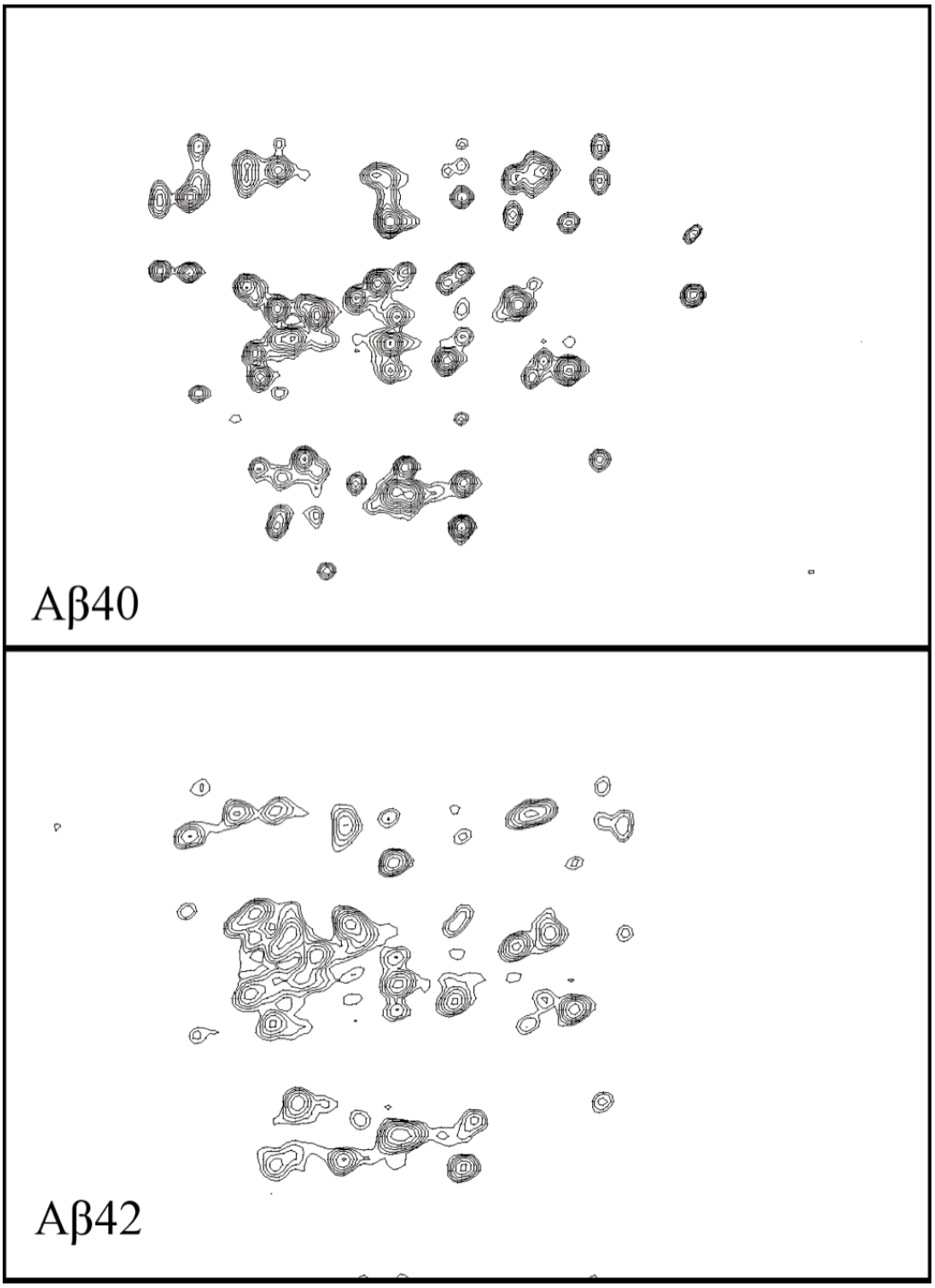


Figure 5.11. *Example NOESY spectra for (a) Aβ40 and (b) Aβ42.* This is a section of the NOESY spectra that contains the NOEs between H_α and H_N atoms. The Aβ40 spectrum and Aβ42 spectrum contain some of the same NOE peaks, but there are also some differences.

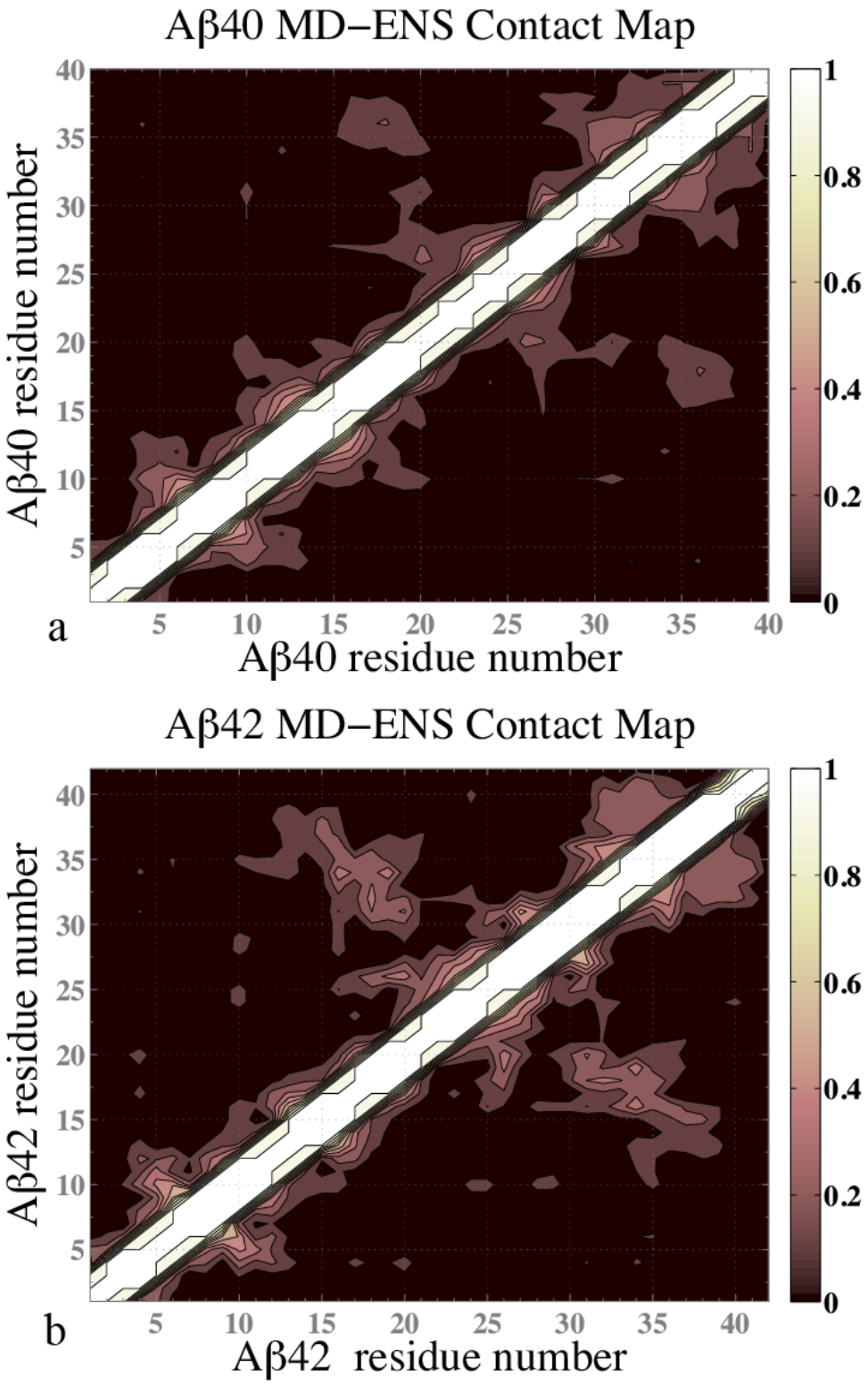


Figure 5.12. *Contact map of the MD simulated ensembles of (a) Aβ40 and (b) Aβ42.* This contact map gives the frequency of interaction between each pair of residues in the peptide *de novo* MD simulated ensembles. White indicates contacts present in 100% of the ensemble and black indicates contacts never seen in the ensemble. We define two residues to be in contact if any of their heavy atoms are within 5 Å of each other.

Chapter 6

Toxicity of an Aβ oligomer studied using human neurons derived from stem cells

While the monomeric form of amyloid-β (Aβ) is not associated with Alzheimer's disease toxicity, small soluble oligomers of amyloid-β 1-42 (Aβ42) are believed to be more toxic even than insoluble Aβ fibrils, and may be the agent of the disease. Because Alzheimer's Disease (AD) is among the most prevalent forms of dementia affecting the aging population, and pharmacological therapies to date have not been successful in preventing disease progression, we need models that enable further investigation of the toxicity of these oligomers and early disease pathology. In particular, disease-relevant models based on human pluripotent stem cells (hPSCs) may be promising approaches to assess the impact of different preparations of Aβ on specific neuronal populations and thereby facilitate the development of novel interventions to avert early disease mechanisms. Our collaborators have implemented an efficient paradigm to convert hPSCs into enriched populations of cortical glutamatergic or GABAergic neurons for this purpose. AD is generally known to be toxic to glutamatergic circuits, yet reports of its effects on GABAergic neurons both *in vitro* and *in vivo* are conflicting. We investigated the sensitivity of both of these cell populations to monomeric Aβ and to an oligomeric pre-fibrillar form of Aβ known as the "globulomer". Globulomers, which have shown strong correlation with the level of cognitive deficits in AD, were administered to human embryonic stem cell (hESC)-derived glutamatergic and GABAergic cells. Toxicity included culture age dependent binding of Aβ and cell death, and interestingly was selective for glutamatergic rather than GABAeric neurons, consistent with previous findings in postmortem human AD brain. This *in vitro* model of cortical neurons thus offers a system for future mechanistic investigation and therapeutic development for AD pathology, and the "globulomer" specifically induces AD-like toxicity on the same human cell types affected by this disease.

Introduction

AD is a neurodegenerative disorder characterized by an abundance of Aβ peptides generated from amyloid precursor abnormal cleavage by membrane-associated secretases (244). The development of AD pathology precedes cognitive symptoms and diagnosis by many years (245), presenting challenges for studying the role of Aβ in early disease stages to aid in the discovery of preventive drugs.

While the intrinsically disordered Aβ monomers are non-toxic, Aβ peptides, especially Aβ42, are extremely aggregation prone, forming large fibrils and plaques in the extra-cellular space of the AD affected brain (3). These plaques are the defining feature of AD; however, amyloid plaque load shows a weak correlation with dementia in AD (246). By comparison, soluble Aβ oligomer levels correlate more closely with AD pathology (12, 13). Thus, a stable oligomer of Aβ42 called the "Aβ globulomer" has been prepared and increasingly studied (68, 247). Human AD brain neuropathology studies show elevated levels of Aβ oligomers surrounding cortical neuronal processes, which may cause synaptic impairment (248). However, the specific neurotoxic effects of different types of Aβ oligomers on different cortical neuronal populations remain to be elucidated. Such results may enhance our understanding of this toxic oligomer's role in the disease process and aid future study of molecular mechanisms for its actions.

In vitro models derived from hPSCs offer strong platforms for basic research and subsequent therapeutic development for early stages of AD if hPSCs are differentiated into neurons affected by AD. In particular, glutamatergic neurons are severely afflicted in the cerebral cortex, and disruption of their circuits is associated with the hallmark memory deficits of AD (249, 250). By comparison, earlier studies of human postmortem AD brain suggest that cortical GABAergic neurons are spared from death (251). Cortical GABAergic neurons have not been broadly studied in animal AD models, but Aβ is neurotoxic to basal forebrain and hippocampal GABAergic neurons in such models (252, 253). To date, the sensitivity of human cortical glutamatergic versus GABAergic neurons to Aβ has not been studied with *in vitro* models. Given the differences in disease symptoms between familial AD and corresponding animal models (254), we have engaged in a collaboration with the Schaffer group (68, 70, 214, 215, 255-257) to apply an hPSC-based system to examine relative Aβ neurotoxicity in enriched populations of human cortical glutamatergic or GABAergic neurons.

The Aβ42 globulomer presents an appealing example of a toxic Aβ oligomer because an oligomer sample can be prepared that is uniform in size and molecular weight, and no aggregation of the sample occurs under physiological conditions over long time scales (68, 247). Globulomer toxicity and its affect on long-term potentiation have been studied *in vitro* and *in vivo*, and it has been shown to modulate calcium currents in rat, Xenopus and human embryonic kidney cells (68, 255-257). Immunochemistry has also shown that oligomers with a similar structure to the globulomer cluster around axon terminals in AD brains (258-260). However, no *in vitro* studies have yet shown that that globulomer is toxic to human neurons of the type involved in AD. Here, we employed human cortical cultures derived in the Schaffer lab and

enriched in glutamatergic or GABAergic neurons as models to investigate the neuronal phenotype selective neurotoxicity of Aβ globulomers in AD pathology. Interestingly, results show that Aβ globulomers exert a distinctive neurotoxicity in glutamatergic neurons in comparison to the GABAergic neuronal populations.

Globulomer Preparation

The Aβ42 globulomer was prepared as described (68, 115). Alkaline pretreatment of Aβ42 and preparation of low molecular weight Aβ by filtration protocols were used before beginning the globulomer preparation as previously described (115). After the 18-20 h incubation, the globulomer sample were concentrated to ~500 μM via centrifugation and dialyzed into PBS before centrifuging the sample at 10,000 g for 10 min to remove aggregates. The supernatant was saved, and the absorbance was measured at 276 nm wavelength to measure the concentration (extinction coefficient = 1390 M^{-1} cm^{-1}).

Aβ42 Monomer Preparation

The alkaline pretreatment of Aβ42 and preparation of LMW Aβ by filtration protocols outlined by Teplow (79) were also used to prepare a monomeric solution of the peptide. This involved dissolving 1 mg of the lyophilized peptide in 2 mL of 2 mM sodium hydroxide, sonicating for 2 minutes, and lyophilizing. This lyophilized peptide was then dissolved in 0.166 mL hexafluoroisopropanol (HFIP) to break any existing hydrogen bonds and prevent aggregation.

Evaluation of human cortical glutamatergic neurons as a model for AD investigation

We generated a soluble and highly stable form of Aβ oligomers termed globulomers, which are excluded from polymerization and assembly into larger fibrils (68, 247). Initially, neurons derived from adult rat hippocampal progenitors were used to determine a concentration range of this species that may be toxic, and globulomer preparations from Aβ peptide concentrations ranging from 1-5 μM were sufficient to induce cell death within 48 hours (Figure 6.1A).

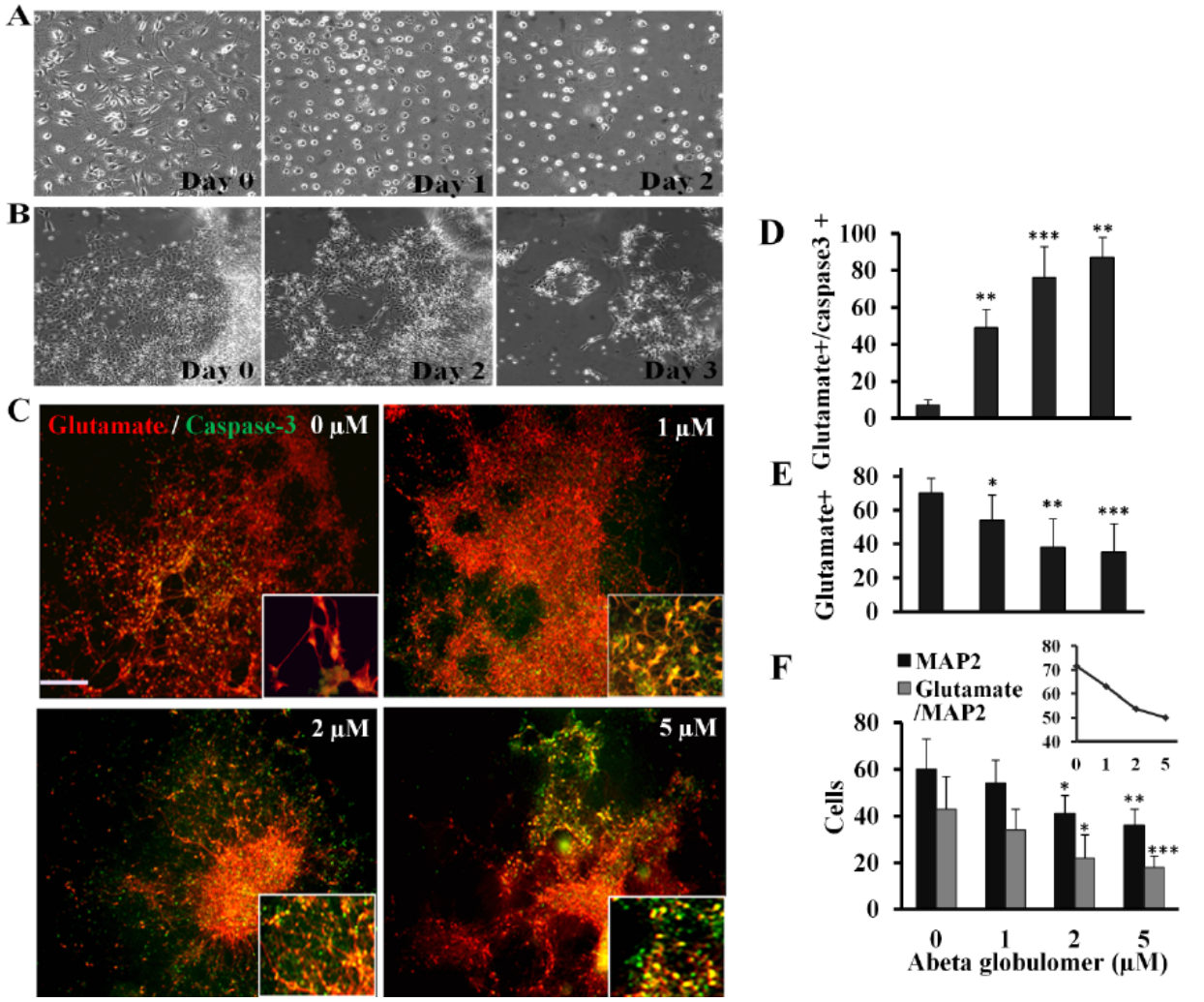


Figure 6.1. *Characteristic response of the neuronal populations to Aβ neurotoxicity.* (A) Representative phase contrast images of (A) adult rat hippocampal neurons and (B) hESC-derived dorsally patterned cortical neurons treated with 2µM Aβ in the globulomer form, exhibiting neuronal cell death after two and three days, respectively. (C) Glutamatergic neurons exhibited progressively higher cell death with increasing concentrations of the globulomeric form of Aβ, as measured by immunofluorescence staining of activated caspase-3. (D) Quantitative analysis demonstrated a concentration dependent onset of apoptosis in glutamatergic neurons treated with Aβ globulomers. As Aβ concentrations increased, (E) the percentages of glutamatergic neurons and (F) the fraction of neurons with a glutamatergic fate decreased. The line graph insert in panel F shows the progressive decrease of percentages of neurons that belonged to a glutamatergic phenotype with increasing Aβ concentrations. The increases in cell death and the decrease in the total percentages of glutamatergic neurons treated with 1, 2, and 5 µM Aβ in the globulomer form were statistically different compared to untreated cultures. The decrease in the number of MAP2 and glutamate expressing cells was significant in conditions treated with 2, and 5 µM Aβ, as compared to culture that were not treated with Aβ. * = $P < 0.005$, ** = $P < 0.001$, *** = $P < 0.0001$. Scale bar: 100 µm.

