

**Biophysical Characterization of Intrinsically Disordered Peptides
Through Molecular Dynamics Simulations and Solution Nuclear
Magnetic Resonance Spectroscopy.**

by

David Jacob Rosenman

A Thesis Submitted to the Graduate
Faculty of Rensselaer Polytechnic Institute
in Partial Fulfillment of the
Requirements for the degree of
DOCTOR OF PHILOSOPHY
Major Subject: Biology

Approved by the
Examining Committee:

Angel E. García, Thesis Co-Advisor

Chunyu Wang, Thesis Co-Advisor

Christopher Bystroff, Member

Wilfredo Colón, Member

Rensselaer Polytechnic Institute
Troy, New York

July 2015
(For Graduation August 2015)

© Copyright 2015
by
David Jacob Rosenman
All Rights Reserved

CONTENTS

LIST OF TABLES.....	viii
LIST OF FIGURES.....	ix
ACKNOWLEDGMENT.....	xiv
ABSTRACT.....	xvii
1. Intrinsically Disordered Peptides and Neurodegenerative Disease.....	1
1.1 Challenges in Studying Intrinsically Disordered Peptides.....	1
1.2 Amyloid β and Alzheimer's Disease.....	5
1.2.1 Societal and Financial Costs of Alzheimer's Disease.....	5
1.2.2 Symptoms of Alzheimer's Disease.....	6
1.2.3 Pathological Hallmarks of Alzheimer's Disease.....	6
1.2.4 Amyloid Cascade Hypothesis.....	8
1.2.5 Processing of A β	9
1.2.6 Aggregates of A β	13
1.2.7 Experimental Properties of A β Monomers and Their Role in Aggregation.....	17
1.2.8 Simulations of A β Monomers.....	20
1.2.9 Familial Alzheimer's Disease and Point Mutations of A β	22
1.3 Other Proteins involved in Neurodegeneration.....	26
1.3.1 α -synuclein.....	26
1.3.2 Cytoplasmic Superoxide Dismutase (SOD1).....	28
2. Replica Exchange Molecular Dynamics Simulations of A β Monomers Using the OPLS-AA/TIP3P Force Field.....	31
2.1 Abstract.....	31
2.2 Introduction.....	32
2.3 Methods.....	34
2.3.1 Simulation Details.....	34

2.3.2	Ensemble Analysis.....	37
2.4	Results.....	38
2.4.1	Convergence.....	38
2.4.2	Secondary Structure.....	39
2.4.3	C α Contact Maps.....	43
2.4.4	Electrostatic Interactions.....	46
2.4.5	RMSD Based Clustering.....	53
2.4.6	Radius of Gyration and End to End Distance	58
2.4.7	Solvent Accessible Surface Area	61
2.4.8	Order Parameters.....	63
2.5	Discussion and Conclusions.....	65
3.	Replica Exchange Molecular Dynamics Simulations of A β Monomers Using the AMBER99sb-ILDN/TIP4P-Ew Force Field	73
3.1	Abstract	73
3.2	Introduction.....	74
3.3	Methods.....	77
3.3.1	Simulation Details.....	77
3.3.2	Ensemble Analysis.....	79
3.4	Results.....	79
3.4.1	Convergence.....	79
3.4.2	Secondary Structure	80
3.4.3	C α Contact Maps.....	84
3.4.4	Electrostatic Interactions.....	87
3.4.5	RMSD Based Clustering.....	91
3.4.6	Radius of Gyration and End to End Distance	95
3.4.7	Solvent Accessible Surface Area	98
3.4.8	Order Parameters.....	100
3.4.9	High Pressure Simulation of A β 40	102

3.5	Discussion and Conclusions.....	104
4.	NMR Experiments on A β and Experimental Validation of REMD Data	111
4.1	Abstract	111
4.2	Introduction	112
4.2.1	Use of Solution NMR in the Study of Intrinsic Disordered Peptides	112
4.2.2	Perturbations and High Pressure NMR	114
4.3	Methods.....	116
4.3.1	Overexpression and Purification of A β Peptides	116
4.3.2	NMR Sample Preparation of A β Peptides	119
4.3.3	NMR Data Acquisition	120
4.3.3.1	NMR Assignment	121
4.3.3.2	NMR Relaxation Measurements.....	122
4.3.3.3	Measurement of J Coupling and Residual Dipolar Coupling Data	122
4.3.3.4	Prediction of NMR Observables from Simulation.....	123
4.3.3.5	Urea Titrations	124
4.3.3.6	High Pressure NMR.....	124
4.4	Results.....	128
4.4.1	Solution NMR of A β Monomers at Ambient Pressure	128
4.4.2	NMR Observables and Simulation Comparison	134
4.4.2.1	Chemical Shifts.....	134
4.4.2.2	J-couplings and Residual Dipolar Couplings.....	139
4.4.2.3	Hydrogen Bonding and Urea Titrations.....	144
4.4.3	High Pressure NMR of A β Monomers	147
4.5	Discussion	162
5.	MD/REMD Simulations of Other Disease-Relevant IDPs: Fragments of the SOD1 Protein and α -synuclein 1-18 A18D	169
5.1	Abstract	169

5.2	Introduction	169
5.2.1	SOD1 Fragments.....	170
5.2.1.1	α -synuclein 1-18 A18D ($N_{\alpha\text{-syn}}$).....	172
5.3	Methods.....	173
5.3.1	Simulation Details.....	173
5.3.1.1	MD and REMD on SOD1 Fragments.....	173
5.3.1.2	REMD on α -synuclein 1-18 A18D ($N_{\alpha\text{-syn}}$).....	176
5.3.2	Ensemble Analysis.....	177
5.3.3	Experiment Details.....	178
5.4	Results.....	180
5.4.1	SOD1 Fragments.....	180
5.4.1.1	Secondary and Tertiary Structure	180
5.4.1.2	RMSD Based Clustering and Trajectory Superposition	184
5.4.1.3	Radius of Gyration, End to End Distance, and Comparison to Experiment.....	187
5.4.2	α -synuclein 1-18 A18D ($N_{\alpha\text{-syn}}$)	191
5.4.2.1	Secondary and Tertiary Structure	191
5.4.2.2	RMSD Based Clustering and Salt Bridges	195
5.4.2.3	Radius of Gyration.....	197
5.4.2.4	Comparison of Chemical Shifts and $^3J_{\text{HNHA}}$ to Experiment.	199
5.5	Discussion	204
6.	Concluding Remarks and Future Directions.....	207
6.1	Concluding Remarks.....	207
6.2	Future Directions.....	208
6.2.1	Simulations.....	208
6.2.2	Experiment.....	216
	REFERENCES	221

7. Appendix.....	251
------------------	-----

LIST OF TABLES

Table 1.1	Phenotypic behavior of point mutations of A β found in human subjects.....	24
Table 2.1	Most commonly populated long distance, inter-residue, backbone hydrogen bonds in the A β ensemble for the OPLS set of simulations	47
Table 2.2	Central region electrostatic interactions of A β involving the side chains of residue 22 and 23 for the OPLS simulation set	50
Table 3.1	Most commonly populated long distance, inter-residue, backbone hydrogen bonds in the A β ensemble for the ILDN set of simulations	88
Table 3.2	Central region electrostatic interactions of A β involving the side chains of residue 22 and 23 for the ILDN simulation set	90
Table 5.1	Contents and properties of the simulation boxes used for SOD1 subpeptide simulations	175
Table 7.1	Force field parameters for L-methionine-(S)-sulfoxide used in the OPLS simulation of A β 42-M35ox(S)	251
Table 7.2	Resonance table for A β 40-D23N and A β 40-A2T based on the assignment experiments performed for each species	253
Table 7.3	$^3J_{\text{HNHA}}$ values for several A β species	262

LIST OF FIGURES

Figure 1.1	Schematics for the free energy landscape of a hypothetical globular protein and an intrinsically disordered peptide	3
Figure 1.2	Diagram illustrating two different pathways for the processing of the amyloid precursor protein	10
Figure 1.3	Schematic illustrating different pathways of aggregation for amyloid β	16
Figure 1.4	Schematic representing the three structural regions of wild type α -synuclein.....	27
Figure 2.1	Content of secondary structure elements versus simulation time for A β runs done with the OPLS condition	40
Figure 2.2	Percent composition of secondary structure for A β , averaged over the production ensemble, for simulations done with the OPLS condition, over all residues and as a function of residue index	41
Figure 2.3	Maps illustrating the probability of contact for C α atoms over the production ensemble for different A β isoforms and mutants derived with REMD under the OPLS condition, and difference contact maps between them	44
Figure 2.4	Comparison of transiently sampled conformations and ensemble biases in monomeric A β REMD simulations to an experimental structure of a trapped A β monomer and intrapeptide models for higher order aggregates	49
Figure 2.5	Central region (A21-G29) in A β 40's third most populated centroid after single linkage clustering with the OPLS condition, illustrating the "crown motif" referred to in the text	52
Figure 2.6	The four highest-populated centroids from single linkage cluster analysis around C α atoms with an RMSD cutoff of 2.0 Å for each A β simulation done with the OPLS condition over the production period	54
Figure 2.7	Comparison of OPLS simulations to an NMR structure of A β 40 monomer trapped by the Z _{Aβ3} Affibody protein	55

Figure 2.8	Other select centroids from single linkage cluster analysis around C α atoms with an RMSD cutoff of 2.0 Å for each A β simulation done with the OPLS condition over the production period	57
Figure 2.9	Radius of gyration and end to end distance for OPLS A β simulations.....	59
Figure 2.10	Solvent accessible surface per residue over the production ensemble of OPLS A β simulations	62
Figure 2.11	S^2 order parameters for all backbone N-H internuclear vectors calculated with the iRED technique over the production period for OPLS simulations.....	64
Figure 3.1	Content of secondary structure elements versus simulation time for A β runs done with the ILDN condition	81
Figure 3.2	Percent composition of secondary structure for A β , averaged over the production ensemble, for simulations done with the ILDN condition, over all residues and as a function of residue index	82
Figure 3.3	Maps illustrating the probability of contact for C α atoms over the production ensemble for different A β isoforms and mutants derived with REMD under the ILDN condition, and difference contact maps between them	85
Figure 3.4	The four highest-populated centroids from single linkage cluster analysis around C α atoms with an RMSD cutoff of 2.0 Å for each A β simulation done with the ILDN condition over the production period	92
Figure 3.5	Other select centroids from single linkage cluster analysis around C α atoms with an RMSD cutoff of 2.0 Å for each A β simulation done with the ILDN condition over the production period	94
Figure 3.6	Comparison of ILDN simulations to an NMR structure of A β 40 monomer trapped by the Z _{Aβ3} Affibody protein	96
Figure 3.7	Radius of gyration and end to end distance for ILDN A β simulations.....	97
Figure 3.8	Solvent accessible surface per residue over the production ensemble of ILDN A β simulations	99
Figure 3.9	S^2 order parameters for all backbone N-H internuclear vectors calculated with the iRED technique over the production period for ILDN simulations.....	101

Figure 3.10	Analysis of A β 40 at 2500 bar simulated with ILDN, using methods covered previously in the chapter	103
Figure 4.1	A simplified flowchart for the A β monomer sample preparation protocol presented in this chapter	117
Figure 4.2	^1H - ^{15}N HSQC spectrum for A β 40-D23N, overlaid with the same spectrum for wild type A β 40	129
Figure 4.3	Chemical shift perturbation in ^1H - ^{15}N HSQC spectra between different A β mutants/isoforms	130
Figure 4.4	Comparison of backbone dynamics of A β 40-A2T versus A β 40, as determined by ^{15}N R_1 , ^{15}N R_2 , and heteronuclear, steady state ^1H - ^{15}N nuclear Overhauser effect, plotted versus residue index	132
Figure 4.5	Correlation and RMSD of predicted chemical shifts with experimentally measured values, with predictions based on frames from the ILDN and OPLS A β ensembles	136
Figure 4.6	Correlation and RMSD of calculated $^3J_{\text{HNHA}}$, $^3J_{\text{NHB}}$, and residual dipolar couplings from ILDN and OPLS A β production ensembles compared to experimentally determined values	141
Figure 4.7	Comparison of the perturbations on chemical shift caused by urea versus hydrogen bond occupancy in simulations between different A β species .	146
Figure 4.8	Effect of pressure on the chemical shifts and peak intensities of monomeric A β	150
Figure 4.9	Solvent exchange rates for each residue as a function of pressure for monomeric A β 40 and A β 42	160
Figure 4.10	Correlation analysis for A β 40 relating β sheet character annotated in REMD simulations with various NMR observables	164
Figure 5.1	Percent composition of secondary structure as a function of residue index for simulations of SOD1 peptides	181
Figure 5.2	Maps illustrating the probability of contact for C α atoms over the production ensemble for different SOD1 peptides derived with REMD and MD	182
Figure 5.3	Representative conformations for simulations of SOD1 peptides	185
Figure 5.4	FRET data for the SOD1 peptides studied in this chapter	188

Figure 5.5	Radius of gyration and end to end distance for simulations of SOD1 peptides.....	190
Figure 5.6	Percent composition of secondary structure as a function of residue index for $N_{\alpha\text{-syn}}$	192
Figure 5.7	Maps illustrating the probability of contact for $C\alpha$ atoms over the production ensemble for the different $N_{\alpha\text{-syn}}$ REMD ensembles	193
Figure 5.8	The four highest-populated centroids from Daura cluster analysis around $C\alpha$ atoms with an RMSD cutoff of 3.0 Å for the $N_{\alpha\text{-syn}}$ REMD ensembles over the production period	196
Figure 5.9	Radius of gyration and end to end distance for simulations of $N_{\alpha\text{-syn}}$	198
Figure 5.10	Correlation and RMSD of calculated chemical shifts and $^3J_{\text{HNHA}}$ couplings for $N_{\alpha\text{-syn}}$ simulations compared to experimentally determined values	201
Figure 7.1	Movement of replicas in both temperature and replica space as a function of simulation time for select replica exchange molecular dynamics simulations described in this thesis	252
Figure 7.2	Assessment of the significance of the PCC using the Δ^3J_{HNHA} between $A\beta_{40\text{-D23N}}$ and $A\beta_{40}$ through 100000 Monte Carlo simulations based on the experimental data	263
Figure 7.3	Temperature calibration for a sample in the 600 MHz Bruker spectrometer inside the high pressure cell used for the high pressure NMR experiments in Chapter 4	264
Figure 7.4	^1H - ^{15}N HSQC spectra for different monomeric $A\beta$ species collected with high pressure NMR, ranging from ambient pressure to 2500 bar at 250 bar increments	265
Figure 7.5	Additional effects of pressure on the chemical shifts and peak intensities of monomeric $A\beta$	267
Figure 7.6	Topology of aggregates of the $A\beta_{40\text{-E22G}}$ sample used for high pressure NMR, as determined by atomic force microscopy	269
Figure 7.7	Solvent exchange rates as a function of time/aggregation for the same sample run at 1 bar	270

Figure 7.8	Analysis of A β 40, recently simulated using REMD with the CHARMM22* force field and the modified CHARMM TIP3P water model	271
Figure 7.9	Assessment of convergence of the OPLS A β 42 simulation by analysis of the distribution of structures with respect to those determined from RMSD-based clustering of earlier parts of the trajectory	274
Figure 7.10	Analysis of A β 40-E22G, recently simulated using REMD with the AMBER99sb-ILDN force field and the TIP4P-Ew water model	275

ACKNOWLEDGMENT

The road to a Ph.D. can often be challenging and lonely. Fortunately, it has been my great privilege to spend the last five years in an environment surrounded by brilliant, hard-working, and supportive people. To properly thank everyone who has contributed to making this dissertation a reality, either directly or indirectly, would require much more than the space allotted here. That said, I will do my best to acknowledge here some of these people who have assisted me and enriched my Ph.D. experience.

First and foremost, I want to offer my sincerest gratitude to Prof. Angel E. García and Prof. Chunyu Wang, my advisors, who have supervised and guided me on the path to becoming a researcher. I am inspired by and indebted to Prof. García for his incredible knowledge, creativity, and passion for science, in tandem with the deep trust and unyielding support he provided at every step. Likewise, Prof. Wang has continuously provided invaluable guidance during my years at RPI, offering many key insights that have directed this work, assisting and encouraging me in my efforts time and time again, and serving as a role model as a patient, committed, and thorough researcher. I will never cease to be grateful to these two outstanding mentors, not just for allowing this work to come to fruition, but for providing a basis for me to mature as a scholar and a human being.

I would like to acknowledge Prof. Christopher Bystroff and Prof. Wilfredo Colón, for taking the time and effort to serve as the other two members of my thesis committee. Both have set great examples as exceptional scientists and provided feedback instrumental to the completion and validation of this dissertation. I cannot thank them enough for their careful reading and review of my work. I also must thank Prof. Bystroff for being a thoughtful and dedicated advisor when I was a rotation student in my first year of graduate school; the environment and direction he provided was instrumental in adapting to life as a graduate student and a researcher.

I wholeheartedly thank my talented and kindhearted colleagues who served as my labmates during my time in the doctorate program. Collectively, they have created an excellent environment for me to pursue my academic and professional goals and taught me important lessons about science and life along the way. Special thanks to current and

former García group members Dr. Charles English, Dr. Christopher Neale, Joel Janke, Jacob Miner, Bryan Makowski, Dr. Alan Chen, Dr. Henry Herce, and Dr. Kun Huang for sharing in many thought provoking discussions, providing invaluable advice and perspectives, and commiserating with me on the all-too-frequent occasion that I had to scream at my computer monitor. I also want to especially recognize Dr. Christopher Neale for leading the effort in writing the successful grant proposals for XSEDE computational resources that were instrumental for my research, and for carefully proofreading and offering invaluable advice to improve this thesis document. Also, I want to thank the system administrator, Robert Healey, for his tireless efforts to maintain the García lab computational resources so crucial to my research. Likewise, I want to personally thank past and present Wang lab members Nicolina Clemente, Jian Xie, Dr. Christopher Connors, and Dr. Wen Chen for teaching me the patience and skills so necessary to being an experimental biologist, collaborating on so many time and effort consuming projects, and for helping me clean up on the all-too-frequent occasion that I would drop a beaker of mysterious fluid on the ground. I must thank Prof. Catherine Royer and Martin Fossat for their continued and enthusiastic assistance in performing high pressure NMR, and Prof. Scott McCallum for his superb maintenance of NMR equipment and troubleshooting of experiments. I also want to thank Dr. Yao-Ming Huang for being my first postdoc mentor at RPI and introducing me to many experimental methods while I was rotating in Prof. Bystroff's lab as an entirely clueless first year graduate student.

I acknowledge all of the sources of funding that this research fundamentally relied on, including the National Institutes of Health-National Institute of General Medical Sciences Biomolecular Science and Engineering Training Grant, the Weissman Family Fellowship, and the National Science Foundation, as well as the Extreme Science and Engineering Discovery Environment (XSEDE) for the generous allocations of computational resources that fueled these projects. I also want to thank Dr. Alexey Gribenko at Pfizer Inc. for accepting me for a summer internship that allowed me to satisfy a requirement for this funding, and for serving as a thoughtful mentor who taught me many valuable experimental techniques while I was there.

Last but definitely not least, I want to dedicate this work to my incredible family – in particular, my father, Dr. Arthur J. Rosenman, my mother Helen, and my sister Michelle. These wonderful people have provided both moral and practical support every day of my doctorate career (and beyond). My father, a doctor who has diligently practiced for over four decades now, is a person I am truly honored to call both "Dad" and "my closest friend;" it is by his example that I strive to push forward as a scholar. My family members are my greatest cheerleaders, and this accomplishment would not be possible without their unquestionable love, kindness, trust, and support at every turn.

ABSTRACT

Representing 20-30% of the human proteome, intrinsically disordered peptides (IDPs) are highly dynamic sequences that do not assume a single, defined native state under physiological conditions(1). These peptides are implicated as causative agents in various human neurodegenerative diseases, primarily due to their susceptibility to misfold and aggregate(2, 3). Further, the ambiguous and diverse free energy landscapes for these systems challenge the validity and applicability of our computational modeling techniques, and preclude the application of many of the experimental methods we currently and proficiently use to characterize well-folded proteins. This thesis defines the ensemble structural biases and individual conformations of the monomer state of various disease-relevant IDPs through the union of simulations and experiments, in order to better understand their role in aggregation and disease pathology. Through their application to these challenging systems, this work also aims to identify previously unknown weaknesses and strengths of these biophysical techniques in order to improve their application to other systems.

The majority of this thesis investigates the amyloid β ($A\beta$) monomer, a prototypic IDP(2-4) whose aggregates are implicated in the pathogenesis of Alzheimer's disease (AD), a neurodegenerative disease that affects 36 million people worldwide(5). Relatively subtle modifications to the peptide, including a two residue change in peptide length ($A\beta_{40}$ vs $A\beta_{42}$) and point mutations in the $A\beta$ sequence that are linked to familial AD (FAD), result in profound changes to the aggregation and toxicity of the peptide(6-8). We hypothesize these changes are fundamentally linked to the intrinsic disorder of monomer state; these disordered ensembles may be easily perturbed by small changes in physicochemistry, allowing them to sample new conformers that can seed different aggregates.

Here, simulations and experiment were used to describe different variants of $A\beta$ and ascertain the nature and extent of these changes at the monomeric level. Numerous computational characterizations of $A\beta$ monomers exist in the literature(9-18), but little consensus exists among these studies. We suspect that simulations of IDPs are particularly sensitive to the conditions of the run; as such, our approach to model $A\beta$

emphasizes explicit representation, broad sampling, and parameter generalizability. Here, the ensembles of A β were simulated using all atom models and explicit water. Further, we employed enhanced sampling through replica exchange molecular dynamics (REMD) using longer timescales than had been employed in the past. Indeed, we observe that these extensive timescales are needed to reach convergence for the emerging β structure of the A β monomer simulations. Further, the A β variants were each simulated using multiple force fields, including the state of the art OPLS-AA(19)/TIP3P(20) and AMBER99sb-ILDN(21)/TIP4P-Ew(22) combinations. We have also collected various experimental data, primarily at the atomistic level using nuclear magnetic resonance (NMR) spectroscopy. These data, including chemical shifts, J-couplings, solvent exchange rates, and pressure coefficients (measures of the sensitivity of chemical shifts to pressure), are not only observables that can be directly compared to values back-calculated from simulation to validate these runs, but also, collectively provide their own description of the secondary structure biases in the A β ensemble that can be compared to that of simulation.

The union between experiment and simulation with multiple force fields reveals a consistent model of A β supported by multiple methods. This model describes wild type A β 40 with antiparallel β -hairpin between L17-A21 and A30-L34, while residues A21-A30 forms an intervening loop region that rarely interacts with the majority of the protein. Meanwhile, A β 42 contributes new β -hairpin motifs involving V40-I41, with a new turn involving residue G37. These structural trends for the A β monomer are of particular interest because they increase the solvent exposure of hydrophobic side chains and because they correlate to intrapeptide models for oligomers and fibrils as determined by solid state NMR(23-27). This suggests that these conformations may represent the seeds of aggregation for these higher order forms.

While these trends were common to our approaches to studying A β , examination at finer levels revealed key inconsistencies. For example, the two force fields used here model the local structure and electrostatics of the A21-A30 central region in very different ways. Unsurprisingly, the FAD mutations we studied which lie in this region also had very different effects on the structural ensembles observed between the

different force fields. Unfortunately, the available experimental data, in general, remain too coarse to inform which model, if any, is correct.

We have further attempted to assess the robustness of our simulation techniques through their application in modeling the conformational landscapes for subpeptides of SOD1 and α -synuclein, involved in the pathogenesis of amyotrophic lateral sclerosis and Parkinson's disease, respectively. The modeled fragments were of comparable length to A β , and some of these are known to be intrinsically disordered. Förster resonance energy transfer (FRET) and small-angle X-ray scattering (SAXS) data were used to inform the ensemble shape of the SOD1 fragments, while NMR measurements were provided for the α -synuclein subpeptide. Unfortunately, comparison to these experimental data yielded poor agreement for these systems. The simulations of SOD1 peptides suggest that our parameters are overly predisposed to sample collapsed states. Meanwhile, simulations between different force fields yielded very different descriptions of the α -synuclein fragment, and neither representation produced observables that agreed well with experimental values. These descriptions inform important weaknesses in our simulation parameters, but also suggest that the concurrences in the characterizations made through simulation and experiment for A β are non-trivial.

The final chapter summarizes our findings and proposes new research directions to expand and improve our characterization of IDPs. Ongoing REMD simulations of A β peptides, including those run with CHARMM22*(28)/TIP3SP(20) parameters are also described. Simulation of A β 40 using this CHARMM combination reduces the compaction of the ensemble, while preserving similar tertiary structure biases as the other force field combinations explored in this thesis and showing good agreement to NMR data.

1. Intrinsically Disordered Peptides and Neurodegenerative Disease

1.1 Challenges in Studying Intrinsically Disordered Peptides

The traditional protein structure-function model holds that the function of a protein is directly dependent upon the specific, well-defined three-dimensional structure that it adopts. However, this long-held paradigm has been challenged with the discovery of a class of highly dynamic peptides and proteins that fail to adopt a stable native state under physiological conditions(1). These intrinsically disordered peptides (IDPs) comprise 20-30% of the sequences in the human proteome and are involved in many biological processes, such as cell signaling, transcription, and chromatin remodeling(29-31). Often, IDPs are well suited to be regulatory targets in the interaction networks behind these biological functions because they can participate in a broader set of protein-protein interactions and are sensitive to structural changes induced by post-translational modifications(32). However, certain IDPs, such as amyloid β ($A\beta$), α -synuclein, PAPf39, and islet amyloid polypeptide (IAPP), play detrimental roles in human pathologies like neurodegenerative diseases, HIV transmission, and diabetes due to their propensity to misfold or aggregate(2, 3).

Most of what is known about disordered proteins comes through comparison of sequence, chemical, and structural properties with those of well-ordered proteins. IDPs, for example, are characterized by high content of polar and charged amino acids and low hydrophobic content, low levels of sequence complexity, and very little secondary structure(1, 33). To better understand the functions of IDPs, we must first understand the biophysical properties of these peptides. However, the structural plasticity of IDPs typically preclude the use of traditional techniques for structure determination.

Portions of this chapter previously appeared as J. Nasica-Labouze *et al.*, Amyloid β protein and Alzheimer's disease: when computer simulations complement experimental studies. *Chem. Rev.* **115**, 3518 (2015). Only content written directly by the author of this thesis is reproduced here, with permission granted by the corresponding author, Prof. Philippe Derreumaux.

Portions of this chapter previously appeared as D. J. Rosenman, C. R. Connors, W. Chen, C. Wang, A. E. Garcia, $A\beta$ monomers transiently sample oligomer and fibril-like configurations: ensemble characterization using a combined MD/NMR approach. *J. Mol. Biol.* **425**, 3338 (2013).

Disordered proteins, by their nature, fail to adopt a single stable native structure and thus continuously undergo transitions between different metastable structural states at timescales much faster than the folded to unfolded transitions sampled by stable proteins (Figure 1.1). These rapid structural interconversions typically prevent IDPs from packing into well-ordered, isomorphous lattices, and thereby prevent structure calculation by X-ray crystallographic methods(34). These same transitions also decrease the dispersion, resolution, and intensity of the signals used to calculate distance and angular constraints with nuclear magnetic resonance (NMR) spectroscopy(34). Even when these methods are viable, only structural information from either a single configuration or a biased/averaged depiction of the ensemble of states that an IDP may sample can be obtained.

Given the difficulties involved in obtaining structural information from experiment for these flexible, polymorphic peptides, investigators often turn to computational simulations to investigate the biophysical properties of IDP ensembles. Over the years, the ability to perform extensive molecular dynamics (MD) simulations has improved. Today, simulations for small IDPs extend over multiple microseconds using explicit and implicit solvent models. Additionally, techniques such as replica exchange MD(35, 36) (REMD) and metadynamics(37, 38) are used to escape local free energy minima and enhance sampling. In general, the results obtained in simulations of intrinsically disordered proteins depend strongly on the set of parameters (such as the force field) used to describe the energy of the peptide and their interactions with the aqueous solvent. Widely used force fields for biomolecular simulations are OPLS-AA(19), AMBER99sb and variants(21, 39, 40), and CHARMM22*(28), while frequently used water models are three-site models TIP3P(20) and SPC/E(41), and the four-site models TIP4P(20) and TIP4P-Ew(42). These force fields have been calibrated against model compounds and peptides, and in many published instances the force fields reproduce folded conformations of small globular, water-soluble proteins with root mean squared deviations (RMSDs) within Angstroms of the experimentally determined structures(43).

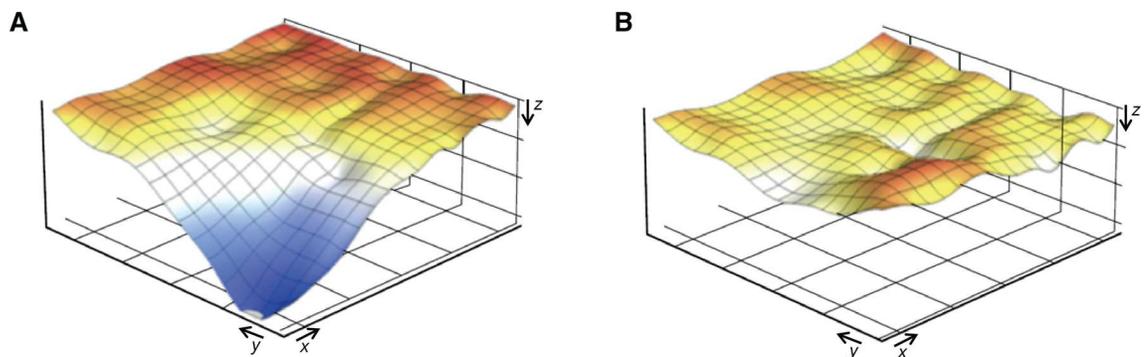


Figure 1.1. Schematics for the free energy landscape of a hypothetical (A) globular protein and an (B) intrinsically disordered peptide (IDP). The xy plane is an abstraction for the coordinate space the peptide can sample, while z represents the respective free energies of those conformations. The globular protein has an unambiguous global energy minimum state whereas the IDP does not. As such, even at equilibrium, an IDP will exist as a mixture of many different conformers. This figure was adapted from Fisher and Stultz(44), published in *Current Opinion in Structural Biology* with permission from © 2011 Elsevier.

However, IDPs, or even the unfolded ensembles of well-folded proteins, remain difficult test cases for our current range of computational techniques because they lack non-ambiguous global energy minima. There are more possible configurations and a greater entropy affiliated with unfolded proteins compared to their folded states(45). Modern force fields have undergone decades of refinement on well-folded proteins, but IDPs and unfolded proteins are not well optimized because of the difficulty in acquiring experimental data and models for these systems, and also because this field of study is still very young. The types of conformations sampled through simulation of unstructured proteins may be much more sensitive to simulation parameters than that of well-folded proteins. For the globular villin headpiece, for example, independent molecular dynamics (MD) simulations using different all-atom force fields were able to recapture the experimental folded structure and folding rate of the protein, but the unfolded states and folding mechanism were highly dependent on force field choice(28). Meanwhile, in the case of two intrinsically disordered proteins (a fifty residue peptide derived from a FG-nucleoporin and a twenty residue RS repeat peptide), μ s length replica exchange MD simulations with four different all-atom force fields were found to adopt substantially different hydrogen bonding, secondary structure tendencies, and radii of gyration (with all simulations underestimating the experimental values of that last parameter to varying degrees)(46, 47). This overestimation of compaction has also been reported for standard MD simulations of Histatin 5, a 24 residue cationic IDP, using different AMBER and GROMOS force fields(48).