When hESC-derived cortical cultures primarily comprised of glutamatergic neurons were treated with 2 μM Aβ as globulomers, clear neuronal cell death initiated after 72 hours (Figure 6.1B). Analysis using caspase-3 as an apoptotic marker showed a progressive increase in glutamatergic neuron death with increasing Aβ globulomer concentration (Figure 6.1C, D), and the fraction of neurons (MAP2+) that were glutamatergic also progressively decreased (Figure 6.1E). Finally, the overall fraction of cells in the culture that were MAP2+ and that were glutamate+/MAP2+ decreased in parallel, suggesting a selective neurotoxicity of Aβ globulomers against these human cortical glutamatergic neurons (Figure 6.1F). By comparison, cultures treated with the same concentrations of the monomeric, non-toxic form of Aβ peptide did not show activation of caspase-3.

We also immunostained for the presence of this pre-fibrillar form of Aβ to further investigate cellular interactions, using an antibody against the Aβ fibrillar oligomeric form that also binds the soluble Aβ globulomers due to some structural similarity (11). In agreement with previous studies (68, 261), globulomer aggregates were tightly associated with the neuronal cell membrane (Figure 6.2A). Similarly, iPSC-derived neurons exhibited Aβ globulomer binding and subsequent apoptosis (data not shown). We also investigated whether neurons that had matured longer in culture may respond differentially to the Aβ globulomers, since organismal aging is the most prominent risk factor for AD, and neuronal lifetime in culture may embody some aspects of aging (262). When hESC-derived cortical neuronal populations were matured for an additional 33 days, for a total of 72 days after the initiation of differentiation, Aβ binding to the plasma membrane and the dendritic spines was significantly increased (Figure 6.2B). Specifically, quantitative assessments of 100 neurons showed an increase in the average number of bound globulomer aggregates per neuron from 3 ± 1 in younger cultures to 8 ± 2 in old ($P < 0.001$), indicating that increased cell culture longevity could offer some parallels to age-related increases in Aβ-induced pathology. Collectively, these findings illustrate that cortical neuronal cultures can be used *in vitro* to emulate and investigate AD phenotypes.

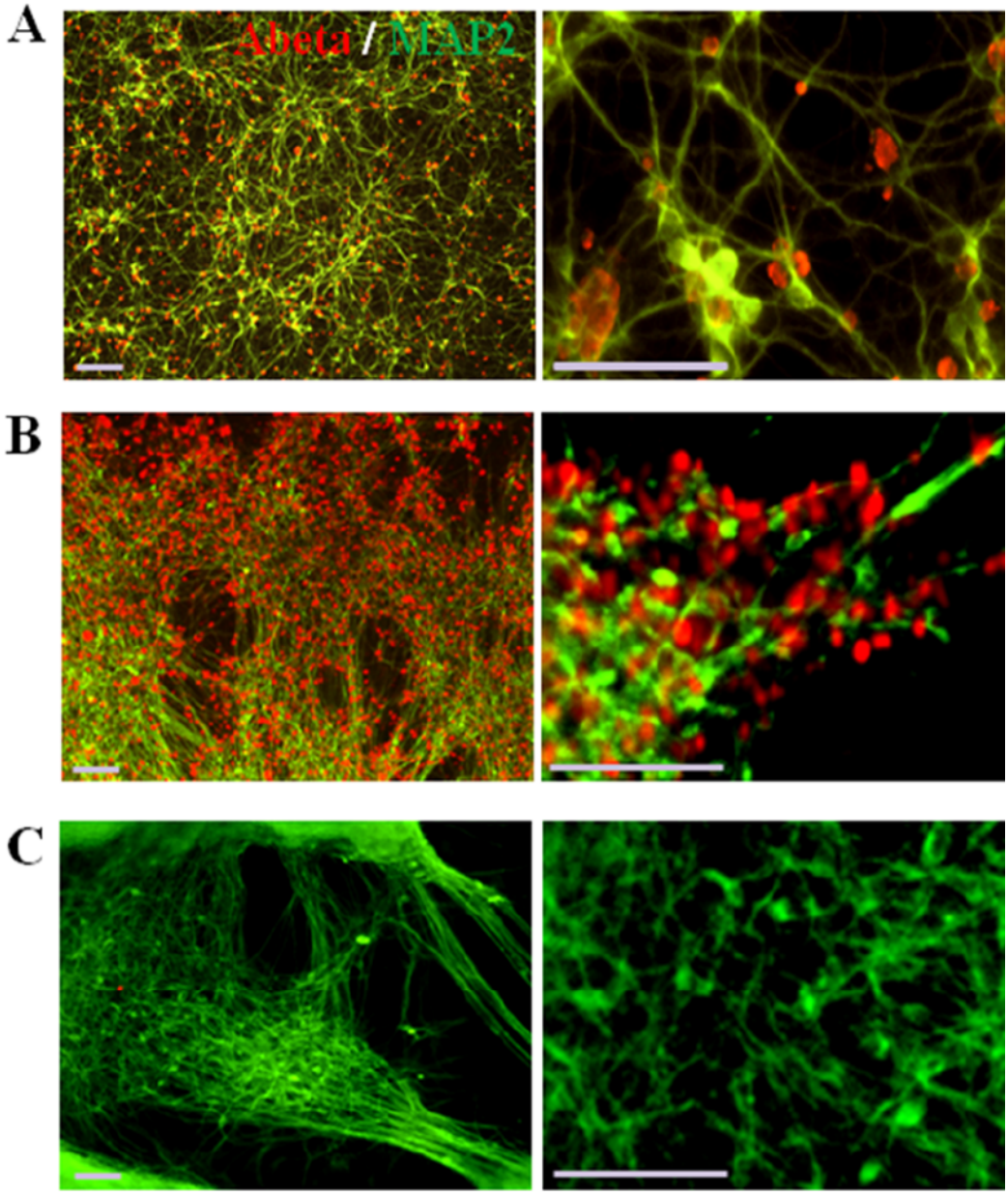


Figure 6.2. *Examination of Aβ binding affinity in aged glutamatergic neurons.* Cortical neurons were treated with 5 μM Aβ in the globulomer form for 24 hr, then immunostained with an anti-Aβ antibody. Binding of Aβ globulomers on the surface of MAP2+ neurons at (A) day 25, as compared to (B) day 58 of neuronal maturation illustrates enhanced interaction and binding of Aβ to aged neurons in culture. (C) The untreated control without Aβ showed no signal. Scale bar, 100 μm.

Examination of Aβ-induced neurotoxicity in glutamatergic and GABAergic populations

The extent to which specific sets of neurons in the cerebral cortex are susceptible to Aβ oligomer binding and toxicity is unknown. To study neuronal phenotype sensitivities to Aβ, we exposed mature hESC-derived cultures primarily comprised of glutamatergic or GABAergic neurons to Aβ globulomer and assessed globulomer binding and neuronal cell death after 72 hr. The levels of globulomer bound to MAP2-expressing neurons was significantly higher in glutamatergic neuronal cultures, differentiated from dorsal NPCs, than primarily GABAergic neurons, derived from ventral NPCs (Figures 6.3A-B, E). Likewise, gluatamergic and GABAergic cultures exhibited substantially different levels of apoptotic cell death (Figure 6.3C-E). In glutamatergic cultures exposed to Aβ, caspase-3 activation occurred in 61% of cells, compared to 33% in untreated conditions, and the fraction of overall MAP2+ neurons decreased from 56 to 31% (representing an overall 55% decrease) in response to Aβ treatment. In contrast, the GABAergic population was more resistant to Aβ, as the fraction of neurons decreased by only 17% after treatment, and neuronal caspase-3 staining was the same in treated vs. untreated cultures (Figure 6.3E).

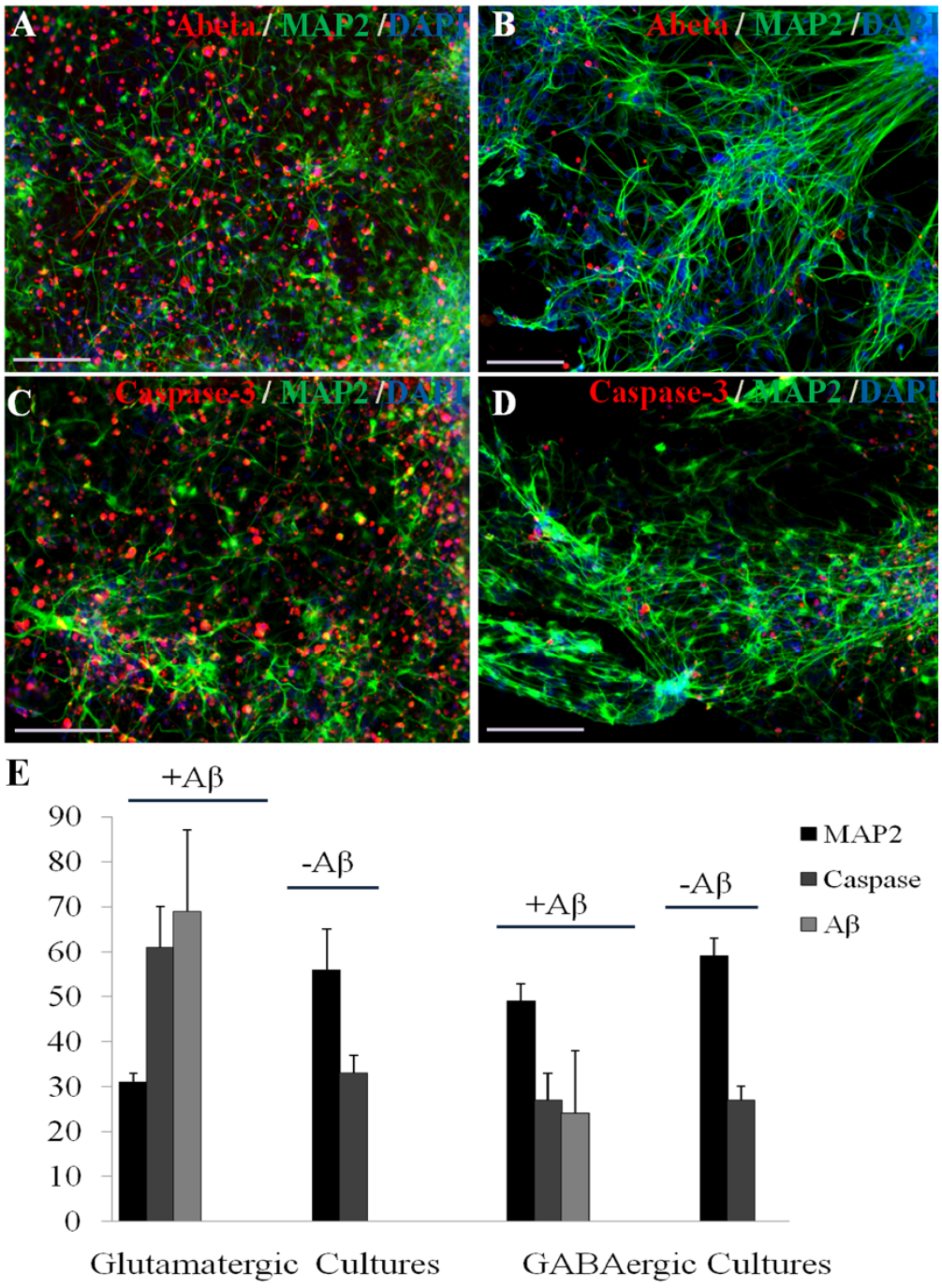


Figure 6.3. *Neuronal phenotype-specific Aβ binding and neurotoxicity.* Aβ and MAP2 staining of cultures derived from dorsal and ventral NPCs mainly giving rise to glutamatergic and GABAergic populations, respectively, exposed to globulomers illustrated enhanced binding of Aβ to the (A) glutamatergic populations as compared to (B) GABAergic cultures. The level of Caspase-3 activations was elevated in (C) glutamatergic versus (D) GABAergic cultures. (E) Quantitative analysis of the percentages of MAP2 expressing neurons, number of globulomer aggregates divided by number of cells stained with DAPI, and percentage of cells containing activated caspase-3 in glutamatergic and GABAergic neuronal populations in the presence or absence of Aβ as globulomers (5 μM) for 72 hr. Scale bar: 100 μm. (E) In conditions treated with 5 μM Aβ, the overall changes in percentages of MAP2+ and activated caspase-3+ neurons in glutamatergic cultures was significantly different from GABAergic cultures, ** = $P < 0.001$.

Discussion

Numerous transgenic animal models of AD have been generated to aid in understanding mechanisms of this human disease (254), and while these models continue to provide valuable insights into disease mechanisms, non-human systems often do not fully emulate human pathophysiology. For example, the need to overexpress multiple pathological proteins to exhibit AD pathology in these animals highlights the value of developing models that could enable investigation in human cells with gene expression closer to endogenous levels. Analogously, overexpressing genes that influence amyloidogenesis in neurons derived from immortalized cell lines, such as the frequently used human neuroblastoma line SH-SY5Y (263), may not represent an accurate model to study AD as these cells are quite different from neurons that degenerate in human AD brain. Finally, tissue from AD patients is both heterogeneous and limited. Such challenges can potentially be addressed by using models of hPSC-derived neuronal phenotypes that undergo degeneration in AD for basic investigation or therapeutic development. Human cortical glutamatergic neurons, which are strongly involved in the neuropathology of AD (249, 250), have remained largely unexplored in biomedical investigations of this disease. In this study, we have begun to evaluate the potential of human glutamatergic or GABAergic neurons as an AD model.

Although growing evidence shows that a soluble oligomer of Aβ is the toxic species associated with AD (12, 13), study of the mechanism of toxicity is still hampered by the polymorphism of Aβ oligomers and fibrils (11). The globulomer is a specific Aβ42 oligomer that has been experimentally characterized and is stable *in vitro* under physiological conditions (68), but has yet to be evaluated in terms of its toxicity towards human neuronal cultures. We used cortical glutamatergic neurons derived from hPSCs by our collaborators in the Schaffer group to study AD pathology *in vitro* by showing severe toxicity for the Aβ42 globulomer (70, 262). We see that this specific Aβ oligomer causes human neuronal cell death at concentrations of 2 μM. Our results add to the existing body of evidence showing that the Aβ42 globulomer is an important example of a toxic AD agent, and an understanding of its structure and mechanism of toxicity is important to developing AD therapies.

Our Aβ neurotoxicity analysis indicated a selective pathology against glutamatergic neurons, as the fraction of neurons of this phenotype decreased with increasing Aβ concentrations. This effect may be related to their capacity to bind the globulomers (Figure 6.3E). In fact, it has been suggested by studies performed in neuronal cell lines and rat primary hippocampal and cortical cultures that differential sensitivity to Aβ is associated with cell membrane Aβ binding (264). However, the binding affinity and selective sensitivity of different type of neurons to neurotoxic effects of Aβ remains a matter of debate and must be studied in greater detail.

Interestingly, the glutamatergic model illustrated enhanced binding of Aβ globulomers to individual neurons with increased time in culture, which may have some parallels with neuronal aging *in vivo* (262). The mechanisms and binding partners involved in the association of Aβ oligomers with cortical neurons have not been identified, though a number of membrane associated proteins – including NMDA receptors, integrins, and proteoglycans – have been identified as potential binding proteins in various types of neurons (261). An alternative

hypothesis for Aβ-induced cytotoxicity involves age-related disruption of the cell surface membrane through alterations of the membrane lipid layer fluidity that eventually lead to apoptotic or necrotic cell death (265, 266). In either case, the model presented in this study may offer a platform for advancing our understanding of age- related association and binding of Aβ to the plasma membranes as a potential point of intervention.

Hippocampal GABAergic neurons are susceptible to Aβ toxicity *in vitro* and decrease in number in the Alzheimer's disease TgCRND8 mouse (253, 267), and small GABAergic neurons were vulnerable to Aβ pentapeptide in primary rat basal forebrain cultures (252). By contrast, cortical GABAergic neurons were relatively spared in a study of postmortem tissue from AD patients (251). Given stark differences in the sensitivity of human and mouse brain to pathological alleles from familial forms of AD, it is important to examine Aβ biology and pathology in the types of human neurons directly impacted by AD, including cortical neurons. To our knowledge, however, the cell-type specific behavior of human cortical neurons has not been explored in an AD neurotoxic environment.

To shed some light on cortical neuronal phenotype-specific vulnerability in AD, we use a human cell model of glutamatergic and GABAergic neuronal cultures to illustrate GABAergic neuronal phenotype comparative immunity against Aβ. These results with human cells agree more closely with the cerebral cortex of postmortem human AD patients (251) than with animal studies conducted to date. This is also the first evidence that the Aβ42 globulomer is less toxic toward GABAergic human neuronal cultures as compared to glutamatergic neuronal cultures. The question of whether other Aβ oligomeric species display this same differential trend in relative toxicity toward different neuronal phenotypes as well as the mechanism of the globulomer toxicity remain to be investigated.