With these findings in mind, the force fields that are capable of reversible folding of globular proteins (AMBER99sb(39) with *(40) and/or ILDN(21) modifications and CHARMM22*(28) for example, in the following studies(43, 49, 50)), may not necessarily be the best at characterizing the metastable states of disordered ensembles. Indeed, even with long timescale simulation using these well-performing CHARMM22* or AMBER99sb-ILDN* force fields, simulations of various unfolded and disordered proteins, including the acid denatured Acyl-CoA binding protein, the natively unfolded zinc-free N-terminal domain of HIV-1 integrase, and the thermally unfolded states of other proteins, gravely underestimated the radii of gyration compared to experimental measurements either directly made for that system or predicted through a model derived

from small-angle x-ray scattering (SAXS) measurements of denatured peptides(51-53). This disparity worsens as the system size increases, and it was further observed that the unfolded states of globular proteins had only marginally increased radii of gyration compared to their folded counterparts(53). Finally, it must also be emphasized that these studies were performed using a force field (CHARMM22*) shown to be least prone to underestimating the radii of gyration for other IDPs(46, 47), so the discrepancy described above may be even worse with other commonly used force fields. That the collapsed states of these ensembles are promoted through these simulations is further evidence of an inherent bias of standard force fields created by their selective optimization against folded, globular proteins. Given these circumstances, extra care must be taken to test computational characterizations of IDPs, both for consistency across multiple simulation studies and through stringent experimental validation. Further, new force field parameters or simulation methods may be needed to better model these peptides; some recent developments in this field are described in Chapter 6.

1.2 Amyloid β and Alzheimer's Disease

A large portion of this thesis focuses on the study of amyloid β ($A\beta$), a peptide intimately implicated in the etiology of Alzheimer's disease (AD). In addition to the connection to neurodegenerative disease, this peptide's monomeric form is both primarily unstructured in water and prone to aggregation(2-4), and thus may serve as a model system for understanding the biophysical properties of IDPs and amyloidogenic proteins and their role in neurodegenerative disease. This section reviews various topics related to AD, with focus on the role, formation, and properties of $A\beta$. For a broader review of these topics and other proteins involved in AD, I also highly recommend Chapter 1 of the thesis(54) written by my colleague, Dr. Christopher Connors, which inspired the writing of this updated, more $A\beta$ -centric review section.

1.2.1 Societal and Financial Costs of Alzheimer's Disease

Alzheimer's disease is an incurable, terminal, neurodegenerative disease estimated to affect 36 million people worldwide(5) and 5.2 million in the United States(55). Though there are early onset forms (described in section 1.2.6), it is a disease that

prevalently affects the elderly; 32% of people age 85 and older in the US have AD(56). Life expectancy once diagnosed with the disease ranges from 7-10 years for patients 60-70 years old, while it is less than 3 years for those over 90(57). Further, the monetary cost of care for AD patients is estimated to be \$604 billion worldwide, representing 1% of the entire global gross domestic product(5).

1.2.2 Symptoms of Alzheimer's Disease

AD is characterized by three major categories of clinical symptoms: 1) cognitive dysfunction, such as memory loss and confusion, problems with language and speech, and loss of higher level planning and coordination abilities; 2) psychological or behavioral problems, such as mood swings, depression, and delusions; 3) difficulty or inability to perform everyday activities(58, 59). AD represents an estimated 60-80% of all diagnosed dementia cases, making it the most frequently reported type of dementia (other categories include vascular dementia and dementia with Lewy bodies)(55). The later stages of AD are characterized by a progressive physical deterioration and loss of body functions, which eventually results in disability and finally, death (most frequently, by aspiration pneumonia)(60). Neurologically, AD is marked by progressive degeneration of the brain and central nervous system (CNS), ultimately culminating in extreme reduction of the cerebral and entorhinal cortices as well as the hippocampus, an increase in ventricular volume, the presence of white matter lesions(60, 61), and the presence of the pathological hallmarks described in the next section.

1.2.3 Pathological Hallmarks of Alzheimer's Disease

Two key histological hallmarks of AD are intracellular neurofibrillary tangles (NFTs) and extracellular senile plaques; both of these features become increasingly ubiquitous as AD progresses, and their content has been directly linked to the severity of dementia in patients(60, 62-64). The NFTs in AD are primarily composed of hyperphosphorylated tau protein, a protein that is normally found in the axons of CNS neurons and responsible for interacting with and stabilizing axonal microtubules(65). The tau protein monomer itself is an aggregation prone IDP, whose sampled conformations and self-association properties were observed to be profoundly affected

by cosolvents(66). There are six tau isoforms in human brains resulting from alternative splicing, the longest of which has 84 potential phosphorylation sites (on serines, threonines, and tyrosines)(67, 68). Phosphorylation of about 30 of these sites can occur in normal tau proteins(69). However, the hyperphosphorylation of tau has been linked to cell toxicity, and is implicated in a category of diseases known as tauopathies. This hyperphosphorylation of tau is thought to result in two consequences: 1) it prevents the tau protein from binding and stabilizing microtubules, leading to the disruption of these structures; 2) the unbound tau is then prone to associating into various aggregates, including NFTs (70, 71). The NFTs in AD are composed of subunits of tau aggregates known as paired and straight helical filaments(72, 73). This said, the significance of NFTs in the etiology of AD and other neurodegenerative diseases remains controversial. Doubts have been raised, for example, through observations that presence of NFTs only represent a small proportion of neuron loss in the brain, that neurons with NFTs have been observed to be viable for decades, and that apparently cognitively healthy individuals have also been observed to have significant content of NFTs in their CNS (71, 74-77).

The second hallmark, the senile plaque, is predominantly composed of β -rich, insoluble, fibrillar deposits of A β protein. More specifically, senile plaques can be composed of various products of amyloid precursor protein (APP) (see section 1.2.4) including full length A β 40, A β 42, A β 43 (the number indicating the length of the peptide, see Section 1.2.6), as well as amino-terminally modified and truncated versions of these peptides(78, 79). In addition, other molecules not derived from APP exist in plaques(80), including proteoglycans such as heparan sulfate(81-83), various metal ions(84, 85), inflammatory agents(86), serum amyloid P(87), and other proteins such as the non-A β component of amyloid plaques (NAC), now known to be aggregates of regions of α -synuclein(88-91). There are 4 categories of senile plaques, defined by the morphological and histological character of the deposits, as well as their interaction with neurites and glia: diffuse, primitive, neuritic or classic, and compact(92, 93). Although mature fibrils of A β (the primary component of certain senile plaques) are neurotoxic(94), senile plaques have been observed in cognitively healthy patients(74, 75), and neuronal death has been observed in AD transgenic mice models independent of the formation of

plaques(95, 96), leading to controversy over the role of these plaques in the etiology of AD (much like NFTs as described in the previous paragraph). Indeed, as discussed in more detail in Section 1.2.5, A β peptides can also form soluble, synaptotoxic oligomers that impair memory function; this type of aggregate is now actually believed to be the proximate neurotoxic factor underlying AD(95, 97-99).

As a final note, although the major AD hallmarks have been discussed individually above, there is a growing interest in studying the interactions between A β and tau in AD(68). First, it has been observed that injection of different aggregates of A β can induce the hyperphosphorylation and loss of microtubule binding of tau, and the formation of NFTs(100-102), while antibodies targeting A β have been observed to clear early hyperphosphorylated tau aggregates(103). On the other hand, tau deficient mice and neurons also have been observed to be protected from A β toxicity(104, 105). With the prevailing explanation behind AD being the amyloid cascade hypothesis (see section 1.2.4), these results in tandem suggest a mediating role for tau in pathologies (at least initially) driven by A β .

1.2.4 Amyloid Cascade Hypothesis

Although other explanations have been proposed in the literature (such as the tau and inflammation hypotheses(106)) and the increase in scrutiny of this hypothesis as of late due to the recent failure of several different anti-A β therapeutics to pass clinical trials(106, 107), the most widely accepted description of the etiology of AD remains the amyloid cascade hypothesis. Initially proposed by Hardy and Selkoe(108), this hypothesis proposes that accumulation of neurotoxic A β 42 aggregates is the chief step in the neuronal death, synaptic failure, and pathogenesis associated with AD. Support for this hypothesis stems from the fact that mutations affiliated with familial Alzheimer's disease (FAD, see section 1.2.9) are located in the amyloid precursor protein or proteins that process it (see section 1.2.5) and consistently result in an larger ratio of A β 42/A β 40 (109-111), as well as the inherent neurotoxicity of A β aggregates (covered in section 1.2.6) and the discovery that A β can induce the hyperphosphorylation of tau (see section 1.2.3), causing further toxicity. Further, patients with Down syndrome (also called trisomy-21), who carry an extra allele for the chromosome 21-encoded amyloid

precursor protein, have both a higher level of A β in their CNS and a higher incidence of AD(112). Indeed, AD has a prevalence rate of 30-75% in people with Down syndrome of age 60 to 69(113), which is substantially larger than the 15% prevalence rate in the general American population between age 65 to 74(55). The amyloid cascade hypothesis, as originally proposed, has since been further revised to incorporate A β 40 as an inhibitor of A β 42 aggregation and AD pathogenesis (see section 1.2.5 for a discussion of the evidence behind this revision)(114), and to shift focus to the importance of A β soluble oligomers as the causative agents of the disease (see section 1.2.6)(115).

1.2.5 Processing of A β

A β is produced by the endoproteolytic cleavage of a transmembrane protein, the amyloid precursor protein (APP), by one of the two β -site APP cleaving enzymes (BACE-1/2) and γ -secretase(116, 117). APP itself is processed by several different enzymes into products of varying length and function (Figure 1.2). APP is an evolutionarily old and well conserved protein(118), expressed in eight isoforms in humans through alternative splicing(119), with the 695/696 residue variants the most common in septal neuronal cells(120). APP is structurally characterized by a number of separate domains, including an extensive luminal region separated into two domains (E1 and E2) connected by an acidic linker region, a much smaller cytosolic domain, and a transmembrane domain that goes through the membrane once(117, 121). The A β sequence within APP goes from the extracellular domain into part of the transmembrane domain; the sites of processing of APP by secretases occur near or within this sequence(117).

Leftward from the center of Figure 1.2 illustrates the “non-amyloidogenic” pathway of APP processing, so-called because A β is not produced through this branch. This pathway is estimated to occur >90% of the time in healthy humans, with various mutations and disease states changing this ratio(122). This branch begins with cleavage by a series of enzymes collectively known as α -secretases, which cuts APP at the α -site, located between residue K16-L17 (numbering matches the A β sequence, we will use this convention throughout this section to maintain consistency), yielding two fragments

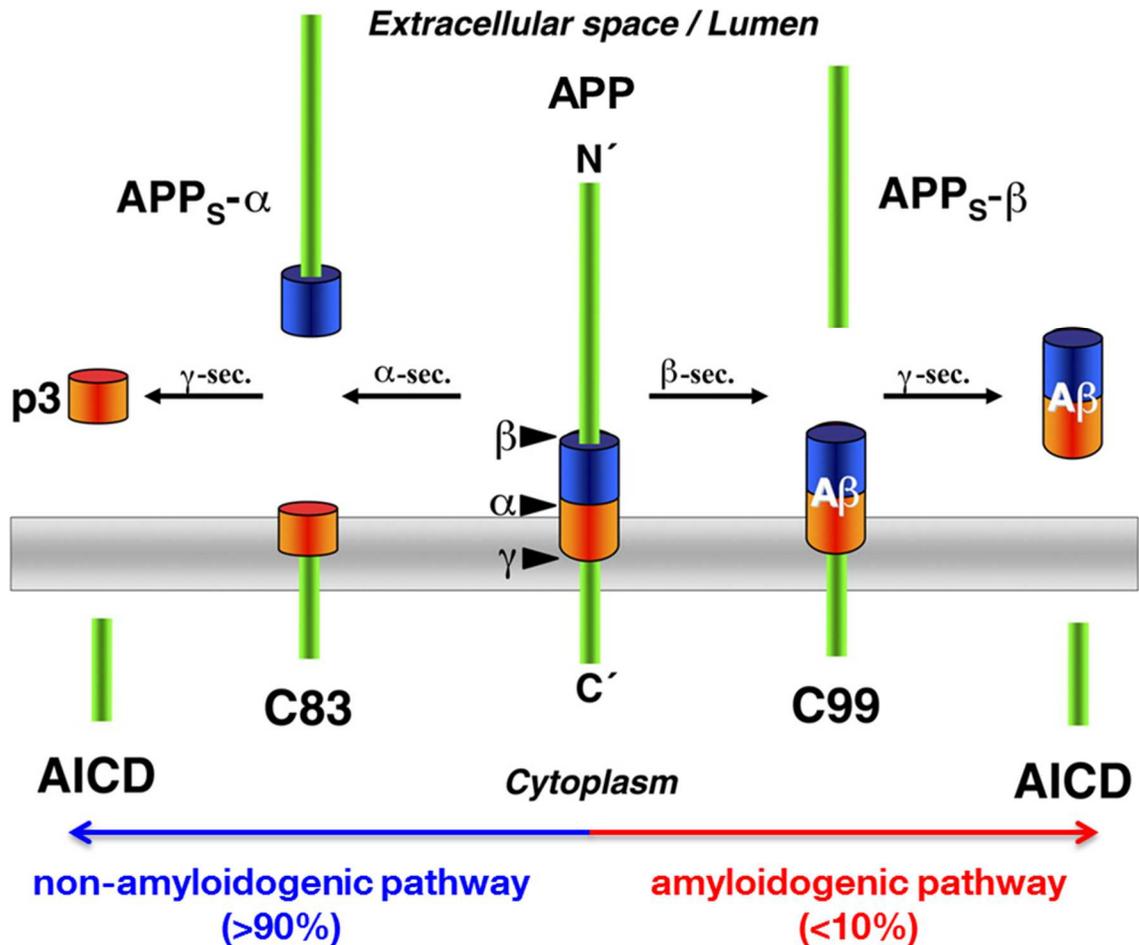


Figure 1.2. Diagram illustrating two different pathways for the processing of the amyloid precursor protein (APP, depicted in the center). The more common route (leftward) begins with α -secretase cleavage and results in the non-amyloidogenic products APP_S- α , C83, p3, and the APP intracellular domain (AICD). The alternate, “amyloidogenic pathway” (rightward) begins with β -secretase cleavage by one of the BACE enzymes, forming APP_S- β and the intermediate product C99. C99 can be further processed by γ -secretase to produce AICD and A β . This cleavage step is variable and can result in A β peptides of different lengths. This figure was adapted from © Kaether and Haass, 2004. Originally published in *The Journal of Cell Biology*. doi: 10.1083/jcb.200410090(123). Permissions for reuse are granted by the “non-commercial third-party reuse” clause specified by the publisher, The Rockefeller University Press.

called APP_s- α and C83(117, 124). C83 may be further processed by γ -secretase (discussed later in the section) to produce the two fragments: p3 and the APP intracellular cytoplasmic (or C-terminal) domain (AICD)(117, 124). Although this is called the “non-amyloidogenic” pathway, the p3 fragment (consisting of residues 17-40 or 17-42 of A β) has no known biological function(124), and there is debate as to whether it can form aggregates or has a pathogenic role (125-128).

The remainder of this section describes the “amyloidogenic” pathway of APP processing responsible for producing A β , which is shown through the rightward arrows starting from the center of Figure 1.2. This pathway begins with β -secretase cleavage, which can be performed by either of two homologous enzymes: BACE-1 and BACE-2. It has been asserted that BACE-1 is the relevant form in AD due to BACE-2’s low levels of expression in the brain and observations the second form predominantly cleaves at a non-A β yielding θ site between F19 and F20(129-131). However, it has also been argued that BACE-2 may be important as BACE-1 deletion only prevents production of A β in neuronal cells of mice and not in the ten times more abundant glial cells of the brain, unlike the double BACE-1/2 knockout(132). β -secretase cleavage can actually happen at two sites: the β site, which occurs right before D1 (in A β numbering) and generates the full N-terminus of A β , and the β' site, which occurs between Y10 and E11, generating an N-terminally shortened peptide(133). Shifts in the preference between these sites in BACE-1 has been linked to familial AD(134). When cleaved at the normal β site, two fragments are produced from APP: APP_s- β and C99(117, 124). Between α -, β -, β' -, and θ -cleavage and other N-terminally truncated A β s, there is a large degree of variability in the generation of the N-terminus of A β and A β -like peptides.

The generation of the C-terminus of A β is also prone to variable cleavage, though the different fragments are generated by a single protease in this case. This enzyme, γ -secretase, is actually a complex of at least 4 individual integral membrane proteins: anterior pharynx-defective 1 (APH-1), presenilin enhancer 2 (PEN-2), nicastrin (NCT), and presenilin (PS)(135). In humans, there are two different homologues of APH-1 (APH-1a and APH-1b) and PS (PS1 and PS2)(135); APH-1a also has two different isoforms through alternative splicing, allowing for up to six different γ -secretase complexes(136). This complexity can increase even further due to the association of

additional proteins, such as the non-essential regulatory subunit CD147(137). The PS subunit is an aspartyl protease responsible for catalysis of APP(138), while the other units have roles in assembling, stabilizing, and/or trafficking the complex(139-141). A full cryo-EM reconstruction of a human γ -secretase complex has recently been resolved to a resolution of 4.5Å; while this model is still insufficient to describe atomic-level details, the overall complex architecture has been clarified, revealing a horseshoe-shaped complex with 19 transmembrane segments(142). Additionally, the structure of a homologue of the presenilin subunit in the archaeon *Methanoculleus marisnigri* JR1 was resolved through X-ray crystallography(143), while the structure of the C-terminal fragment of human presenilin was also solved through NMR spectroscopy in SDS micelles(144).

Ultimately, γ -secretase processes C99 in the transmembrane region of the fragment to produce a mature A β peptide and an AICD fragment(117, 124). However, variability in the exact location of the γ -secretase cleavage site results in the production of different A β alloforms, ranging in length from 34 to 43 residues(145). The most predominant of these species are A β 40 (~90% of A β content secreted in the brain) and A β 42 (~10%), the latter of which is characterized by higher amyloidogenicity and toxicity(99, 146, 147). As such, there are two major γ sites (between V40 and I41 and between A42 and T43). It is the A β 42 form that deposits first and makes up the primary component of AD senile plaques(78). A β 40 and A β 42 species demonstrate distinct mechanisms of oligomerization(6, 7), and A β 40 slows the aggregation of A β 42 monomers in a A β 42/A β 40 ratio-dependent manner(114, 148).

The region where γ -secretase begins its initial cleavage of C99 (and other membrane anchored APP fragments) is the ϵ -site, occurring between T48 and L49 or between L49 and V50(145, 149, 150). This site is located in the α -helical, homodimeric, transmembrane domain of APP (APPTM)(151). The successive cleavage mechanism of γ -secretase releases tripeptides or tetrapeptides along each α -helical turn of APPTM(145). Starting from an ϵ -cleavage that yields A β 48, primarily A β 45, A β 42 and A β 38 are produced (the A β 42 product line), while A β 49 begets A β 46, A β 43, and A β 40 (the A β 40 product line)(145, 152). Thus, the variable cleavage of APP by γ -secretase (and, thus, the existence of different C-terminal alloforms of A β) is linked to the

successive nature by which γ -secretase cleaves A β and variability in the location of the ϵ -site. Interestingly, structural studies of APPTM reveal that certain mutations linked to familial AD in this region (but outside the A β sequence) may promote the accessibility of the A β 48 affiliated ϵ -site relative to the A β 49-producing site, which may explain why these mutations increase the A β 42/A β 40 ratio(151).

As a final note, the functional roles of A β peptides, or even APP, in the CNS remains a subject of controversy in the literature(153). Disruption of APP in mice has been linked to learning, anatomical, and behavior abnormalities, suggesting that it plays a role in the long term potentiation of neurons(154-156). APP itself has been implicated in the control of neural plasticity(157), functioning of synapses(158), axonal vesicle transport(159), and regulation of iron transport(160). Even blocking or disruption of A β has been linked to impaired learning in mice(161), and roles have been suggested for A β as a controller of synaptic vesicle release(162), regulator of ion(163) and cholesterol(164) transport, an antioxidant(165), a signaling molecule(166), and even as an antimicrobial peptide(167). Meanwhile, other products of APP, such as APP_s- α , are implicated in many neuroprotective functions(168), and supplementing APP_s- α in APP deficient mice has been observed to rescue the aforementioned defects of APP-disrupted mice(169). The AICD fragment is also believed to be involved in signaling and gene regulation through its interaction with the Fe65 adapter protein(170, 171). Lastly, A β has been observed to act as a prion, wherein injection of A β aggregates was observed to induce the widespread deposition of A β in the brains of live, transgenic mice expressing APP(172).

1.2.6 Aggregates of A β

As mentioned in section 1.2.2, A β can aggregate to form the fibrils that make up the primary component of senile plaques. Kinetic studies of A β fibril formation are typically done using the dye thioflavin T, which exhibits enhanced fluorescence upon preferential binding to the cross β structure of amyloid fibrils(173). These assays reveal that wild type A β aggregation is characterized by a nucleation/lag phase with little change in fluorescence with time, followed by rapid increases in fluorescence during an elongation/growth step(174). Another property of interest in investigating the fibrillation

process is the critical concentration, the initial soluble peptide concentration above which fibrils are the dominant species once equilibrium is reached and below which lower molecular weight forms are preferred(175).

As the final, thermodynamically favored form of wild type A β under most sample preparations, these amyloid fibrils have been characterized through several structural studies *in vitro*, primarily using solid state NMR. In general, these studies reveal a characteristic in-register, intermolecular parallel β sheet architecture for both A β 40 and A β 42(23, 24, 26, 27, 176-179). These typically feature residue E11-D23 and A30-V40/A42 in intermolecular β interactions, while the intervening residues (V24-G29) form a turn region, creating an intrapeptide “horseshoe” topology whose side chains pack in the center (though the patterns of these side chain interactions are debated)(27). The N-terminal region is believed to form an isolated, unstructured tail located outside the main fibril strand and is often not modeled in solid state structures of full, wild type fibrils(23, 24). This region only encompasses D1-S8 in certain models(23, 24), though others have suggested that this N-terminal disordered region extends until residue L17(26, 176). It has alternatively been argued that the N-terminus forms an independent region of parallel β interactions instead(27). Either way, in each of these fibril models, a series of the central region to C-terminal β_1 -turn- β_2 motifs stack in a parallel way to form a fibril strand. These filaments also interact with each other polymorphically, forming different quaternary structures. The first way is where the fibril strands pair in a 2-fold 180° symmetric pattern along the long axis of the filament, creating a “striated ribbon” morphology(23, 177). The second morphology is called a “twisted pair”, and it involves filaments interacting in a 3-fold rotationally symmetric way(24, 177). The cross β components of these two fibrils are similar, but feature some differences in side chain interactions and packing(177). In particular, the turn region of striated ribbon fibrils (but not twisted pair fibrils) feature a salt bridge between D23-K28(23, 177) that is believed to be key to fibril nucleation(180). As a final note, brain-derived A β 40 fibrils – that is, aggregates extended with isotopically labeled peptide *in vitro* using amyloid fibrils from brain tissue of two different AD patients as seeds – have also been characterized with solid state NMR(178). Structural data from fibrils derived from these seeds suggest these fibrils are roughly organized as parallel strand-bend-strand motifs like synthetic fibrils,

but these data also reveal that the model derived from one patient shows a distinct molecular structure compared to either data from the other patient or *in vitro* fibrils(178, 181). In particular, the brain derived model features a twist in F19-D23, a kink located at residue G33, and a bend that spans G37-G38, all of which contrast with the unbroken β sheets observed in synthetic fibrils(178, 181).

In addition to fibrils, A β can form various types of oligomers and other aggregates; a schematic illustrating the web of different aggregation pathways the peptide may adopt is illustrated in Figure 1.3. Unlike mature fibrils, these alternative or intermediary aggregates are often very hard to characterize because they are often inhomogeneous or unstable, and the type of aggregates formed are highly sensitive to peptide concentration, sample conditions, temperature, and time. Further, while soluble oligomers of A β as a collective are implicated as the causative agents of disease in AD, the actual identity (or identities) of these aggregates has not yet been defined(115).

En route to fibril formation are A β 42-specific pentameric or hexameric paranuclei, which are abolished with the toxicity reducing, aggregation-slowing M35 oxidation modification(7, 182). A neurotoxic A β 42 pentamer that may match the identity of this paranucleus was studied and analyzed with solid state NMR; this prefibrillar aggregate is associated with enhanced toxicity compared to other forms and is capable of converting to fibrils(26). This NMR data revealed a more loosely aggregated structure compared to fibril models that protects the C-terminus from solvent exchange, and a unique intrapeptide model characterized by at least three solvent accessible turns at H13-Q15, G25-G29, and G37-G38 as opposed to only one affiliated with fibrils (G25-G29)(26). Different small molecular compounds have also been used to remodel A β 42 into different oligomers, the most neurotoxic of which were prefibrillar(183, 184). Further bridging the gap between earlier “on pathway” oligomers and fibrils are protofibrils, the soluble penultimate intermediate of insoluble A β fibrils(185-187). Oligomers, protofibrils, and fibrils are all neurotoxic forms that correlate with dementia in patients(62-64, 94, 188).

Besides aggregates like protofibrils and paranuclei, which are both believed to be “on pathway” for aggregation into fibrils(189), various other A β 42 oligomers have been

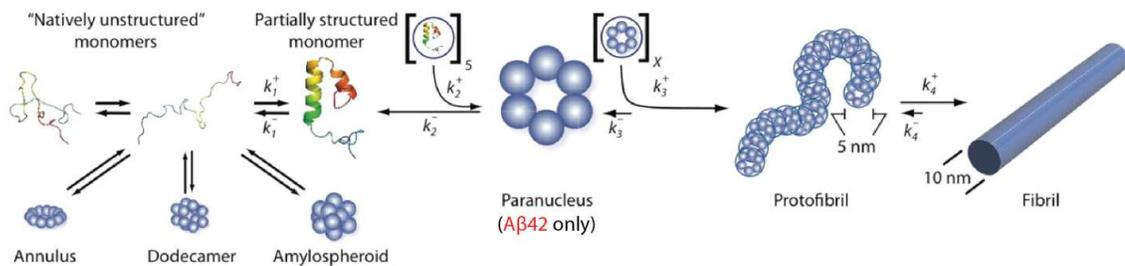


Figure 1.3. Schematic illustrating different pathways of aggregation for amyloid β ($A\beta$). “On pathway” aggregation refers to the pathway that ultimately results in the formation and elongation of characteristic amyloid fibrils. One species involved in fibril growth is called a paranucleus, a pentameric or hexameric oligomer characteristic of $A\beta_{42}$ self-association only. Besides “on pathway” aggregation, there are many different “off pathway” complexes formed by $A\beta$, such as annuli, dodecamers, octadecamers, and amylospheroids. In monomeric form, $A\beta$ is an intrinsically disordered peptide that samples many conformers on a fast timescale. The monomer is a building block common to the complex scheme of $A\beta$ aggregation mechanisms, and its conformational flexibility likely underlies the broad range of pathways the peptide can self-associate through. This figure was originally published in *The Journal of Biological Chemistry*. R. Roychaudhuri, M. Yang, M. M. Hoshi, D. B. Teplow, Amyloid β -Protein Assembly and Alzheimer Disease. *Journal of Biological Chemistry*. 2008; 284:4749. © 2008 American Society for Biochemistry and Molecular Biology(189). Permissions for reuse are granted by the “For other parties using material for noncommercial use” clause specified by the publisher.

identified, most of which exhibit synaptotoxicity, including (in size order): dimers(190), trimers(191), a tetrameric preglobulomer(25), three types of dodecamers (A β -derived diffusible ligands (ADDLs)(192), A β *56(193), and globulomers(194, 195)), octadecamers(196), and amylospheroids(197). Experimental structural characterization of these alternate forms is relatively sparse due to their inhomogeneity. Solid state NMR was used to characterize SDS-trapped A β 42 preglobulomers and globulomers, revealing a mixture of intramolecular antiparallel β interactions between V18-D23 and K28-I32 and intermolecular parallel β interactions involving G33-V40 in the repeating dimeric unit of the tetrameric preglobulomer(25). Similar NMR techniques were also used to characterize a different, 150 kDa SDS-trapped A β 42 off-pathway oligomer (roughly ~33 A β molecules), revealing that this form possesses intermolecular, antiparallel β sheets that differ from the in-register, parallel sheets seen in fibrils(198). While A β monomers themselves are nontoxic (199-201), they directly form aggregation nuclei and pathogenic oligomers for many of the species described above. Characterization at the monomer level may reveal key structural features that serve as determinants of A β self-assembly. Manipulation of these processes could block the formation of the toxic species involved in AD.

1.2.7 Experimental Properties of A β Monomers and Their Role in Aggregation

Characterizing the monomeric state of A β under physiological conditions may be key to understanding how A β assembles into disease-related fibrils and oligomers because they represent a base state common to all aggregation pathways(108). This knowledge can potentially be used to rationally develop therapeutics that prevent nontoxic monomers from progressing into toxic species; indeed, this has been defined as one of the fundamental strategies in the ongoing effort to treat Alzheimer's disease(189). As discussed in detail in this chapter (for example, see sections 1.2.4 and 1.2.9), A β self-assembly can be profoundly influenced by very subtle chemical changes, ranging from the two-residue difference between A β 40 and A β 42(6, 7), to the point mutations that are linked to familial Alzheimer's disease (FAD)(8), and even to the single atom chemical modification caused by oxidation of the M35 side chain(202-204). The polymorphism of monomeric A β under physiological conditions may underlie this relationship. In the

absence of unambiguous, stable native states, simple chemical modifications could have a profound effect on the types of ensembles sampled by that particular A β peptide. This intrinsic disordered property, in addition to the high aggregation propensity, has frustrated experimental efforts to characterize A β peptides. Nevertheless, there is a breadth of evidence highlighting the importance of critical regions in A β monomeric folding and aggregation propensity.

A number of studies reveal the importance of the central hydrophobic cluster (CHC, residues L17-A21) and the C-terminus of A β peptides. These hydrophobic regions have been well established to form intermolecular β sheets in mature fibrils and other higher order aggregates (as discussed in section 1.2.6)(23-27, 176), but their role in the self-aggregation and monomeric ensemble of A β is just beginning to become clear. Incubation of A β 40 fragments with the full length peptide show enhanced fibrilization rates only for the fragments containing residues L17-F20 or A30-M35(205). Proline mutations of residues in the L17-F20 or A30-M35 regions in A β 40 and A β 42 are more disruptive to fibrilization than mutations in other regions(206-208). There is also evidence suggesting a bias toward β -strand character in these regions in A β monomers. Solution NMR studies targeting the monomeric state best characterize A β with a collapsed coil ensemble and not by a unique structure(209, 210). Nevertheless, backbone H α , C α , and C β chemical shift indices suggest β -strand propensities in the CHC, I31-V36, and V39-I41, as well as turn character in D7-E11 and F20-S26 in A β 42(4). Residues V18-F20 (in A β 40, A β 42, and A β 42-M35ox) and V39-I41 (in A β 42 and A β 42-M35ox) also possess experimental $^3J_{\text{HNHA}} > 7.5$ Hz, indicative of a bias toward ϕ dihedrals characteristic of β -strands in these regions(211, 212). Tethered A β 40 demonstrated decreased spin mobility at regions H14-V18, G29-A30, and G38-V40 when investigated with electron paramagnetic resonance spectroscopy, indicating increased order in these regions(213). Far UV circular dichroism spectra for A β 40 and A β 42 monomers are dominated by random coil, but also suggest substantial β -strand content(7, 214). Although these biases may characterize the general A β ensemble in aqueous conditions, other individual structures may be possible. For example, an NMR structure of A β 40 was reported by Vivekanandan *et al.* that forms a ten residue α helix from H13 to D23(215). This contrasts with other studies that were unable to determine

sufficient medium or long range nuclear Overhauser effect (NOE) restraints to gauge secondary or tertiary structure through NMR(216); Vivekanandan *et al.* cite the additional salt content (50 mM NaCl) in their sample conditions as a possible contributor to the stabilization of their determined structure(215).