Conclusions

We have examined the effects of a pre-fibrillar form of Aβ knows as the “globulomer” on human pluripotent stem cell derived glutamatergic and GABAergic neurons, phenotypes that comprise a large fraction of the human cortical regions that undergo severe degeneration in AD. Our study clearly shows a human cortical neuronal phenotype-dependent binding of and susceptibility to Aβ globulomers. The ability to reproducibly generate large quantities of different human neuronal subtypes will help investigations of cellular and biochemical dynamics during early to late stages of AD- induced neurodegeneration, and further study of the globulomer will aid in identifying possible disease mechanisms and thereapies.

Chapter 7

Molecular dynamics simulations of a toxic Aβ42 oligomer

Soluble oligomers of amyloid-β 1-42 (Aβ42), such as the "globulomer", a globular oligomer consisting of 12 Aβ42 monomers with β secondary structure, are believed to be the toxic agent involved in Alzheimer's disease. Although one NMR study by Yu and coworkers (12, 13, 72) has resulted in a model for a globulomer dimer repeat, little structural characterization of the globulomer or evaluation of different possible models has been attempted thus far. Here, we compare to possible models for the globulomer structure using all-atom explicit solvent molecular dynamics simulations. The Yu *et al.* dimer model, based on chemical shift and NOE data collected from a globulomer precursor, the "preglobulomer", is the first of these two competing models. The other model is derived from coarse grain simulations conducted by Fawzi and coworkers on Aβ40 dodecamer protofibrils (69). We evaluate the stability of the Aβ42 β-sheet structure for two models during molecular dynamics simulations and calculate NMR observable values for each, which we can then compare quantitatively to current and future NMR experiments.

Introduction

Soluble amyloid-β (Aβ) oligomers have recently been identified as the probable toxic agent in Alzheimer's Disease (AD) (12, 13, 268). This conclusion is a result of growing evidence that the presence of large, insoluble Aβ fibrils does not correlate well with disease symptoms (246). Because the monomer is known to be nontoxic (268), an intermediate between the monomeric and fibril forms of the peptide is likely responsible for toxicity. This oligomer could be an early form of the larger fibril or a different structure that forms off-pathway to fibrilization. One soluble Aβ oligomer that has been recently prepared and characterized is the Aβ42 "globulomer" (68). This oligomer is a promising target for structural and mechanistic studies of AD pathology because is can be prepared consistently, has a somewhat uniform size, shape, and structure, and does not aggregate, even after long periods of time under physiological conditions (68, 247). Studies in animal models have suggested that the globulomer is toxic to the brain (255, 257, 269) and our work with human neuronal cultures (Chapter 6) (68, 70) have confirmed similar toxicity

patterns to those observed in AD patients. To understand the mechanism of this toxicity, and how it might be prevented, this globulomer species must be structurally characterized.

Initial structural characterization of the Aβ42 globulomer by size-exclusion chromatography has shown that it is made of approximately 12-16 Aβ42 monomer chains and exists in a single globular structure (68). Although the globulomer is not as compact as a typical globular protein, circular dichroism shows that it has stable β-sheet secondary structure. This structure most probably occurs at the C-terminal portion of the Aβ42 chains since the protease digestion results indicate that the N-termini are largely disordered and solvent exposed (68). Antibody binding assays have also been conducted on the globulomer, and while specific antibodies show higher binding to the Aβ42 globulomer than to Aβ42 fibril aggregates, other antibodies developed to bind Aβ fibrils also bind the globulomer, indicating some structural similarities to the fibril structure (68). Without atomic level structural information, however, it is difficult to identify the similarities and differences between the Aβ42 fibril and globulomer conformations. NMR has been a useful technique in our characterization of the Aβ monomers (48, 64, 65) because it reveals residue specific structural and dynamic information without subjecting the sample to disrupting experimental conditions. For these same reasons, NMR is an ideal method for building a detailed model of the toxic Aβ42 globulomer.

One NMR study of the globulomer has been published thus far by Yu *et al.* at Abbott Laboratories, who propose a globulomer structural model from their NMR data (115). Although the goal of their NMR study is the characterization of the Aβ42 globulomer, a dodecamer, they found that their preparations of N-Met-Aβ42 globulomers produced NMR spectra not suitable for structural studies. As a result, a large amount of their analyzed NMR data reports not on the globulomer, but on a smaller oligomer known as the "preglobulomer" which forms as a precursor to the globulomer during preparation and consists of ~4 Aβ42 chains stabilized in 0.2% SDS solution. Yu *et al.* report amide exchange data that indicate similar protection levels at each residue in the Aβ42 preglobulomer to those of the globulomer. NOE and chemical shift data from ^{15}N and ^{13}C labeled samples were then used to construct a structural model of the preglobulomer based on a peptide dimer repeating unit. Preparation of a mutant Aβ42 globulomer with a designed disulfide bridge indicates that the model proposed for the preglobulomer may also be present in the larger globulomer.

The NMR data and globulomer structure proposed by Yu *et al.* leave several open questions that must be addressed in order to have a complete structural picture of this toxic Aβ species. The most important question regards the similarities between the globulomer and preglobulomer structures. Although the amide exchange data for the two species appears to be similar, it is not identical and a quantitative comparison was not presented. In fact, any proposed model must necessarily contain differences between a tetrameric species stabilized by SDS (the preglobulomer) and a dodecamer with less than one SDS molecule present per complex. Although the same disulfide bond may be compatible with in both structures, there will be structural differences in the arrangement of Aβ chains, which has yet to be investigated. The model proposed by Yu *et al.* does not give a full picture the preglobulomer structure either. They propose a repeating dimer unit and build a structure consisting of only two Aβ42 chains using restraints from NMR data, however it has been shown that the dimer is not the most stable form of Aβ42 (270, 271). To form the preglobulomer tetramer, two of the modeled structures would

need to come together, but the resulting configuration is unclear. The dimer model presented is not symmetrical, however Yu *et al.* do not report two sets of chemical shifts for any of the residues, which would be expected if the monomer chains were not in identical environments. Finally, the construction of the Aβ42 preglobulomer dimer model from NMR restraints assumes that all of these restraints are satisfied simultaneously. Because the Aβ42 monomer is extremely disordered, switching between different conformational states on timescale faster than the NMR data collection (48, 65), it may be the case that the preglobulomer and globulomer are also somewhat disordered and flexible and that the NMR data only report on an average of an ensemble of structures that form. Such an ensemble would be difficult to construct from NMR data alone, but molecular dynamics simulations could allow characterization of several different Aβ42 globulomer structures.

A competing model to the Yu *et al.* Aβ42 globulomer structure could be a fibril-like oligomer. Despite the fact that the globulomer is off-pathway to fibril formation, many of its structural properties are consistent with the Aβ fibril structure. Both consist of β-sheet secondary structures with ordered and buried C-termini and flexible N-termini. Some globulomer antibodies have been shown to bind the Aβ fibril as well, which has led to classification of the globulomer as a fibrillar oligomer (11). Short segments of the Aβ fibril, protofibrils, may serve as a good starting model for investigating the Aβ42 globulomer structure. The structure and dynamics of protofibrils have been studied previously by Fawzi *et al.* with MD simulations using a coarse grain protein model (69). These simulations started from a Aβ40 fibril structure from solid-state NMR (21, 272) with varying numbers of monomer chains to simulate different lengths of Aβ protofibrils. The coarse grain model allowed a relatively large system consisting of multiple Aβ40 chains to be simulated over long timescales, revealing many possible conformational states. While protofibrils consisting of only 4 monomer chains were unstable and dissociated during the coarse grain simulations, larger protofibrils (> 14 chains) maintained fibril order over long timescales (69). At intermediate protofibril sizes, many structures in the simulation lose fibril order (i.e. cannot form a nucleus for fibrilization), but do not dissociate. Instead they form stable β-sheet structures which do not have a fibril-like cross section. This is the case for the protofibril dodecamer, which is close to the size determined as the critical nucleus for fibril formation (69). The dodecamer protofibril structures sampled in this study provide an alternative model to the NMR-based preglobulomer model, which represents all 12 chains of the globulomer and is stable in coarse grain simulations.

Here we investigate both the NMR-based preglobulomer model and the coarse grain simulation-based protofibril model as possible Aβ42 globulomer structures. Initially we have run atomic resolution *de novo* molecular dynamics simulations of the preglobulomer dimer structure generated by Yu *et al.* to assess its stability. We find that it does not retain its initial β-strand structure during the simulation in physiological conditions. The NMR-based preglobulomer model may be more stable as a tetramer, and we are currently constructing a tetramer model from the dimer structure, which will be tested for stability in future work. We have also created an atomic resolution Aβ42 dodecamer from one structure from the coarse grain Aβ40 protofibril dodecamer simulations by Fawzi *et al*. Based on initial simulations in physiological conditions, this model appears retain stable β-sheet structure. We are also interested in back-calculating NMR observables from these all-atom simulations. These can be used to compare with the currently available experimental data from Yu *et al.* as well as data collected in the future. We

can also see which types of NMR data are useful in distinguishing between the different globulomer models and design future experiments accordingly. We have already calculated amide exchange protection factors for the preglobulomer dimer structure from simulation and we have calculated chemical shifts and J-coupling constants from the protofibril dodecamer simulations. This work is ongoing.

Methods

Preglobulomer Simulation

Dimer structure preparation from NMR model. The starting Aβ42 dimer structure was taken from Yu *et al.* who have collected the only published NMR data on the globulomer and preglobulomer (115). NMR chemical shifts and NOE data were used to construct the dimer model. NOE contacts between amides were obtained by using ^{15}N-labeled peptide. Interchain and intrachain contacts were distinguished by mixing labeled peptide with an unlabeled sample in a 1:1 ratio and comparing to a homogeneous labeled mixture. The NOE's primarily seen in the homogeneous mixture indicate interchain interactions. Distance restraints were created based on ^{13}C- and ^{15}N-resolved NOE contacts. Chemical shifts were also used to determine the secondary structure and local environment of each residue in the peptide. Finally, the proposed preglobulomer dimer model of the soluble N-Met-Aβ(1-42) was generated using CNX by determining the lowest energy structure to satisfy all NOE distance restraints and φ dihedral angle restraints based secondary structure from chemical shifts. Although the preglobulomer is believed to be a tetramer, Yu *et al.* created a model that only represents two interacting Aβ42 chains, the assumption being that the preglobulomer is symmetrical (115).

After applying the restraints, Yu *et al.* came up with an ensemble of Aβ dimers, modeling residues 15-42 for each chain, and analyzed the 10 lowest energy models consistent with the NMR data (115). They omitted the first 14 residues since NOE restraints were not available for these residues, which indicates that they were not structured. Their structure was consistent with the NOE data in that there were intrachain antiparallel β-strands between residues 18-33 and interchain parallel strands at residues 34-40, with a β-turn between residues 23 and 30. We were given one of their Aβ15-42 dimer models and used Modeler (273-275) to construct the omitted N-terminus residues. The resulting Aβ42 dimer was the starting structure for our molecular dynamics simulations.

Dimer simulation. We performed MD simulations on the proposed preglobulomer dimer from Yu *et al.* to determine whether this structure is stable under physiological conditions and to compare NMR observables calculated from this model to the experimental NMR data. The peptide was modeled using the Amber ff99SF force field (23) and aqueous solvent represented by the TIP4P-Ew water model (25). To calculate equilibrium ensemble averages in the NVT ensemble, we used an Andersen thermostat, a leapfrog integrator with a 2.0 fs time step, and periodic boundary conditions. Our initial structure was built by using modeler to add residues 1-14 to Yu *et al.*'s low energy NMR structure. This structure was solvated in a rectangular water box such that there were 8.0 angstroms of water surrounding the peptide on all sides and six Na^+ ions to neutralize the charge of the peptide. The structure was minimized and equilibrated with a constant volume while the temperature was increased to 298 K and then equilibrated for 1.0 ns

under a constant pressure of 1 bar, maintained with a Berendsen barostat, to achieve the correct density.

Distance and angle restraints pertaining to β-strands were determined for the equilibrated structure based on the hydrogen bonds using ptraj. Due to the intrinsically disordered nature of Aβ42, it would not be satisfactory to look at our starting structure as a representation of the protein. Intrinsically disordered proteins are known to quickly sample a large number of conformations, so multiple starting structures that still obey our restraints must be included. In an attempt to increase conformational diversity of the Aβ42 dimer model, we heated the structure to 600 K in order to overcome energy barriers, applied our determined distance and angle restraints, and cooled the dimer at different points over the course of the high temperature simulation. This allowed for multiple variations on the starting structure that still followed the secondary structure of the original model. During the heating step the system was raised to 600 K, where the peptide adopted an extended conformation, from its starting temperature at 298K. Structures were then selected at 1 ns intervals and cooled while applying distance and angle restraints. The cooling with restraints was performed gradually over 1.0 ns with a timestep of 1.0 fs. This process was repeated until 11 different structures were obtained. We then equilibrated each structure at 298 K for a total of 20 ns.

Tetramer preparation. Construction of a tetramer model began by analyzing the distances between dimers in the Aβ40 fibril structure from a solid-state NMR study by Petkova *et al.* (272). We used VMD in order to determine the closest atomic distances at the interface of the two C-terminal β-strands that are stacked on top of each other in the fibril structure. We then used Packmol in order to orient two dimers so that their interchain parallel β-strands were aligned in an anti-parallel orientation with the odd numbered amino acid side chains of one dimer facing the odd number side chains of the other dimer. We then applied distance restraints based the fibril structure (272) to bring the two identical dimers together.

Secondary structure and stability analysis. The secondary structure and backbone stability of the preglobulomer simulated structures were analyzed using ptraj. β-strands were identified according to the DSSP definitions (97) and compared to the β-strand content determined from the NMR experiments. The backbone RMSD from the starting NMR model was determined for each snapshot in the simulation to asses backbone stability. We also performed RMSD analysis of residues 19-41 only since they were shown to form β-strands, while the N-terminus lacks any stable structure.

Amide exchange prediction. The amide protection factors were determined by 2-D $^{15}N/^{1}H$ HSQC spectra (115). The peptide was grown in an H_2O medium, then placed in a solution containing $^{2}H_2O$. Samples were flash-frozen with liquid nitrogen after different lengths of time in the $^{2}H_2O$ buffer, and 2-D $^{15}N/^{1}H$ HSQC spectra were taken as time progressed. These spectra were compared to an H_2O-only control and hydrogen exchange rates indicated the solvent exposure for each residue.

To investigate whether the proposed preglobulomer model is consistent with experimental amide protection factors, we predicted amide exchange protection from the MD simulations of the preglobulomer. The level of amide protection for each residue was determined by analyzing the amount of time a hydrogen bond occurred between two residues during the last

9 ns of each of 11 simulations. Hydrogen bonds for the backbone were defined using a 3.5 Å distance cutoff and a 60° angle cutoff. Residues that spend the most time engaged in hydrogen bonding with other Aβ residues are less able to undergo amide exchange with water and therefore are more protected.

Protofibril Simulation

Dodecamer structure preparation from coarse-grain simulation model. In previous work, Fawzi et. al showed that the critical nucleus for Aβ40 fibril formation is 12, and that dodecamers which have values of the fibril similarity parameter (χ) greater than 0.7 result in the formation of stable fibrils, while those less than 0.7 are disordered and do not form stable fibrils. We choose these disordered dodecamer structures for our simulations and postulate that these dodecamers are a good starting model for Aβ42 globulomer structure.

To predict the NMR observables based on this Aβ42 oligomer model, we needed to run MD simulations of the structures in atomistic detail. Therefore, we converted the coarse-grained protofibril structure to an all-atom representation. The first step was to build in backbone atoms based on the coarse-grain residue coordinates using Perl scripts. Then Modeller 9.10 (273-275) was used to add Ile41 and Ala42 to the C-terminus of each Aβ chain and also to optimize the backbone coordinates. Scwrl4 (276) was used to add the side chains and optimize their rotamers.

Dodecamer simulation. The all-atom Aβ42 dodecamer structure was solvated using TIP4P-Ewald water model (25). The minimization and equilibration of the structure was done using the PMEPD program as part of the Amber 11 software package (111) and the Amber ff99SB force field (23). First, the structure was minimized with 2000 steps of steepest descent followed by 2000 steps of conjugate gradient minimization, with 500.0 kcal/mol restraints on the protein, solvent, and ions Then we minimized further for 2000 steps of steepest descent followed by 2000 steps of conjugate gradient, with 50.0 kcal/mol restraints only on the protein. Next we equilibrated the structure at 10 K for 20 ps with 25.0 kcal/mol restraints on the protein complex , followed by heating the system from 10 to 298 K with 10.0 kcal/mol restraints over 40 ps and 20 ps further equilibration at 298 K. Next we gradually decreased the restraints to 1.0 kcal/mol over 40 ps and then to 0.1 kcal/mol restraints over another 40 ps. The last step was a 5.5 ns equilibration in the NPT ensemble. A 2 fs step size was used for the equilibration and production runs.

Secondary structure and stability analysis. RMSD (with reference to the pre-minimization and equilibration structure) and secondary structure of simulated Aβ42 protofibrils were analyzed using the ptraj module of Amber. The fibril similarity parameter, χ, was calculated using the method described by Fawzi *et al.* (69) for terminal B. The reference structure for χ calculation is the -2 staggered structure determined by Petkova *et al.* (272) from solid state NMR experiments. The summation,

$$\chi = \frac{1}{M}\sum_{a=1}^{3}\sum_{b=i+1}^{4}\sum_{i=1}^{10}\sum_{j=1}^{10} h\left(\varepsilon - \left|r_{a,i:b,j} - r^{0}_{a:i,;b,j}\right|\right), \tag{7.1}$$

is done for pairs of the 4 outermost monomer (*a* and *b*) on the terminal B, and *i* and *j* are the residues 17-21 and 30-34. M (=600) is the total number of pairs involved in the summation. r and r^0 are the distance between the C_α atoms for the dodecamer model and the reference structure. The tolerance is set by ε (0.5 Å in our case) and h is the heavyside function.

J-coupling constant prediction. J- coupling constants, $J_{H^N H^\alpha}$, were calculated using the Karplus equation,

$$J(\phi) = A\cos^2(\phi - 60) + B\cos(\phi - 60) + C, \tag{7.2}$$

with coefficients A = 6.51, B = −1.76, and C = 1.60 given by Vuister and Bax (93). It gives information about the dihedral angles. For glycine residues, the calculation was different because it has two H-atoms bonded to the backbone N;

$$J(\phi) = 0.5*[A\cos^2(\phi - 60) + B\cos(\phi - 60) + C + A\cos^2(\phi + 60) + B\cos(\phi + 60) + C]. \tag{7.3}$$

Chemical shift prediction. We used SHIFTX for to calculate chemical shifts for C_α, C_β, H_α, and H_N atoms. The chemical shift gives a measure of electron density, electronegativity of neighboring atoms and magnetic field effects, which also depend on the secondary structure of the residue.

Results

Preglobulomer Simulation

Structure and stability. Aβ42 is known to convert to its insoluble fibrillar form in most conditions. These fibrils have been structurally characterized and shown to have cross-β structure with in-register parallel β-sheets (21, 109, 272, 277-281). In addition, the diverse set of amyloid intermediates which correlate with AD symptoms also contain ordered β-sheet structure (11, 22). Therefore, any good model for the globulomer structure must contain stable β secondary structure. We monitored the β-strand structure of the preglobulomer dimer during our MD simulations (Figure 7.1 & 7.2). The amount of parallel and antiparallel β-strand structures decreased as the simulations progressed, but many of them disappeared and reappeared before eventually being lost. This may be because the β-strand interactions were still in place when fluctuations occurred that deformed the β-strand away the typical β-structure. The loss of β structure was more evident for the parallel β-strands than the anti-parallel β-strands. The RMSD of the backbone from the starting structure was higher than the data gathered by Yu *et al.* by a factor of 10 (Table 7.1). Even when comparing only those residues known to participate in β-strand formation and excluding the N-terminus, as Yu *et al.* did, the RMSD values were still large. Residues 19-32 did have a lower RMSD compared to the full structure, but not enough to indicated that the dimer structure is stable. The data gathered from these simulations reveal that the NMR-based model is not stable as a dimer. The formation of a tetramer structure may stabilize the β-strand structure observed experimentally.

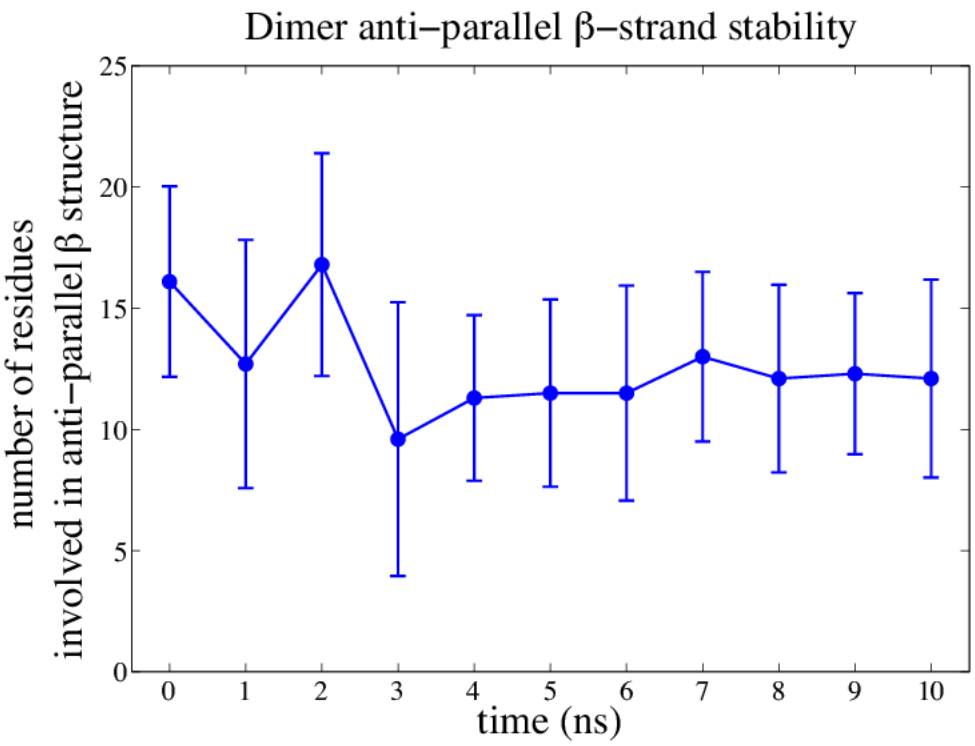


Figure 7.1. *The average number of residues characterized as part of an anti-parallel β-strand in 1 ns intervals during the 10 ns runs calculated over all residues in the dimer ensemble.*

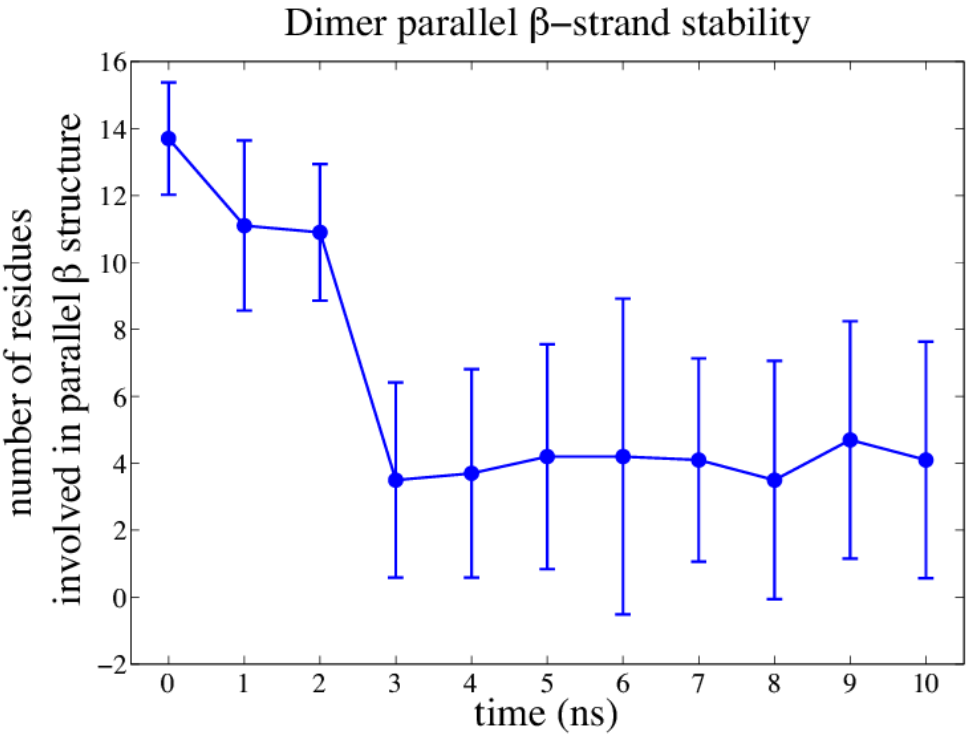


Figure 7.2. *The average number of residues characterized as part of a parallel β-strand in 1 ns intervals during the 10 ns runs calculated over all residues in the dimer ensemble.*

RMSD (Å)						
	residues 19-41		residues 19-32		residues 34-41	
Structure	**Post Restraints**	**Post 10ns**	**Post Restraints**	**Post 10ns**	**Post Restraints**	**Post 10ns**
1.0	10.7	9.7	8.6	7.4	6.0	4.8
2.0	12.5	10.8	11.1	10.3	6.5	5.8
3.0	10.5	9.5	7.0	7.5	6.8	5.9
4.0	10.2	9.9	9.3	9.2	6.4	5.9
5.0	10.3	11.7	8.8	11.4	6.8	6.9
6.0	10.6	10.9	11.7	12.0	6.2	5.8
7.0	10.2	9.1	10.6	9.0	5.6	4.9
8.0	10.1	9.3	11.3	10.1	6.5	6.0
9.0	10.0	9.9	11.1	11.3	6.1	5.4
10.0	10.5	9.9	11.5	10.6	6.5	5.6
control	12.2	10.1	8.7	8.2	6.0	5.1
average	**10.6**	**10.1**	**10.1**	**9.9**	**6.3**	**5.7**
Standard deviation	0.7	0.8	1.6	1.6	0.4	0.6
experimental		**2.4**		**1.1**		**0.5**

Table 7.1. The average backbone RMSD in Å of the dimer ensemble (residues 19-41) for the 11 starting structures before and after applying restraints with heating simulation was analyzed as a whole. In an effort to minimize these values we excluded the unstructured N-terminus and only exclude the residues that are known to adopt a β-strand structure.

Amide exchange prediction. The experimental amide exchange data show that the same residues were buried in the preglobulomer and globulomer C-termini. This may indicate that the globulomer consists of multiple preglobulomer subunits. The presentation of the amide exchange protection, however, is unclear and many of the exchange rates for the preglobulomer and globulomer residues could not be measured.

Hydrogen bond analysis of the final 9 ns of each of our 20 ns runs showed high variation (Figure 7.3). Due to the parameters used to defined a hydrogen bond, more than one oxygen atom was able to hydrogen bond to the same amide at once, causing the percentage of time a single residue spent hydrogen bonded to exceed 100%. However, this was seen in <1% of the residues examined. The large error bars indicate that our 11 different simulations had high variation in terms of the amount of time spent hydrogen bonding. Longer simulation times are required to achieve converged hydrogen bonding statistics. Even so, it can be observed that the average time spent in a hydrogen bond is quite low, often less than 60%, even near the C-terminus where the residues are involved in β-strands. This means that dimer model may be too disordered and solvent exposed to account for the amide exchange protection observed experimentally for the globulomer and preglobulomer. Although our hydrogen bond data is not yet converged, it does show a trend along the Aβ42 sequence, and so we have classified each residue of the Aβ42 chain into fast, medium, or slow exchange categories in order to compare to the experimental data reported by Yu *et al*. (72, 160) (Table 7.2). We find that even if we pick

the cutoffs for each of these classifications to best match the experimental data, there are still many discrepancies between the calculated and experimental protection factors for the dimer model. It may be that simulations with a tetrameric model will result in better agreement with experimental amide exchange data.

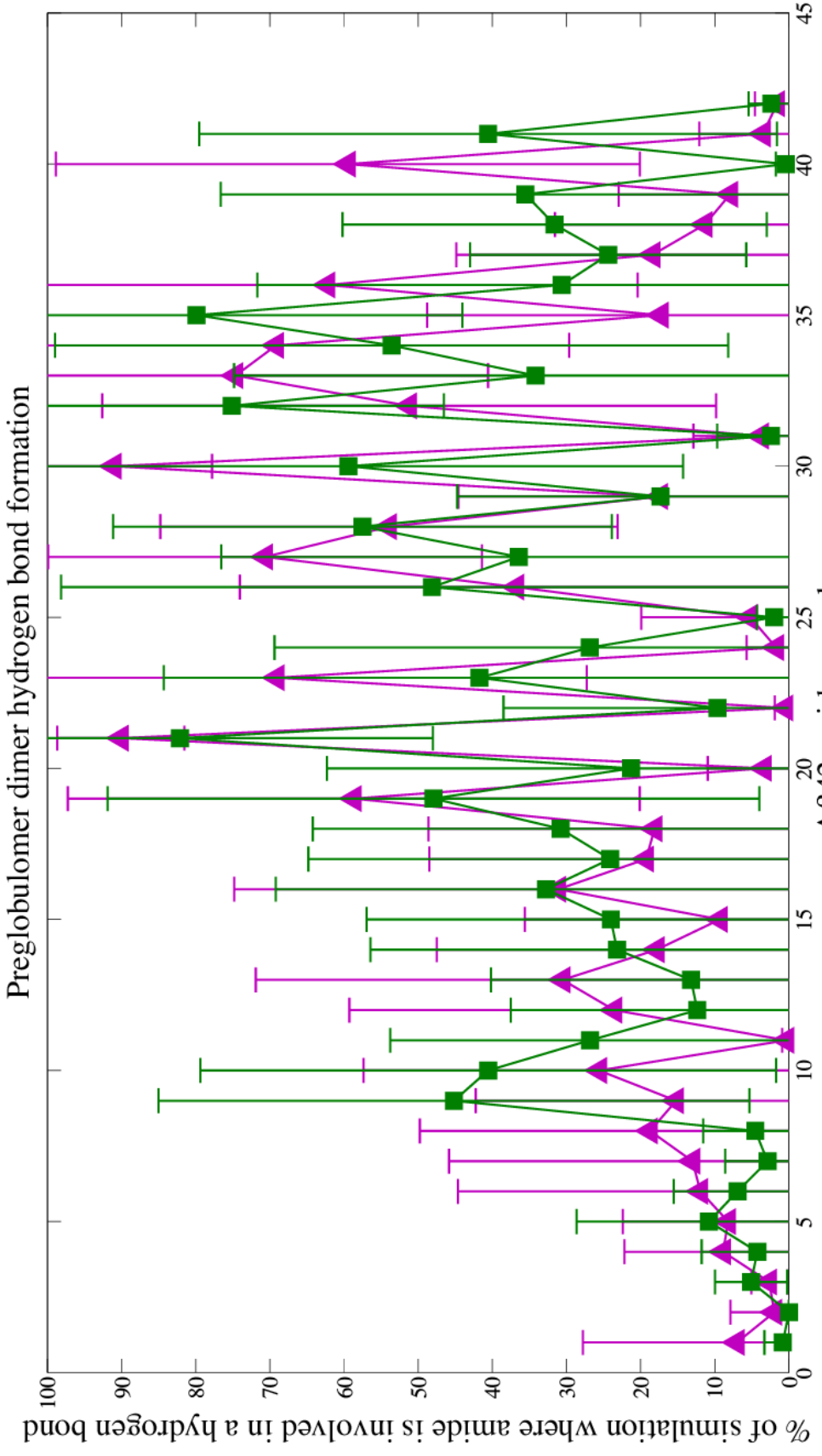


Figure 7.3. *The average percent of time a backbone nitrogen forms a hydrogen bond with a backbone oxygen over the last 9 ns of dimer simulation. A 3.5 Å maximum distance cutoff 60° angle cutoff were applied in order to define a H-bond.*

Residue number	Globulomer experiment	Preglobulomer experiment	Dimer simulation
1	x	f	f
2	f	f	f
3	f	f	f
4	f	f	f
5	f	x	f
6	f	f	f
7	x	f	f
8	x	f	m
9	f	f	s
10	f	x	s
11	g	x	m
12	f	x	m
13	f	x	m
14	x	x	m
15	f	x	m
16	f	f	s
17	f	s	m
18	g	s	m
19	m	f	s
20	f	f	m
21	f	f	s
22	f	x	f
23	g	x	s
24	f	f	m
25	f	f	f
26	f	x	s
27	f	x	s
28	f	x	s
29	f	f	m
30	f	m	s
31	s	s	f
32	g	s	s
33	s	s	s
34	s	s	s
35	s	s	s
36	g	x	s
37	s	s	m
38	s	s	m
39	s	s	m
40	s	x	m
41	g	m	m
42	f	m	f

Table 7.2. Comparison of calculated backbone amide exchange protection for the dimer chains and Yu *et al.*'s preglobulomer and globulomer data. The backbone amides that exhibited slow, moderate, and fast exchange are indicated by s, m, and f, respectively , below the residue number. Residues for which exchange rates could not be measured are indicated by an x. The sequence positions of the three pairs of overlapping cross-peaks of amides in slow exchange in globulomer are indicated by g. For the amide protection calculated from simulation, residues involved in hydrogen bonds less than 10% of the time were considered to be in fast exchange, 10-30% in medium exchange, and greater than 30% in slow exchange. These cutoffs were chosen to give the best possible agreement with the experimental data, although the true exchange chemistry would probably lead toward even worse agreement with the experimental data.

Protofibril Simulation

Structure and stability. After the minimization and equilibration protocol, we compared the RMSD, secondary structure and χ values for each step to figure out whether the structure remains stable in simulation (Figure 7.4). For the structure obtained from the coarse-grain model and Modeller, the RMSD is close to zero, showing that the backbone and side chain optimization steps do not change the backbone configuration significantly. The RMSD gradually increases with step number as we decrease the restraints on the protein complex and there is a sudden jump in the values after simulation in the NPT ensemble, where we lift off the restraints and start equilibration in the NPT ensemble. This is because the starting structure was obtained from a coarse grained model and without side chains the C_α atoms were much closer to each other than is possible in an all-atom representation. Also, the C-terminus of the Aβ peptide is more structured than the N-terminus and is involved in parallel-β sheet formation. We expect the N-terminus to be more disordered because it can move freely. Hence, the RMSD without the N-terminus is less than that with all the residues. Figure 7.5 shows that in the last 1 ns simulation in NPT ensemble the RMSD values become steady, indicating that the structure has equilibrated.