Further evidence has helped clarify the role of the additional C-terminal residues in A β 42-specific aggregation. NMR relaxation data reveal that A β 42 demonstrates more rigidity at the C-terminus than A β 40, both in terms of side chain and backbone dynamics(217, 218), suggesting residual structure formation. This is consistent with the observation that a hairpin promoting motif at the C-terminus (G33V, V36P, G38V; the “VPV” motif) was found to increase the A β aggregation rate and promote higher order oligomers, while inclusion of a hairpin breaking motif (V36D-Pro, G37L-Pro; the “pP” motif) substantially slowed the A β 42 aggregation kinetics and changed the oligomer size distribution to one more characteristic of A β 40(219). Oxidation of the M35 side chain also abolished A β 42 paranucleus formation, while point mutations of residue I41 and A42 either promoted or disrupted A β 42 characteristic aggregates, depending directly on the hydrophobicity of the mutant side chain(202-204). These findings suggest that the formation of an additional β topology sampled at the C-terminus, driven by hydrophobic side chain interactions, may be responsible for A β 42’s unique assembly properties.

The central region, consisting of the more hydrophilic, intervening residues A21-A30, has also been implicated for its unique properties and effects on A β assembly. This region was identified due to its inherent resistance to proteolysis, which is maintained when the A β (21-30) fragment is isolated. Solution NMR of this fragment reveals that V24-K28 samples a turn-like structure that may be critical in the folding of the monomer(216). Further, substitution of contiguous pairs of residues in the V24-N27 region with a turn-nucleating D-ProGly motif accelerated fibril self-assembly of A β 40(220), while turn-altering chemical modifications G25-S26 increased the oligomerization order of A β 42 (221). A β monomers readily exchange all amide protons with solvent in mass spectrometry and NMR amide hydrogen/deuterium (H/D) exchange experiments measured on the minutes timescale(222). However, data for A β have been collected using CLEANEX-PM with fast ¹⁵N HSQC (fHSQC) detection(223, 224), a pulse sequence well suited for IDP solvent exchange measurements because it employs a

fast mixing time (up to the millisecond timescale) and because signal is actually increased upon heightened exchange with water rather than lost (as is the case with more commonly used amide H/D exchange experiments). The CLEANEX-PM-fHSQC experiments on A β indicate residues in the central region, such as D23 and S26-G29 are particularly susceptible to exchange(225, 226), consistent with the characterization of the region as a solvent exposed turn. Lastly, charge-altering point mutants of residues E22 and D23 that are implicated in familial Alzheimer's disease (FAD) and cerebral amyloid angiopathy (see section 1.2.9) also demonstrate increased oligomerization orders and fibrilization rates when introduced into A β 40(227). Further, the oligomerization propensities of each of these mutants in full length A β are directly correlated to both susceptibility of trypsin proteolysis and instability of the V24-K28 turn for the A β (21-30) fragment(227). These observations suggest that these mutants may destabilize turn like structures in the central region, possibly changing the ensemble and allowing the monomer to seed different types of aggregates.

1.2.8 Simulations of A β Monomers

The challenges and limitations inherent to the current set of experimental techniques for studying these polymorphic, aggregation-prone A β monomers have encouraged many groups to use a wide variety of computational techniques to more thoroughly investigate the biophysical properties of these peptides. A β was previously characterized through all-atom REMD simulations with OPLS-AA(19)/TIP3P(20) and AMBER99sb(39)/TIP4P-Ew(42) that compared well with NMR observables, revealing conformations where A β was mostly flexible but possessed some structured segments; in particular, β -hairpins populated the C-terminus in A β 42 but not A β 40(9, 11). A different approach was taken by Ball *et al.*(13), who used 100 ns multiple reservoir replica exchange (MRRE) simulations with AMBER99sb(39)/TIP4P-Ew(42) to determine that A β 42 is mostly disordered, with significant α -helical character in residues Y10-F19 and E22-N27 and little to no β -sheet content. The same group then reported their analysis of multiple trajectories acquired with the same simulation method, this time processing the resulting ensemble using the ENSEMBLE package(228-230) to select structures that best match experimental chemical shifts, residual dipolar couplings, $^3J_{\text{HNHA}}$ couplings,

and ^1H - ^{15}N NOEs(17). After refinement with ENSEMBLE, the aforementioned α -helical content is de-emphasized, and antiparallel β hairpins between the K16-A21 and G29-V36 regions are promoted in A β 42. A β 40 after ENSEMBLE refinement, in contrast, is characterized by reduced C-terminal β propensity and sampling of a hairpin between the CHC and residues G9-H13. The data suggest that the extra two residues in A β 42 primarily promote hydrophobic clustering that increases the β content in the CHC and G29-V36 region, rather than the direct formation of additional secondary structure, such as a C-terminal hairpin like discussed above. Finally, the proclivity for A β 42 to form pockets that can bind small molecules was investigated by Zhu *et al.*(18) The 100 ns/replica REMD simulation with AMBER99sb(39)/TIP3P(20) conducted by this group also revealed that A β 42 formed contacts between L17-A21 and I31-V36 with intervening turn regions, consistent with hairpin formation between these regions. Cluster centroids of the simulation were analyzed with FTMap(231) and FRED(232), which revealed that the N-terminal part of the molecule (up to residue 20) had a relatively higher probability of binding small molecules. In contrast, the central region (particularly D23-G25) had a much lower tendency to form binding hotspots. This reduced probability agrees with the reduced experimental ^1H - ^{15}N HSQC chemical shift perturbation of central region residues when certain small compounds are titrated into A β monomer sample(233-235).

Other approaches have also been used to investigate the properties of A β . Different A β alloforms and mutants were also explored using the Folding@home platform, using thousands of MD trajectories simulated with AMBER99sb(39)/TIP3P(20), each of average length \sim 30 ns, for each species(14, 15). Both wild type A β s are described as mostly disordered ensembles, with some α helical character from residue Y10-F20 and almost no β -content. Of the β content that exists, β sheet propensity near the C-terminus is notably less in A β 40 compared to A β 42. Granata *et al.*(16, 236) investigated the A β 40 ensemble with NMR guided metadynamics, which uses experimental data as collective variables to drive metadynamics calculations rather than using them purely for simulation validation or as hard structural restraints. Simulations were carried out with CHARMM22*(28)/TIP3P(20) at 350 K for 8 replicas for 310 ns/replica. Each replica was biased by a history-dependent potential acting on a different collective variable,

including two variables for the difference between predicted and experimental chemical shifts. The resulting ensemble was represented by an “inverted” free energy landscape compared to globular proteins, where progressively misfolded/unfolded states are more energetically favorable and native structure is only transiently explored. In terms of structured metastable states, only short β sheet structures are favored, but besides short α helices, there is also an energy minimum represented by structures with long α helical content that are similar to the Vivekanadan *et al.*(215) structure. A β monomers have also been investigated with extensive simulations using coarse grained models(12) and Monte Carlo simulated annealing(10). The former revealed A β ensembles that contained mostly turn structure but possessed substantial β sheet propensity in the N-terminus, while the latter characterized A β 42 as possessing strong β probability in many of the residues throughout the peptide.

Simulation of A β is prone to the difficulties discussed in Section 1.1. While the interface between these most recent and large scale studies have started to bear some consistent depictions of the properties of the A β ensemble, most characterizations based on computational studies of these peptides remain highly divergent. These variations may arise from differences in simulation parameters, extent of sampling, or trajectory analysis.

1.2.9 Familial Alzheimer’s Disease and Point Mutations of A β

While there are genetic risk factors underlying Alzheimer’s disease in general, there is a rare subset of cases in which mutations in known genes directly cause the disease. This classification is termed familial Alzheimer’s disease (FAD) and is estimated to describe less than 5% of all cases(55), while the other designation, sporadic AD, represents the vast majority of patients. FAD is further distinguished by its early onset compared to sporadic AD: the former generally occurs in patients 30-65 years, while the latter usually affects people older than 65. Further, patients with FAD are particularly prone to other symptoms such as cerebral amyloid angiopathy (CAA, the buildup of amyloid plaques in blood vessels of the brain), stroke, and hemorrhaging(237, 238). Despite the rarity of the disease, it can be particularly devastating to those affected by it due to its early onset. Further, FAD represents an important model for better

understanding the role of the genes and proteins associated with Alzheimer's disease in general.

FAD is caused by mutations, usually autosomal dominant, of APP, γ -secretase, and other proteins that process APP. Many of the mutations in FAD involve the production of A β peptides. Among these mutations is a common theme where the ratio of production of A β 42 to A β 40 is increased (109-111). The effects of this increased ratio may be an important key to the pathogenicity of these mutations; changes in A β 42/A β 40 have been linked to increased A β aggregation and neurotoxicity(239-241). Increasing A β 42 directly favors the creation of a faster aggregating, more neurotoxic A β species(99, 146), while disfavoring A β 40 reduces the production of a peptide that has been observed to reduce the aggregation of A β 42 monomers *in vitro* (114, 148) (see section 2.1.3).

Another category of FAD mutants are those mutants that exist in the A β sequence itself; these will be the focus of the remainder of this section. While some of these mutants may also affect the processing of A β , they could have an alternate or additional effect in changing the aggregation properties of the A β in the brain. Table 1.1 summarizes the effects of point mutations of A β previously reported in human subjects; indeed, these mutations are often associated with changes in A β production and the types of aggregates formed, but the effects and clinical manifestations of these mutants are also substantially different from each other. Observing the sequence location of these variants reveals certain hotspots in which FAD mutants are clustered, such as H6-D7 and A21-D23. The latter region contains the most mutations and is a site where two anionic side chains are adjacent in peptide sequence (E22, D23). All of the mutations in this hotspot are affiliated with increased fibril formation except A21G and E22G(242-251). Further, all of the mutants therein except A21G remove one of the two negative charges in the region. E22 and D23 have been implicated as salt bridge partners for K28 across a central turn region, for example, in the intrapeptide solid state NMR-derived models of wild type A β fibrils(23, 24, 26, 27); disruption of these interactions involving these residues may underlie the predisposition of mutations in this region to FAD pathogenesis.

Table 1.1. Phenotypic behavior of point mutations of A β found in human subjects. These mutations, with the exception of the neuroprotective mutant A2T, are known to cause familial Alzheimer's disease, with varying sets of neuropathological and clinical phenotypes. Further, these mutations are autosomal dominant, except the recessive mutant A2V(252, 253). The E11K(254) and E22 Δ (255) mutations may have incomplete penetrance. Most of the mutations listed here are also known to either affect A β aggregation and/or production. --- = unknown, N/A = not applicable. Table partially adapted from Tomidokoro et al.(238)

Kindred	A β Mutation	Onset (yrs)	Clinical Phenotype		Neuropathology		
			Cognitive Impairment	Stroke / Hemorrhage	Plaques	Neurofibrillary Tangles (NFTs)	Cerebral Amyloid Angiopathy (CAA)
Icelandic	A2T	N/A	N/A	N/A	N/A	N/A	N/A
Italian	A2V	36	Yes	No	---	---	---
British	H6R	55	Yes	No	---	---	---
Taiwanese	D7H	50-65	Yes	No	---	---	---
Tottori	D7N	60	Yes	No	---	---	---
Leuven	E11K	50	Yes	No	---	---	---
German	K16N	53	Yes	No	---	---	---
Flemish	A21G	35-60	Yes	Yes	Perivascular	Yes	Massive
Dutch	E22Q	30-40	Yes	Yes	Diffuse	Rare	Massive
Italian	E22K	50-60	Mild	Yes	Diffuse	No	Massive
Arctic	E22G	55-60	Yes	No	Compact	Yes	Scarce
Japanese	E22 Δ	36-62	Yes	No	---	---	---
Iowa	D23N	50-66	Yes	Rare	Diffuse	Yes	Massive
Piedmont	L34V	50-70	Yes	Yes	No	No	Massive
Italian/Spanish	A42T	58-68	Yes	Yes	Compact	Yes	Massive

Kindred	A β Mutation	Change in A β Aggregation	Change in A β Production	Reference
Icelandic	A2T	---	↓ A β 40 and A β 42	(256-260)
Italian	A2V	↑ fibril	↑ A β 40 and A β 42	(252, 253, 257, 259, 260)
British	H6R	↑ fibril, ↑ oligomers	None	(261-263)
Taiwanese	D7H	↑ fibril, ↑ oligomers	↑ A β , with ↑ A β 42/A β 40	(264)
Tottori	D7N	↑ fibril, ↑ oligomers	None	(262, 263, 265)
Leuven	E11K	None	↑ A β 40 and A β 42	(254)
German	K16N	↑ toxic heteromeric oligomers, ↓ fibril	↑ A β 40 and A β 42	(266)
Flemish	A21G	↓ fibril, ↓ protofibril	↑ A β , with ↑ A β 42/A β 40	(267-272)
Dutch	E22Q	↑ fibril, ↑ protofibril	↑ A β , with ↓ A β 42/A β 40	(242-244, 246, 270, 273, 274)
Italian	E22K	↑ fibril	↓ A β , with ↓ A β 42/A β 40	(244, 270, 275)
Arctic	E22G	↑ oligomers, ↑ protofibrils	↓ A β , with ↓ A β 42/A β 40	(199, 270, 276-278)
Japanese	E22 Δ	↑ fibril, with unique morphology	↓ A β 40 and A β 42	(248-250, 255, 279, 280)
Iowa	D23N	↑ fibril, can form fibrils with antiparallel β arrangement	↑ A β 40 and A β 42	(245, 247, 251, 281-283)
Piedmont	L34V	---	---	(284)
Italian/Spanish	A42T	---	None	(285-287)

Some of these FAD-linked mutations have been observed to form unique morphologies not yet observed during wild type aggregation. First, it has been observed that the Iowa mutation (D23N) in the context of A β 40 forms fibrils with faster kinetics and substantially more polymorphism at the molecular level than wild type(247). Further, the majority of A β 40-D23N's fibrils possess an antiparallel cross- β topology; this is currently the only A β variant observed to adopt this form (247, 251). The D23N mutant can also form in register parallel β fibrils more typical of wild type A β 40, but this may only represent a minor population of aggregates initially formed by this variant(247). Both antiparallel and parallel forms are toxic to neurons, but the antiparallel architecture is thermodynamically metastable and eventually exchanges to the parallel motif when both forms are initially mixed(251). This is because, while antiparallel fibrils are more efficient to nucleate, parallel structures show superior propagation properties; thus, antiparallel structures eventually dissolve while the thermodynamically favored parallel forms extend(251, 282). Due to this, the antiparallel aggregate has been later designated as a protofibril(179). It is suggested that the D23N mutation allows the antiparallel motif to become observable by disfavoring the nucleation rate of the parallel architecture, which in turn may be caused by the loss of the D23-K28 salt bridge that normally stabilizes wild type, parallel fibril structures(251).

The newly discovered Osaka (E22 Δ) polymorphism of A β , only reported in the literature as early as 2008(255), also demonstrates novel aggregation properties. The latest studies report that this deletion mutant forms fibrils with a unique morphology, with electron microscopy revealing that A β 40-E22 Δ forms a new type of short, single-stranded fibril(248-250), differing from the two-stranded striated ribbon model for wild type fibrils discussed in Section 1.2.6. A model of these E22 Δ characteristic fibrils is not yet available, but solid state NMR and X-ray fibril diffraction data confirm a distinct morphology that is not compatible with parallel β sheets characteristic of wild type fibrils(248). These E22 Δ fibrils have been reported to demonstrate low thioflavin T (ThT) fluorescence compared to wild type(248), commonly used as a measure of amyloid content. In spite of this, E22 Δ mutants forms β -sheet secondary structure 400-fold faster than wild type(250), and the reason for the low ThT response was discovered to be due to the novel architecture of the E22 Δ fibril itself, which exhibits both reduced

affinity and number of binding sites compared to wild type for the amyloid-binding dye(248). Indeed, E22 Δ aggregates are formed very rapidly in the context of A β 40 and A β 42, requiring no lag phase and a substantially reduced critical concentration than wild type(248, 250). Interestingly, mixing equal volumes of wild type A β 40 and E22 Δ A β 40 results in aggregates where the Osaka mutant's characteristic assemblies dominate(248). This said, adding wild type A β can also be used to slow the kinetics of E22 Δ 's accelerated aggregation (combining wild type and mutant A β 40 or A β 42 in a 9:1 ratio slows the aggregation to a rate in between that of pure wild type and pure mutant)(250). Unlike wild type, E22 Δ fibrils were also more resistant to disaggregation with cosolvents like HFIP and TFE(250). Finally, A β 40-E22 Δ is also associated with a greater toxicity to rat primary neuronal cell culture compared to wild type(249).

1.3 Other Proteins involved in Neurodegeneration

1.3.1 α -synuclein

α -synuclein is a 140 residue protein that is intrinsically disordered as a monomer in solution(89). It is ubiquitous in the brain(288, 289) and aggregates into the primary component of intraneuronal Lewy bodies that are a pathological hallmark of Parkinson's disease and Lewy body dementia(289-291). In addition to the predominant A β content, α -synuclein is also a precursor to the non-A β component of amyloid plaques (NAC) of the senile plaques in Alzheimer's disease(88-90).

α -Synuclein is primarily composed of three regions, summarized with the schematic in Figure 1.4. The N-terminal (residues M1-K60) is a positively charged region which possesses four 11-residue amphipathic repeats (including a KTKEGV consensus sequence) that forms an extended α -helix upon binding to the negatively charged surfaces of synthetic membranes, detergent micelles, and lipid vesicles(292-295). It was found that amino acids D2-A11 of the sequence are necessary to bind vesicles in yeast and form an α -helical structure *in vitro*(91). Indeed, in the absence of lipids/detergents and/or these N-terminal residues, the monomeric protein is natively unfolded(89, 91, 294). Correspondingly, deletion of the D2-A11 residues dramatically increased cell viability of yeast cells exposed to the peptide; indeed, as a general rule,

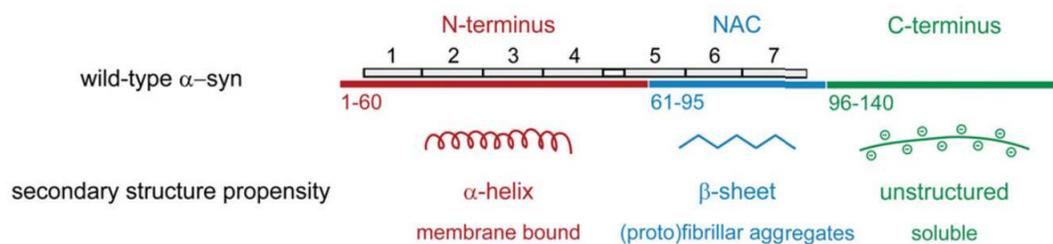


Figure 1.4. Schematic representing the three structural regions of wild type α -synuclein. The N-terminal 60 residues are disordered in water but form α -helical character after binding to membrane. The central section is the “non- $A\beta$ component of amyloid plaques” or NAC region, characterized by high hydrophobicity and a tendency to form β sheet aggregates. Finally, the C-terminus is more soluble, unstructured, and possesses several negatively charged residues and prolines. The N-terminus and NAC contain seven 11-residue repeats (containing consensus sequence KTKEGV) involved in helix formation. This figure was adapted from Vamvaca et al.(91), published in *The Journal of Molecular Biology* with permission from © 2009 Elsevier.

propensity of the variant to form α -helix and bind membranes directly correlated with cytotoxicity(91). That said, the same deletion of the very same residues in the context of human neuroblastoma SHSY-5Y cells did not affect membrane binding or peptide toxicity(296). The middle section (E61-V95) of α -synuclein is known as the non-A β component of amyloid plaques, or NAC, region(88-90); this part of the protein is prone to forming β -sheet secondary structures and fibrils(91). This region is clearly important to the aggregation of this protein, however, it has been found that cytotoxicity was not correlated to fibril formation among different variants of α -synuclein in yeast(297). The NAC sequence alone is mostly non-toxic, and deletion or duplication of the NAC region in the context of the full protein slightly reduced the toxicity in yeast, but to a significantly lesser extent than modifications that deleted the extreme N-terminus of the peptide(91). Lastly, the third C-terminal region (K96-A140) is rich in prolines as well as anionic residues, has no affiliated secondary structure, but is also not aggregation-prone. Deletion of residues in this C-terminal region had a negligible effect on cytotoxicity(91).

1.3.2 Cytoplasmic Superoxide Dismutase (SOD1)

The human cytoplasmic copper/zinc superoxide dismutase, known as SOD1, is an abundant cytoplasmic enzyme that, like all superoxide dismutases, reduces superoxide radicals (O_2^-) into hydrogen peroxide (H_2O_2)(298-300). Note that residue numbering in our discussion of this protein has been adjusted to match the indexing in PDB:1SOS(301) for the sake of consistency. Spinal cord associated, cytosolic aggregates of SOD1 are implicated in amyotrophic lateral sclerosis (ALS), an incurable motor degeneration disorder(302-307). ALS occurs in 2-3 of 100,000 people, and is often called Lou Gehrig's disease in the United States in memorial of the renowned baseball player who suffered from this disorder in the 1930s(308). ALS typically strikes its victims when they are 40-70 years old; patients with the disease undergo progressive paralysis and typically only survive 1-5 years after diagnosis(308, 309). SOD1 aggregates appear early in the progression of the disease, only small amounts of the SOD1 protein are needed to be pathological, and aggregates are exclusively in the affected neurons(310).

The molecular mechanism of SOD1's involvement in this disease remains unclear, but it appears that this aggregation is likely driven by familial ALS associated point mutations located throughout the SOD1 peptide, of which, more than 125 variants have been implicated spanning all five exons that encode the protein(311-313). 5-10% of ALS cases are familiarly-linked in an autosomally recessive nature(308), and of these, 20-25% are linked to these mutations in SOD1(313, 314); this said, familiar and sporadic ALS are virtually identical in terms of pathological profile(313). A gain-of-function mechanism for these mutations in familiar AD is widely accepted because certain variants do not abolish the metal binding properties or enzymatic function of the protein (such mutations include A1V, G37R, L38V, G85R, and G93A)(315-318), these mutations manifest in distinctive clinical manifestations and phenotypes(313, 319-321), and mice engineered to lack the *SOD1* gene do not develop ALS (though other age-related neuronal and muscular degenerative symptoms develop as a consequence)(322, 323).

Unlike A β and α -synuclein which are inherently disordered and aggregation prone, SOD1 is actually a very stable, well folded protein that is likely destabilized by familiar mutations and/or other chemical modifications, promoting an amyloidogenic state. The structure of properly folded, metal-bound, wild type human SOD1 protein has already been resolved(301, 324). The protein normally exists as a homodimer. Each monomer consists of 154 residues, adopts a β -sandwich fold with 8 antiparallel strands, and contains a highly conserved single intramolecular disulfide bond between two of the four cysteines in the sequence (C57 and C146) which is required to stabilize mature SOD1(325). Disulfide-reduced versions of SOD1 show a dramatic increase in the ensemble population of unfolded species(326), and several familiar ALS mutants of SOD1 destabilize this bond and create oligomers that rearrange the formation of disulfide bonds both intra- and intermolecularly(327). Each SOD1 monomer binds one zinc and one copper ion; the two metal binding sites are close to each other spatially, coordinated by histidines in the structure (copper binds H46, H48, H63, and H120, while zinc contacts H63, H71, H80(301, 324)). Mutations of SOD1 that disrupt either or both of these metal binding regions are linked to ALS(328, 329). Lastly, the SOD1 structure contains a highly charged loop between strand 7 and strand 8 (residues E121-R143) that

is also important for metal binding(318) and whose increased destabilization is associated with ALS linked mutants(330). Indeed, two regions of the protein: this electrostatic loop, as well as the zinc binding loop (residues R69-E100) are intrinsically disordered in the absence of metal binding and the C57-C146 disulfide bond(331, 332). Further, deficiency of metal binding (particularly copper) is linked to a loss of kinetic stability of the protein, even in wild type SOD1(333). In tandem, this suggests that *apo* SOD1 in general may be more inherently prone to aggregation and may have a role in disease progression.

2. Replica Exchange Molecular Dynamics Simulations of A β Monomers Using the OPLS-AA/TIP3P Force Field

2.1 Abstract

Amyloid β (A β) peptides are a primary component of fibrils and oligomers implicated in the etiology of Alzheimer's disease (AD). However, the intrinsic flexibility of these peptides has frustrated efforts to investigate the secondary and tertiary structure of A β monomers, whose conformational landscapes directly contribute to the kinetics and thermodynamics of A β aggregation. Here, *de novo* replica exchange molecular dynamics (REMD) simulations on the μ s/replica timescale are used to characterize the structural ensembles of different isoforms (A β 42, A β 40, and M35-oxidized A β 42) and familial AD-linked mutants (E22 Δ , D23N, and E22K in A β 40) with substantially different aggregation properties. Chemical shift and J-coupling data calculated from these REMD trajectories were compared to corresponding NMR-derived values, revealing that all simulations converge toward ensembles that yield good agreement with experiment. Though all A β species adopt highly heterogeneous ensembles, these are considerably more structured compared to simulations on shorter timescales. Prominent in the C-terminus are antiparallel β -hairpins between L17-A21, A30-L36, and (for A β 42) V39-I41, similar to oligomer and fibril intrapeptide models, which expose these hydrophobic side chains to solvent and may serve as hotspots for self-association. Compared to reduced A β 42, the absence of a second β -hairpin in A β 40 and the sampling of alternate β topologies by M35-oxidized A β 42 may explain the reduced aggregation rates of these forms. A persistent V24-K28 bend motif, observed in all three species, is stabilized by buried backbone to side chain hydrogen bonds with D23 and a cross-region salt bridge between E22 and K28. Further, mutations of E22/D23 disrupt the character of this bend and increase the polymorphism of hydrogen bonding across the central hairpin. These characterizations help illustrate the conformational landscapes of A β monomers at atomic resolution and provide insight into the early stages of A β aggregation pathways.

Portions of this chapter previously appeared as D. J. Rosenman, C. R. Connors, W. Chen, C. Wang, A. E. Garcia, A β monomers transiently sample oligomer and fibril-like configurations: ensemble characterization using a combined MD/NMR approach. *J. Mol. Biol.* **425**, 3338 (2013).

2.2 Introduction

Alzheimer's disease (AD) is characterized by senile plaques composed of β -rich, insoluble, fibrillar deposits of amyloid β (A β)(334). A β can also form oligomers that are soluble and neurotoxic(95, 146, 193). The amyloid cascade hypothesis involves aggregates of A β such as these in neural damage associated with the loss of memory in AD(108). Characterizing the monomeric state of A β can be key to understanding how A β assembles into disease-related fibrils and oligomers because they represent a base state common to all aggregation pathways(108), but many difficulties exist toward characterizing these peptides due to their intrinsic disorder property. These difficulties, along with the findings of numerous experimental and computational investigations to the structural properties of A β monomers, are described in detail in Chapter 1.

A β peptides are formed by the successive cleavage of the amyloid precursor protein (APP) by both β - and γ -secretases, generating the N-terminus and C-terminus of the peptide, respectively. Various alloforms of the A β peptide are formed by variability in the location of the the γ -secretase cleavage site; the most predominant of these species are A β 40 and A β 42, the latter of which is characterized by greater toxicity and faster aggregation kinetics(99, 146). The side chain of methionine 35 in both of these isoforms may be oxidized by reactive oxygen species *in vivo*. Oxidized A β s have been demonstrated to comprise 10-50% of A β content in postmortem senile plaques(335). The sulfoxide form of A β 42, A β 42-M35ox, demonstrates diminished aggregation and delayed protofibril formation compared to its reduced form(202), with experimental aggregation kinetics ranked A β 42 > A β 40 > A β 42-M35ox(4). M35-oxidation abolishes A β 42's characteristic paranuclei formation and results in the assembly of oligomers indistinguishable from those formed by A β 40.(203, 204) Further, oxidation of the M35 side chain to sulfoxide induces A β 40-like changes in structure and dynamics of the V18 and C-terminal region compared to reduced A β 42(211, 218). A number of rare mutations in the APP and the enzymes that process this protein predispose to familial AD (FAD) (as opposed to the more common sporadic AD form) and cerebral amyloid angiopathy (CAA). Some of these mutations occur in the A β sequence; one hotspot for these mutations occurs with the negatively charged residues E22 and D23. This cluster is far in sequence from the termini of the peptide where proteolytic cleavage occurs, and

these mutants result in increased A β 40 oligomerization orders or fibrilization rates *in vitro*(8). Further, the oligomerization propensity of these mutants has been correlated to both susceptibility to trypsin proteolysis and the magnitude of destabilization of the V24-K28 turn region in the A β (21-30) fragment(227). The processes underlying the drastic changes induced by these chemical modifications are not yet fully understood at the atomic level.

In this study, we employed a combined MD/NMR approach toward investigating the structural ensembles of the three physiologically relevant A β monomer species: A β 40, A β 42, and A β 42-M35ox(S), as well as three FAD mutants of A β 40 linked to faster aggregation kinetics and higher neurotoxicity: E22 Δ , D23N, and E22K(8, 248, 336). Employing a REMD protocol with similar parameters to the simulations by Sgourakis and coworkers(9) that produced good correlation to experimental data, we observe that long simulation times on the order of hundreds of ns/replica are required to reach convergence. This is the first study to sample A β extensively on the μ s/replica timescale using unbiased, all-atom, explicitly solvated REMD simulations. Our simulation data are validated by quantitative comparison of calculated three-bond J-couplings, residual dipolar couplings (RDCs), and chemical shifts with their experimental counterparts (see Chapter 4). These NMR data are derived from a single set of resonances that almost entirely (>90%) represent the monomeric state in solution(337).

Although our simulations reveal that all three species share similar regional organization and secondary structure tendencies, we have identified key differences in tertiary structure and hydrogen bonding patterns that may underlie the differential aggregation rates between these A β isoforms. Among the diverse set of topologies sampled by A β , our data reveal that there is a substantial proclivity to adapt β -hairpins encompassing the the central hydrophobic cluster (CHC, residues L17-A21) and 30-36 in all 3 species. Further, residues 39-41 in A β 42 and A β 42-M35ox(S) contribute to additional β -hairpin motifs absent in A β 40. Contact maps and hydrogen bond populations suggest prevalent sampling of a β -meander (two successive antiparallel β -hairpins) composed of three antiparallel strands (CHC, 30-36, and 39-41) in A β 42. Reduced sampling of this motif by A β 42-M35ox(S) in favor of increased contact

between CHC and 39-41 may be responsible for the decrease in aggregation propensity of this form. The transiently-sampled β topologies of A β 40 and A β 42 are comparable to hypothesized intramolecular models for fibrils and "on-pathway" intermediate oligomers(23, 24, 26, 27); they expose hydrophobic side chains to solvent and may serve as hotspots for self-association. These β structures are also the slowest to converge in our simulations, which may explain their absence in other computational studies that rely on characterizations made on shorter timescales.