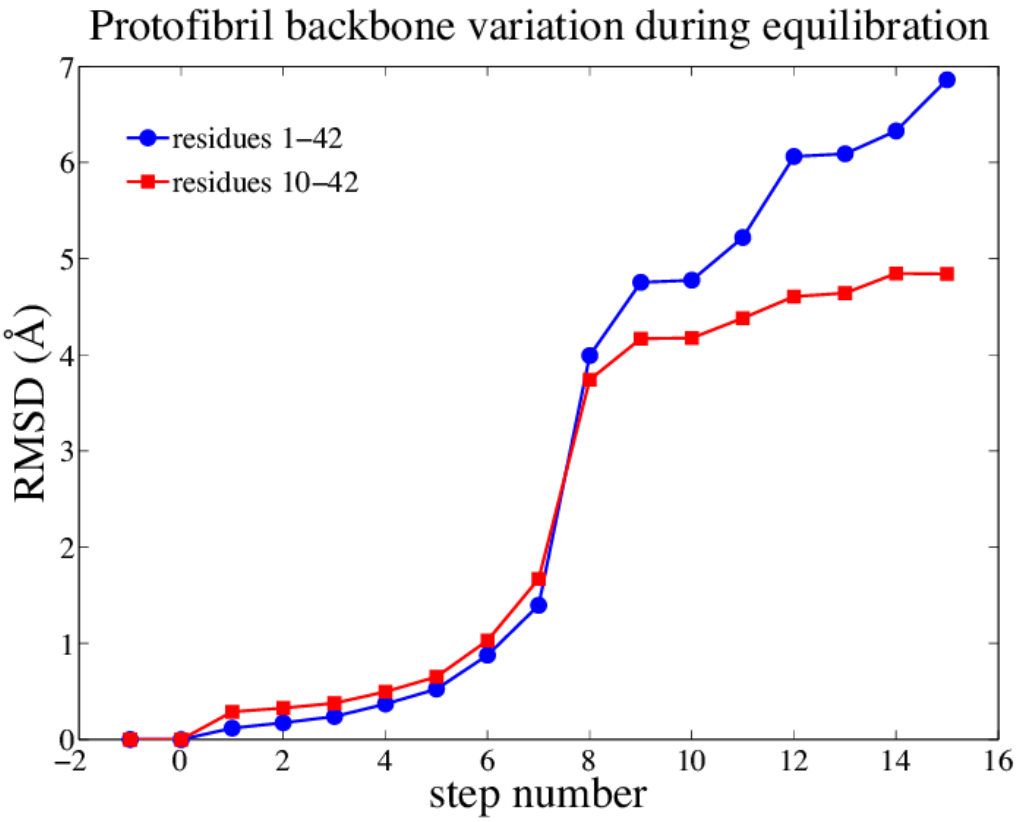


Figure 7.4. *Plot of the backbone RMSD (with reference to the Scrwl4-optimized step 0 structure) versus equilibration and minimization step number.* Step -2: Coarse Garin Structure, Step -1: Structure obtained from Modeller, Step 0: Structure obtained from Scrwl4, Step 1- Minimization of protein hydrogen atoms, ligand, solvents and ions for 2000 steps of steepest descent followed by 2000 steps of conjugate gradient, with 500.0 kcal/mol restraints on the protein heavy atoms, Step 2 -Minimization of solvents and ions for 2000 steps of steepest descent followed by 2000 steps of conjugate gradient, with 50.0 kcal/mol restraints on the protein-ligand complex , Step 3 - 20 ps equilibration at 10 K, with 25.0 kcal/mol restraints on the protein complex, Step 4 - 40 ps heat-up from 10 to 298 K followed by 20 ps equilibration at 298 K, Step 5- 40 ps equilibration at 298 K, with 10.0 kcal/mol restraints on protein complex, Step 6 - 40 ps equilibration at 298 K, with 1.0 kcal/molrestraints on protein complex, Step 7 - 40 ps equilibration at 298 K, with 0.1 kcal/mol restraints on protein complex, Step 8-16: The system was further equilibrated by running it in the NPT ensemble for 0.5 ns at each step.

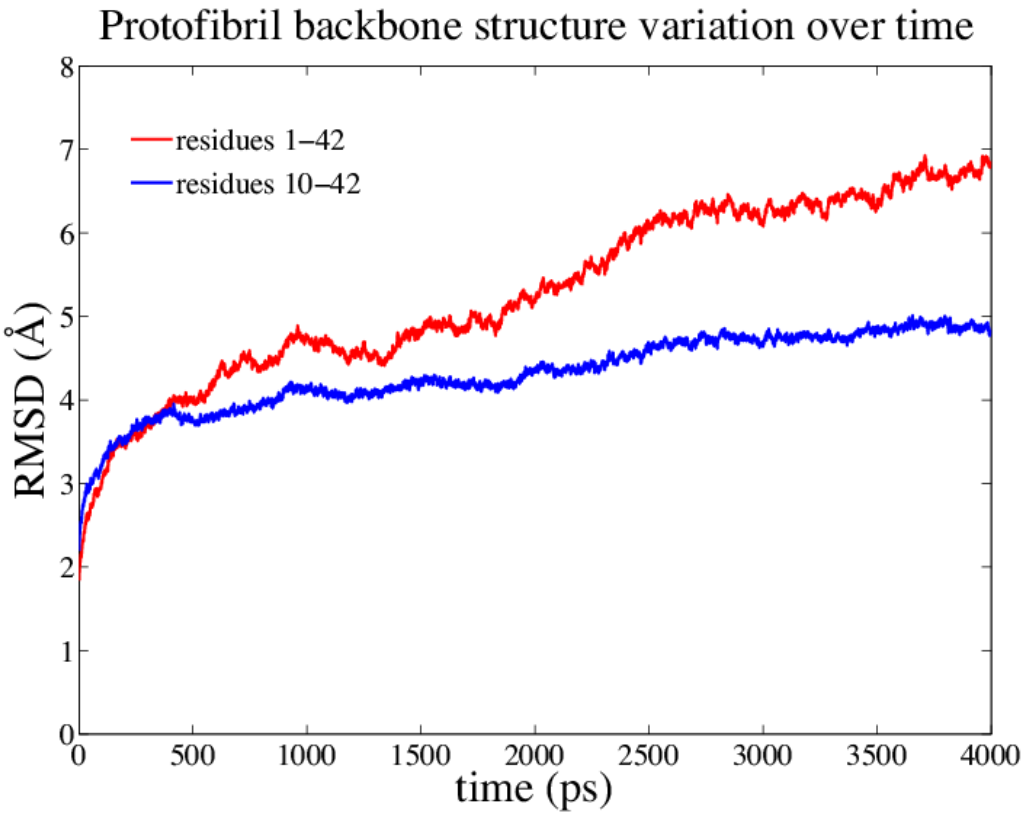


Figure 7.5. *Plot of backbone RMSD verses time for the NPT ensemble equilibration (steps 8-15).*

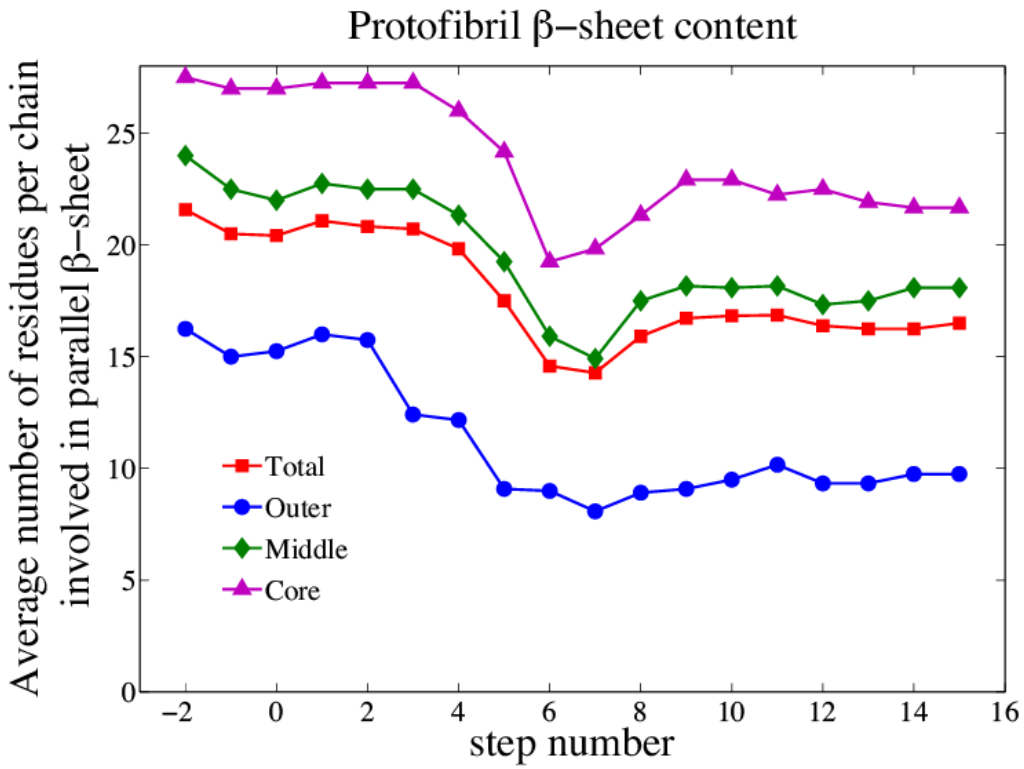


Figure 7.6. *Plot of parallel β-sheet content versus step number.* Total β is the average number of residues involved in β sheets in all 12 monomers, Outer corresponds to the average number of residues involved in β-sheets for the outermost 4 monomers, Middle for the 4 monomers present in the next inner layer, Core for the 4 innermost monomers, Step -2: Coarse Garin Structure, Step -1: Structure obtained from Modeller, Step 0: Structure obtained from Scrwl4, Step 1- Minimization of protein hydrogen atoms, ligand, solvents and ions for 2000 steps of steepest descent followed by 2000 steps of conjugate gradient, with 500.0 kcal/mol restraints on the protein heavy atoms, Step 2 -Minimization of solvents and ions for 2000 steps of steepest descent followed by 2000 steps of conjugate gradient, with 50.0 kcal/mol restraints on the protein-ligand complex, Step 3 -20 ps equilibration at 10 K, with 25.0 kcal/mol restraints on the protein complex , Step 4 - 40 ps heat-up from 10 to 298 K followed by 20 ps equilibration at 298 K, Step 5- 40 ps equilibration at 298 K, with 10.0 kcal/mol restraints on protein complex, Step 6 - 40 ps equilibration at 298 K, with 1.0 kcal/molrestraints on protein complex, Step 7 - 40 ps equilibration at 298 K, with 0.1 kcal/mol restraints on protein complex, Step 8-16: The system was further equilibrated by running it in the NPT ensemble for 0.5 ns at each step.

Figure 7.6 shows the average number of residues involved in parallel β-sheets after each of the minimization and equilibration steps. The core Aβ chains (the 4 innermost monomers) have the highest amount of β-sheet content, followed by the middle (the next layer of 4 monomers), and then by the outermost 4 chains. In the beginning of the simulation the structure looses its β-sheet content, possibly due to conversion from the coarse-grain model. However, the β-sheet content increases during simulation in the NPT ensemble, showing that the structure is stable and the individual Aβ chains do not disassociate from the protofibril. In the last 1.5 ns of NPT equilibration (steps 14-16), the β-sheet content remains almost constant, indicating that the structure has equilibrated. The fibril similarity parameter, χ, also decreases slightly with the minimization and equilibration steps (1-7) and then increases during NPT equilibration with no restraints (steps 8-15) and remains fairly constant in the last few steps of the simulation (Figure 7.7). Thus we can see from our analysis that the structure has equilibrated and is able to retain

stable β secondary structure. We simulate this structure in NVT ensemble for calculating the NMR observables.

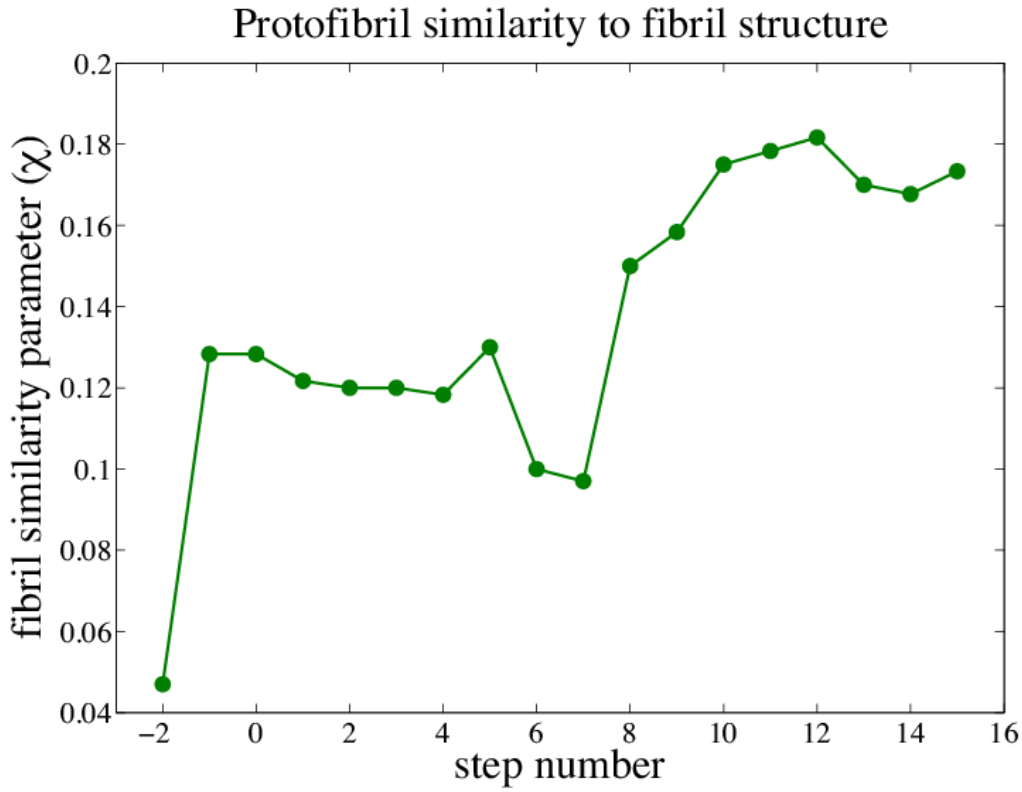


Figure 7.7. *Plot of Fibril Similarity Parameter (χ) versus step number.* Step -2: Coarse Grain Structure, Step -1: Structure obtained from Modeller, Step 0: Structure obtained from Scrwl4, Step 1: Minimization of protein hydrogen atoms, ligand, solvents and ions for 2000 steps of steepest descent followed by 2000 steps of conjugate gradient, with 500.0 kcal/mol restraints on the protein heavy atoms, Step 2 -Minimization of solvents and ions for 2000 steps of steepest descent followed by 2000 steps of conjugate gradient, with 50.0 kcal/mol restraints on the protein-ligand complex , Step 3 -20 ps equilibration at 10 K, with 25.0 kcal/mol restraints on the protein complex , Step 4 - 40 ps heat-up from 10 to 298 K followed by 20 ps equilibration at 298 K, Step 5- 40 ps equilibration at 298 K, with 10.0 kcal/mol restraints on protein complex, Step 6 - 40 ps equilibration at 298 K, with 1.0 kcal/molrestraints on protein complex , Step 7 - 40 ps equilibration at 298 K, with 0.1 kcal/mol restraints on protein complex, Step 8-16: The system was further equilibrated by running it in the NPT ensemble for 0.5 ns at each step.

Chemical shift prediction. From Figure 7.8 a and b we can see that the H_α and H_N chemical shifts for the Aβ42 protofibril vary significantly from residue to residue unlike monomeric Aβ42 for which they are almost constant (48, 64). The values are higher in the β-sheet region and decrease slightly around the loop region (residues 25-29), consistent with the fact that average H_α and H_N β-sheet chemical shifts are higher than random coil values (160). Negative C_α secondary chemical shifts correspond to β-strand (161), and we do see that our C_α shifts in Figure 7.8c are on average lower than random coil, although the trend is small. However, from 7.8d, we can see grouping of positive chemical shifts in the β-sheet region, which correctly represents the NMR structure predictions. Overall, the protofibril chemical shifts generally reflect known trends for β-sheet secondary structure, although for each residue the correlation is not perfect. We do not see a large variation in chemical shift between the different strands making up the protofibril.

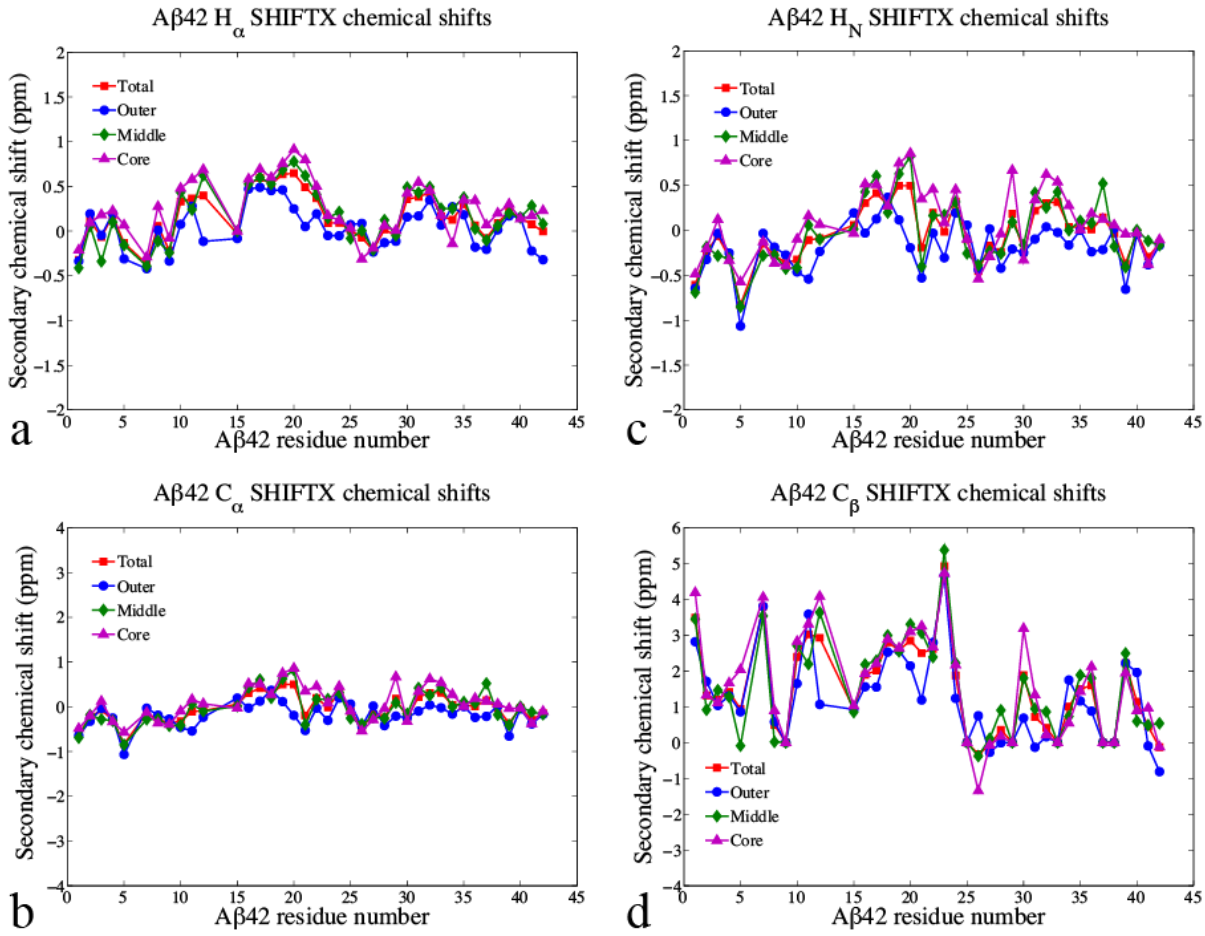


Figure 7.8. *H_α (a), H_N (b), C_α (c), and C_β (d) Chemical shifts calculated for the protofibril globulomer model versus residue number.* Random coil specific values are subtracted from the simulations value. Total is the calculated value averaged over all 12 monomers, Outer corresponds to the average over the outermost 4 monomers, Middle for the 4 monomers present in the next inner layer, Core for the 4 innermost monomers.

J-coupling constant prediction. In Figure 7.9, we can see that the J-coupling values are higher than the random coil or average value (around 5 Hz), as expected for β-sheet structure. Also, the values are higher for the β-sheet regions and drop down around the turn region (residues 23-28). The values for the different chains compare to each other in the β-sheet region, however differ in the termini, which is expected as the termini of the outer monomers have more freedom to move, followed by the middle, and the core monomers are the most ordered with little freedom to move. Overall, the J-coupling constants do not vary much over the different chains in the globulomer, indicating that this model would be consistent with uniform secondary structure for each of the Aβ chains.

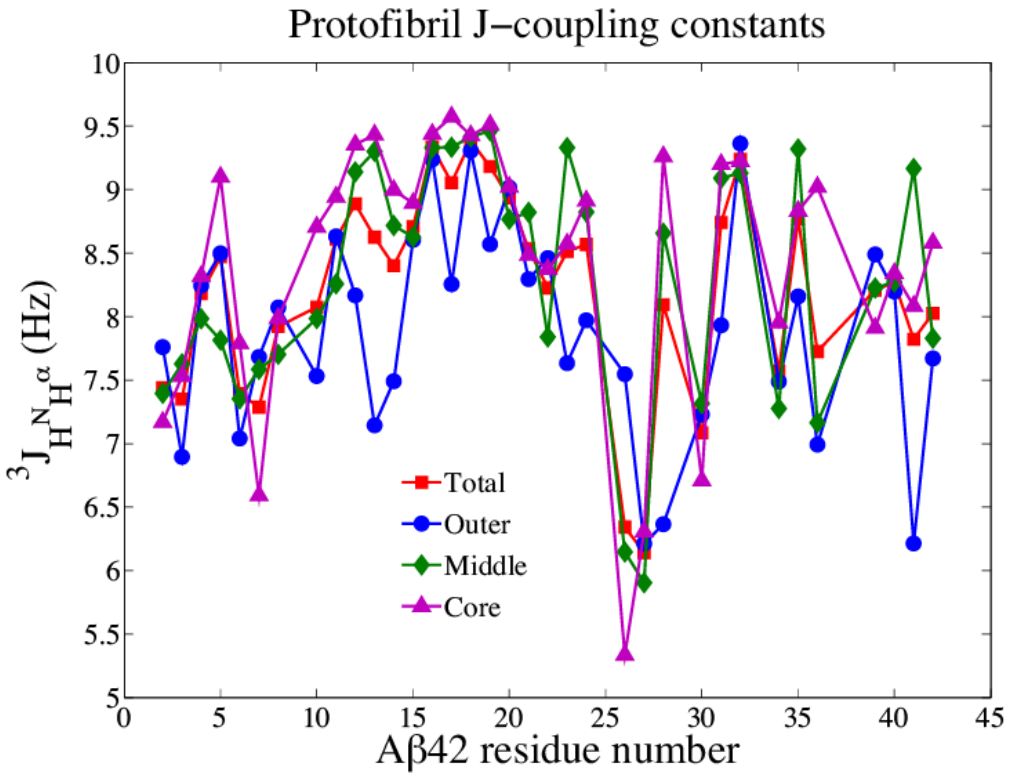


Figure 7.9. *J-coupling constants calculated for the protofibril globulomer model versus residue number.* Total is the calculated value averaged over all 12 monomers, Outer corresponds to the average over the outermost 4 monomers, Middle for the 4 monomers present in the next inner layer, Core for the 4 innermost monomers.

Discussion

We have done preliminary work to characterize the stability of two alternate structural models of the globulomer, a toxic Aβ42 oligomer. One model is based on NMR data collected on a globulomer precursor (72), the preglobulomer, which is thought to have similar structure to the globulomer. Our explicit water, all-atom MD simulations of the dimer repeat unit proposed in this model indicate that the structure is not stable. We are currently investigating whether the tetramer structure, consisting of two dimer repeats will have greater stability during simulations. The tetramer model, however, is still not a symmetrical model of the preglobulomer, which contradicts NMR data showing that each chain in the preglobulomer is in the same chemical environment. The competing model for the globulomer that we evaluate is based on coarse grain simulations of Aβ40 protofibrils (69), which indicate that the dodecamer protofibril adopts structures that will not nucleate extended fibrils. When we simulate an all-atom Aβ42 dodecamer structure derived from one of these protofibrils, we find that it is quite stable and retains β secondary structure over the course of the simulation. There are many protofibril structures from the coarse grain simulations that we could use as globulomer models, which have varying levels of order. It is likely that the globulomer samples many of these states, and we plan to create an ensemble of potential globulomer structures in the future, from which we can assess the average level of order for the model.

We have done initial calculations of NMR observables based on the all-atom simulations of both models. These calculated data can be quantitatively compared to existing and future NMR experiments. Differences between the calculated experimental observables from each model will allow us to differentiate between the two models experimentally. For the NMR-based preglobulomer dimer model we have examined the percent of the simulation time where backbone amide hydrogen atoms for each residue are present in order to gauge the amount of protection each residue has from solvent exchange. We find that both chains have similar levels of protection, with more exposed N-termini and protected C-termini, since only the C-termini are involved in β-strands. The high level of variation between our 11 independent simulations indicates that an ensemble of structures must be used to calculate any observable quantities for this preglobulomer model, since even structures with the same β-strand conformations can yield quantitatively varied data. We also find that we are unable to achieve a good match between the experimentally reported amide exchange rates and those calculated from the dimer model simulations, which may mean that this model needs to be modified, perhaps with the construction of a tetramer. However, it is difficult to do a careful comparison of the calculated and NMR data, since the amide exchange rates were not reported in a quantitative manner (68, 72). Further hydrogen exchange experiments would be beneficial for an effective assessment of amide protection in the various Aβ42 globulomer models.

From the dodecamer protofibril model we have calculated chemical shifts and $J_{H^N H^\alpha}$ J-coupling constants for each residue. The J-coupling constants report on the dihedral angles sampled by the protein, and are therefore directly related to secondary structure. We see that the J-coupling values are increased for residues participating in β-sheets and drop off for residues without β structure. This means that a J-coupling experiment would be useful in distinguishing between two possible globulomer models that contain β-sheet structure at different points in the chain sequence. However, J-coupling constants may not distinguish well between the NMR-based preglobulomer model and the protofibril model, because both models contain β-sheet structure at residues 18-23 and residues 28-40. One area that might distinguish between the two models is residues 13-17, which are involved in β-sheets in the protofibril structure, increasing their J-coupling values. In the NMR preglobulomer model, these residues are unrestrained and not designated as β-sheet residues, indicating that their J-coupling values will probably be lower. The J-coupling values may also help to distinguish the protofibril model from a model where each of the Aβ42 chains samples a different type of secondary structure. The protofibril model produces similar calculated J-coupling values for each of the chains, since they all adopt essentially the same secondary structure, with the end chains having slightly lower values due to the loss of β-sheet structure on the ends. Calculated chemical shifts for the protofibril model are also very similar for each monomer chain, indicating that experimentally the monomers could be indistinguishable and only set of Aβ42 chemical shifts would be observed experimentally. Although we also see a trend in the secondary chemical shift values corresponding to the stretches of sequence with β-sheet structure, this would be less useful in distinguishing between the protofibril and NMR preglobulomer models since the trend is not clearly defined for each residue.

In future work, we will calculate amide exchange data for the dodecamer protofibril model and chemical shifts and J-coupling values for the preglobulomer dimer structure as well as the tetramer. We can also calculate NOE intensities for both models. These values have been

measured for the preglobulomer by Yu *et al.* (48, 64, 65, 72), and therefore could be quantitatively compared to their experimental data as well as future experiments. After calculating expected values for all of these different types of NMR observables for both our alternative globulomer models, we will be able to identify the NMR experiments that can best distinguish between the two models. We can then design specific experiments using labeled Aβ42 peptides to collect data that will help us refine our globulomer model.

Conclusions

The investigation of Aβ oligomer structure is an important part of ongoing Alzheimer's disease research, given that a small soluble oligomer is believed to be responsible for AD toxicity. This is a challenging task because there are many possible polymorphic arrangements of the Aβ monomers which could compose such an oligomer, and many oligomers which can be studied experimentally. The Aβ42 globulomer is one oligomer that is experimentally shown to be relatively consistent in terms of size and structure, however, even it may also incorporate a considerable level of disorder and heterogeneity. Size exclusion chromatography data suggest that the globulomer is more loosely packed than globular folded proteins, much like a molten globule state (68). This indicates that a single, static model of the globulomer may be insufficient to describe its structure, as we know is the case for the Aβ42 monomer, which is an intrinsically disordered protein. Although the globulomer is not completely disordered and seems to have stable β secondary structure, it may be best represented by an ensemble of structures, which could be effectively generated from MD simulations.

These computationally generated Aβ oligomer structures should always be well validated with experimental evidence, as we have done with the Aβ monomer ensembles (48, 64, 65). As we discussed in Chapter 4, there are many ways of incorporating NMR data into the construction of an IDP ensemble, and the same methods should, in principle, apply to an ensemble of structures representing a disordered Aβ oligomer. It is likely that a similar fusion of the *de novo* MD and knowledge-based approaches that we used to construct the Aβ40 and Aβ42 monomer ensembles could be adapted to the Aβ42 globulomer. After using MD to generate a large number of possible globulomer structures, the available NMR data could be used to select the best ensemble of globulomer structures. Use of coarse grained simulations, as in the construction of our dodecamer protofibril model could also be very useful in the sampling of possible globulomer structures, since extensive MD sampling for a large oligomer would be very difficult using only all-atom MD simulations.

After using NMR data to validate a set of globulomer structures, a detailed analysis of the oligomer conformations can be carried out in the context of Alzheimer's disease and the formation of this toxic species. It will be interesting to determine whether any of the conformations we have already characterized in the Aβ monomer ensembles will appear in the context of the globulomer. If so, this indicates that globulomer formation involves a process of conformational selection whereby oligomerization occurs when Aβ42 monomers that are already adopting a globulomer-like structure come into contact, stabilizing the conformation through intermolecular interactions.

The comparison of the globulomer structure to the Aβ monomer ensembles may also lead to a better understanding of why only Aβ42, and not Aβ40, appears to form the toxic globulomer. One possible explanation lies in the major Aβ42 monomer β-hairpin formation that we describe in Chapter 5 (65). This structure includes a β-turn at residues 26-27, that is also present in the NMR-based preglobulomer model presented by Yu *et al.* (115). This turn is not present in the Aβ40 monomer ensemble, one possible explanation for why Aβ40 does not form the globulomer structure. Yu *et al.* also argue that the presence of this β-turn, rather than a larger bend, is the key factor that leads to globulomer, rather than fibril, formation for Aβ42. This leads to the question of how different polymorphic Aβ structures are formed, another question that we hope to address using oligomer models in combination with our well-characterized Aβ monomer ensembles. Conformational selection of different monomer structures may well be part of the explanation for the wide array of oligomer and fibril polymorphs observed for Aβ experimentally.

A complete structural picture of the Aβ monomeric and oligomeric populations will contribute to a broad understanding of the molecular aspects of Alzheimer's disease, providing a better guide for efforts to diagnose patients and prevent or treat disease symptoms.

Bibliography

1. Chiang PK, Lam MA, Luo Y (2008) The many faces of amyloid beta in Alzheimer's disease. *Curr Mol Med* 8:580–584.

2. Cavallucci V, D'Amelio M, Cecconi F (2012) Aβ toxicity in Alzheimer's disease. *Mol Neurobiol* 45:366–378.

3. Minati L, Edginton T, Bruzzone M, Giaccone G (2008) Current Concepts in Alzheimer's Disease: A Multidisciplinary Review. *American journal of Alzheimer's disease and other dementias*.

4. Dahlgren KN et al. (2002) Oligomeric and fibrillar species of amyloid-beta peptides differentially affect neuronal viability. *J Biol Chem* 277:32046–32053.

5. Gravina SA et al. (1995) Amyloid beta protein (A beta) in Alzheimer's disease brain. Biochemical and immunocytochemical analysis with antibodies specific for forms ending at A beta 40 or A beta 42(43). *J Biol Chem* 270:7013–7016.

6. Suzuki N et al. (1994) An increased percentage of long amyloid beta protein secreted by familial amyloid beta protein precursor (beta APP717) mutants. *Science* 264:1336–1340.

7. Iwatsubo T et al. (1994) Visualization of A beta 42(43) and A beta 40 in senile plaques with end-specific A beta monoclonals: evidence that an initially deposited species is A beta 42(43). *Neuron* 13:45–53.

8. Younkin S (1995) Evidence that Aβ42 is the real culprit in Alzheimer's disease. *Ann Neurol*.

9. Jarrett J, Berger E (1993) The carboxy terminus of the. beta. amyloid protein is critical for the seeding of amyloid formation: Implications for the pathogenesis of Alzheimer's disease. *Biochemistry*.

10. Hasegawa K, Yamaguchi I, Omata S, Gejyo F, Naiki H (1999) Interaction between A beta(1-42) and A beta(1-40) in Alzheimer's beta-amyloid fibril formation in vitro. *Biochemistry* 38:15514–15521.

11. Glabe CG (2008) Structural classification of toxic amyloid oligomers. *J Biol Chem* 283:29639–29643.

12. Selkoe DJ (2001) Alzheimer's disease: genes, proteins, and therapy. *Physiol Rev* 81:741–766.

13. Lambert MP et al. (1998) Diffusible, nonfibrillar ligands derived from Abeta1-42 are potent central nervous system neurotoxins. *Proc Natl Acad Sci USA* 95:6448–6453.

14. Hou L et al. (2004) Solution NMR studies of the A beta(1-40) and A beta(1-42) peptides establish that the Met35 oxidation state affects the mechanism of amyloid formation. *J Am Chem Soc* 126:1992–2005.

15. Uversky VN (2009) Intrinsic disorder in proteins associated with neurodegenerative diseases. *Front Biosci* 14:5188–5238.

16. Fisher CK, Stultz CM (2011) Constructing ensembles for intrinsically disordered proteins. *Current Opinion in Structural Biology* 21:426–431.

17. Wright PE, DYSON H (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293:321–331.

18. Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27:527–533.

19. Hilbich C, Kisters-Woike B, Reed J, Masters CL, Beyreuther K (1991) Aggregation and secondary structure of synthetic amyloid beta A4 peptides of Alzheimer's disease. *J Mol Biol* 218:149–163.

20. Riek R, Güntert P, Döbeli H, Wipf B, Wüthrich K (2001) NMR studies in aqueous solution fail to identify significant conformational differences between the monomeric forms of two Alzheimer peptides with widely different plaque-competence, A beta(1-40)(ox) and A beta(1-42)(ox). *Eur J Biochem* 268:5930–5936.

21. Petkova AT et al. (2002) A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci USA* 99:16742–16747.

22. Chimon S et al. (2007) Evidence of fibril-like β-sheet structures in a neurotoxic amyloid intermediate of Alzheimer's β-amyloid. *Nat Struct Mol Biol* 14:1157–1164.

23. Hornak V et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65:712–725.

24. Wickstrom L, Okur A, Simmerling C (2009) Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophys J* 97:853–856.

25. Horn HW et al. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* 120:9665.

26. Fawzi NL et al. (2008) Structure and dynamics of the Abeta(21-30) peptide from the interplay of NMR experiments and molecular simulations. *J Am Chem Soc* 130:6145–6158.

27. Bowman GR, Voelz VA, Pande VS (2011) Taming the complexity of protein folding. *Current Opinion in Structural Biology* 21:4–11.

28. Piana S, Lindorff-Larsen K, Shaw DE (2013) Atomic-level description of ubiquitin folding. *Proc Natl Acad Sci USA* 110:5915–5920.

29. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314:141–151.

30. Okur A, Roe DR, Cui G, Hornak V, Simmerling C (2007) Improving convergence of replica-exchange simulations through coupling to a high-temperature structure reservoir. *J Chem Theory Comput* 3:557–568.

31. Ruscio JZ, Fawzi NL, Head-Gordon T (2010) How Hot? Systematic Convergence of the Replica Exchange Method Using Multiple Reservoirs. *J Comput Chem* 31:620–627.

32. Kirschner DA et al. (1987) Synthetic peptide homologous to beta protein from Alzheimer disease forms amyloid-like fibrils in vitro. *Proc Natl Acad Sci USA* 84:6953–6957.

33. Halverson K, Fraser PE, Kirschner DA, Lansbury PT (1990) Molecular determinants of amyloid deposition in Alzheimer's disease: conformational studies of synthetic beta-protein fragments. *Biochemistry* 29:2639–2644.

34. Fraser PE et al. (1991) Morphology and antibody recognition of synthetic beta-amyloid peptides. *J Neurosci Res* 28:474–485.

35. Spencer RG et al. (1991) An unusual peptide conformation may precipitate amyloid formation in Alzheimer's disease: application of solid-state NMR to the determination of protein secondary structure. *Biochemistry* 30:10382–10387.

36. Barrow C, Yasuda A, Kenny P (1992) Solution conformations and aggregational properties of synthetic amyloid [beta]-peptides of Alzheimer's disease:: Analysis of circular dichroism spectra. *Journal of molecular ….*

37. Zagorski MG, Barrow CJ (1992) NMR studies of amyloid beta-peptides: proton assignments, secondary structure, and mechanism of an alpha-helix----beta-sheet conversion for a homologous, 28-residue, N-terminal fragment. *Biochemistry* 31:5621–5631.

38. Fraser PE et al. (1992) Fibril formation by primate, rodent, and Dutch-hemorrhagic analogues of Alzheimer amyloid beta-protein. *Biochemistry* 31:10716–10723.

39. Lansbury P, Costa P, Griffiths J (1995) Structural model for the β-amyloid fibril based on interstrand alignment of an antiparallel-sheet comprising a C-terminal peptide. *Nature Structural & ….*

40. Malinchik SB, Inouye H, Szumowski KE, Kirschner DA (1998) Structural analysis of Alzheimer's beta(1-40) amyloid: protofilament assembly of tubular fibrils. *Biophys J* 74:537–545.