We also observe that V24-K28 forms a persistent bend motif in all wild type species that is directly nucleated by buried backbone to side chain hydrogen bonds with D23 and further stabilized by a long distance salt bridge between E22 and K28. Mutants of these residues were observed to fundamentally destabilize the electrostatics of this region, and increase the freedom of β hairpin hydrogen bonding pairs across this region. Indeed, the D23N mutation promoted configurations that reduced the expanse of the bend motif by disrupting D23 specific interactions, biasing hydrogen bonding to form between more N-terminal partners. In contrast, E22 Δ (and to a lesser extent E22K) was biased toward broader bend regions, promoting more C-terminal hydrogen bonding. These observations provides insight into the role of the side chain identities of FAD-linked E22 and D23 in A β monomer structure, and the possible effects mutants of these residues have in generating aberrant aggregates leading to disease phenotypes. This study has yielded rigorous, experimentally-consistent representations of different intrinsically disordered A β monomer ensembles, allowing us to identify key structural features and properties that may serve as determinants of A β self-assembly.

2.3 Methods

2.3.1 Simulation Details

In this study, we have investigated the structural ensembles of A β 42, A β 40, A β 42-M35ox(S), and three point mutations of A β 40 (E22 Δ , D23N, and E22K) using the REMD simulation protocol with the OPLS-AA/L force field(19) and the TIP3P water model(20), a combination which has been previously shown to yield good correlation to experimental observables for these systems(9). REMD is a generalized ensemble method(35, 36), wherein multiple initially identical copies of the system (replicas) are

simulated concurrently over a set of different temperatures. Trials to swap the temperature between neighboring replica pairs occur frequently through the simulation, each with a probability of success governed by a Metropolis Monte Carlo criterion. All molecular dynamics calculations were performed with the GROMACS 4.5.3 or 4.5.5 package(338). In general, many simulation parameters were kept as close to possible as a previous investigations performed by the García group (9, 11), but they are described in their entirety below for the sake of completeness.

The A β 42-M35ox(S) system features L-methioinine-(S)-sulfoxide (MTO) in place of L-methionine at the residue 35 position. From what is currently known, oxidation of the prochiral sulfur atom on methionine by typical reactive oxidation species is believed to occur racemically between R and S diastereomers(339), so our choice of the S conformation here was made arbitrarily. To accommodate this new residue, we used a modified OPLS-AA/L force field for this system only. Parameters in the OPLS force field are, in general, optimized to fit experimental thermodynamic properties for pure liquids of a set of small, model compounds comprising many key functional groups(340). These parameters have therefore been designed to be highly transferable among molecules containing similar chemical groups. The $-\text{CH}_2\text{-SO-CH}_3$ moiety in methionine sulfoxide is chemically similar to the DMSO ($\text{CH}_3\text{-SO-CH}_3$) molecule already parameterized in OPLS(341), so the nonbonded and bonded parameters have been transferred from DMSO to the corresponding group in L-methioinine-(S)-sulfoxide (with a modification on the CG partial charge to preserve overall charge neutrality) to complete the definition of this residue. The modifications to the force field are further described in Table 7.1 in the Appendix.

All three peptide systems were constructed from a fully extended structure of the full length peptide with zwitterionic termini and side chain protonation states appropriate for pH 7. The sequences are:

- A β 42:
 $^1\text{DAEFRHDSG}^{10}\text{YEVHHQKLVF}^{20}\text{FAEDVGSNKG}^{30}\text{AI IGLMVG}^{40}\text{GV}^{40}\text{VIA}$
- A β 40:
 $^1\text{DAEFRHDSG}^{10}\text{YEVHHQKLVF}^{20}\text{FAEDVGSNKG}^{30}\text{AI IGLMVG}^{40}\text{GV}^{40}\text{V}$

- A β 42-M35ox(S):
¹DAEFRHDSG¹⁰YEVHHQKLVF²⁰FAEDVGSNKG³⁰AI IGLXVGGV⁴⁰VIA
 where X = L-methioinine-(S)-sulfoxide
- A β 40-E22 Δ :
¹DAEFRHDSG¹⁰YEVHHQKLVF²⁰FA-DVGSNKG³⁰AI IGLMVGGV⁴⁰V
 (The hyphen in this sequence represents the deleted 22nd residue. Here, the numbering for this peptide keeps the residue numbers consistent with the other peptides so index 22 is skipped; ie, A21 is followed by D23)
- A β 40-D23N:
¹DAEFRHDSG¹⁰YEVHHQKLVF²⁰FAENVGSNKG³⁰AI IGLMVGGV⁴⁰V
- A β 40-E22K:
¹DAEFRHDSG¹⁰YEVHHQKLVF²⁰FAKDVGSNKG³⁰AI IGLMVGGV⁴⁰V

Each of these systems were then constructed using identical protocols with identical parameters. First, a 1-ns MD simulation of the peptide was run *in vacuo* at a high temperature (700 K) to collapse each extended system. Each system was then energy minimized and solvated to produce a 5.4 nm cubic simulation box with 4947 TIP3P water molecules. An equilibration simulation was run for each of these solvated systems at a constant temperature (300 K) and pressure (1 bar), both coupled with the standard Berendsen method(342), for 1 ns and using an integration time step of 1 fs. For all three systems, the simulation box dimensions only changed marginally after the NPT simulation, remaining 5.4 nm in each dimension.

From the resulting configuration, 52 replicas were generated to optimally encompass the temperature range 270.0 K to 601.2 K with an average exchange ratio of 15%, as calculated in a previous study(11). Figure 7.1 in the Appendix illustrates that this distribution promoted sufficient random walks of replicas through temperature space during the simulation. Constant volume REMD simulations were run using these replicas for a total 1000 ns/replica for each of the systems (cumulative simulation time of 52 μ s per system). These extensive simulations are required to properly sample the equilibrated state, as discussed in the Convergence subsection (2.4.1) in this chapter. Exchanges between adjacent replicas in temperature space were attempted every 4 ps throughout the simulation. In this stage, bond lengths were constrained with the

LINCS(343) and SETTLE(344) algorithms, permitting an integration step of 2 fs. A 1 nm cutoff for Lennard-Jones interactions was imposed, while the neighbor list was updated every 10 steps. Electrostatic interactions were calculated with the particle mesh Ewald method(345) with a 54 unit cubic grid and fourth order extrapolation. Explicit ions were not added to the solvent; instead, charge neutrality of each system was established by use of the Ewald method for calculating long-range electrostatics. These conditions imitate the minimal salt buffers (20 mM potassium phosphate with no other buffer salt) used in the NMR experiments presented for comparison to this study (see Chapter 4). The system was thermalized with a Nose-Hoover heat bath(346), thereby ensuring constant temperatures for each trajectory of the REMD ensemble in temperature space. The REMD simulations were each performed using different domain decomposition assortments on either 208 or 416 CPUs of a Linux cluster based at Rensselaer Polytechnic Institute or the Stampede cluster at the Texas Advanced Computing Center at the University of Texas (part of the Extreme Science and Engineering Discovery Environment (XSEDE)).

2.3.2 Ensemble Analysis

Simulated conformations were analyzed over the 400-1000 ns/replica time interval over 7 temperatures composing the range 289.2 K to 310.7 (here, this is referred to as the "production period"). Multiple temperatures were used to improve the sampling statistics. Analysis was done using the GROMACS 4.5.3 or 4.5.5 suite(338) and Python/shell scripts. Secondary structures were annotated over the ensemble using the DSSP method(347, 348). Tertiary structure was studied via clustering, contact maps, and populations for key hydrogen bonds and salt bridges. The single linkage(349), Jarvis-Patrick(350), and Daura(351) clustering methods implemented in GROMACS's `g_cluster` tool were each employed to cluster around $C\alpha$ atoms over the production period for every 100 ps-th frame, grouping 42000 conformations for each monomer. Each of these identified similar dominant centroids, but single linkage was chosen as the method best suited toward both categorizing the well separated, highly diverse set of morphologies in the $A\beta$ ensemble with fewer, more highly representative structures and yielding output that is minimally dependent on the cutoff parameter. Given this, the

clusters presented in the RMSD-Based Clustering subsection (2.4.5) were calculated with the single linkage clustering scheme with a root mean square deviation (RMSD) cutoff of 2.0 Å. Contact maps are based on the probability of contact over the production period, with two residues defined to be in contact if their C α - C α distance is ≤ 8 Å(352). Hydrogen bonds between donor atom (D), hydrogen atom (H), and acceptor atom (A) were identified when D to A distance ≤ 3.5 Å and an interior H-D-A angle $\leq 60^\circ$, while salt bridges were detected using a 4.0 Å distance cutoff between any pair of oppositely charged atoms. S^2 order parameters for all backbone nitrogen-hydrogen internuclear vectors (excluding the N-terminus) were calculated with the iRED technique(353). Finally, to complement this analysis, we also measured solvent accessible area(354) per residue and histograms of radius of gyration for the production period. Histograms of C α RMSD between residues 16-35 for each frame in the production period compared to Chain C, Model 1 from an NMR structure of A β 40 monomer trapped by an Affibody protein(355) (PDB: 2OTK) were also determined.

2.4 Results

2.4.1 Convergence

Simulation convergence (steady state) was assessed by three criteria: 1) The time required for the simulations to equilibrate is first determined by the convergence of computationally derived couplings to their experimentally determined values as judged by Pearson correlation coefficient (PCC). These data are not shown, but is presented for some of the simulations in Rosenman et al.(212). Based on the apparent transition of these data from a phase where PCC generally improves to a phase where oscillations about a steady state occur, we judge equilibration for $^3J_{\text{HNHA}}$ to occur at ~ 200 ns/replica for A β 42-M35ox(S) and A β 40-D23N, ~ 300 ns/replica for A β 42, ~ 350 ns/replica for A β 40, and ~ 600 ns/replica for A β 40-E22 Δ . On the other hand, while the block average PCCs for both $^3J_{\text{NHB}}$ and RDCs generally appear to improve over time, it is difficult to make out distinct phases in the time progression of these values, and so we don't consider these values in judging convergence. We refer the reader to Chapter 4 for more information about these NMR observables, and overall comparisons of these values between simulation and experiment. 2) C α average distance maps over 100 ns/replica

windows throughout the simulation serve as another reporter on equilibration time (*data not shown, presented for some of the simulations in Rosenman et al.(212)*). We determine distance map equilibration to occur at 100-200 ns/replica for A β 40-D23N/A β 40-E22 Δ , 200-300 ns/replica for A β 42, 300-400 ns/replica for A β 42-M35ox(S)/A β 40-E22K, and 500-600 ns/replica for A β 40, because at these points, the sum over the absolute value of differences between the pairwise distances in the map of the current interval and that of the previous interval is no greater than 50 nm for all subsequent time windows. 3) Finally, the content of secondary structure over the whole peptide as determined by DSSP(347, 348) over 100 ns/replica windows is used as a third check on simulation equilibration. Beginning at the following intervals – 200-300 ns/replica for A β 42-M35ox(S), 300-400 ns/replica for A β 42/A β 40-D23N, 500-600 ns/replica for A β 40-E22K, and 600-700 ns/replica for A β 40 and A β 40-E22 Δ , each ensemble has sampled to a state where the composition of each DSSP element varies by no more than 5% over all subsequent time windows, implying convergence at these junctures (Figure 2.1).

Overall, this evidence illustrates that the A β simulations are very dynamic even on the 100 ns/replica timescale, stressing the necessity for long timescales in simulations of these systems, at least using the parameters we have chosen. For A β 42, A β 40-M35ox(S), and A β 40-D23N, the three methods reveal that our choice of the analysis period to be 400-1000 ns/replica successfully captures an equilibrated segment of the simulation. Although these methods reveal that the other ensembles may converge only later in the simulation, the analysis period for was also chosen for these simulations to be 400-1000 ns/replica so that we acquire sufficient data for statistical significance and consistently analyse the same duration of time for each simulation.

2.4.2 Secondary Structure

Figure 2.2A illustrates the secondary structure propensity for each of the A β species averaged over all production period structures for all residues for the OPLS condition, as determined by the DSSP(347, 348) program. In general, all A β ensembles possess similar overall secondary structure distributions. Propensity to form β is the most

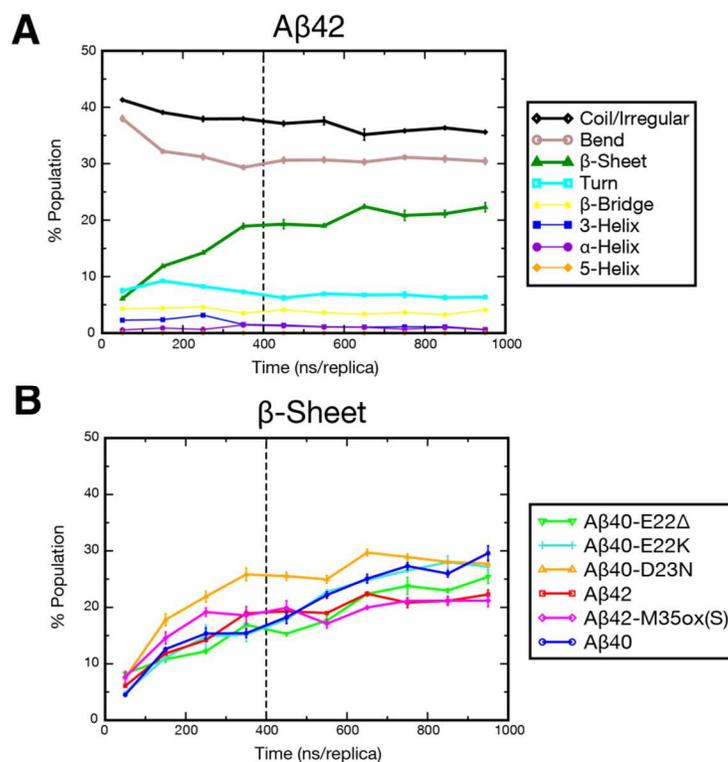


Figure 2.1. Content of secondary structure elements versus simulation time for $A\beta$ runs done with the OPLS condition, as determined by DSSP(347, 348) over 100 ns/replica windows. (A) Profile for all DSSP elements for an example $A\beta$ variant, $A\beta_{42}$. Dominant secondary structure elements in the ensemble coil/irregular, bend, β -sheet, and turn are denoted with thick lines. (B) Graph for the most widely varying element with time, β -sheet, for all species.

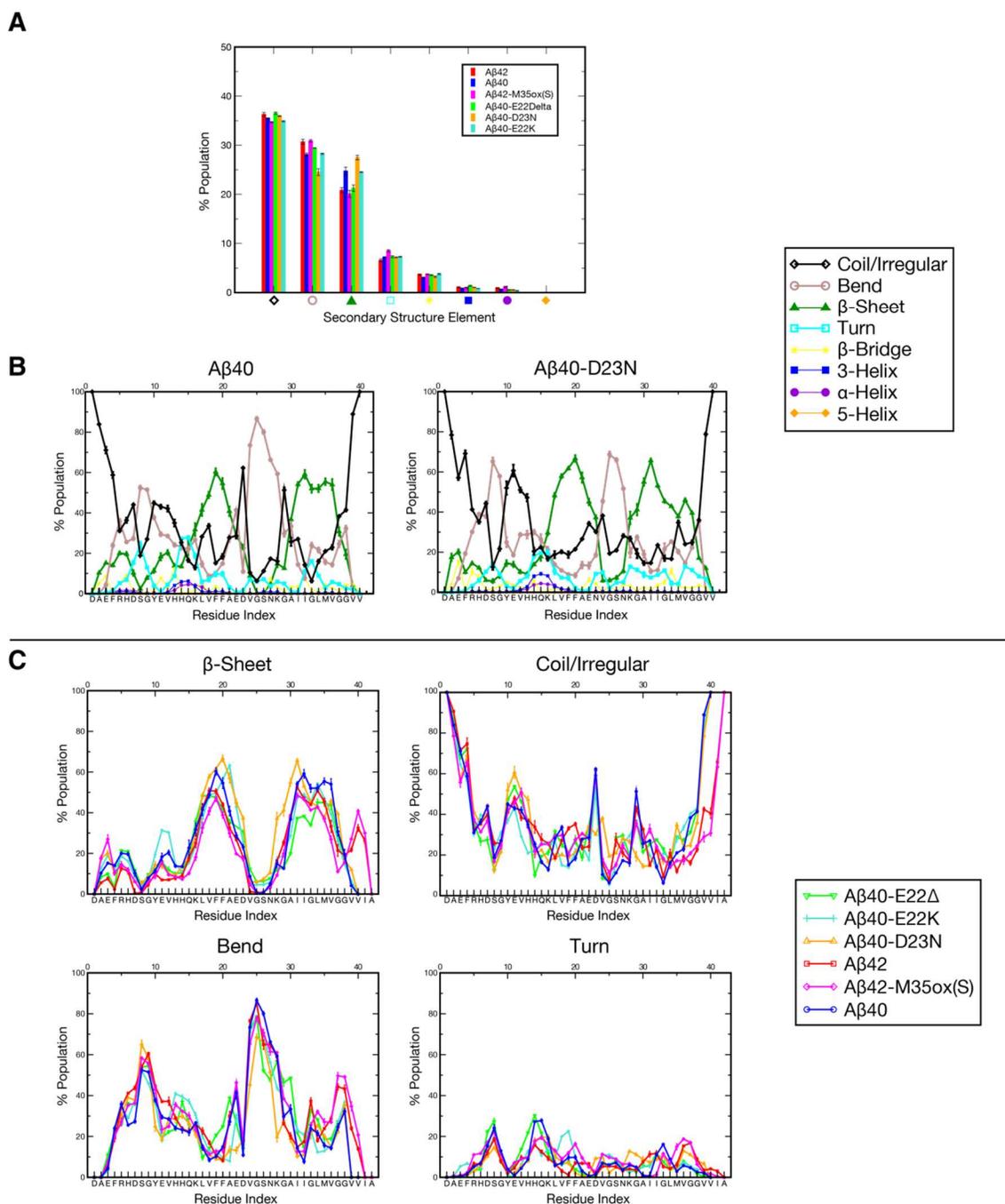


Figure 2.2. Percent composition of secondary structure as determined by DSSP(347, 348) for $A\beta$, averaged over the production ensemble, for simulations done with the OPLS condition (A) over all residues, (B and C) as a function of residue index. (B) Profile for all DSSP elements for two example $A\beta$ variants, $A\beta 40$ and $A\beta 40$ -D23N. Of particular importance are the dominant secondary structure elements in the ensemble, denoted with thicker lines in the graph: coil/irregular in black, bend in grey, β -sheet in green, and turn in cyan. (C) Plot comparing all of the $A\beta$ variants studied here for these four dominant DSSP elements. The sequence of wild type $A\beta 42$ is presented in the lower x -axis.

variable element, having a ranking $A\beta_{40}\text{-D23N} > A\beta_{40} > A\beta_{40}\text{-E22K} \gg A\beta_{40}\text{-E22}\Delta > A\beta_{42} > A\beta_{42}\text{-M35ox(S)}$, with other secondary structures like coil and bend compensating for the difference between these species. These compositions, in general, compare well to values for monomeric $A\beta_{40}$ as determined by a recent CD experiment (8.7% α -helix, 24.0% β -sheet, 67.3% statistical coil)(214), though the composition of α -helix is underestimated.

Illustrated in Figure 2.2B and 2.2C are the ensemble-averaged secondary structure populations per residue for each $A\beta$ variant. A common theme in β character among the species is robust β propensity in residues L17-A21 and A30-V36, as well as very pronounced bend character in residues V24-K28, indicating the sampling of a β -hairpin like motif. However, the different FAD mutants perturb this profile. $A\beta_{40}\text{-D23N}$ is the most disruptive mutant, significantly increasing β content in residues F19-D/N23 and K28-I31, while the content of bend is decreased throughout the V24-K28 region, particularly in residue K28. $A\beta_{40}\text{-E22K}$ increases β propensity in A21-D23, while $A\beta_{40}\text{-E22}\Delta$ reduces β content in the CHC region. Both E22 mutants, particularly E22 Δ , reduce β character in A30-V36, particularly among the earlier residues of that group. Consequentially, bend character is increased in G29-A30, opposite the effect of D23N. $A\beta_{42}$ and $A\beta_{42}\text{-M35ox(S)}$ somewhat weaken these features (particularly the latter) but also possess substantial β character in V39, V40, and I41 separated by a bend centered around glycines 37-38, suggesting a second β -hairpin topology between the residues A30-V36 and V39-V41 in these species.

The N-terminus is generally disordered in each ensemble, but can be characterized by up to $\sim 30\%$ β character in certain regions in $A\beta_{40}\text{-E22K}$ and $A\beta_{42}\text{-M35ox(S)}$. α -helical character is virtually non-existent in all variants, except some minor populations at residue H13-L17 ($< 20\%$, even combining 3-helix and α -helix). Other regions of strong bend/turn character in $A\beta_{40}$ include residues H6-G9 and H14-K16. These observations match the positions of β proclivity inferred by chemical shift indices(4) and the position of the V24-K28 turn region that nucleates monomer folding described by NMR(216). Regions strongly characterized as bend or turn here correlate well with fast HD exchanging regions in $A\beta_{40}$, while those strongly affiliated with β character in our simulations generally undergo slower exchange(225, 226). The V18-F20 and V39-I41

(in A β 42/42-M35ox(S)) regions annotated as β -dominant by DSSP also coincide with regions of high experimental $^3J_{\text{HNHA}}$ (> 7.5 Hz) (Figure 4.6), further validating these characterizations.

As discussed in the previous section (2.4.1), hundred ns/replica timescales or larger are necessary to reach steady state values for secondary structure composition. Figure 2.1 illustrates that the β content is the slowest to converge, which may explain why other computational studies indicate lower values (for example, (9, 11-15)).

2.4.3 C α Contact Maps

The C α contact map analysis in (Figure 2.3, upper corner) illustrates a characteristic pattern of tertiary structure shared by the equilibrated A β populations, wherein structures adopt three regions: N-terminal (residues 1-20), central (residues 21-30), and C-terminal (residues 31-40/42). Intra-region contacts are ample for all 3 regions, except the C-terminus in A β 40 and its point mutants. While the N-terminal and C-terminal regions have close mutual contacts, interactions between the central region and the other two regions are significantly limited in wild type A β , except at the very extreme N-terminus (residues F4-H6). The restriction of this central region from the rest of the protein corresponds to the fact the composite residues align with a bend-dominated population bridging two β -compatible regions, as discussed in the Secondary Structure section (2.4.2).

Robust anti-diagonal contacts exist between Q15-G25 and N27-V39 for A β 40 and are particularly well-populated between K16-F20 and A30-M35, indicative of antiparallel β hairpin interactions between these regions. Difference contact maps (Figure 2.3, lower corner) reveal important discrepancies in the nature of this antiparallel binding among the different mutants and alloforms of A β . A common theme among the FAD mutants of A β is a reduction of contacts between these regions on diagonal compared to wild type, but an increase of contact probabilities in diagonals flanking the center diagonal (manifesting as a blue-red-blue stripe pattern in the difference maps). This suggests that FAD mutants increase the polymorphism in possible binding partners across this diagonal. Further, D23N, in particular, preferably enhances the contacts

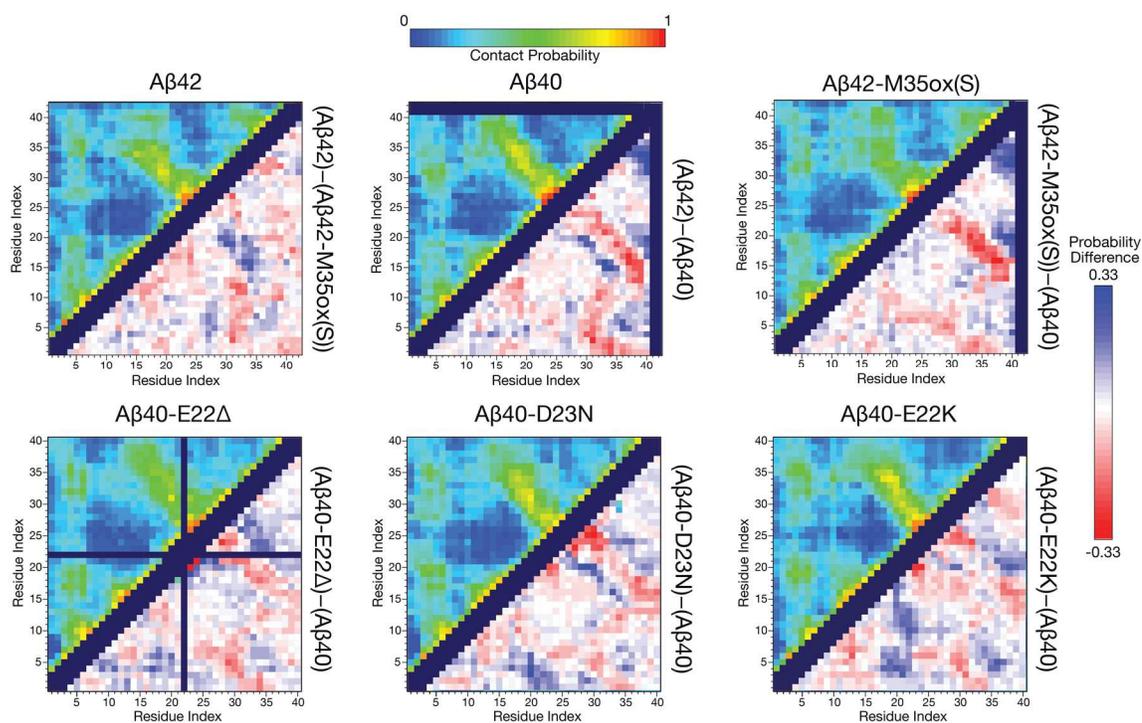


Figure 2.3. (upper corner) Maps illustrating the probability of contact for $C\alpha$ atoms over the production ensemble for different $A\beta$ isoforms and mutants derived with REMD under the OPLS condition. Here, two residues are defined to be in contact if their $C\alpha$ - $C\alpha$ distance is $\leq 8 \text{ \AA}$ (352). (lower corner) Difference between these contact probability maps for selected $A\beta$ mutants/isoforms with the OPLS condition, as listed in the title of each map. A red color indicates a reduced likelihood of contact while a blue value indicates increased contact probability for the minuend species compared to the subtrahend species. Self or sequential contacts, such as between (i, i) , $(i, i+1)$, and $(i, i+2)$, and invalid squares on the map were not computed and are represented by the dark purple regions.

where the C-terminal binding partner is N-terminal of the main wild type diagonal, while the two E22 mutants favor more C-terminal binding diagonals. This corresponds to the breadth of the bend regions as discussed in the Secondary Structure section (2.4.2), with D23N possessing a narrower bend region and the two E22 mutants broadening the bend to include G29-A30.

Similar antiparallel contacts also exist in A β 42, but occur between residues Q15-G25 and N27-V36, representing a truncation of the C-terminal partner compared to A β 40. A β 42 possesses a second set of anti-diagonal contacts that occur between I31-V36 and G37-I41. This suggests a second β -hairpin motif between these residues. The difference maps further reveal a register shift in the antiparallel contacts in A β 40 versus A β 42 among the first β -hairpin motif, with residues in the CHC contacting residues further C-terminal in A β 40. These two sets of A β 42-like antiparallel contacts exist but are less robust in A β 42-M35ox(S), and the pattern of cross-central domain contacts seems to follow the A β 40 register in terms of contact partners. Conversely, there are greater contact probabilities observed between the CHC and the V39-I41 region in A β 42-M35ox(S) compared to A β 42. In tandem with the secondary structure characterization of these regions, this suggests the oxidized form adopts alternate CHC to 39-41 β contacts that compete with the sampling of the aforementioned β -meander that is prevalent in the reduced form.

In all simulations, residues E3-D7 readily contact the rest of the protein, including the central region. Contacts are particularly strong with both the CHC and G29-V36. These long range contacts are, in general, promoted by all of the FAD mutants, particularly E22K, which strongly promotes contacts to V18-E/K22 by most residues of the N-terminal domain (E3-L17), and D23N, which promotes antidiagonal contacts between the extreme N-terminus and F20-G25. A β 40-E22 Δ , in contrast, primarily enhances contacts between the N-terminus and other residues in the N-terminal region and CHC, while the wild type favors interactions with G29-I31. These features, in tandem with the increased breadth of binding partners across the main diagonal, represents increased long-range contact of the central region to the N-terminal and C-terminal regions permitted by FAD mutants.

When comparing A β 42 with A β 40, note that there is a much smaller difference between the ensembles in how the N-terminal half of the protein interacts with itself compared to how the C-terminus acts with itself and the rest of the protein. This matches the observation that only C-terminal residues have significant chemical shift differences when comparing the respective ^1H - ^{15}N HSQCs for each alloform(211). Further, while the dissimilarities between ensembles noted here may serve as key structural hallmarks for distinguishing A β species, it is important to emphasize that these differences are relatively subtle, with no average pairwise distance between C α atoms changing by more than 0.57 nm (0.33 nm excluding residues 38-42) between any two of the ensembles, also in accordance with subtle changes in ^1H - ^{15}N HSQC patterns between the species.

2.4.4 Electrostatic Interactions

The most commonly populated long distance backbone hydrogen bonds in each of the simulations are listed in Table 2.1. Prevalent in A β 42 is a $i+2:j-2$ register defined by hydrogen bonds between V18/G33 and F20/I31, while A β 40 has a dominant bonding register defined by hydrogen bonds between V18/M35 and F20/G33, yielding a shift of 2 residues for the C-terminal partner between the two species. Interestingly, two of the FAD mutants (A β 40-D23N and A β 40-E22K) favor the A β 42-like pattern. These are the preferred registers for these systems, but other cross-hairpin bonds outside the register are possible (even substantially populated) due to the intrinsic flexibility of the system. However, these register hydrogen bonds are surprisingly well-populated for a disordered system, reaching populations up to $\sim 30\%$ of the ensemble. A β 40-E22 Δ and A β 42-M35ox(S) show meager cross-hairpin populations in the first place, though of the two registers, A β 42-M35ox(S) is interestingly biased toward the A β 40 pattern. A β 42 is also characterized by high hydrogen bond populations between residues A30/G37 and L34/V41, suggesting a second hairpin structure. These two interactions are depleted for A β 42-M35ox(S), while long-range interactions with the C-terminus such as K16/V41 and A30/V41 are more populated. This supports the idea that A β 42-M35ox(S) more frequently samples alternate binding patterns that compete with the A β 42-like double hairpin. Consistent with reduced hairpin content, A β 42-M35ox(S) is characterized by

Table 2.1. Most commonly populated long distance, inter-residue, backbone hydrogen bonds in the A β ensemble for the OPLS set of simulations. All entries are backbone-backbone hydrogen bonds separated by at least two residues and populated $\geq 10\%$. Bonds whose acceptor and donor are separated by more than 5 residues are bolded. Hydrogen bonds across the central bend that follow the V18/G33, F20/I31 $i+2:j-2$ register are colored in red, while those that follow the V18/M35, F20/G33 register are colored blue.

% for OPLS A β 42		% for OPLS A β 40		% OPLS A β 42-M35ox(S)	
BB H-Bond (Donor–Acceptor)	% Pop	BB H-Bond (Donor–Acceptor)	% Pop	BB H-Bond (Donor–Acceptor)	% Pop
K16 – H13	19	K16 – H13	22	K16 – H13	21
K16 – M35	11	L17 – H13	10	L17 – H13	15
L17 – H13	11	L17 – H14	10	E22 – I31	11
L17 – H14	11	V18 – M35	19	G33 – F20	11
V18 – H14	11	F19 – V36	11	A42 – D1	11
V18 – G33	29	F20 – G33	22		
F19 – G37	12	E22 – I31	23		
F20 – I31	28	D23 – I31	13		
A21 – A30	10	I31 – D23	15		
E22 – G29	11	G33 – V18	10		
G29 – E22	12	G33 – F20	23		
A30 – G37	10	L34 – I31	12		
I31 – F20	17	M35 – V18	21		
I32 – F19	10	V36 – F19	12		
I32 – M35	11	G37 – K16	16		
G33 – V18	24				
L34 – I41	16				
M35 – K16	28				
G37 – A30	12				
I41 – L34	17				
% for OPLS A β 40-E22 Δ		% for OPLS A β 40-D23N		% for OPLS A β 40-E22K	
BB H-Bond (Donor–Acceptor)	% Pop	BB H-Bond (Donor–Acceptor)	% Pop	BB H-Bond (Donor–Acceptor)	% Pop
H6 – G9	11	K16 – H13	18	H13 – F20	14
G9 – H6	18	L17 – H13	12	V18 – G33	12
K16 – H13	13	L17 – H14	16	F19 – G33	12
L17 – R5	10	V18 – G33	14	A21 – I31	11
V18 – L34	12	F20 – I31	19	K22 – E11	11
A21 – L34	12	E22 – G29	16	D23 – A30	11
		V24 – N27	24	V24 – N27	11
		G29 – E22	24	G29 – E22	11
		I31 – F20	21	G33 – V18	11
		I32 – M35	14	G33 – F19	11
		G33 – V18	14	M35 – K16	12
		M35 – K16	13		
		M35 – L17	14		
		G37 – A30	14		

significantly weaker cross-region interactions. Interestingly, OPLS simulations possess very few medium range, helical-like hydrogen bonds (ie, 3 or 4 residue separation); those that exist primarily involve histidines in the N-terminal region. Only A β 40-D23N and, to a far lesser extent, A β 40-E22K, sample a short range bond in the central region (V24-N27); these mutations disrupt the bulge-like “crown” motif (described below) present in the other variants that likely prevents short range hydrogen bonding from occurring.

Based on these prevalent interactions in the ensemble, in tandem with the secondary and tertiary structure characterizations made in the previous chapters, schematics for the ensemble biases of A β 40 and A β 42 and their relation to other experimental models in the literature are depicted in Figure 2.4. Interestingly, the A β 40 bonding pattern and topology matches that of an Affibody-trapped monomeric A β 40 hairpin solved with NMR by Hoyer *et al.*(355) (PDB: 2OTK). Further, the side chain contacts in the intramolecular model of A β 40 fibrils proposed by Bertini *et al.*(27) are analogous to the backbone hydrogen bonding of this hairpin model; both models are related by a 90 degree rotation about the axes of both β strands(27). Meanwhile, the double beta hairpin structure of A β 42 is comparable to the intrapeptide topology for a toxic A β 42 oligomer proposed by Ahmed *et al.*(26). Lastly, it is important to emphasize that, while the models proposed in Figure 2.4 are based on ensemble-averaged structural biases and hydrogen populations, they also match actual representative centroid structures sampled during the simulations (discussed in Section 2.4.5), such as the third and fourth centroid of A β 40 and the eighth centroid of A β 42.

Charge-altering substitutions in E22 and D23 implicated in FAD result in increased A β 40 oligomerization order and fibrilization rates(8), linked to destabilization of a turn in the A β (21-30) fragment.(227) Observation of the interactions involving the side chains of these residues in our monomer simulations (Table 2.2) indicates that D23 often forms side chain to backbone hydrogen bonds with each of the amides of residues V24 through K28, with certain populations exceeding 60% of the ensemble in A β 40, particularly with G25 and S26’s amide. These interactions are quite prevalent in all OPLS simulations, with the notable exception of A β 40-D23N. In this mutant, these

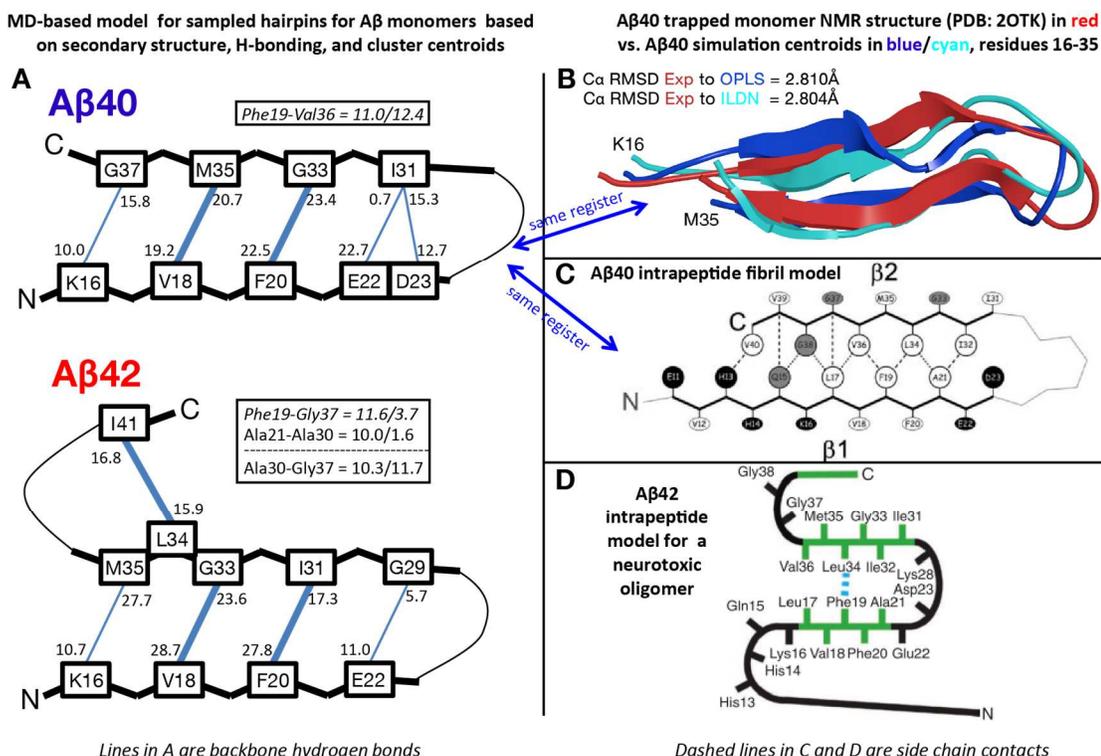


Figure 2.4. Comparison of transiently sampled conformations and ensemble biases in monomeric A β REMD simulations to an experimental structure of a trapped A β monomer and intrapeptide models for higher order aggregates. (A) β -hairpin models for REMD-derived A β 42 and A β 40 monomers with the OPLS model, based on the most populated cross-region backbone hydrogen bonds and secondary structure proclivities in the ensemble. The donor-acceptor and acceptor-donor populations of these hydrogen bonds are listed such that the number is closest to the donor residue of that bond. Residues that have a population > 15% for both donor to acceptor and acceptor to donor backbone hydrogen bonds are illustrated with a bold line in the figure. Sampled conformations matching these models exist as high ranking centroid structures. (B) Ribbon overlay of residues 16-35 for centroid 3 derived from clustering analysis of the OPLS A β 40 simulation (in blue), centroid 1 for ILDN A β 40 simulations (in cyan; simulation discussed in Chapter 3) and the NMR-derived structure of monomeric A β 40 in complex with a phage selected Affibody published by Hoyer et al.(355) (in red) (C) Intrapeptide model for A β 40 fibrils derived from solid state NMR constraints. This panel was adapted from Bertini and coworkers(27), published in *The Journal of the American Chemical Society* with permission granted by © 2011 American Chemical Society. (D) Intrapeptide model for A β 42 “on pathway” pentameric oligomers based on solid state NMR data. This panel was adapted from Ahmed et al.(26), published in *Nature Structural & Molecular Biology* with permission granted by © 2010 Nature Publishing Group. This full figure was published by Nasica-Labouze et al.(356), published in *Chemical Reviews* with permission granted by © 2015 American Chemical Society. Permissions for reuse as an author of the work are granted by “ACS Thesis/Dissertation Policy” clause specified by the publisher, ACS Publications.

Table 2.2. Central region electrostatic interactions of A β involving the side chains of residue 22 and 23 for the OPLS simulation set. Listed populations represent average bond frequencies for equivalent atoms in the side chain (for example, "D23O γ " is the average of interactions involving the D23O γ_1 and D23O γ_2 atoms). For A β 40-D23N only, populations for the interaction between K28HZ and N23's O δ in the final row follow the hydrogen bond criteria, not salt bridge criteria.

Donor – Acceptor Atoms	% for OPLS A β 42	% for OPLS A β 40	% for OPLS A β 42-M35ox(S)
<i>Hydrogen Bonds</i>			
V24NH – D/N23O δ	27	32	25
G25NH – D/N23O δ	57	61	51
S26NH – D/N23O δ	59	65	54
S26H γ – D/N23O δ	32	37	32
N27NH – D/N23O δ	31	36	29
K28NH – D/N23O δ	17	25	18
<i>Salt Bridges</i>			
K28NZ – E/K22O ϵ	10	23	20
K28NZ – D/N23O δ	4	2	5

Donor – Acceptor Atoms	% for OPLS A β 40-E22A	% for OPLS A β 40-D23N	% for OPLS A β 40-E22K
<i>Hydrogen Bonds</i>			
V24NH – D/N23O δ	29	7	25
G25NH – D/N23O δ	57	26	52
S26NH – D/N23O δ	61	4	50
S26H γ – D/N23O δ	39	0	32
N27NH – D/N23O δ	33	4	25
K28NH – D/N23O δ	14	0	13
<i>Salt Bridges</i>			
K28NZ – E/K22O ϵ	N/A	8	N/A
K28NZ – D/N23O δ	2	2	5

interactions are all greatly diminished, but local interactions involving the amides of V24 and G25 are still present (7% and 24%); whereas the interactions involving farther residues, including S26-K28's amides and S26's side chain are considerably rarer or not observed. In wild type A β , E22 frequently samples a salt bridge with the side chain of K28 across the bend region; this interaction is significantly more populated in A β 40 and A β 42-M35ox(S) (20% and 18% respectively) than A β 42 (9%), inversely correlating with the aggregation proclivity of the species. Further, this cannot be formed in the E22 mutants, while it is rare in the D23N variant (8%). Salt bridges of K28 involving residue 23 are rare (5% or less in all simulations), suggesting this residue does not compensate for the role of E22 (or lack thereof in the different mutants and alloforms).

Direct observation of V24-K28 turn in hairpin-forming centroid structures of A β 40 (Figure 2.5) suggest that the amide groups of the turn point inward to form the aforementioned hydrogen bonds with a buried D23 side chain, exposing all of the negatively polarized carbonyl groups of the turn to solvent; this forms a large, bulge-like bend structure is referred to as the “crown” motif in this paper. The electrostatic interactions involving D23's side chain are likely to be essential toward establishing bend structure in this region, while the long range salt bridge formed by E22 and K28 is also likely to be important toward both burying D23 and stabilizing this “crown” motif. The A β 40-D23N simulation severely reduces the bonds involving residue 23, even after considering interactions with the still polar O δ present in the asparagine variant. Further, none of its centroid structures are observed to have “crown” features in the central turn, instead forming alternate bend structures that are shorter in breadth, as further discussed in RMSD Based Clustering section (2.4.5). In contrast, neither of the E22 mutants can form the E22 to K28 salt bridge, yet they sample centroids with crown features. A β 40-E22K introduces a new cation in the central region, yet it does not interact with D23 (population < 1%); among its possible partners, it interacts with D1 and D7 the most (population ~6% and ~4%, respectively). These ionic interactions very likely drive the increased interaction of E22K's CHC with the N-terminus. All of this suggests the amino acid identity of D23 is critical to forming the crown, and only the D23N mutation strongly reduces the tendency of crown formation.

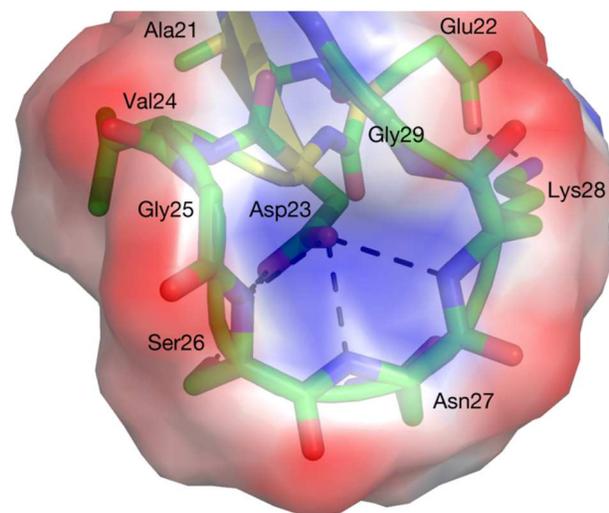


Figure 2.5. Central region (A21-G29) in A β 40's third most populated centroid after single linkage clustering (see Figure 2.6) with the OPLS condition, illustrating the "crown motif" referred to in the text. For all OPLS variants but A β 40-D23N, the side chain of D23 is buried and forms hydrogen bonds with the inwardly pointing backbone amide groups of V24-K28 with very high fidelity (some bonds present in >60% of the A β 40 ensemble), while E22 (in all variants but A β 40-E22 Δ and A β 40-E22K) forms a persistent salt bridge with K28 (present in ~20% of the ensemble). This motif presents a negatively charged external surface created by radially exposed backbone carbonyl groups and D23's side chain, with a central indentation of positive charge aligning with the backbone amide groups. The electrostatic surface was calculated with ABPS(357) and PDB2PQR(358, 359) using default parameters. Figure rendered in Pymol(360).

2.4.5 RMSD Based Clustering

Representatives of the populations for each A β variant are depicted in Figure 2.6, revealing conformations that are topologically diverse but nonetheless share several common structural motifs. Although both mostly random coil and partially β -structured conformations are observed, all of the centroids are collapsed. The central region (residues 21-30) aligns with yellow in the rainbow spectrum coloring of the backbone ribbon; as expected from the previous analyses, this region is consistently isolated from the rest of the protein and characterized by loop structure in all simulations. The C-terminal region (in orange) frequently folds back on the rest of the protein to interact with N-terminal region residues in all three species, particularly at the CHC (in green) to form an antiparallel β -hairpin. Indeed, the centroids are almost exclusively represented by a mixture of loop and β secondary structures, with α character very rare. It should be further noted that, even with single linkage clustering scheme using a relatively large 2.0 Å cutoff, only 30-50% of the structures of the production period are represented by the top 10 centroids for all species, indicative of highly diverse simulated ensembles.

A β 40's third and fourth centroid, together representing 8.15% of the ensemble, each possess a nearly identical extended C-terminal β hairpin topology (residues 16-35 have a C α RMSD of 0.583 Å). This topology is similar to the Affibody-trapped A β monomer hairpin solved by Hoyer and coworkers(355). C α RMSD of residues 16-35 between the third and fourth centroid structure and this NMR-derived conformation are 2.810 Å and 2.718 Å, respectively. Indeed, 13.3% of the entire A β 40 REMD ensemble has a C α RMSD \leq 3Å compared to this NMR structure for this region (Figure 2.7). Further, these two centroids both possess cross-region hydrogen bonds between V18/M35, F20/G33, E22/I31(donor/acceptor), and D23/I31(acceptor/donor) characteristic of the A β 40 bonding register derived from ensemble hydrogen bond populations. As seen in Figure 2.7B, A β 40 has the highest population of its ensemble as structures with low RMSD to the trapped monomer; some species also sample similar configurations at a lower population, while other species like A β 42-M35ox(S) and A β 40-E22 Δ , which have overall weak central hairpin interactions, almost never sample these conformations.

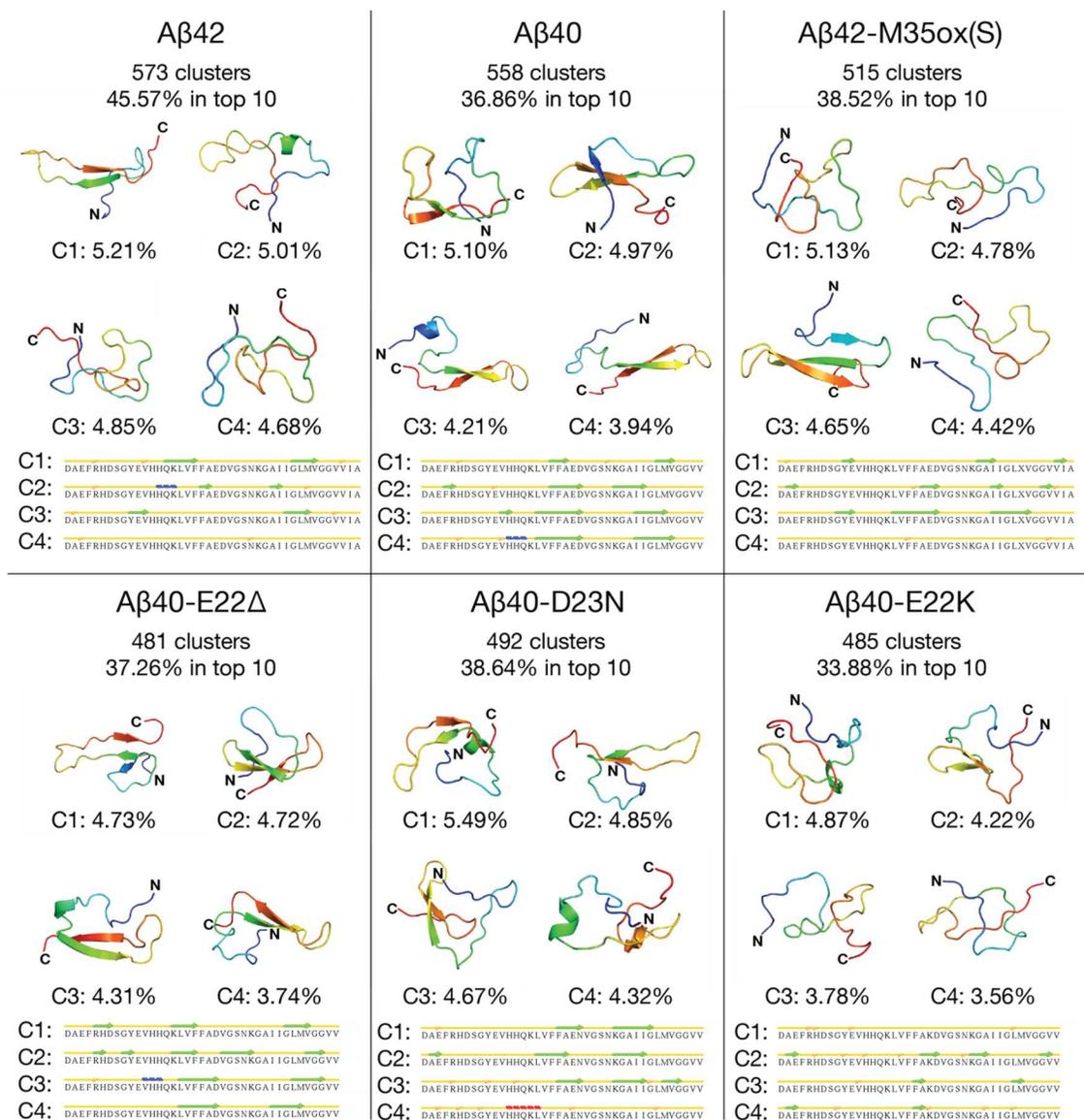
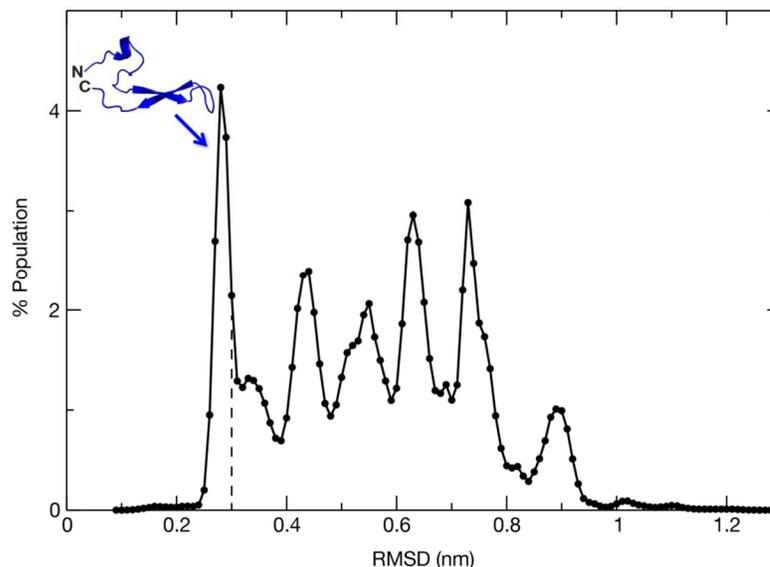


Figure 2.6. The four highest-populated centroids from single linkage cluster analysis(349) around $C\alpha$ atoms with an RMSD cutoff of 2.0 Å for each A β simulation done with the OPLS condition over the production period. Ribbons are colored by a blue-to-red rainbow spectrum from the N-to-C terminus. Secondary structure strips for each centroid, as calculated by STRIDE(361), are also shown. Figure rendered in Pymol(360).

A**B**

RMSD to Trapped Monomer	% for OPLS Aβ42	% for OPLS Aβ40	% for OPLS Aβ42-M35ox(S)
Res 16-35, % within 3Å	0	13	0
Res 16-35, % within 4Å	22	24	0
Res 16-40, % within 3Å	0	1	0

RMSD to Trapped Monomer	% for OPLS Aβ40-E22Δ	% for OPLS Aβ40-D23N	% for OPLS Aβ40-E22K
Res 16-35, % within 3Å	0	3	2
Res 16-35, % within 4Å	8	14	19
Res 16-40, % within 3Å	0	1	0

Figure 2.7. Comparison of OPLS simulations to Chain C, Model 1 from an NMR structure of Aβ40 monomer trapped by the Z_{Aβ3} Affibody protein(355) (PDB: 2OTK). (A) Histogram of Cα RMSDs to this structure for the OPLS Aβ40 simulation over the production period using 0.01 nm-wide bins. Only residues 16-35 for both the reference structure and the simulation were considered for this analysis. Dotted lines guide the eye toward a cutoff of RMSD ≤ 3Å. The top centroid structure after single linkage clustering (0.2 nm cutoff) of structures within this cutoff is illustrated. (B) Table listing the % of frames of OPLS simulations that fall within certain RMSD cutoffs to the NMR structure, over different criteria. 16-35 is the dominant selection of residues for comparison here, in order to discount the effect of the highly flexible C-terminus.

A β 42's most populated centroid also represents an extended β -hairpin, this time possessing long distance hydrogen bond pairs for K16/M35 and V18/G33 characteristic of the A β 42 ensemble hydrogen bonding register. A double hairpin between the CHC, the 31-36 region, and residues 39-42 is directly observed in A β 42's eighth centroid (representing 4.33% of the ensemble, Figure 2.8B) as well as the two A β 42 centroids with back-calculated $^3J_{\text{HNHA}}$ values that have the highest PCC and lowest RMSD, respectively, to the average experimental J couplings (Figure 2.8A). In all three of these conformations, the first hairpin has similar K16/M35, V18/G33, and F20/I31 backbone hydrogen bond pairs, while the second hairpin is stabilized by a pair of L34/V41 hydrogen bonds; the bonds and secondary structure align well with our ensemble-based model.

A β 42-M35ox(S)'s third centroid also possesses a similar CHC to C-terminal β -strand-turn- β -strand motif, but is unique from the other centroids discussed above in that the N-terminus interacts with the motif with residues G9-E11 contributing an additional β -strand. Here, the cross region hydrogen bonds are K16/G37 (acceptor/donor), V18/MTO35, and F20/G33, aligning with the A β 40-characteristic bonding pattern (MTO = methionine sulfoxide). Note that the two A β 42-M35ox(S) centroids with $^3J_{\text{HNHA}}$ values that have the highest PCC and lowest RMSD compared to experiment (Figure 2.8A) do not form double hairpins like A β 42; neither 34-36 nor 39-41 form β -interactions and instead tend to interact with more N-terminal parts of the protein.

As discussed in Section 2.4.2, the overall content of β -interactions is reduced in A β 40-E22 Δ . This said, the top centroids of A β 40-E22 Δ are characterized with longer β strands than those in other species. In particular, residues closer to the C-terminus are involved in binding in these hairpins, corresponding to the increased preference of C-terminal registers by this mutant as seen in other analyses. Interestingly, centroid 2 places residues of the normally isolated central region (including G25-N27) in the center of a β strand, something not seen in any of the other simulations. The remaining three centroids have more typical central hairpins, but the bonding registers of these centroids are different (defined by V18/L34, F19/V36, and V18/G33 respectively), this variability is consistent with the broadened anti-diagonal seen in the contact maps. In contrast,

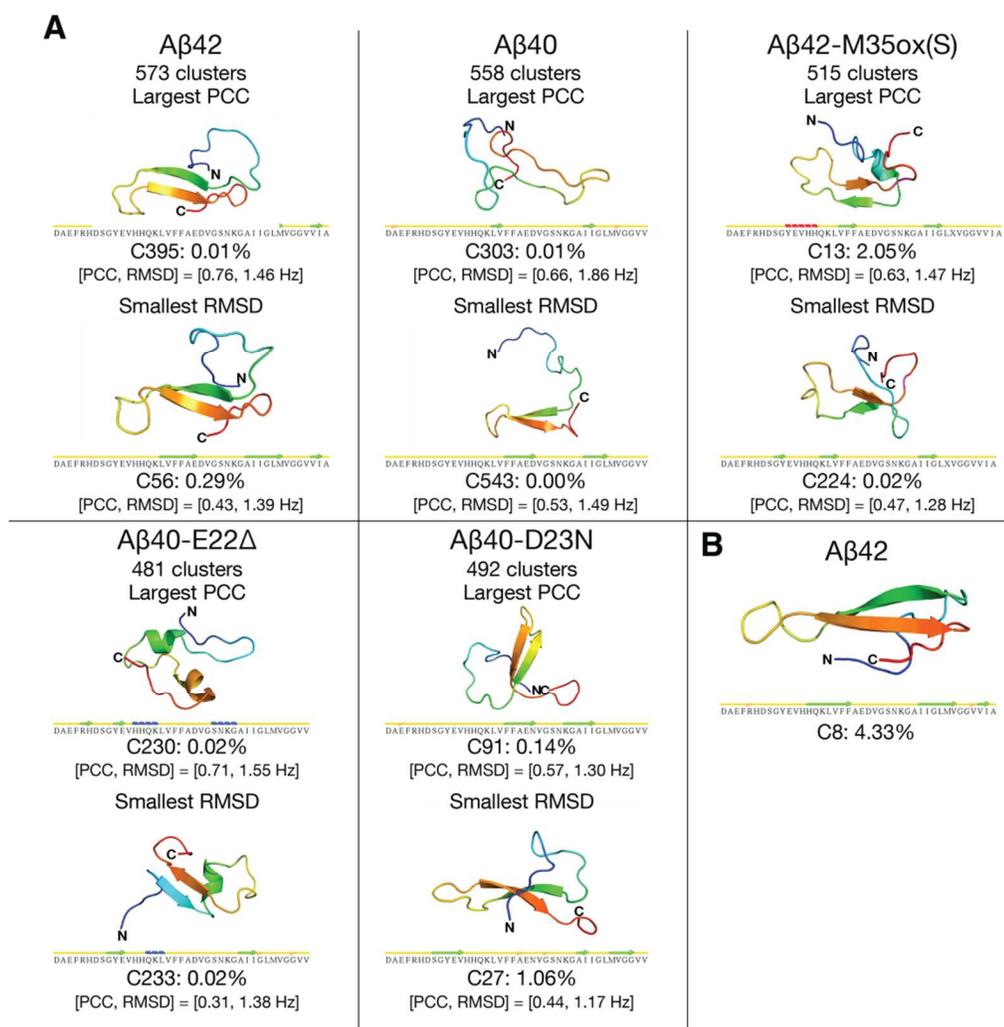


Figure 2.8. Other select centroids from single linkage cluster analysis(349) around $C\alpha$ atoms with an RMSD cutoff of 2.0 Å for each A β simulation done with the OPLS condition over the production period. (A) Clusters with back-calculated $^3J_{HNHA}$ couplings (using Vuister and Bax's parameter set(362)) demonstrating the largest PCC and smallest RMSD, respectively, to the inclusive average of the corresponding experimental data sets for each species as discussed in Chapter 4. (B) Centroid 8 from the A β 42 ensemble, demonstrating a similar double β -hairpin topology and electrostatic interactions as diagrammed in Figure 2.4. Ribbons are colored by a blue-to-red rainbow spectrum from the N-to-C terminus. Secondary structure strips for each centroid, as calculated by STRIDE(361), are also shown. Figure rendered in Pymol(360).

β -hairpins that do exist in A β 40-D23N and A β 40-E22K's top centroid structures are fairly short, despite an overall increased proportion of β annotated secondary structure. Centroid 1 and 3 for D23N follow A β 42's bonding register, while D23N's centroid 2 has an even earlier register (V18/I31). Both D23N's centroid 1 and 3, as well as E22K's centroid 2 involve residues in the central region in their hairpins.

For each species except A β 40-D23N, a common theme observed in the centroid structures is D23's side chain being buried while binding to multiple inwardly pointing amide groups in the central region (V24-K28). Indeed, only a few of the top 4 centroid structures for these species do not feature this "crown" motif (A β 42 centroid 4, A β 42-M35ox(S) centroid 1, and A β 40-E22 Δ centroid 2). These interactions are associated with a bulge in the backbone of the peptide in this region, often inducing a high curvature bend structure. These features are absent in A β 40-D23N; neither the negatively charged E22 nor polar N23 take the mutant's place (N23 does interact with V24's amide in centroid 1, but the side chain is solvent exposed). Bends do populate A β 40-D23N's centroids 1-3 at V24-N27, but these do not feature amides pointing inward or a "bulge" structure. Another somewhat rarer feature that accompanies the "crown" motif in wild type A β is a salt bridge between E22-K28 (present in A β 40 centroids 3 and 4, A β 42 centroid 1, A β 42-M35ox(S) centroid 2 and 3). This ionic association is also present in A β 40-D23N's fourth centroid, despite lacking D23 "crown" interactions. E22 Δ or E22K mutants prevent this interaction from occurring; no substitute interaction involving D23 is observed in any centroid structure of these mutants either. Interestingly, all of the wild type A β centroids (except A β 42-M35ox(S) centroid 2) that involve D23 in "crown" interactions and E22 in a salt bridge also involve the CHC and C-terminus β -hairpin structures, suggesting these interactions support hairpin formation.