41. Benzinger TL et al. (1998) Propagating structure of Alzheimer's beta-amyloid(10-35) is parallel beta-sheet with residues in exact register. *Proc Natl Acad Sci USA* 95:13407–13412.

42. Balbach JJ et al. (2002) Supramolecular structure in full-length Alzheimer's beta-amyloid fibrils: evidence for a parallel beta-sheet organization from solid-state nuclear magnetic resonance. *Biophys J* 83:1205–1216.

43. Sawaya MR et al. (2007) Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* 447:453–457.

44. Rodziewicz-Motowidło S et al. (2007) Conformational solution studies of the SDS micelle-bound 11-28 fragment of two Alzheimer's beta-amyloid variants (E22K and A21G) using CD, NMR, and MD techniques. *Biopolymers* 87:23–39.

45. Grant MA et al. (2007) Familial Alzheimer's disease mutations alter the stability of the amyloid beta-protein monomer folding nucleus. *Proc Natl Acad Sci USA* 104:16522–16527.

46. Takano K (2008) Amyloid beta conformation in aqueous environment. *Current Alzheimer research* 5:540–547.

47. Istrate AN et al. (2012) NMR solution structure of rat aβ(1-16): toward understanding the mechanism of rats“ resistance to Alzheimer”s disease. *Biophys J* 102:136–143.

48. Ball KA et al. (2011) Homogeneous and heterogeneous tertiary structure ensembles of amyloid-β peptides. *Biochemistry* 50:7612–7628.

49. Roher AE et al. (1993) beta-Amyloid-(1-42) is a major component of cerebrovascular amyloid deposits: implications for the pathology of Alzheimer disease. *Proc Natl Acad Sci USA* 90:10836–10840.

50. Schneider R et al. (2012) Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Mol Biosyst* 8:58–68.

51. Vise P, Baral B, Stancik A, Lowry DF, Daughdrill GW (2007) Identifying long-range structure in the intrinsically unstructured transactivation domain of p53. *Proteins* 67:526–530.

52. Marsh JA et al. (2007) Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. *J Mol Biol* 367:1494–1510.

53. Allison JR, Várnai P, Dobson CM, Vendruscolo M (2009) Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. *J Am Chem Soc* 131:18314–18326.

54. Marsh JA, Forman-Kay JD (2009) Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *J Mol Biol* 391:359–374.

55. Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM (2005) Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J Am Chem Soc* 127:476–477.

56. Richter B, Gsponer J, Várnai P, Salvatella X, Vendruscolo M (2007) The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J Biomol NMR* 37:117–135.

57. Lindorff-Larsen K et al. (2004) Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme a binding protein. *J Am Chem Soc* 126:3291–3299.

58. Krzeminski M, Marsh JA, Neale C, Choy W-Y, Forman-Kay JD (2013) Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* 29:398–399.

59. Huang A, Stultz CM (2008) The effect of a DeltaK280 mutation on the unfolded state of a microtubule-binding repeat in Tau. *PLoS Comput Biol* 4:e1000155.

60. Yoon M et al. (2008) Residual structure within the disordered C-terminal segment of p21(Waf1/Cip1/Sdi1) and its implications for molecular recognition. *Protein Sci*.

61. Salmon L et al. (2010) NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc* 132:8407–8418.

62. Jensen MR, Salmon L, Nodet G, Blackledge M (2010) Defining Conformational Ensembles of Intrinsically Disordered and Partially Folded Proteins Directly from Chemical Shifts. *J Am Chem Soc* 132:1270–.

63. Choy WY, Forman-Kay JD (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J Mol Biol* 308:1011–1032.

64. Ball KA, Wemmer DE, Head-Gordon T (2013) Determining the Structural Ensemble of Intrinsically Disordered Disease Peptides using Computation and Experiment. *J Am Chem Soc*.

65. Ball KA, Phillips AH, Wemmer DE, Head-Gordon T (2013) Differences in β-strand Populations of Monomeric Aβ40 and Aβ42. *Biophys J.*

66. Yan Y, Wang C (2006) Abeta42 is more rigid than Abeta40 at the C terminus: implications for Abeta aggregation and toxicity. *J Mol Biol* 364:853–862.

67. Yan Y, McCallum SA, Wang C (2008) M35 oxidation induces Abeta40-like structural and dynamical changes in Abeta42. *J Am Chem Soc* 130:5394–5395.

68. Barghorn S et al. (2005) Globular amyloid beta-peptide oligomer - a homogenous and stable neuropathological protein in Alzheimer's disease. *J Neurochem* 95:834–847.

69. Fawzi NL, Okabe Y, Yap E-H, Head-Gordon T (2007) Determining the critical nucleus and mechanism of fibril elongation of the Alzheimer's Abeta(1-40) peptide. *J Mol Biol* 365:535–550.

70. Vazin T et al. Efficient derivation of cortical glutamatergic and GABAergic neurons from human pluripotent stem cells: A model system to study neurotoxicity in Alzheimer's disease.

71. Dunker AK et al. (2001) Intrinsically disordered protein. *J Mol Graph Model* 19:26–59.

72. Liping Yu et al. Structural Characterization of a Soluble Amyloid β-Peptide Oligomer.

73. Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. *Current Opinion in Structural Biology* 18:756–764.

74. Uversky V, Gillespie J, Fink A (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41:415–427.

75. Uversky VN, Dunker AK (2010) Understanding protein non-folding. *Biochim Biophys Acta* 1804:1231–1264.

76. Romero P et al. (2001) Sequence complexity of disordered protein. *Proteins* 42:38–48.

77. Dunker AK, Obradovic Z (2001) The protein trinity--linking function and disorder. *Nat Biotechnol* 19:805–806.

78. Lu J-X, Yau W-M, Tycko R (2011) Evidence from solid-state NMR for nonhelical conformations in the transmembrane domain of the amyloid precursor protein. *Biophys J* 100:711–719.

79. Teplow D (2006) Preparation of Amyloid β-Protein for Structural and Functional Studies. *Methods in Enzymology* 413:20–33.

80. Keller R (2004) *The Computer Aided Resonance Assignment Tutorial*.

81. Bax A, Davis D (1985) Practical aspects of two-dimensional transverse NOE spectroscopy. *J Magn Reson* 63:207–213.

82. Liu M, Mao X, Ye C, Huang H, Nicholson J (1998) Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy. *Journal of Magnetic ….*

83. Dhalluin C, Wieruszeski J (1995) Use of a water flip-back pulse in the homonuclear NOESY experiment. *J Biomol NMR* 5:327–331.

84. Piotto M, Saudek V, Sklenár V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2:661–665.

85. Sklenar V, Piotto M, Leppik R (1993) Gradient-tailored water suppression for 1H-15N HSQC experiments optimized to retain full sensitivity. *Journal of Magnetic ...*.

86. Best RB, Buchete N-V, Hummer G (2008) Are current molecular dynamics force fields too helical? *Biophys J* 95:L07–9.

87. Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* 113:9004–9015.

88. Nerenberg PS, Head-Gordon T (2011) Optimizing Protein-Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *J Chem Theory Comput* 7:1220–1230.

89. Sgourakis NG, Yan Y, McCallum SA, Wang C, Garcia AE (2007) The Alzheimer's peptides Abeta40 and 42 adopt distinct conformations in water: a combined MD / NMR study. *J Mol Biol* 368:1448–1457.

90. Sgourakis NG et al. (2011) Atomic-level characterization of the ensemble of the Aβ(1-42) monomer in water using unbiased molecular dynamics simulations and spectral algorithms. *J Mol Biol* 405:570–583.

91. Xu XP, Case DA (2001) Automated prediction of 15N, 13Calpha, 13Cbeta and 13C' chemical shifts in proteins using a density functional database. *J Biomol NMR* 21:321–333.

92. Karplus M, Grant D (1959) A criterion for orbital hybridization and charge distribution in chemical bonds. *Proc Natl Acad Sci USA* 45:1269–1273.

93. Vuister G, Bax A (1993) Quantitative J correlation - A new approach for measuring homonuclear 3-bond J(H(N)H(alpha) coupling-constants in N-15-enriched proteins. *J Am Chem Soc* 115:7772–7777.

94. Peter C, Daura X, van Gunsteren WF (2001) Calculation of NMR-relaxation parameters for flexible molecules from molecular dynamics simulations. *J Biomol NMR* 20:297–310.

95. Mallows C (1973) Some comments on C p. *Technometrics*.

96. Zweckstetter M, BAX A (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *J Am Chem Soc* 122:3791–3792. Available at: http://www.cs.duke.edu/~brd/Teaching/Bio/asmb/Papers/NMR/Residual-dipolar-coupling/bax-zwecksetter-JACS00.pdf.

97. Kabsch W, Sander C (1983) Dictionary of protein secondary structure - pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.

98. Merutka G, Dyson H, Wright P (1995) Random coil H-1 chemical-shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J Biomol NMR* 5:14–24.

99. Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. *J Biomol NMR* 26:215–240.

100. Ulrich EL et al. (2007) BioMagResBank. *Nucleic Acids Res* 36:D402–D408.

101. Cavanagh J (2007) Protein NMR spectroscopy: principles and practice.

102. Wang C private communication.

103. Tycko R (2011) Solid-state NMR studies of amyloid fibril structure. *Annu Rev Phys Chem* 62:279–299.

104. Massi F, Peng JW, Lee JP, Straub JE (2001) Simulation study of the structure and dynamics of the Alzheimer's amyloid peptide congener in solution. *Biophysj* 80:31–44.

105. Tarus B, Straub JE, Thirumalai D (2006) Dynamics of Asp23-Lys28 salt-bridge formation in Abeta10-35 monomers. *J Am Chem Soc* 128:16159–16168.

106. Flory PJ (1989) *Statistical Mechanics of Chain Molecules*.

107. Kohn JE et al. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* 101:12491–12496.

108. Fitzkee NC, Rose GD (2004) Reassessing random-coil statistics in unfolded proteins. *Proc Natl Acad Sci USA* 101:12497–12502.

109. Lührs T et al. (2005) 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc Natl Acad Sci USA* 102:17342–17347.

110. Case DA et al. (2008) AMBER 10.

111. Case DA et al. (2010) AMBER 11.

112. Hukushima K, Nemoto K (1996) Exchange Monte Carlo method and application to spin glass simulations. *J Phys Soc Jpn* 65:1604–1608.

113. Glenner GG, Wong CW (1984) Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochemical and Biophysical Research Communications* 120:885–890.

114. Masters CL et al. (1985) Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc Natl Acad Sci USA* 82:4245–4249.

115. Yu L et al. (2009) Structural characterization of a soluble amyloid beta-peptide oligomer. *Biochemistry* 48:1870–1877.

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

117. Bernstein SL et al. (2009) Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat Chem* 1:326–331.

118. Fisher CK, Stultz CM (2011) Protein structure along the order-disorder continuum. *J Am Chem Soc* 133:10022–10025.

119. Lim KH, Collver HH, Le YTH, Nagchowdhuri P, Kenney JM (2007) Characterizations of distinct amyloidogenic conformations of the Abeta (1-40) and (1-42) peptides. *Biochemical and Biophysical Research Communications* 353:443–449.

120. Melquiond A, Dong X, Mousseau N, Derreumaux P (2008) Role of the region 23-28 in Abeta fibril formation: insights from simulations of the monomers and dimers of Alzheimer's peptides Abeta40 and Abeta42. *Current Alzheimer research* 5:244–250.

121. Baumketner A et al. (2006) Structure of the 21-30 fragment of amyloid beta-protein. *Protein Sci* 15:1239–1247.

122. Teplow DB et al. (2006) Elucidating amyloid beta-protein folding and assembly: A multidisciplinary approach. *Acc Chem Res* 39:635–645.

123. Yang M, Teplow DB (2008) Amyloid beta-protein monomer folding: free-energy surfaces reveal alloform-specific differences. *J Mol Biol* 384:450–464.

124. Ma B, Nussinov R (2002) Stabilities and conformations of Alzheimer's beta-amyloid peptide oligomers (A beta(16-22 ‘) A beta(16-35 ’) and A beta(10-35)): Sequence effects. *Proc Natl Acad Sci USA* 99:14126–14131.

125. Yan Y, Liu J, McCallum SA, Yang D, Wang C (2007) Methyl dynamics of the amyloid-beta peptides Abeta40 and Abeta42. *Biochemical and Biophysical Research Communications* 362:410–414.

126. Dunker AK, Gough J (2011) Sequences and topology: intrinsic disorder in the evolving universe of protein structure. *Current Opinion in Structural Biology* 21:379–381.

127. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* 37:215–246.

128. Wright PE, Dyson HJ (2009) Linking folding and binding. *Current Opinion in Structural Biology* 19:31–38.

129. Chen J (2012) Towards the physical basis of how intrinsic disorder mediates protein function. *Archives of Biochemistry and Biophysics* 524:123–131.

130. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z (2002) Intrinsic disorder and protein function. *Biochemistry* 41:6573–6582.

131. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* 323:573–584.

132. Uversky VN, Oldfield CJ, Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit* 18:343–384.

133. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197–208.

134. Xie H et al. (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res* 6:1882–1898.

135. Oldfield CJ et al. (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* 9 Suppl 1:S1.

136. KENDREW JC et al. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181:662–666.

137. Henderson R et al. (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J Mol Biol* 213:899–929.

138. Lange OF, Schäfer LV, Grubmüller H (2006) Flooding in GROMACS: accelerated barrier crossings in molecular dynamics. *J Comput Chem* 27:1693–1702.

139. Mittag T et al. (2008) Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc Natl Acad Sci USA* 105:17772–17777.

140. Esteban-Martín S, Fenwick RB, Salvatella X (2010) Refinement of ensembles describing unstructured proteins using NMR residual dipolar couplings. *J Am Chem Soc* 132:4626–4632.

141. Marsh JA, Baker JMR, Tollinger M, Forman-Kay JD (2008) Calculation of residual dipolar couplings from disordered state ensembles using local alignment. *J Am Chem Soc* 130:7804–7805.

142. Montalvao RW, De Simone A, Vendruscolo M (2012) Determination of structural fluctuations of proteins from structure-based calculations of residual dipolar couplings. *J Biomol NMR* 53:281–292.

143. Felitsky DJ, Lietzow MA, Dyson HJ, Wright PE (2008) Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc Natl Acad Sci USA* 105:6278–6283.

144. Ganguly D, Chen J (2009) Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states. *J Mol Biol* 390:467–477.

145. Fawzi NL, Ying J, Torchia DA, Clore GM (2010) Kinetics of amyloid beta monomer-to-oligomer exchange by NMR relaxation. *J Am Chem Soc* 132:9948–9951.

146. Vendruscolo M (2007) Determination of conformationally heterogeneous states of proteins. *Current Opinion in Structural Biology* 17:15–20.

147. Lawrence CW, Bonny A, Showalter SA (2011) The disordered C-terminus of the RNA polymerase II phosphatase FCP1 is partially helical in the unbound state. *Biochemical and Biophysical Research Communications* 410:461–465.

148. Marsh JA, Forman-Kay JD (2011) Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. *Proteins*.

149. Marsh JA et al. (2010) Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. *Structure* 18:1094–1103.

150. Prompers JJ, Brüschweiler R (2002) General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J Am Chem Soc* 124:4522–4534.

151. Zandarashvili L, Li D-W, Wang T, Brüschweiler R, Iwahara J (2011) Signature of mobile hydrogen bonding of lysine side chains from long-range 15N-13C scalar J-couplings and computation. *J Am Chem Soc* 133:9192–9195.

152. Case DA (2002) Molecular dynamics and NMR spin relaxation in proteins. *Acc Chem Res* 35:325–331.

153. Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* 319:209–227.

154. López-Méndez B, Güntert P (2006) Automated protein structure determination from NMR spectra. *J Am Chem Soc* 128:13112–13122.

155. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160:65–73.

156. Schwieters CD, Kuszewski JJ, Clore GM (2006) Using Xplor-NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy* 48:47–62.

157. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44:213–223.

158. Fisher CK, Huang A, Stultz CM (2010) Modeling intrinsically disordered proteins with bayesian statistics. *J Am Chem Soc* 132:14919–14927.

159. Click TH, Ganguly D, Chen J (2010) Intrinsically disordered proteins in a physics-based world. *Int J Mol Sci* 11:5292–5309.

160. Wishart DS, Sykes BD (1994) Chemical shifts as a tool for structure determination. *Methods in Enzymology* 239:363–392.

161. Wishart DS, Sykes BD (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR* 4:171–180.

162. Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* 15:2795–2804.

163. Camilloni C, De Simone A, Vranken WF, Vendruscolo M (2012) Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry* 51:2224–2231.

164. Lim KH, Henderson GL, Jha A, Louhivuori M (2007) Structural, dynamic properties of key residues in Abeta amyloidogenesis: implications of an important role of nanosecond timescale dynamics. *Chembiochem* 8:1251–1254.

165. Mohana-Borges R, Goto NK, Kroon GJA, Dyson HJ, Wright PE (2004) Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *J Mol Biol* 340:1131–1142.

166. Mukrasch MD et al. (2007) Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. *J Am Chem Soc* 129:5235–5243.

167. Sitkoff D, Case DA (1997) Density Functional Calculations of Proton Chemical Shifts in Model Peptides. *J Am Chem Soc* 119:12262–12273.

168. Clore GM, Gronenborn AM, Tjandra N (1998) Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J Magn Reson* 131:159–162.

169. Pinheiro AS, Marsh JA, Forman-Kay JD, Peti W (2011) Structural signature of the MYPT1-PP1 interaction. *J Am Chem Soc* 133:73–80.

170. Fenwick RB, Esteban-Martín S, Salvatella X (2011) Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur Biophys J* 40:1339–1355.

171. Valafar H, Prestegard JH (2004) REDCAT: a residual dipolar coupling analysis tool. *J Magn Reson* 167:228–241.

172. Plotkin SS, Onuchic JN (2000) Investigation of routes and funnels in protein folding by free energy functional methods. *Proc Natl Acad Sci USA* 97:6509–6514.

173. Chavez LL, Onuchic JN, Clementi C (2004) Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J Am Chem Soc* 126:8426–8432.

174. Nerenberg PS, Jo B, So C, Tripathy A, Head-Gordon T (2012) Optimizing solute-water van der Waals interactions to reproduce solvation free energies. *J Phys Chem B* 116:4524–4534.