2.4.6 Radius of Gyration and End to End Distance

Figure 2.9A is a histogram of the radii of gyration over the A β ensembles. We can compare these values to the model proposed by Kolinski and Skolnick(363) for single domain globular proteins derived from a statistical analysis of known protein structures, which states that the mean squared radius of gyration $\langle R_g^2 \rangle$ relates to n , the number of residues, as:

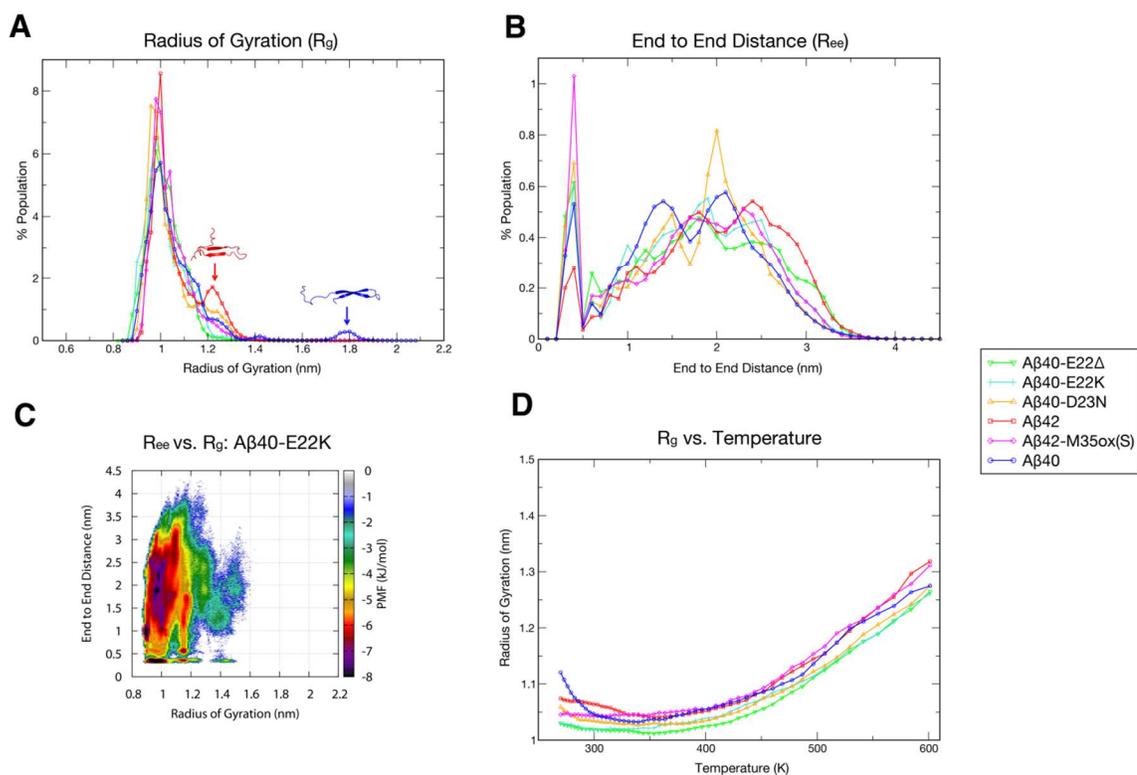


Figure 2.9. Radius of gyration (R_g) and end to end distance (R_{ee}) for OPLS A β simulations. (A) Histogram for R_g over the production ensemble, using 0.02 nm-wide bins. Depicted over high R_g basins for A β 42 and A β 40 are the highest populated centroid of the frames that compose that basin; these basins are discussed in more detail in the text. (B) Histogram for R_{ee} , using 0.1 nm-wide bins. (C) Free energy surface along the coordinates of R_g and R_{ee} in units of kJ/mol for the production ensemble of one of the A β peptides, A β 40-E22K, using 0.02 nm binning in both dimensions. (D) Average R_g as a function of temperature over all frames in the timeframe 400–1000 ns/replica.

$$\langle R_g \rangle = 0.22n^{0.38} \quad (2.1)$$

We therefore expect a radius of gyration of 0.89 nm and 0.91 nm for our 40 and 42-residue alloforms respectively if they behave exactly like fully collapsed, globular proteins. All of the OPLS simulations dominantly sample radii of gyration from 0.90-1.20 nm, consistent with a mostly collapsed ensemble. These values are consistent with the measurement of the hydrodynamic radius of HFIP-treated, monomeric, rhodamine-labeled A β 40/A β 42 (estimated to be \sim 0.9 nm for both peptides) by fluorescence correlation spectroscopy(364).

Some of the ensembles sample minor secondary peaks indicative of more extended structure. A β 42 samples populations (17.3% of the ensemble) with a radius of gyration from 1.18-1.38 nm; clustering analysis of the frames that fall in this basin indicate this region is characterized by extended β -topologies between the CHC and 30-36 region. While A β 42 is the most prominent, all of the species do significantly sample configurations with these R_g s, with the notable exception of A β 40-E22 Δ , which is in general slightly biased toward lower R_g structures. A β 40 also samples a small population of frames from 1.68-1.88 nm. Frames in this region are similarly characterized by an extended β hairpin structure, but in this basin, there are periodic boundary contacts between the farthest edges of the N-terminus and the central region. This causes the N-terminus to be positioned away from the rest of the residues in its own image, accounting for the large radius of gyration. Given the involvement of crystal contacts and minor population (2.7% of the total frames) in this ensemble, we do not assign any significance to this peak in our analysis of A β 40 monomers.

The peptides can adopt a wide variety of end to end distances, as seen in Figure 2.9B. A basin is observed at the lowest end to end distances (0.25-0.5 nm) indicative of interactions between the zwitterionic termini, strongest in A β 42-M35ox(S) and weakest in A β 42. Indeed, there is a general bias toward larger end to end distances in A β 42, possibly caused by the C-terminus's involvement in more local β interactions. Further, direct end to end interactions were seen in A β 42-M35ox(S)'s ensemble, such as the centroids in Figure 2.6; the presence of these global interactions may contribute to the reduction in C-terminal β -meanders sampled in this ensemble. Meanwhile, wild type A β 40 has a tendency to sample smaller end to end distances. Figure 2.9C is an example

of a free energy landscape calculated over radius of gyration and end to end distance for A β over the production ensemble. In general, the landscapes for A β peptides are very rough, possessing many different basins. Variation of average radius of gyration with temperature was also assessed in Figure 2.9D. The data show that at our highest temperature (~600K), average R_g value does not reach a plateau, indicating that REMD continues to enhance sampling efficiency up to the highest simulated temperatures.

2.4.7 Solvent Accessible Surface Area

Backbone and normalized side chain solvent accessible surface areas (SAS) over each simulation are illustrated in Figure 2.10. In general, these SAS profiles are very similar for all species and, like radius of gyration data discussed in the previous subsection (2.4.6), are consistent with partially collapsed structures. For all species except A β 40-D23N, residues E/K22-G25) form a very characteristic alternating backbone and side chain SAS pattern highly conserved in all three species. In A β 40-E22 Δ , this pattern is present, but A21 assumes the role of missing residue E22. This pattern matches the direct observations made in centroid structures (Figure 2.6). The burial of the D23 side chain appears to be due to its hydrogen bonding interactions with backbone amides of V24-K28 that establish the crown motif in this region, as discussed in Section 2.4.4. A β 40-D23N disrupts these interactions, and produces a pattern where residue F20-E22 are more solvent buried, while residues N23 and G25 are more backbone exposed and residues S26-K28 are more side chain exposed. A β 40-E22K follows the zig-zag pattern in this region, but also strongly buries residue A21 and K22's backbone and side chains compared to wild type. Also worth noting is the exposure at residues I31-I32 and burial at residues G33-M35 of A β 42 of the backbone (and side chains, to a more limited extent) compared to A β 40. It is possible that the N-terminal shift in hairpin hydrogen bonding register may force these residues to be more exposed in A β 42. Interestingly, A β 42-M35ox(S)'s backbone SAS profile aligns with A β 40 much more than A β 42 in this region, correlating with the similarity in hydrogen bonding patterns of these species. Unsurprisingly, both 42 residue alloforms also bury the backbones and side chains of residue V39 and V40 much more than the 40 residue

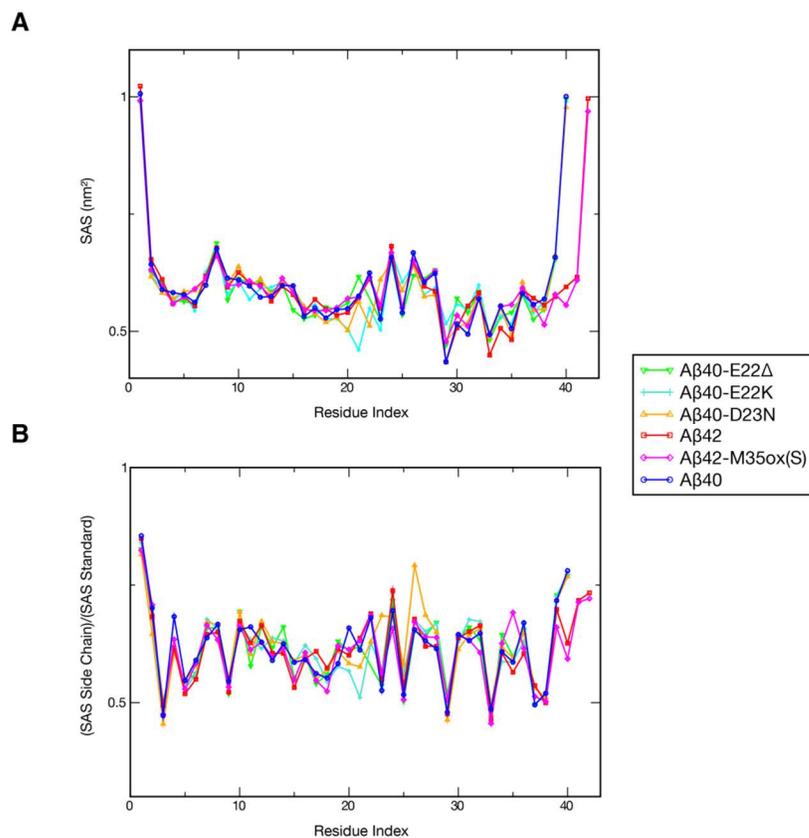


Figure 2.10. Solvent accessible surface (SAS) per residue(354) over the production ensemble of OPLS A β simulations, for the (A) backbone atoms and (B) side chain atoms. The latter is normalized by the standard side chain solvent accessibility for each residue defined by G-X-G tripeptides(365). Mto35 in A β 42-M35ox(S) is divided by the G-M-G standard, since no standard exists for methionine sulfoxide. Error bars are not visible here because error values calculated for these graphs for all points are all very close to zero.

length species. Besides glycines, surprisingly, the side chains that are significantly hidden from solvent in general are very polar, including residues E3, R5, and D23; as seen through other analyses, these residues are responsible for coordinating intramolecular interactions in the peptide.

2.4.8 Order Parameters

We want a measure that can be used to characterize the fluctuations of the ensemble given the lack of a dominant reference structure. The isotropic reorientational eigenmode dynamics (iRED) method proposed by Prompers and Brüschweiler(353) permits the calculation of modes of correlated motion and their associated order parameter without the need of an overall alignment frame, a condition that is especially unlikely to be satisfied in long timescale simulations of IDPs. Order parameters (S^2) derived for backbone amide groups in this system by this method describe the angular rigidity of the internuclear vector for each group (Figure 2.11). The obtained order parameters for all amides fall within the range 0 to 0.7, as expected for a highly flexible system. Nevertheless, there are substantial residue-specific differences for this quantity. For all species, the CHC, I31-V36, and (in A β 42/42-M35ox(S)) 40-41 regions possess significantly larger order parameters, correlating to the locations of β -character of the systems. FAD mutants like A β 40-D23N and A β 40-E22K have a broader distribution of high S^2 , while G29 and A30 have reduced order parameter for A β 40-E22 Δ but larger rigidity, consistent with the changes in hairpin register seen in these species. The H6-S8 section also demonstrates substantial rigidity among the simulations; this region is weakly structured but is observed to make significant cross-region interactions in all three contact maps. Low order parameters, in contrast, are generally observed at the termini as well as the mostly disordered 9-15 and the 25-26, 28, and (in A β 42/42-M35ox(S)) 37-38 regions that are dominantly characterized by DSSP as bend secondary structure. Curiously, A β 40 and A β 40-E22 Δ demonstrate higher order parameters for 23-24 and residue 27 compared to the other species; these may have to do with interactions of the “crown” motif, particularly the greater prevalence of D23 side chain to 27-28 backbone hydrogen bonding in A β 40 (Table 2.2). Note that these order parameters are

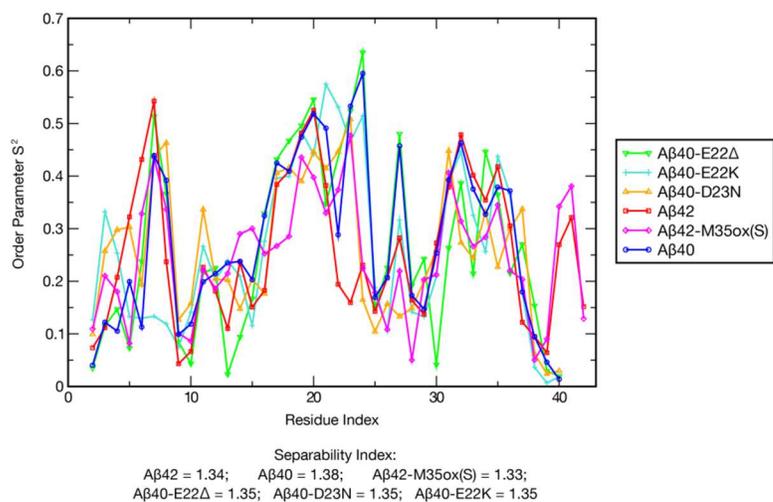


Figure 2.11. S^2 order parameters for all backbone N-H internuclear vectors (excluding the N-terminus), calculated with the iRED technique(353) over the production period for OPLS simulations. The Prompers-Brüschweiler separability index between the internal modes and global modes is also provided for each species; compare to an index of 5.39 for ubiquitin in bulk water(366).

very low for A β 40-D23N, which lacks these interactions. A β 42-M35ox(S) demonstrates reduced rigidity in the CHC and 31-36 regions compared to the other alloforms; this may be attributed to weakening of hydrogen bonding interactions between these regions. Finally, it is important to note the very low Prompers-Brüschweiler separability indices of 1.3-1.4 for these ensembles, compared to a value of 5.39 for well folded ubiquitin in bulk water.⁽³⁶⁶⁾ This is indicative of disordered dynamics where the global and internal motions are highly intertwined. Unfortunately, due to the low separability of the disordered A β ensembles, the standard model-free formalism proposed by Lipari and Szabo⁽³⁶⁷⁾ cannot be used to fit order parameters from NMR relaxation data, prohibiting accurate comparison of these data to experiment.

2.5 Discussion and Conclusions

Key to understanding and manipulating the self-assembly processes that naturally yield the higher order A β assemblies intimately associated with Alzheimer's disease lies in properly characterizing the A β monomers that comprise the building blocks of these conformers. While there have been many attempts to characterize A β monomers through NMR, CD, and computational simulations (outlined in Chapter 1), the results of these studies present an overall incomplete and inconsistent depiction of the system. This is largely due to the highly disordered and aggregative nature of these peptides, which preclude the use of many traditional experiments, frustrate sample preparation, and increase the roughness of the free energy landscape, thereby complicating folding simulations of the system. To help address these issues, we have performed REMD simulations on the μ s/replica timescale for each of the full length A β peptides while closely monitoring trajectory convergence and comparing calculated J-couplings and RDCs with their experimental counterparts. Further, the secondary structure character and sampled topologies of A β determined from analysis of these trajectories are consistent with several previously published experimental findings about the system. Here, we will summarize our results and compare them to those of previous experimental and computational investigations.

Our simulations suggest a general characterization of wild type A β with frequent sampling of β -hairpin topologies between the CHC (residues L17-A21) and residues

A30-V36, with the V24-K28 region composing the intermediary bend region. These secondary structure proclivities are supported by chemical shift indices(4), and it is recognized experimentally (by the effects of proline mutations as well as co-incubation with protein fragments) that the exact regions characterized as β dominant above are critical for A β fibrilization(205-208). Further, the β -hairpin structure of monomeric A β 40 in complex with a phage selected Affibody determined by solution NMR(355) (PDB: 2OTK) aligns with the central β topology in the free monomer simulations presented here, with the backbone hairpin hydrogen bonding pattern of the NMR structure exactly matching the dominant bonding register of our A β 40 simulation. This robust agreement lends support to the hypothesis presented by the investigators that this hairpin represents a transiently populated conformation accessible to A β monomers.

Extended hairpin structures such as these are likely to expose the dominantly hydrophobic faces of the β sheet residues to solvent, which may permit them to act as seeds for self-association. (Note that an exception to this would occur in the environment of an oxidized methionine 35 side chain.) We refer the reader to Chapter 1 for further discussion of the A β aggregates mentioned herein. Hoyer and coworkers further posited an aggregation mechanism wherein A β stacks along these hydrophobic interfaces to form soluble oligomers, which then undergo a concerted 90° rotation about the axes of both β strands of the hairpin to form fibril seeds characterized by intermolecular parallel β sheets(355). This hypothesis is reinforced by attenuated total reflectance Fourier-transform infrared (ATR-FTIR) studies, which consistently report a peak at ~1695 cm⁻¹ indicative of antiparallel β character for soluble oligomers but not for fibrils(368, 369). In agreement with this hypothesis, the generally unstructured N-terminus and the C-terminal β -strand-turn- β -strand motif we report in our simulations shares important structural characteristics with different intrapeptide models for A β fibrils and "on-pathway" intermediate oligomers(23, 24, 26, 27). In particular, the locations of secondary structure correlate well to these models, and reported side chain contacts in these intrapeptide models are often analogous to the most populated backbone hydrogen bonds in our monomer simulations. This is consistent with a putative 90° rotation in the axis of backbone hydrogen bonding from perpendicular to parallel to the fibril axis in the pathway between monomers and fibrils. Lending further possible insight into this so-

called “Venetian blinds” theory(355) is another solid state NMR model, this time of a soluble, oligomeric preglobulomer of A β 42, which showed in its repeating dimer subunit intramolecular antiparallel β pairing between V18-D23 and K28-I32 and intermolecular parallel β pairing with G33-V40(25). The aforementioned antiparallel interactions in this model map to the monomer structures seen through our simulations, and though globulomers are observed to be off-pathway to fibril formation(195), one could speculate this model may nonetheless bear resemblance to a transitional form in the conversion between antiparallel and parallel character in on-pathway A β aggregation.

Further analysis of the A β 42 and A β 42-M35ox(S) simulations reveal that a second hairpin motif involving residues V39-I41 is formed, with G37-G38 constituting a second bend-dominated region. This is consistent with the measurement of greater C-terminal rigidity in A β 42 compared to A β 40(217) by NMR, as well as the observation of similar hairpin structures in simulations of A β 42 C-terminal fragments(370). It is also reminiscent of the β -meander topology in the component monomer model for toxic A β 42 pentameric, "on pathway" oligomers derived from solid state NMR data by Ahmed and coworkers(26). Based on the dimensions and shape of these oligomers as determined by size exclusion chromatography and atomic force microscopy, Ahmed *et al.* further propose a model wherein the C-terminal components of the meander associate to form a core, a structure that would be less likely to form with the less rigid C-terminus of A β 40. This may explain the presence of paranuclei in A β 42-like aggregation but lack thereof in A β 40-like aggregation (see Chapter 1). This hypothesis matches experimental findings: substitution of a hairpin promoting motif at the C-terminus was found to increase A β aggregation rate and promote higher order oligomers, while inclusion of a hairpin breaking motif substantially disrupted the A β 42 aggregation kinetics and changed the oligomer size distribution to one more characteristic of A β 40(371).

If this is true, however, the inability of A β 42-M35ox to form paranuclei would not be explained directly by the secondary structure proclivities of our simulation, which still suggest a prevalent sampling of C-terminal β character in this alloform. Nevertheless, it is observed in our simulations that antiparallel contacts involved in β -meander formation (CHC to I31-V36 and I31-V36 to V39-I41) are significantly weaker in A β 42-M35ox(S), while there is a greater tendency to form CHC to 39-41 contacts.

Likewise, hydrogen bonds across the β -meander, in general, are considerably weaker than the other alloforms, while V39-I41 contacts with other parts of the protein (like K16/V41 and A30/V41) are better populated. Reduced sampling of a double hairpin topology in favor of alternate β -hairpin binding patterns may therefore account for A β 42-M35ox's altered aggregation properties.

Solution NMR studies reveal that V24-K28 forms a turn that is suggested to nucleate folding of the monomer(216). The very high prevalence of bend character in these simulations and the quicker equilibration of bend compared to β sheet in our simulations seem to support this hypothesis (though it is important to note that REMD simulations will not reflect the true folding dynamics of the system). We report that this region typically forms all wild type A β simulations, wherein D23's side chain is positioned in the core of the turn and interacts with the inwardly-pointed amide groups of the composite residues of the turn. The maintenance of this "crown" motif structure throughout the simulation appears to be reliant on electrostatic interactions involving the negatively charged side chains of the FAD-linked residues E22 and D23. The latter appears to position the turn residues and the former seems to position D23 itself and forms a long distance salt bridge with K28 across the region. It is very probable that the FAD mutations that disrupt the capacity of these side chains as hydrogen bond and salt bridge partners would also disrupt the turn structure, aligning with experimental observations of this phenomenon(227). The structure of the central region and electrostatic involvement of E22 and D23 reported in our simulations nearly perfectly match the characterizations of the highest populated cluster in REMD simulations of the A β (21-30) fragment using a similar OPLS-AA/TIP3P combination(372). This suggests that the fold of the A21-A30 region is not affected appreciably by the rest of the protein, supporting the idea that it folds autonomously and nucleates the folding of the monomer. Indeed, the frequency of these interactions in the ensemble and consistency of their presence among the top centroid structures suggests that D23's side chain to backbone hydrogen binding may be the first step in turn nucleation in the central region, while the E22-K28 long distance salt bridge caps off the domain allowing even further long distance hydrogen bonds to form between the β -regions. The structural features of the turn region reported here are also reminiscent of other reported computational ensembles

derived with different simulation parameters for both the A β (21-30) fragment(373-375) and the central region of full length A β monomers(376). It is important to note, however, that despite a central region turn motif being common to the Affibody-trapped monomer(355) and oligomer/fibril models(23-27), the roles of E22 and D23 are observed or are hypothesized to be substantially different. In the former, E22 and D23 appear to interact with the solvent and Affibody protein rather than bonding to the 24-28 region, while in the latter, it is typically observed that D23 actually forms a salt bridge with the K28 side chain.

Our analyses of the familial AD associated mutants A β 40-E22 Δ , A β 40-D23N, A β 40-E22K more explicitly describe what occurs to the system when the chemistry surrounding these key residues is changed. A common theme in investigating these mutants is that each alters the central region's charge-charge interactions and structural biases, which in turn permit different cross-bend hydrogen bonding patterns. The way they each do this is quite different among the different species. In general, mutants of the E22 reduce but do not remove the prevalence of the D23 side chain to backbone interactions (ie, populations of key interactions are still 50% or greater in both mutants), but trivially nullify the E22-K28 salt bridge, with no anion in the central region replacing this interaction. The primary effect of this modification in E22 Δ is a broader distribution of bend character in the central region, with reduced character in apex residues V24 and G25 but increased bend character in C-terminal boundary residues G29 and A30. In E22K, ionic and β sheet interactions before the bend with the N-terminus are encouraged, while the bend character is generally reduced. A β 40-D23N, on the other hand, has the most profound effect on central region electrostatics observed. Substitution of the less charged side chain greatly reduces both E22 and D23 characteristic interactions, and disrupts sampling of bulge "crown" features in favor of narrower bends. The distribution of bend secondary structure is both reduced and narrowed, while residues flanking the bend (F20-D/N23 and K28-I31) are greatly increased in β character. Consequentially, all FAD mutants broaden the distribution of C-terminal partners in the CHC to I31-V36 main antidiagonal, but the two E22 mutants particularly bias bonding toward more C-terminal registers, while D23N prefers registers more N-terminal to that of wild type. As covered in Chapter 1, aggregates formed by FAD

mutants may be different from their wild type counterparts. For example, A β 40-E22 Δ uniquely forms fibrils with a filamentous morphology (250), while A β 40-D23N is the only A β variant observed thus far that can be induced to adopt antiparallel cross β structure (251). If transient sampling of the central hairpin in the monomeric form truly acts as a seed for A β aggregation as the evidence seems to suggest, then we speculate that the increased polymorphism in the types of cross-bend H-bonding patterns seen through REMD simulation may permit the adoption of unique aggregates and new, faster aggregation pathways. As an interesting note, a comparison of the intramolecular coordinates from the solid state structures for the antiparallel A β 40-D23N fibril(251) (PDB ID: 2LNQ) with those from 2-fold symmetric parallel A β 40 fibrils(23) (PDB ID: 2LMO and 2LMN) reveal a narrower central bend (V24-K28 as opposed to V24-G29), greater distance between the two opposite strands, and a shift in side chain contacts toward a more N-terminal register (defined by the side chain of F19 being closer to I32 in 2LNQ, as opposed to F19 being opposite L34 in 2LMO/2LMN). Parallels to these observations can be drawn for both the reduced extent of the central bend and the 2-register shift seen by backbone hydrogen bonding between A β 40-D23N and A β 40 at the monomer level in the REMD simulations.

The central region “crown” motif present in our simulations exposes the carbonyl groups radially outward into the solvent, creating a negatively dominated external surface with a central, positively charged indentation (Figure 2.5). This distribution of charge, in tandem with the preferred geometries of the protein, may explain why this bend region is restricted from interacting with all of the other (mostly hydrophobic) segments of protein, with residues 4-6 (containing a positively charged R5) being the only exception observable in the contact maps (Figure 2.3) presented in this study. We speculate that this may present a link to the H6R and D7N FAD mutations, which both increase positive charge in this region. It is possible they promote the N-terminus to interact more persistently with the negative side chains and carbonyls in the region, competing with the electrostatic interactions that nucleate the turn and potentially destabilizing the motif. Indeed, this phenomenon was already observed here with the A β 40-E22K mutation, which induced the A21-D23 region to form ionic and β interactions with the N-terminus. The negative character of the loop surface could also

explain certain experimental findings for A β , such as the increased resistance of this region to proteolytic attack not accounted for by sequence-specific variation(216) and the reduced ^1H - ^{15}N HSQC chemical shift perturbation of these residues compared to the rest of the protein when titrating A β monomers against negatively polarized or hydrophobic dominated compounds(21, 233-235). While formation of the V24-K28 turn correlates with the formation of cross-region β -hairpin character that may form aggregation-prone hydrophobic interfaces, one could also speculate that the negative surface presented by this turn could also have a protective effect against this hydrophobic-dominated self-assembly. This hypothesis may explain why disruption of the turn region is associated with higher orders of oligomerization(227).

Finally, differences in the backbone hydrogen bonding patterns that form the central hairpins represent another significant change between these species. The C-terminally shifted pattern of contacts between the CHC to 30-36 in A β 40 compared to A β 42 may be preferred in that species because it can better accommodate the interactions that form the central region. For example, the E22-K28 salt bridge is able to form with significantly higher frequency in A β 40 and A β 42-M35ox(S) compared to A β 42 or any of the FAD mutants (Table 2.2). Indeed, the fact that both A β 40-D23N and A β 40-E22K, two mutants that disrupt this salt bridge, are dominated by the A β 42 register supports this hypothesis. A β 42 and these FAD mutants also uniquely feature E22/G29 as a significantly populated backbone hydrogen bond (Table 2.1). N-terminal shifting of cross-hairpin contacts in A β 42 may be induced by the formation of hairpin character between 30-36 and 39-41 (in particular, the strongly populated hydrogen bonds between A30/G37 and L34/V41). If this is the case, then the weakening of these interactions in A β 42-M35ox(S) may explain the weak preference for this species to actually adopt A β 40-like bonds across the CHC to 30-36 motif rather than those of its reduced counterpart.

In this study, we report computationally-derived ensembles for wild type and FAD mutant monomers that reproduce experimental NMR observables with reasonable fidelity and match experimental trends observed for the system by other studies. The high proclivity of A β monomers to establish hairpin-like conformations that resemble oligomer and fibril models hints at the importance of these motifs in A β self-assembly.

Indeed, transient sampling of these structurally similar conformations may govern aggregation of these peptides, as suggested by other investigations of aggregation-prone (N*) states (22, 377). We further report the enduring presence of a bend region nucleated by interactions with the side chain of D23 that is disrupted by the charge-altering FAD mutants of these residues. This suggests a mechanism that explains the electrostatic involvement of FAD-linked residues in A β structure. A β monomers are a building block in the formation of both "on pathway" intermediate structures in fibril formation and "off pathway" oligomers. The major conformations calculated in this study may be useful for drug discovery targeting this basic form. Further, the detailed biophysical characterization of A β monomers presented here, if experimentally verified, is expected to provide valuable predictions for understanding and manipulating the mechanisms that yield the pathogenic assemblies involved in AD.

3. Replica Exchange Molecular Dynamics Simulations of A β Monomers Using the AMBER99sb-ILDN/TIP4P-Ew Force Field

3.1 Abstract

Amyloid β (A β) monomers represent the base state in the pathways of aggregation that result in the fibrils and oligomers involved in Alzheimer's disease (AD). The structural properties of these intrinsically disordered peptides remain unclear despite extensive efforts to resolve these through experiment and computation. To assess the generalizability of our simulations, we compare the ensembles described in the previous chapter (with OPLS-AA/L forces and TIP3P water) with all-atom, explicitly solvated simulations of different isoforms (A β 42 and A β 40) and point mutants (E22 Δ , D23N, E22K, and A2T in A β 40) of A β using a different force field (AMBER99sb-ILDN) and water model (TIP4P-Ew); this field can reversibly fold globular proteins *de novo*(43, 50). Despite this fundamental change in simulation parameters, we find that these ensembles demonstrate a strong convergence in structural properties and good agreement with experimental NMR observables. In particular, antiparallel β -hairpin structure between L17-A21 and A30-L34 is prevalent in ensembles of A β 40 in both force fields, while residues A21-A30 forms an interceding region that rarely interacts with the majority of the protein. Further, A β 42 contributes new β -hairpin motifs involving V40-I41, regardless of force field. However, the structural flexibility of the central region and the electrostatic interactions that characterize it are notably different between the two conditions. Further, the effects of FAD-linked mutations in this region are much more subtle in the ILDN set. These mutations primarily encourage more global contacts involving the N-terminus and the central region, and promote the formation of new β topologies that may seed different aggregates involved in disease phenotypes. Subtle differences aside, the large degree of agreement between these simulation sets across force fields provides a generalizable characterization of A β that is also highly consistent with experimental data and models.

Portions of this chapter previously appeared as D. J. Rosenman, C. R. Connors, W. Chen, C. Wang, A. E. Garcia, A β monomers transiently sample oligomer and fibril-like configurations: ensemble characterization using a combined MD/NMR approach. *J. Mol. Biol.* **425**, 3338 (2013).

3.2 Introduction

As reviewed in Chapter 1, there have been many efforts to determine the structural properties of A β monomers experimentally and computationally; however, progress has been frustrated both by the high aggregation propensity and inability of A β monomers to establish a stable native state under physiological conditions (intrinsic disorder). While a number of computational investigations have been published to investigate full-length A β monomeric ensembles (summarized in Chapter 1, examples include (9-18)), there is a disparity in the secondary and tertiary structural properties of the simulated ensembles reported in these studies. These discrepancies may arise from (among other factors) differences in simulation parameters, insufficient sampling, and differences in simulation analysis protocols. Among the first category, inconsistencies in force field have been documented to lead to widely varying results in simulating unfolded and intrinsically disordered ensembles(28, 46-48). To test which conclusions made in the previous chapter regarding the effect of sequence length and point mutations linked to familial Alzheimer's disease (FAD) on the ensemble of A β monomers are robust to these changes, we have simulated these peptides once more under a fundamentally different force field/water model combination. One additional point mutant of A β 40 is studied here, the rare Icelandic mutation A2T, the first A β substitution found to protect against Alzheimer's disease and mental decline in the elderly(257), interesting considering that alanine to threonine substitutions are often affiliated with familial amyloidogenic diseases(378). Although this particular A2T mutant (unlike the others simulated here) is proximate to the N-terminal cleavage site and has been observed to reduce both A β production and β cleavage activity by BACE1(257), the effects of this mutant on A β structure, dynamics, and aggregation may also be implicated in its phenotype of improving cognitive health in advanced age. Nevertheless, studies of the effects of the A2T substitution on A β aggregation remain controversial, with certain ThT fluorescence studies showing enhanced(379), similar(259), or reduced(260) rates of aggregation in the context of A β 40, and either similar(260) or reduced(259) rates of aggregation in the context of A β 42.

Here, we compare unbiased, all-atom, explicitly solvated REMD simulations performed by our lab over hundreds of ns/replica using different force field and water

model combinations: the runs discussed in the previous chapter generated with the OPLS-AA/L(19) force field and TIP3P water model(20) (referred to as the "OPLS" condition here) and new simulations run with AMBER99sb-ILDN(21) and TIP4P-Ew(22) (herein called the "ILDN" set). The philosophy of design of OPLS and AMBER type force fields are comparable; in particular, both fields employ: a nearly identical functional form, very similar bonded parameters (peptide parameters for bond angles and dihedrals were originally adapted from the same force field published by Weiner *et al.*(380) for both OPLS and AMBER fields), similar fitting of Lennard-Jones parameters to match the experimental heat of vaporization for model liquid compounds, and a similar method (though employing different weights) for the rescaling of 1-4 non-bonded interactions(381-385). The primary difference between the two force fields is in the way partial charges are calculated. For new molecules, OPLS charges are usually derived using parameters transferred from previously fitted model compounds that are then either coparameterized with Lennard-Jones to reproduce the experimental properties of liquid compounds and/or adjusted based on chemical intuition; either way, this method often requires no new explicit quantum mechanical (QM) fitting(381, 383, 385). In contrast, AMBER charges are explicitly derived for each new compound by fitting single residue fragments *in vacuo*; this process entails optimizing the restricted electrostatic potential once the geometry of the molecule has been refined with *ab initio* QM methods(382, 384, 386).

Besides changing the force field, we are also employing a different water model in the two sets of simulations discussed here. The main difference is that TIP3P is a three site model (one for each atom of water), while TIP4P-Ew has four sites(20, 22). The new fourth site is introduced to move the center of the charge density away from the center of the Lennard-Jones interaction for the oxygen atom: this entails introducing a negatively charged dummy atom located near the oxygen that bisects the normal H-O-H bond angle, as well as adjusting the partial charges of the oxygen and hydrogen atoms(20, 22). Although this extra site adds extra computational cost, its addition has been observed to better model the electrostatic potential of water, as observed through neutron diffraction(20). Besides changes in partial charge, the TIP3P and TIP4P-Ew possess identical bond length and bond angle properties for its hydrogens and oxygen, while

both the Lennard-Jones repulsive and attractive terms are roughly ~10% weaker in TIP3P(20, 22).

Unlike OPLS-AA/L (to our knowledge), the AMBER99sb-ILDN force field has been shown to be capable of reversibly folding multiple globular proteins, along with other iterations of the same force field (AMBER99sb, AMBER99sb*, and AMBER99sb*-ILDN)(43, 50). This lends further importance to validating our conclusions with simulations using this second parameter set. Here, multiple alloforms (A β 42 and A β 40) and point mutants of A β 40 (E22 Δ , D23N, E22K, and A2T) were investigated with this ILDN set. Remarkably, the OPLS and ILDN ensembles both demonstrate a remarkable similarity in terms of secondary and tertiary structural properties and good agreement with experimental nuclear magnetic resonance (NMR) observables, despite the use of a different force field and water model. Both sets demonstrate antiparallel β contacts among residues L17-A21, A30-L34, and, for A β 42, V40-I41, consistent with transient sampling of a β sheets between these regions. Likewise, the central region (21-30) rarely interacts with the rest of the protein except at the N-terminus and forms a loop like structure in both simulations.

Interestingly, the two simulations vary on the structural and electrostatic nature of this intervening region, with OPLS strongly favoring high curvature bend secondary structure while ILDN samples a mixture of bend and turn structures in this region, involving more helical hydrogen bonds. While the aforementioned central hairpin is prevalent in ILDN, the increased flexibility of this central region permits ILDN to sample more diverse combinations of β sheets as well. The nature of familial AD-linked residues E22 and D23 is also quite different between the ensembles, with the OPLS combination strongly populating the side chains of these residues in bend-promoting electrostatic interactions, while the ILDN combination favors a more diverse role for these side chains and often has these interacting with solvent. As such, the effect of FAD linked mutations of these residues is quite different in the ILDN case. OPLS implicates these mutants in affecting the physicochemistry of the central bend and bonding across the central hairpin, but ILDN simulations indicate that the point substitutions investigated primarily promote the formation of both local and global β hairpins involving the N-terminus, at the cost of interactions of this region with the central

hydrophobic cluster (L17-A21, CHC) more prevalent in wild type. The contribution of new β topologies with N-terminal residues seen in A β point mutants may be able to seed different pathogenic aggregates involved in familial AD. Lastly, both simulation sets are compared here for agreement to NMR observables (discussed in Chapter 4); while both yield comparable agreement through direct comparison of experimental chemical shifts and scalar couplings, ILDN shows better agreement when the changes in these values between different A β forms are compared. This seems to be largely due to inconsistencies caused by the “crown” motif characteristic of OPLS simulations. This said, the overall consensus between the extensively simulated, experimentally-validated ensembles reported in this study provides a highly generalizable characterization of the A β ensemble. Further, the specific differences reported between the ensembles here provide insight into the differences between force fields in the simulation of intrinsically disordered proteins that may facilitate force field refinement.

3.3 Methods

3.3.1 Simulation Details

This study compares different simulations generated with the replica exchange molecular dynamics (REMD) protocol, each using identical simulation parameters except water model, force field, and simulation duration. The REMD simulation protocol(35, 36) begins with several initially identical copies (called replicas) that are each simulated over a set of increasing temperatures. Replicas that neighbor each other in temperature are exchanged with a probability governed by the Metropolis Monte Carlo criterion throughout the simulation. The “OPLS” simulations discussed in this chapter use the OPLS-AA/L force field(19) and the TIP3P water model(20) and are described in detail in the previous chapter of this thesis. The “ILDN” simulations are new to this chapter, and are coupled to the AMBER99sb-ILDN force field(21) and the TIP4P-Ew water model(22). This combination is similar to one used in a previous study(11) (the only difference being the latter uses the older AMBER99sb(377) force field) that yielded an ensemble able to reproduce experimental NMR observables with reasonable accuracy. We chose nonetheless to use AMBER99sb-ILDN rather than the original AMBER99sb force field because the ILDN side chain torsion modifications

have been shown to considerably improve agreement with both side chain and backbone related NMR data for microsecond-timescale simulations(21). We rely on these criteria for the choice of AMBER99sb-ILDN as the force field for this study because the NMR data we have collected for A β , used for both the checks in the study mentioned above(11) or elsewhere in this thesis (i.e., Chapter 4), is, in general, not sensitive enough to report on the side chains.

Besides the choice of force field and water model, the “ILDN” simulation was run in an identical fashion as the “OPLS” simulations, as described in the previous chapter, using the same number of water molecules, box size, equilibration protocol, and other REMD simulation parameters. Recall that these simulations involve 52 replicas, spanning the temperatures 270.0 K to 601.2 K. Mobility of replicas through temperature space for these ILDN simulations was comparable to the OPLS runs (Figure 7.1 in the Appendix). These simulations were also run to 1000 ns/replica for all systems (cumulative simulation time of 52 μ s per system), necessary to reach an equilibrated ensemble. A β 42-M35ox(S) was not investigated under the ILDN condition, while a new point mutation of A β 40, A2T, was simulated only with ILDN parameters. The sequence for this peptide is:

- A β 40-A2T:

¹DTEFRHDSG¹⁰YEVHHQKLVF²⁰FAEDVGSNKG³⁰AIIGLMVGGV⁴⁰V

One exception to this construction for ILDN simulations is A β 42. For this run only, we used the final configurations for each of the 52 replicas of the simulation described by Sgourakis *et al.*(11) as the starting state. The AMBER99sb/TIP4P-Ew combination of the Sgourakis *et al.* simulation is very similar to the AMBER99sb-ILDN/TIP4P-Ew combination used for our ILDN run for this peptide, and no evidence of simulation instability was observed upon switching the force field. This A β 42 ILDN system alone was simulated to 750 ns/replica under the AMBER99sb-ILDN force field. All other REMD simulation conditions were identical to those used for the other simulations discussed in this chapter (and to those in the Sgourakis *et al.* predecessor).

The effects of pressure on the A β ensemble were studied using another A β 40 construct. This peptide simulated under identical conditions with the ILDN combination, but during the NPT equilibration simulation phase (described in Chapter 2), the pressure

was set to 2500 bar (still coupled using the standard Berendsen method⁽³⁴²⁾, with a temperature of 300 K), producing a box with sides 5.2 nm in length (as opposed to 5.4 nm at 1 bar). Use of a smaller box here may promote the formation of periodic boundary contacts, however, only 0.9% of frames of the analysed production period (see next section, 3.3.2) were characterized as being less than or equal to 0.5 nm away from its periodic image. Although this value is higher than the $\sim 0.001\%$ representation observed for the same system in the original 5.4 nm box, the percent of the ensemble represented by these contacts in the high pressure system is still mostly negligible. Further, this smaller box is needed to represent the higher pressure system while keeping the other simulation parameters constant, permitting a fair comparison to the 1 bar system.

3.3.2 Ensemble Analysis

Analysis was done for each of the ILDN simulations over the following time intervals (referred to as the "production period" in the rest of the chapter): 400-750 ns/replica for A β 42 and 400-1000 ns/replica for the other ILDN simulations. The enhanced sampling provided by the REMD protocol in tandem with the fact that analysis only begins after several hundred ns/replica into the run both mitigate the bias that comes from using the Sgourakis *et al.* starting conformation for the A β 42 simulation. All of these ensembles were analysed over 7 temperatures comprising the range 289.2 to 310.7 K; multiple temperatures were used to enhance sampling statistics. Analysis of the ILDN simulations were done in an identical manner as the OPLS runs, as described in the previous chapter. For A β 42, clustering was still done for every 100 ps-th frame despite the smaller ensemble, encompassing 24500 structures (in contrast to the 42000 conformations grouped for the other simulations).

3.4 Results

3.4.1 Convergence

We can apply the same methods we used to assess convergence in Chapter 2 for OPLS simulations to those of the new ILDN simulations presented here. C α average distance maps over 100 ns/replica time blocks were again investigated and the points where in the simulation the sum over the absolute value of pairwise distance differences

between the maps calculated over the current and previous time intervals does not exceed 50 nm for all subsequent time windows (*data not shown*) is: 100-200 ns/replica for A β 40-A2T/A β 40-E22 Δ , 200-300 ns/replica for A β 40, 300-400 ns/replica for A β 42/A β 40-D23N, and 500-600 ns/replica for A β 40-E22K, implying convergence at these segments. Likewise, the content of secondary structure over the entire sequence was assessed over 100 ns/replica blocks using DSSP(347, 348) and illustrated in Figure 3.1; the segments in which the population of each DSSP type does not change by more than 5% over all following time intervals is: 300-400 ns/replica for A β 40-D23N/A β 40-E22 Δ , 400-500 ns/replica for A β 42, 500-600 ns/replica for A β 40-A2T, and 600-700 ns/replica for A β 40. A β 40-E22K apparently does not converge by this criterion, progressively sampling increasing β content throughout the simulation and experiencing a large increase from 26 to 32% β character between 800-900 and 900-1000 ns/replica time steps.

In general, this analysis confirms that the A β ensemble is very prone to directional changes in secondary structure content even with enhanced sampling on the 100 ns/replica timescale. Further, ILDN simulations were even less prone to converge in terms of secondary and tertiary structural content than the OPLS simulation set. This said, 400-1000 ns/replica (400-750 ns/replica for A β 42) was still chosen as our period of time for ensemble analysis here for ILDN simulations. We believe this cutoff still represents a good compromise between excluding unequilibrated data while providing sufficient frames to draw statistically significant conclusions. Further, keeping a consistent cutoff means that a similar duration of time is analysed within and between the ILDN and OPLS simulation sets, removing this possible confounding variable for the ensemble comparison.

3.4.2 Secondary Structure

Reported in Figure 3.2A is the overall secondary structure content based on DSSP(347, 348) annotation for the production period of each A β ILDN simulation. These A β ensembles are even more similar in terms of secondary structure profile than the OPLS ensemble (compare to Figure 2.2A). Ranking for β structure is A β 40-E22K >

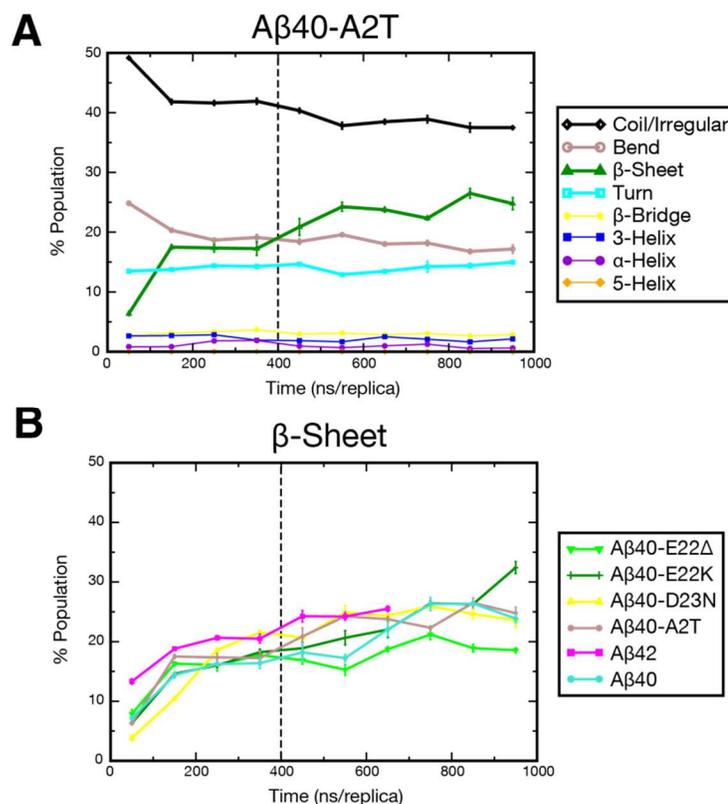


Figure 3.1. Content of secondary structure elements versus simulation time for $A\beta$ runs done with the ILDN condition, as determined by DSSP(347, 348) over 100 ns/replica windows. (A) Profile for all DSSP elements for an example $A\beta$ variant, $A\beta_{40-A2T}$. Dominant secondary structure elements in the ensemble coil/irregular, bend, β -sheet, and turn are denoted with thick lines. (B) Graph for the most widely varying element with time, β -sheet, for all species.

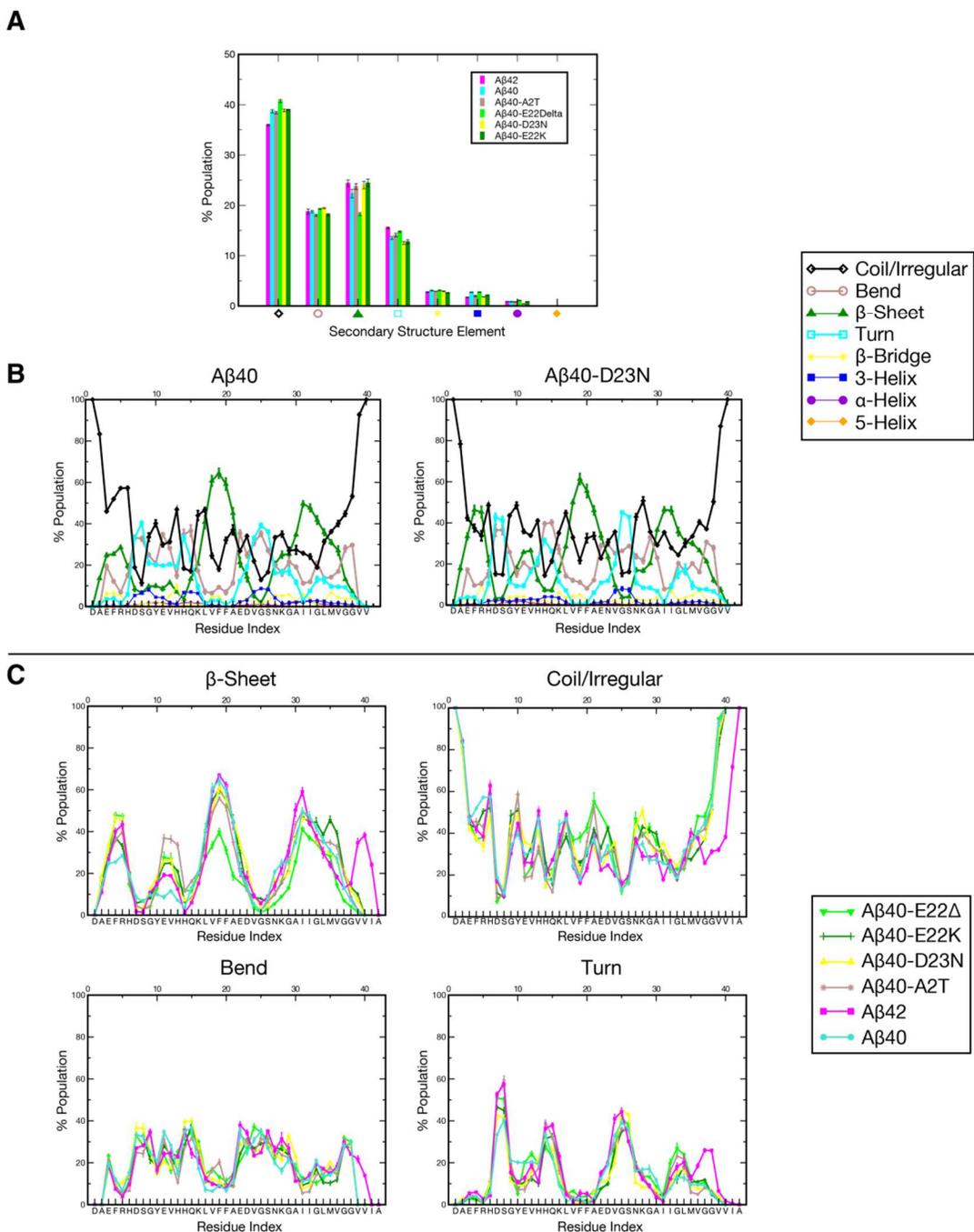


Figure 3.2. Percent composition of secondary structure as determined by DSSP(347, 348) for $A\beta$, averaged over the production ensemble, for simulations done with the ILDN condition (A) over all residues, (B and C) as a function of residue index. (B) Profile for all DSSP elements for two example $A\beta$ variants, $A\beta 40$ and $A\beta 40$ -D23N. Of particular importance are the dominant secondary structure elements in the ensemble, denoted with thicker lines in the graph: coil/irregular in black, bend in grey, β -sheet in green, and turn in cyan. (C) Plot comparing all of the $A\beta$ variants studied here for these four dominant DSSP elements. The sequence of wild type $A\beta 42$ is presented in the lower x -axis.

A β 42 > A β 40-D23N > A β 40-A2T > A β 40 >> A β 40-E22 Δ , with coil, bend, and turn structures making up the difference between these species. The most striking difference in secondary structure distributions between ILDN and OPLS conditions are an exchange of bend structure in the latter for turn character in the former. Bend, turn, and irregular each represent loop structures in DSSP nomenclature; however, DSSP defines bends to possess high curvature (that is, a direction change around the annotated residue of more than 70°), DSSP turns also possess high curvature but are also characterized by the presence of helical-like hydrogen bonds (that is, a hydrogen bond from carbonyl oxygen of residue i to the amide proton of residue $i+x$, where $n=[3,4,5]$), while regions annotated as DSSP irregular are loops that lack either of these features(347). In our discussion, we will define a “loop”-like structure in this paper to encompass both bend and turn secondary structures. Finally, like OPLS, the secondary structure compositions for ILDN simulations are consistent with those determined by CD for monomeric A β 40(214).

Figure 3.2B and 3.2C depict the secondary structure propensity as a function of residue index over each ILDN ensemble. Like OPLS (compare to Chapter 2, Figure 2.2B and 2.2C), ILDN simulations are characterized by strong β character in the CHC and A30-L34 regions separated by a loop like motif, with comparable population percentages between the force fields. Among the wild type simulations, the equivalent of that A30-L34 β region is actually considerably wider in OPLS, possessing two maxima at I31-I32 and L34-M35 as opposed to ILDN’s single maximum at I31. The FAD mutations largely do not perturb these peaks under this condition, with notable exception of A β 40-E22 Δ which strongly reduces the β content of both regions, and A β 40-E22K, which expands β content to include M35 and V36. A β 42 actually promotes β character in these regions, and, like OPLS, promotes a new maximum involving V39-I41 with an intervening loop region dominated by G37 and G38, suggesting the creation of new C-terminal β topology.

Some of the ILDN simulations are also characterized by considerably more β character in the N-terminus than the relatively less structured equivalents in OPLS, with substantial content at residues E3-H6 and Y10-H13. For ILDN, all of the point mutations of A β 40, in particular, enhance β content in these regions and turn content in the intervening residues. In contrast, A β 40 is more likely to sample turn and bend in this

region. This suggests that FAD mutants may promote the involvement of the N-terminus in new β topologies not observed in the wild type.

The nature of regions D7-G9, A21-G29, and G37-G38 are also different between the force fields. Both OPLS and ILDN conditions characterize these regions as loop-like, but OPLS is characterized by dominantly bend motifs while ILDN contains significantly more turn and irregular character in these regions. The increased heterogeneity of structures over the ensemble suggests that these regions are considerably more disordered in nature in the ILDN force field set. Unlike OPLS, in general, the profile of the central region is not strongly perturbed by mutations in ILDN, though subtle changes are present. For example, there is reduced turn content in A21-D/N23 and bend in E/K22 in all FAD mutants compared to wild type A β 40 and A β 42, in exchange for increased β and coil character. A β 40 also samples the largest turn character in K28-A30 of the different species, while A β 40-D23N samples reduced turn content (but more bend and coil content) in this region. The effect is that the distribution of the turn character is narrower in FAD mutants, with the distribution for A β 40-D23N having the most obvious difference.

Like OPLS, α -helical character is sparse in ILDN, except for a small tendency (<10%) to form 3-helices in D7-G9, H14-K16, and V24-N27. The commonalities ILDN simulations have with OPLS, such as similar locations of β and loop character, are also the attributes that are consistent with experimental data such as solvent exchange measurements(225, 226), chemical shifts(4), and J-couplings; this interesting concurrence is further elaborated in Chapter 4 of this thesis.

Finally, as discussed in more detail in Section 3.4.1, ILDN simulations were also similar to OPLS simulations in the observation that secondary structure composition continues to change even on the hundred ns/replica timescales, particularly for β content (Figure 3.1). This may explain why other simulation studies with less extensive sampling report lower contributions of β character to the A β ensemble(9, 11, 13, 18).

3.4.3 C α Contact Maps

C α contact maps (Figure 3.3, upper corner) and difference contact maps (Figure 3.3,

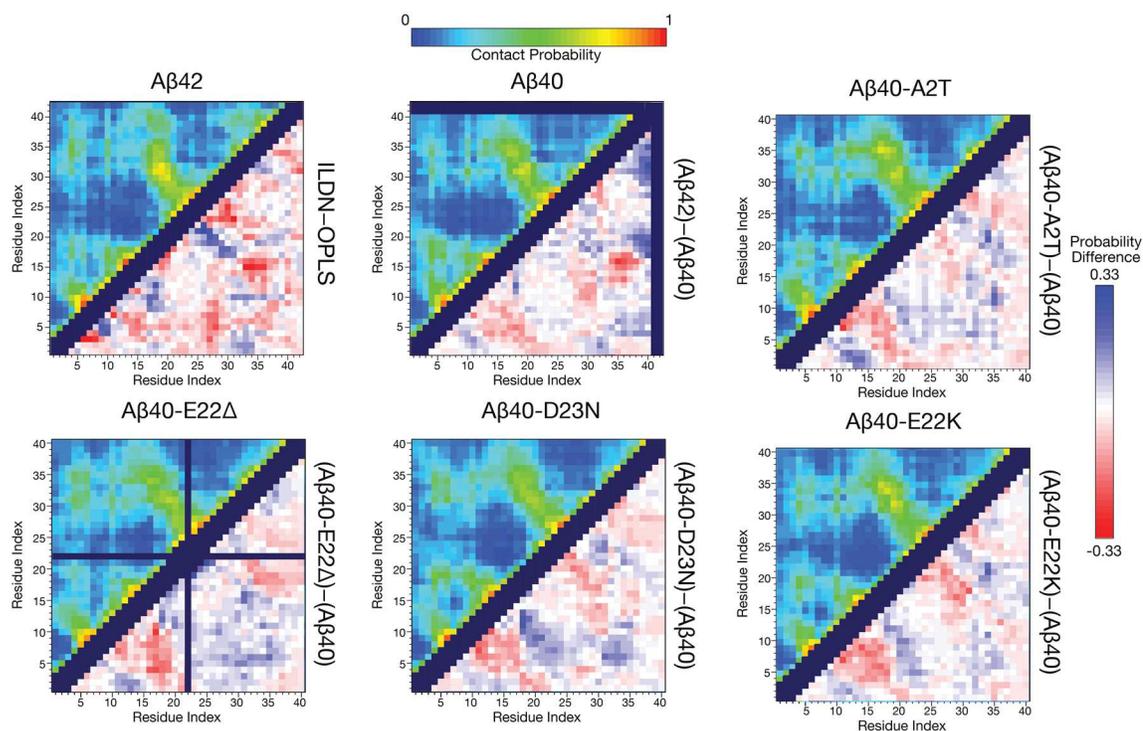


Figure 3.3. (upper corner) Maps illustrating the probability of contact for $C\alpha$ atoms over the production ensemble for different $A\beta$ isoforms and mutants derived with REMD under the ILDN condition. Here, two residues are defined to be in contact if their $C\alpha$ - $C\alpha$ distance is $\leq 8 \text{ \AA}$ (352). (lower corner) Difference between these contact probability maps for selected $A\beta$ mutants/isoforms with the ILDN condition, as listed in the title of each map. The difference map for $A\beta 42$ is a comparison between the OPLS and ILDN condition for this alloform. A red color indicates a reduced likelihood of contact while a blue value indicates increased contact probability for the minuend species compared to the subtrahend species. Self or sequential contacts, such as between (i, i) , $(i, i+1)$, and $(i, i+2)$, and invalid squares on the map were not computed and are represented by the dark purple regions.

lower corner) were calculated for all ILDN simulations, revealing biases in tertiary structure of the A β ensembles very similar to the OPLS condition in nature (compare to Figure 2.3). Common to both force field conditions, three regions of the contact map emerge: N-terminal (residues 1-20), central (residues 21-30), and C-terminal (residues 31-40/42), wherein the central region very rarely contacts the rest of the protein in wild type A β . Changes in contact probabilities across the different mutants and isoforms of A β are subtle with the ILDN condition, even more so than those than those noted for OPLS, but grant important insight into the structural differences between the simulated ensembles.

Characteristic of ILDN simulations are prevalent antidiagonal contacts between K16-G25 and N27-I32. Residues G33-G37 also interact with the CHC region, but contacts that extend the antidiagonal farther to involve residues N-terminal of L17 or C-terminal of V36 are much sparser than in the OPLS condition. FAD mutants are also far less prone to disrupt the nature of binding across this antidiagonal in the ILDN condition. Each of the point mutants of A β 40 studied here have a common effect of promoting a register shift for the main antidiagonal where more C-terminal binding partners are favored, except E22 Δ , which merely reduces the contact probability of this hairpin in general. This register shift, however, is much less prevalent than the shifts in diagonal binding discovered in OPLS simulations. These antidiagonal contacts also exist in A β 42 but are curtailed to L17-G25 and N27-I32, with G33-V36 interacting as a vertical extension. There is no register shift in the contacts of the main antidiagonal compared to A β 40. Analogous to OPLS, A β 42 possesses a second set of antidiagonal contacts that occur between I31-V36 and G37-I41, though these are somewhat more robust in the OPLS condition. Also prevalent are contacts between the CHC and the extreme C-terminus, with higher probability than OPLS. Subtle differences aside, the two force fields share a strong commonality in that they both involve the extreme terminus in a second hairpin that can join with the central hairpin to form a β -meander (that is, 2 subsequent antiparallel β -hairpins).

Like OPLS, the N-terminus also plays an active role in contacting the rest of the peptide for the ILDN condition, particularly at the CHC and the I31-V36 region. The N-terminus is less prone to contact the central region (particularly in wild type A β), while it

is more prone to form antidiagonals between A/T2-D7 and S8-H14 and between V12-K16 and the CHC. The point mutants of A β 40 studied here have a strong effect on the interactivity of this region. In common, all mutants reduced contact between D1-Y10 and the CHC, but increased contact between the extreme N-terminal residues and the following regions: Y10-Q15 (strongest increase in A2T), F20-G24 (strongest increase in D23N and E22K), and I31-M35 (strongest increase in D23N and E22 Δ). This suggests that these mutants are more likely to form additional β topologies involving the extreme N-terminus, in contrast to local, turn-like structures in wild type A β 40. Further, the increased contact of the N-terminus to F20-G24, in tandem with the increased likelihood to sample more C-terminal binding partners across the main diagonal, represent ways in which these point mutants promote long range contacts to the normally isolated central region.

3.4.4 Electrostatic Interactions

To complement the analysis done for OPLS simulations, Table 3.1 lists the most populated long distance backbone hydrogen bonds of the ILDN ensembles. Unlike the OPLS condition (compare to Table 2.1), clear $i+2:j-2$ patterns are, for the most part, not present across the central hairpin, and we cannot define species-specific registers as we have done for OPLS in the previous chapter. The only clear pattern is seen in wild type A β 40, where bonds defined by V18/M35, F20/G33, and E22/A30 are strongly populated. This register matches that also assumed by wild type A β 40 in the OPLS condition. However, this register isn't assumed by any other species in the ILDN set; while there are interactions between the CHC and 30-32 regions like OPLS, these are generally single interactions and are different across the different simulations. This said, the L34/V41 interaction suggestive of a second hairpin formation in A β 42 is present in both OPLS and ILDN condition, suggesting this is a generalizable interaction that is unique to A β 42.