175. Best RB, Mittal J (2010) Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J Phys Chem B* 114:14916–14923.

176. Lindorff-Larsen K, Trbovic N, Maragakis P, Piana S, Shaw DE (2012) Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J Am Chem Soc* 134:3787–3791.

177. Terakawa T, Takada S (2011) Multiscale ensemble modeling of intrinsically disordered proteins: p53 N-terminal domain. *Biophys J* 101:1450–1458.

178. Cornell WD et al. (1995) A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J Am Chem Soc* 117:5179–5197.

179. Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118:11225–11236.

180. MacKerell AD et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* 102:3586–3616.

181. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107:8183–8188.

182. Uversky V (2002) Natively unfolded proteins: A point where biology waits for physics. *Protein Science* 11:739–756.

183. Ponder JW, Case DA (2003) Force fields for protein simulations. *Adv Protein Chem* 66:27–.

184. Patel S, MacKerell AD, Brooks CL (2004) CHARMM fluctuating charge force field for proteins: II - Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J Comput Chem* 25:1504–1514.

185. Halgren TA, Damm W (2001) Polarizable force fields. *Current Opinion in Structural Biology* 11:236–242.

186. Kaminski GA, Stern HA, Berne BJ, Friesner RA (2004) Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry. *Journal of Physical Chemistry A* 108:621–627.

187. Harder E, Kim BC, Friesner RA, Berne BJ (2005) Efficient simulation method for polarizable protein force fields: Application to the simulation of BPTI in liquid. *J Chem Theory Comput* 1:169–180.

188. Lamoureux G, Harder E, Vorobyov IV, Roux B, MacKerell AD (2006) A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem Phys Lett* 418:245–249.

189. Schnieders MJ, Fenn TD, Pande VS, Brunger AT (2009) Polarizable atomic multipole X-ray refinement: application to peptide crystals. *Acta Crystallogr D Biol Crystallogr* 65:952–965.

190. Ponder JW et al. (2010) Current status of the AMOEBA polarizable force field. *J Phys Chem B* 114:2549–2564.

191. Ren P, Wu C, Ponder JW (2011) Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J Chem Theory Comput* 7:3143–3161.

192. Leone V, Marinelli F, Carloni P, Parrinello M (2010) Targeting biomolecular flexibility with metadynamics. *Current Opinion in Structural Biology* 20:148–154.

193. Voter AF (1997) A method for accelerating the molecular dynamics simulation of infrequent events. *Journal of Chemical Physics* 106:4665–4677.

194. Voter AF (1997) Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys Rev Lett* 78:3908–3911.

195. Wu C, Shea J-E (2012) in *Computational Modeling of Biological Systems: From Molecules to Pathways*, Biological and Medical Physics, Biomedical Engineering. ed Dokholyan N (Springer US, Boston, MA), pp 215–227.

196. Friedrichs MS et al. (2009) Accelerating Molecular Dynamic Simulation on Graphics Processing Units. *J Comput Chem* 30:864–872.

197. Pierce LCT, Salomon-Ferrer R, de Oliveira CAF, McCammon JA, Walker RC (2012) Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J Chem Theory Comput* 8:2997–3002.

198. Pande VS et al. (2003) Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers* 68:91–109.

199. Lin Y-S, Bowman GR, Beauchamp KA, Pande VS (2012) Investigating How Peptide Length and a Pathogenic Mutation Modify the Structural Ensemble of Amyloid Beta Monomer. *Biophys J* 102:315–324.

200. Shaw DE et al. (2009) Millisecond-scale molecular dynamics simulations on Anton. *Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis (SC09)*.

201. Sivakolundu SG, Bashford D, Kriwacki RW (2005) Disordered p27(Kip1) exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. *J Mol Biol* 353:1118–1128.

202. Chen Y-X et al. (2006) Alternative O-GicNAcylation/O-phosphorylation of Ser(16) induce different conformational disturbances to the N terminus of murine estrogen receptor beta. *Chemistry & Biology* 13:937–944.

203. Ma K, Forbes JG, Gutierrez-Cruz G, Wang K (2006) Titin as a giant scaffold for integrating stress and Src homology domain 3-mediated signaling pathways: the clustering of novel overlap ligand motifs in the elastic PEVK segment. *J Biol Chem* 281:27539–27556.

204. Wu K-P, Weinstock DS, Narayanan C, Levy RM, Baum J (2009) Structural reorganization of alpha-synuclein at low pH observed by NMR and REMD simulations. *J Mol Biol* 391:784–796.

205. Higo J, Nishimura Y, Nakamura H (2011) A free-energy landscape for coupled folding and binding of an intrinsically disordered protein in explicit solvent from detailed all-atom computations. *J Am Chem Soc* 133:10448–10458.

206. Hazy E et al. (2011) Distinct hydration properties of wild-type and familial point mutant A53T of α-synuclein associated with Parkinson's disease. *Biophys J* 101:2260–2266.

207. Wostenberg C, Kumar S, Noid WG, Showalter SA (2011) Atomistic simulations reveal structural disorder in the RAP74-FCP1 complex. *J Phys Chem B* 115:13731–13739.

208. Savard P-Y et al. (2011) Structure and dynamics of Mycobacterium tuberculosis truncated hemoglobin N: insights from NMR spectroscopy and molecular dynamics simulations. *Biochemistry* 50:11121–11130.

209. Shan B, Li D-W, Brüschweiler-Li L, Brüschweiler R (2012) Competitive binding between dynamic p53 transactivation subdomains to human MDM2 protein: implications for regulating the p53·MDM2/MDMX interaction. *Journal of Biological Chemistry* 287:30376–30384.

210. Kumar P et al. (2012) Multisite phosphorylation disrupts arginine-glutamate salt bridge networks required for binding of cytoplasmic linker-associated protein 2 (CLASP2) to end-binding protein 1 (EB1). *Journal of Biological Chemistry* 287:17050–17064.

211. Knott M, Best RB (2012) A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: evidence from molecular simulations. *PLoS Comput Biol* 8:e1002605.

212. Feldman HJ, Hogue CWV (2002) Probabilistic sampling of protein conformations: new hope for brute force? *Proteins* 46:8–23.

213. Feldman HJ, Hogue CW (2000) A fast method to sample real protein conformational space. *Proteins* 39:112–131.

214. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292:195–202.

215. Buchan DWA et al. (2010) Protein annotation and modelling servers at University College London. *Nucleic Acids Res* 38:W563–8.

216. Goedert M, Spillantini MG (2006) A Century of Alzheimer's Disease. *Science* 314:777–781.

217. Antzutkin ON, Leapman RD, Balbach JJ, Tycko R (2002) Supramolecular structural constraints on Alzheimer's beta-amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance. *Biochemistry* 41:15436–15450.

218. Stromer T, Serpell LC (2005) Structure and morphology of the Alzheimer's amyloid fibril. *Microsc Res Tech* 67:210–217.

219. Goldschmidt L, Teng PK, Riek R, Eisenberg D (2010) Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc Natl Acad Sci USA* 107:3487–3492.

220. Kim W, Hecht MH (2005) Sequence determinants of enhanced amyloidogenicity of Alzheimer A{beta}42 peptide relative to A{beta}40. *J Biol Chem* 280:35069–35076.

221. Weinreb P, Jarrett J, Lansbury P (1994) Peptide models of a hydrophobic cluster at the C-terminus of the beta-amyloid proteinN. *J Am Chem Soc* 116:10835–10836.

222. Morimoto A et al. (2004) Analysis of the secondary structure of beta-amyloid (Abeta42) fibrils by systematic proline replacement. *J Biol Chem* 279:52781–52788.

223. Irie K et al. (2005) Structure of beta-amyloid fibrils and its relevance to their neurotoxicity: implications for the pathogenesis of Alzheimer's disease. *J Biosci Bioeng* 99:437–447.

224. Maji SK et al. (2009) Amino acid position-specific contributions to amyloid beta-protein oligomerization. *J Biol Chem* 284:23580–23591.

225. Lin Y-S, Bowman GR, Beauchamp KA, Pande VS (2012) Investigating How Peptide Length and a Pathogenic Mutation Modify the Structural Ensemble of Amyloid Beta Monomer. *Biophysj* 102:315–324.

226. Dyson HJ, Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Current Opinion in Structural Biology* 12:54–60.

227. Ferrone F (1999) Analysis of protein aggregation kinetics. *Methods in Enzymology* 309:256–274.

228. Naiki H, Nakakuki K (1996) First-order kinetic model of Alzheimer's beta-amyloid fibril extension in vitro. *Lab Invest* 74:374–383.

229. Kusumoto Y, Lomakin A, Teplow DB, Benedek GB (1998) Temperature dependence of amyloid beta-protein fibrillization. *Proc Natl Acad Sci USA* 95:12277–12282.

230. Urbanc B et al. (2004) In silico study of amyloid beta-protein folding and oligomerization. *Proc Natl Acad Sci USA* 101:17345–17350.

231. Olubiyi OO, Strodel B (2012) Structures of the amyloid β-peptides Aβ1-40 and Aβ1-42 as influenced by pH and a D-peptide. *J Phys Chem B* 116:3280–3291.

232. Yun S et al. (2007) Role of electrostatic interactions in amyloid beta-protein (A beta) oligomer formation: a discrete molecular dynamics study. *Biophys J* 92:4064–4077.

233. Triguero L, Singh R, Prabhakar R (2008) Comparative Molecular Dynamics Studies of Wild-Type and Oxidized Forms of Full-Length Alzheimer Amyloid beta-Peptides Abeta(1-40) and Abeta(1-42). *J Phys Chem B* 112:7123–7131.

234. Vitalis A, Caflisch A (2010) Micelle-like architecture of the monomer ensemble of Alzheimer's amyloid-β peptide in aqueous solution and its implications for Aβ aggregation. *J Mol Biol* 403:148–165.

235. Lee C, Ham S (2011) Characterizing amyloid-beta protein misfolding from molecular dynamics simulations with explicit water. *J Comput Chem* 32:349–355.

236. Straub JE, Thirumalai D (2011) Toward a molecular theory of early and late events in monomer to amyloid fibril formation. *Annu Rev Phys Chem* 62:437–463.

237. Ladiwala ARA et al. (2012) Conformational differences between two amyloid β oligomers of similar size and dissimilar toxicity. *Journal of Biological Chemistry* 287:24765–24773.

238. Kłoniecki M et al. (2011) Ion Mobility Separation Coupled with MS Detects Two Structural States of Alzheimer's Disease Aβ1-40 Peptide Oligomers. *J Mol Biol* 407:110–124.

239. Wiberg H et al. (2010) Separation and characterization of aggregated species of amyloid-beta peptides. *Anal Bioanal Chem* 397:2357–2366.

240. Narayanan S, Reif B (2005) Characterization of chemical exchange between soluble and aggregated states of beta-amyloid by solution-state NMR upon variation of salt conditions. *Biochemistry* 44:1444–1452.

241. LeVine H (2004) Alzheimer's beta-peptide oligomer formation at physiologic concentrations. *Anal Biochem* 335:81–90.

242. Laurents DV (2004) Alzheimer's A 40 Studied by NMR at Low pH Reveals That Sodium 4,4-Dimethyl-4-silapentane-1-sulfonate (DSS) Binds and Promotes -Ball Oligomerization. *Journal of Biological Chemistry* 280:3675–3685.

243. Fawzi NL, Ying J, Ghirlando R, Torchia DA, Clore GM (2011) Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. *Nature* 480:268–272.

244. Murphy MP, LeVine H (2010) Alzheimer's disease and the amyloid-beta peptide. *J Alzheimers Dis* 19:311–323.

245. Lazarczyk MJ, Hof PR, Bouras C, Giannakopoulos P (2012) Preclinical Alzheimer disease: identification of cases at risk among cognitively intact older individuals. *BMC Med* 10:127.

246. Näslund J et al. (2000) Correlation between elevated levels of amyloid beta-peptide in the brain and cognitive decline. *JAMA* 283:1571–1577.

247. Gellermann G et al. (2008) Aβ-globulomers are formed independently of the fibril pathway. *Neurobiology of Disease* 30:212–220.

248. Viola KL, Velasco PT, Klein WL (2008) Why Alzheimer's is a disease of memory: the attack on synapses by A beta oligomers (ADDLs). *J Nutr Health Aging* 12:51S–7S.

249. Greenamyre JT, Maragos WF, Albin RL, Penney JB, Young AB (1988) Glutamate transmission and toxicity in Alzheimer's disease. *Prog Neuropsychopharmacol Biol Psychiatry* 12:421–430.

250. Francis PT, Sims NR, Procter AW, Bowen DM (1993) Cortical pyramidal neurone loss may cause glutamatergic hypoactivity and cognitive impairment in Alzheimer's disease: investigative and therapeutic perspectives. *J Neurochem* 60:1589–1604.

251. Rossor MN et al. (1982) A post-mortem study of the cholinergic and GABA systems in senile dementia. *Brain* 105:313–330.

252. Pakaski M et al. (1998) Vulnerability of small GABAergic neurons to human beta-amyloid pentapeptide. *Brain Res* 796:239–246.

253. Krantic S et al. (2012) Hippocampal GABAergic neurons are susceptible to amyloid-β toxicity in vitro and are decreased in number in the Alzheimer's disease TgCRND8 mouse model. *J Alzheimers Dis* 29:293–308.

254. Götz J, Ittner LM (2008) Animal models of Alzheimer's disease and frontotemporal dementia. *Nat Rev Neurosci* 9:532–544.

255. Nimmrich V et al. (2008) Amyloid beta oligomers (A beta(1-42) globulomer) suppress spontaneous synaptic activity by inhibition of P/Q-type calcium currents. *J Neurosci* 28:788–797.

256. Mezler M, Barghorn S, Schoemaker H, Gross G, Nimmrich V (2012) A β-amyloid oligomer directly modulates P/Q-type calcium currents in Xenopus oocytes. *Br J Pharmacol* 165:1572–1583.

257. Hermann D et al. (2013) Synthetic Aβ oligomers (Aβ1-42 globulomer) modulate presynaptic calcium currents: Prevention of Aβ-induced synaptic deficits by calcium channel blockers. *Eur J Pharmacol* 702:44–55.

258. Kokubo H, Kayed R, Glabe CG, Yamaguchi H (2005) Soluble Abeta oligomers ultrastructurally localize to cell processes and might be related to synaptic dysfunction in Alzheimer's disease brain. *Brain Res* 1031:222–228.

259. Kokubo H et al. (2005) Oligomeric proteins ultrastructurally localize to cell processes, especially to axon terminals with higher density, but not to lipid rafts in Tg2576 mouse brain. *Brain Res* 1045:224–228.

260. Noguchi A et al. (2009) Isolation and characterization of patient-derived, toxic, high mass amyloid beta-protein (Abeta) assembly from Alzheimer disease brains. *Journal of Biological Chemistry* 284:32895–32905.

261. Verdier Y, Zarándi M, Penke B (2004) *Amyloid beta-peptide interactions with neuronal and glial cell plasma membrane: binding sites and implications for Alzheimer's disease.*

262. Lesuisse C, Martin LJ (2002) Long-term culture of mouse cortical neurons as a model for neuronal development, aging, and death. *J Neurobiol* 51:9–23.

263. Belyaev ND et al. (2010) The transcriptionally active amyloid precursor protein (APP) intracellular domain is preferentially produced from the 695 isoform of APP in a {beta}-secretase-dependent pathway. *Journal of Biological Chemistry* 285:41443–41454.

264. Simakova O, Arispe NJ (2007) The cell-selective neurotoxicity of the Alzheimer's Abeta peptide is determined by surface phosphatidylserine and cytosolic ATP levels. Membrane binding is required for Abeta toxicity. *J Neurosci* 27:13719–13729.

265. Igbavboa U, Avdulov NA, Schroeder F, Wood WG (1996) Increasing age alters transbilayer fluidity and cholesterol asymmetry in synaptic plasma membranes of mice. *J Neurochem* 66:1717–1725.

266. Kremer JJ, Pallitto MM, Sklansky DJ, Murphy RM (2000) Correlation of beta-amyloid aggregate size and hydrophobicity with decreased bilayer fluidity of model membranes. *Biochemistry* 39:10309–10318.

267. Pike CJ, Cotman CW (1993) Cultured GABA-immunoreactive neurons are resistant to toxicity induced by beta-amyloid. *Neuroscience* 56:269–274.

268. Benilova I, Karran E, De Strooper B (2012) The toxic Aβ oligomer and Alzheimer's disease: an emperor in need of clothes. *Nature neuroscience*.

269. Hillen H et al. (2010) Generation and therapeutic efficacy of highly oligomer-specific beta-amyloid antibodies. *J Neurosci* 30:10369–10379.

270. Urbanc B et al. (2004) Molecular dynamics simulation of amyloid beta dimer formation. *Biophys J* 87:2310–2321.

271. Moore BD, Rangachari V, Tay WM, Milkovic NM, Rosenberry TL (2009) Biophysical analyses of synthetic amyloid-beta(1-42) aggregates before and after covalent cross-linking. Implications for deducing the structure of endogenous amyloid-beta oligomers. *Biochemistry* 48:11796–11806.

272. Petkova AT, Yau W-M, Tycko R (2006) Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils. *Biochemistry* 45:498–512.

273. Martí-Renom MA et al. (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29:291–325.

274. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815.

275. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9:1753–1773.

276. Krivov GG, Shapovalov MV, Dunbrack RL (2009) Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* 77:778–795.

277. Olofsson A, Sauer-Eriksson AE, Ohman A (2006) The solvent protection of alzheimer amyloid-beta-(1-42) fibrils as determined by solution NMR spectroscopy. *J Biol Chem* 281:477–483.

278. Makin O (2005) Structures for amyloid fibrils. *FEBS Journal*.

279. Eisenberg D et al. (2006) The structural biology of protein aggregation diseases: Fundamental questions and some answers. *Acc Chem Res* 39:568–575.

280. Ferguson N et al. (2006) General structural motifs of amyloid protofilaments. *Proc Natl Acad Sci USA* 103:16248–16253.

281. Kirschner DA, Abraham C, Selkoe DJ (1986) X-ray diffraction from intraneuronal paired helical filaments and extraneuronal amyloid fibers in Alzheimer disease indicates cross-beta conformation. *Proc Natl Acad Sci USA* 83:503–507.