Interestingly, ILDN simulations are much more prone to sample helical-like hydrogen bonds in turn regions such as H6/G9 or Y10, D (or N) 23/S26, and V24/N27. There is also a consistent helical hydrogen bond between I31/L34 or M35 in all of the

Table 3.1. Most commonly populated long distance, inter-residue, backbone hydrogen bonds in the A β ensemble for the ILDN set of simulations. All entries are backbone-backbone hydrogen bonds separated by at least two residues and populated $\geq 10\%$. Bonds whose acceptor and donor are separated by more than 5 residues are bolded.

% for ILDN A β 42		% for ILDN A β 40		% for ILDN A β 40-A2T	
BB H-Bond (Donor–Acceptor)	% Pop	BB H-Bond (Donor–Acceptor)	% Pop	BB H-Bond (Donor–Acceptor)	% Pop
H6 – G9	11	Y10 – D7	16	F4 – V12	17
H6 – Y10	21	V12 – G9	16	H6 – Y10	30
G9 – H6	45	H13 – G9	12	G9 – H6	50
E11 – F4	10	Q15 – V12	13	V12 – F4	27
V12 – F4	16	K16 – H13	26	H14 – T2	14
V12 – G9	15	L17 – S8	19	K16 – H13	22
H13 – G9	13	V18 – M35	12	V18 – A30	10
K16 – H13	28	F20 – G33	12	A21 – A30	13
L17 – H13	15	E22 – A30	10	V24 – A21	11
A21 – A30	13	S26 – D23	17	G25 – A21	10
V24 – A21	12	N27 – D23	12	S26 – D23	14
G25 – A21	11	N27 – V24	20	N27 – V24	23
S26 – D23	20	G29 – E22	11	K28 – G25	12
N27 – V24	16	A30 – N27	17	I32 – F19	14
I32 – F19	13	I32 – F20	13	L34 – I31	15
I41 – L34	10	M35 – V18	13	M35 – I31	13
		G37 – K16	12		
% for ILDN A β 40-E22A		% for ILDN A β 40-D23N		% for ILDN A β 40-E22K	
BB H-Bond (Donor–Acceptor)	% Pop	BB H-Bond (Donor–Acceptor)	% Pop	BB H-Bond (Donor–Acceptor)	% Pop
F4 – V12	17	G9 – H6	33	H6 – Y10	16
H6 – Y10	25	V12 – F4	13	G9 – H6	38
G9 – H6	40	Q15 – V12	12	V12 – F4	11
V12 – F4	20	K16 – H13	16	K16 – H13	26
H13 – Y10	17	S26 – N23	17	F20 – A30	13
Q15 – V12	10	N27 – V24	26	S26 – D23	20
K16 – H13	18	I31 – A21	10	N27 – D23	13
N27 – V24	14	L34 – I31	10	N27 – V24	17
A30 – F19	11	M35 – I31	11	K28 – G25	14
A30 – F20	11			A30 – F20	11
L34 – I31	20			I32 – V18	11
M35 – I31	19			L34 – I31	14

FAD mutants. Each of these sections, incidentally, cross a glycine (G9, G25, G33), and the first two also involve serines (S8 and S26), two residues that are highly likely to be involved in loop structure. The presence of these hydrogen bonds is consistent with these regions being defined as turn-prone over the ensemble by DSSP for this force field, as opposed to predominantly bend in the OPLS condition. Besides these short range hydrogen bonds, ILDN simulations are also characterized by long range hydrogen bonds at the N-terminus not present in OPLS simulations, in particular F4/E11 or V12. These bonds are a prevalent feature of the A β 40-A2T and A β 40-E22 Δ ensembles, and are conspicuously reduced in character in wild type A β 40. These interactions are likely affiliated with the N-terminal antiparallel character present in mutants but not in wild type A β 40 (see Section 3.4.3). Besides the conversion to a residue that intrinsically favors β character(387, 388), the A2T mutant introduces a hydroxyl group that interacts with the backbones of D1, A2, and E3 (4%, 28%, and 5% of the time respectively). In contrast, wild type A β 40 populates its N-terminus (D7-K16) with many more medium-range, helical backbone hydrogen bonds. This is in accordance with this variant's higher bend and turn structure, and increased tendency to form more local contacts.

Table 3.2 illustrates the populations of electrostatic interactions involving FAD-linked residue 22 and 23's side chains in each of the ILDN simulations. Compared to the OPLS simulations where these were highly prevalent, ILDN simulations have a much smaller likelihood for D23 to interact with V24-K28 through buried side chain to backbone hydrogen bonds and for E22 to interact with K28 through a cross-region salt bridge. These interactions do not appear to be supplemented with interactions involving other parts of the peptide; instead, these residues' side chains interact with solvent with a much higher proclivity. That said, these interactions are not completely abolished even in these ILDN simulations; indeed, local interactions between D/N23's side chain O δ and V24 and G25's amide are populated 7-13% in all simulations, while interactions involving S26's amide and side chain are present to 3-6% except in A β 40-D23N. Further, salt bridges with E22 or D23 as the anion and K28 as the cation are present in wild type A β simulations up to 5%. FAD mutants reduce the prevalence of these salt bridges to varying degrees. Mutants of E22 trivially eliminate the E22-K28 interaction,

Table 3.2. Central region electrostatic interactions of A β involving the side chains of residue 22 and 23 for the ILDN simulation set. Listed populations represent average bond frequencies for equivalent atoms in the side chain (for example, "ASP23O γ " is the average of interactions involving the D23O γ 1 and D23O γ 2 atoms). For A β 40-D23N only, populations for the interaction between K28HZ and N23's O δ in the final row follow the hydrogen bond criteria, not salt bridge criteria.

Donor – Acceptor Atoms	% for ILDN A β 42	% for ILDN A β 40	% for ILDN A β 40-A2T
<i>Hydrogen Bonds</i>			
V24NH – D/N23O δ	9	8	7
G25NH – D/N23O δ	9	10	10
S26NH – D/N23O δ	4	5	3
S26H γ – D/N23O δ	3	6	4
N27NH – D/N23O δ	4	3	1
K28NH – D/N23O δ	2	1	1
<i>Salt Bridges</i>			
K28NZ – E/K22O ϵ	4	5	5
K28NZ – D/N23O δ	4	5	4

Donor – Acceptor Atoms	% for ILDN A β 40-E22 Δ	% for ILDN A β 40-D23N	% for ILDN A β 40-E22K
<i>Hydrogen Bonds</i>			
V24NH – D/N23O δ	12	13	11
G25NH – D/N23O δ	11	11	11
S26NH – D/N23O δ	4	2	4
S26H γ – D/N23O δ	5	0	6
N27NH – D/N23O δ	5	1	2
K28NH – D/N23O δ	5	1	2
<i>Salt Bridges</i>			
K28NZ – E/K22O ϵ	N/A	2	N/A
K28NZ – D/N23O δ	3	2	2

but the D23-K28 interaction is also reduced in each case (to 3% in E22K and 2% in E22Δ), while Aβ40-D23N demonstrates reduced E22-K28 salt bridge occupancy, as well as low populations for the interaction with between K28's HZ and N23's Oδ (2% in each case). Unlike OPLS, residue 22 of the E22K variant interacts with D23 and does so for about 5% of the ensemble; no other significantly populated anionic partners exist for this lysine. In general, the reduction of central region electrostatic interactions permits the increased helical hydrogen bonding and greater secondary structure diversity seen in all simulations of the ILDN condition at the central region.

3.4.5 RMSD Based Clustering

Figure 3.4 depicts the top 4 representative structures for each of the ILDN simulations, determined by single linkage clustering with a 2.0 Å root mean square deviation (RMSD) cutoff. Like OPLS, all ILDN ensembles are represented by structurally diverse, collapsed structures. In the large majority of the centroids, the central region (residues 21-30), aligning with yellow in the rainbow coloring scheme of the backbone ribbons, is generally isolated from the rest of the protein, except for interactions with the extreme N-terminus; this is consistent with the characterization of the region made from our analysis of Cα contact maps made above. This said, the structure and orientation of the central region is much more variable in the ILDN data set. Similar to OPLS, there are several centroids where the CHC (K16-A21, colored in green) forms β sheets with C-terminal residues (particularly A30-M35, colored with orange). Beyond this common hairpin, however, the top centroids are much more prone to being partially β structured than their OPLS counterparts, particularly at the N-terminus, with β interactions frequently forming between A/T2-R5 (in blue) and G9-H13 (in cyan). Note that α character is present in some centroids, but is very uncommon. Lastly, only 27-38% of the structures are represented by the top 10 centroids despite the somewhat lenient clustering method (see Chapter 2, Methods for more information) used here, indicative of an even more diverse set of ensembles than the OPLS condition (which had values that ranged from 33-46%, and with each value uniformly greater for OPLS than ILDN for each species).

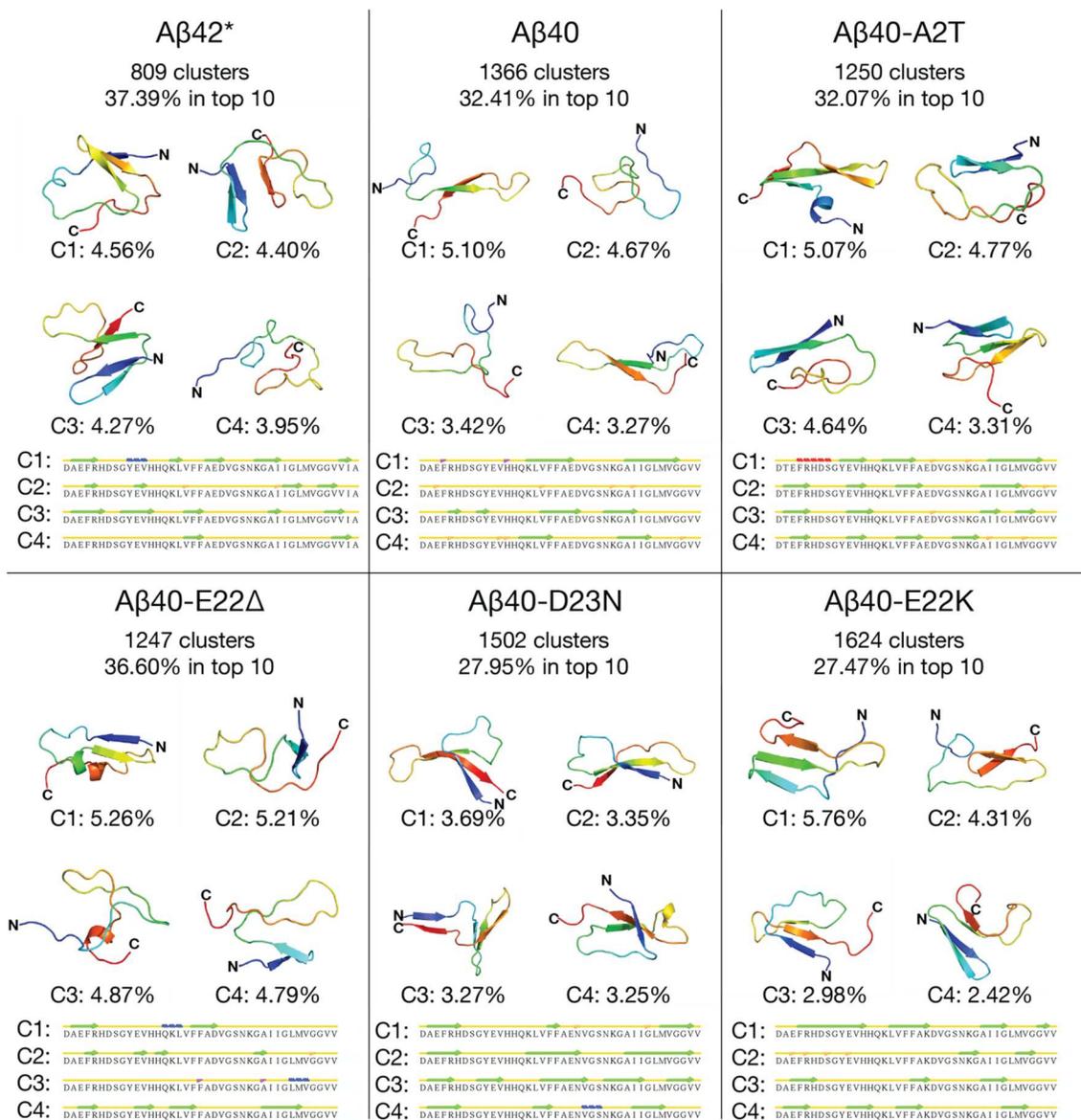


Figure 3.4. The four highest-populated centroids from single linkage cluster analysis(349) around $C\alpha$ atoms with an RMSD cutoff of 2.0 Å for each A β simulation done with the ILDN condition over the production period. Ribbons are colored by a blue-to-red rainbow spectrum from the N-to-C terminus. Secondary structure strips for each centroid, as calculated by STRIDE(361), are also shown. Figure rendered in Pymol(360). The clusters for A β 42's ensemble were calculated over fewer structures (24500 vs. 42000) due to the reduced simulation time, this is denoted with an asterisk.

Wild type A β 40 very rarely involves the N-terminus in β interactions; a feature that is otherwise shared by all of the other ILDN ensembles. Instead, this terminus is isolated from the rest of the peptide in three of the top four centroids. As such, A β 40's structures appear the most similar in character to their OPLS counterparts than any of the other species. Centroid 1, representing 5.10% of the ensemble, is highly reminiscent of the CHC to C-terminal β model derived with the other force field and represented by centroid 3 and 4 of that simulation (C α RMSD for residue 16-35 to centroid 3 is 3.682 Å and to centroid 4 is 3.701 Å).

Further, C α RMSD of residues 16-35 between centroid 1 for ILDN A β 40 and the Affibody-trapped A β 40 monomer fragment determined through solution NMR by Hoyer *et al.* (355) is 2.804 Å, meaning the ILDN force field is also well populated with structures that closely match this experimentally solved topology. Finally, OPLS and ILDN's centroids and the Hoyer *et al.* model possess the same V18/M35, F20/G33 hydrogen bond register across the hairpin. The primary difference is the nature of the intervening central region; ILDN's E22 and D23 side chains interact with solvent, instead of the intrapeptide interactions that stabilize the formation of a crown motif (see Chapter 2, section 2.4.4). Like OPLS, however, A β 40 samples conformations with low RMSD to the Hoyer *et al.* hairpin with ILDN parameters for a significant portion of its ensemble (6.62% within 3Å of residue 16-35)(Figure 3.5). A β 40-D23N, A β 40-E22K, and A β 40-A2T are also prone to sampling these conformations, but A β 42 and A β 40-E22 Δ rarely sample such configurations, in agreement with our observations that these ensembles are less likely to sample extended hairpins at the central region.

In examining the other top centroids for each ensemble, there are other key biases in structure between the different species. A β 42, notably, implicates the extreme C-terminal residues G39-I41 in β strand for all 4 of its top centroid structures, something either relatively uncommon or completely unobserved among A β 40 and its mutants. This permits a complex variety of β strand combinations seen in the centroids, including sheets with three or four strands, seen in centroid 1-3. Additionally, a double hairpin akin to those seen in the OPLS condition existing between the CHC, the A30-M35 region, and residues V39-V40 is present in the 5th most populated centroid

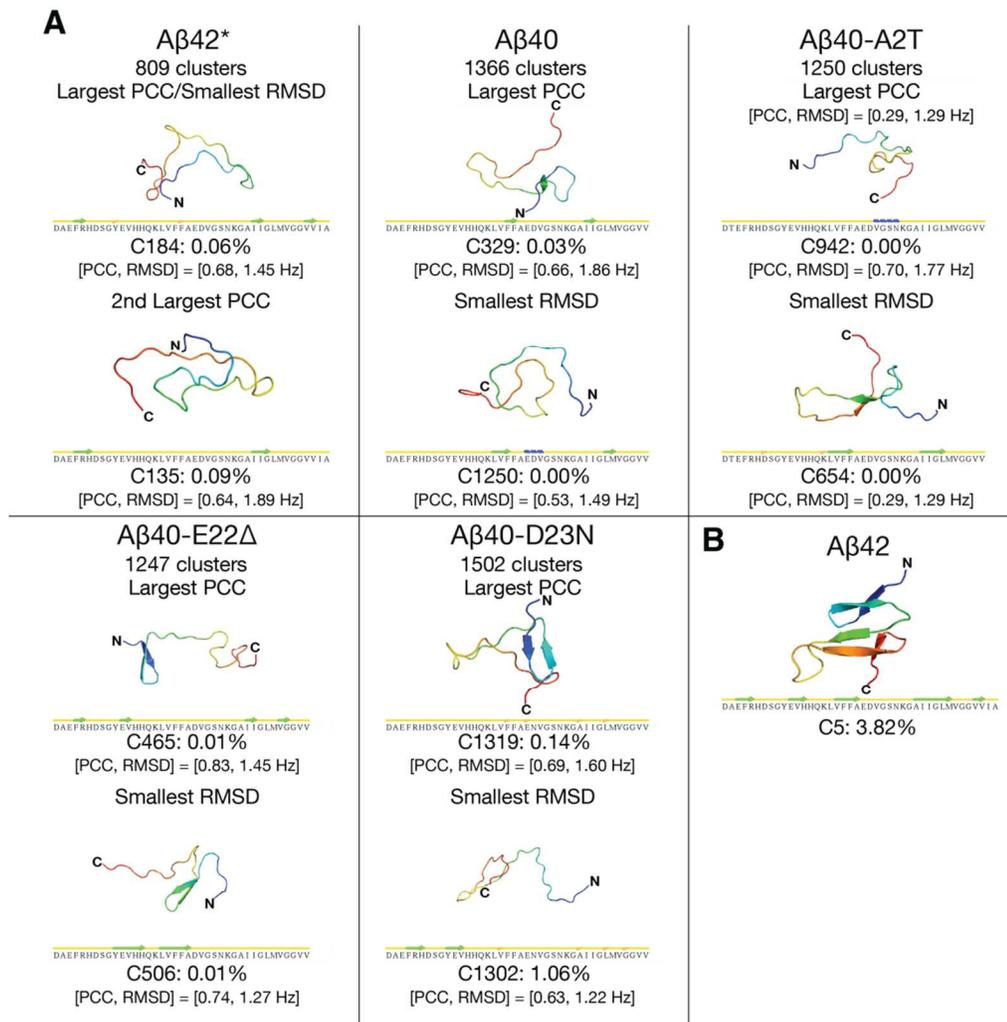


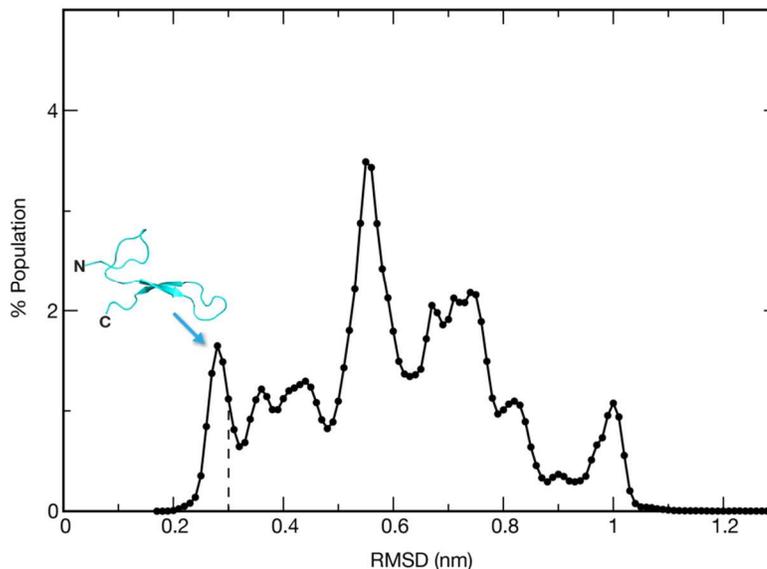
Figure 3.5. Other select centroids from single linkage cluster analysis(349) around $C\alpha$ atoms with an RMSD cutoff of 2.0 Å for each A β simulation done with the ILDN condition over the production period (A) Clusters with back-calculated $^3J_{HNHA}$ couplings (using Vuister and Bax's parameter set(362)) demonstrating the largest PCC and smallest RMSD, respectively, to the inclusive average of the corresponding experimental data sets for each species as discussed in Chapter 4. (B) Centroid 5 from the A β 42 ensemble, demonstrating a double β -hairpin topology involving the C-terminus. Ribbons are colored by a blue-to-red rainbow spectrum from the N-to-C terminus. Secondary structure strips for each centroid, as calculated by STRIDE(361), are also shown. Figure rendered in Pymol(360). The clusters for A β 42's ensemble were calculated over fewer structures (24500 vs. 42000) due to the reduced simulation time, this is denoted with an asterisk.

(Figure 3.6B). A β 40-A2T has a higher likelihood than other species to adopt extended hairpins involving N-terminal residues; these are observed in centroids 2-4. This peptide's top ranked centroid uniquely involves the extreme N-terminus in an α -helix among the centroid structures being discussed here, while all 4 top centroids of A β 40-A2T possess β character at E11-H13. A β 40-D23N and A β 40-E22K have the longest β strands among the top centroid structures, and also have several centroids that adopt β sheets with three strands. Further, three of four of A β 40-D23N's most populated centroid conformations join the extreme N- and C- terminus (blue and red in the ribbon diagram) through a direct β interaction. This N-terminal β strand is present among A β 40-E22 Δ 's centroids, but this species instead lacks the other common A β feature, extended β sheets between the CHC and C-terminus.

ILDN's centroids are, in summary, characterized by a far greater diversity in possible β binding partners beyond the direct CHC to C-terminal combination that dominates the OPLS set, while the structure of the central region is also much more diverse in the types of structures adopted. As such, prevalent electrostatic interactions (such as the role of E22/D23) and patterns of hairpin association (i.e., hydrogen bonding register) are far less consistent among ILDN's centroids than those of OPLS, and we cannot meaningfully generalize those belonging to the former to the same extent as we did for the latter in Chapter 2.

3.4.6 Radius of Gyration and End to End Distance

We observe through the histograms depicted in Figure 3.7A that A β simulations in the ILDN set are, in general, dominated by configurations that range in radius of gyration (R_g) from 0.9-1.3 nm; the distribution has a leftward skew with a long tail on the right side indicative of occasional sampling of more extended structures. This main distribution is considerably broader than the OPLS values. As discussed in Chapter 2, given the peptide length, we expect a radius of gyration of \sim 0.9 nm based on the Kolinski-Skolnick model(363) if A β behaves like a fully collapsed, globular peptide (see Equation 2.1). We conclude that A β monomers sample a mixture of collapsed and extended states. This said, unlike the OPLS set, these distributions are mostly

A**B**

RMSD to Trapped Monomer	% for ILDN Aβ42	% for ILDN Aβ40	% for ILDN Aβ40-A2T
Res 16-35, % within 3Å	0	7	3
Res 16-35, % within 4Å	11	16	12
Res 16-40, % within 3Å	0	0	1

RMSD to Trapped Monomer	% for ILDN Aβ40-E22Δ	% for ILDN Aβ40-D23N	% for ILDN Aβ40-E22K
Res 16-35, % within 3Å	0	5	9
Res 16-35, % within 4Å	7	17	20
Res 16-40, % within 3Å	0	1	1

Figure 3.6. Comparison of ILDN simulations to Chain C, Model 1 from an NMR structure of Aβ40 monomer trapped by the Z_{Aβ3} Affibody protein(355) (PDB: 2OTK). (A) Histogram of C_α RMSDs to this structure for the ILDN Aβ40 simulation over the production period using 0.01 nm-wide bins. Only residues 16-35 for both the reference structure and the simulation were considered for this analysis. Dotted lines guide the eye toward a cutoff of RMSD ≤ 3Å. The top centroid structure after single linkage clustering (0.2 nm cutoff) of structures within this cutoff is illustrated. (B) Table listing the % of frames of ILDN simulations that fall within certain RMSD cutoffs to the NMR structure, over different criteria. 16-35 is the dominant selection of residues for comparison here, in order to discount the effect of the highly flexible C-terminus.

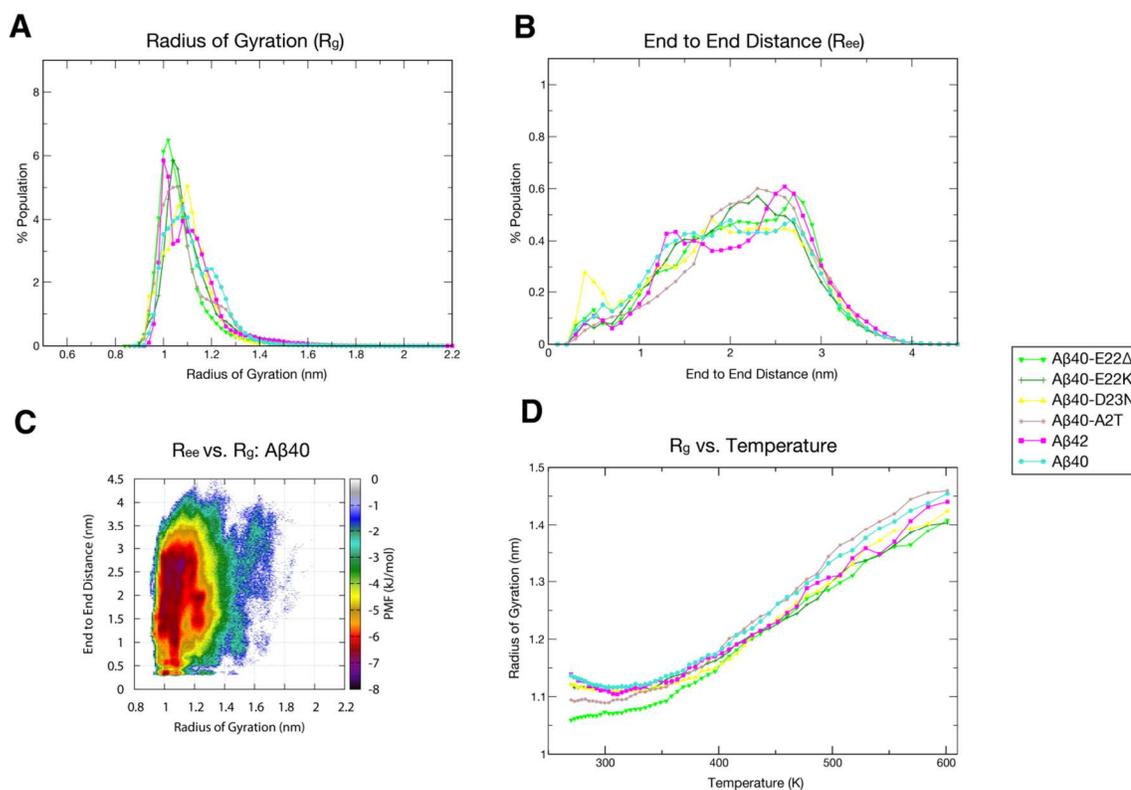


Figure 3.7. Radius of gyration (R_g) and end to end distance (R_{ee}) for ILDN A β simulations. (A) Histogram for R_g over the production ensemble, using 0.02 nm-wide bins. (B) Histogram for R_{ee} , using 0.1 nm-wide bins. (C) Free energy surface along the coordinates of R_g and R_{ee} in units of kJ/mol for the production ensemble of one of the A β peptides, A β 40, using 0.02 nm binning in both dimensions. (D) Average R_g as a function of temperature over all frames in the timeframe 400-1000 ns/replica.

continuous, with no isolated basins. We note that A β 40-E22 Δ has a tendency to sample lower R_g values throughout its whole distribution compared to other species (i.e., a leftward shift in the mean/median of the distribution), while A β 40 and A β 40-A2T possess a similar mode to other species but tend to sample a slightly larger right tail that extends to higher R_g .

ILDN simulations also have a tendency to sample larger end to end distances than their OPLS counterparts, as observed in Figure 3.7B compared to Figure 2.9B. The basin of lowest end to end separations, representing direct interactions between the termini, is much more sparsely populated in the ILDN set. Of the different species in this set, A β 40-D23N notably has the largest propensity to sample this basin; recall, centroids with direct N- to C-terminal interactions are observed for this species alone among the top 4 centroids of each species (Figure 3.4). A β 42's distribution, interestingly, is bimodal, and the larger end to end distances in the second mode may be due C-terminus's involvement in local β hairpins. There is also a relative skew toward larger end to end distances seen in A β 40-A2T and A β 40-E22 Δ . An example free energy landscape using radius of gyration and end to end distance as order parameters is shown in panel 3.7C. In general, ILDN's surfaces are highly reminiscent of those populated in OPLS (see Figure 2.9C), but are somewhat broader in the types of configurations sampled; the most populated basins are also considerably shallower in the former. Lastly, figure 3.7D depicts the relation of average radius of gyration with temperature. Like OPLS, the curve has a sigmoidal character but the temperature only begins to level at the highest temperatures, suggesting the suitability of the REMD temperature range used here. ILDN samples higher R_g at all values, along with a broader range of R_g values compared to OPLS: these are approximately 0.05 nm greater for ILDN at the lowest temperatures, and \sim 0.1 nm greater at the highest temperatures.

3.4.7 Solvent Accessible Surface Area

Figure 3.8 shows ensemble-averaged solvent accessible surface areas (SAS) for the backbone and side chain of each simulation in the ILDN set; the latter has been normalized by the experimentally observed side chain solvent accessibility in model

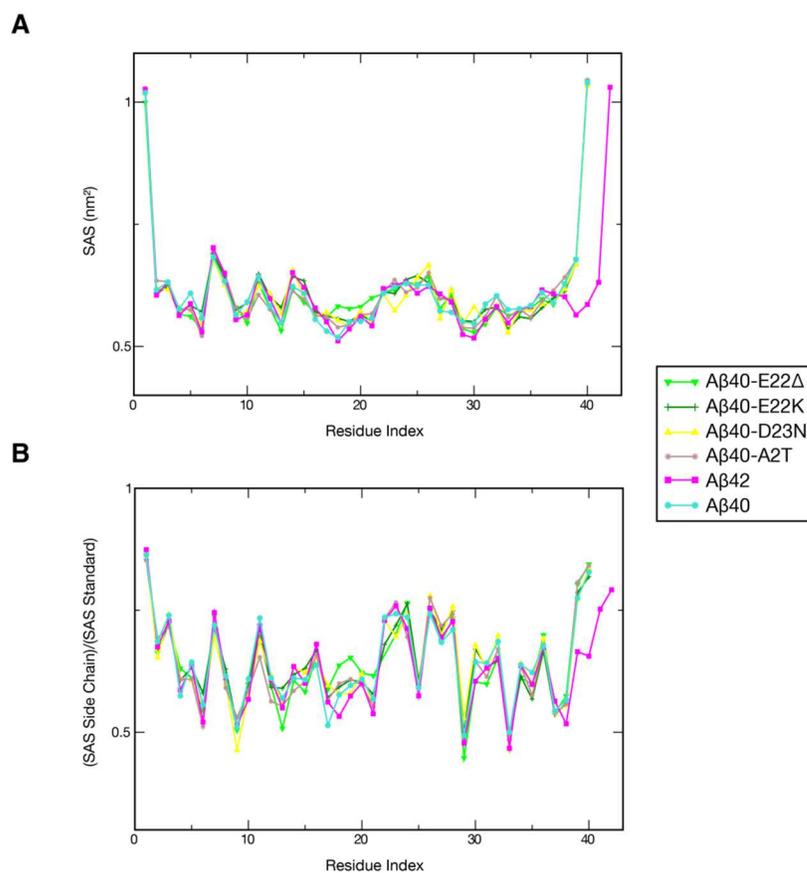
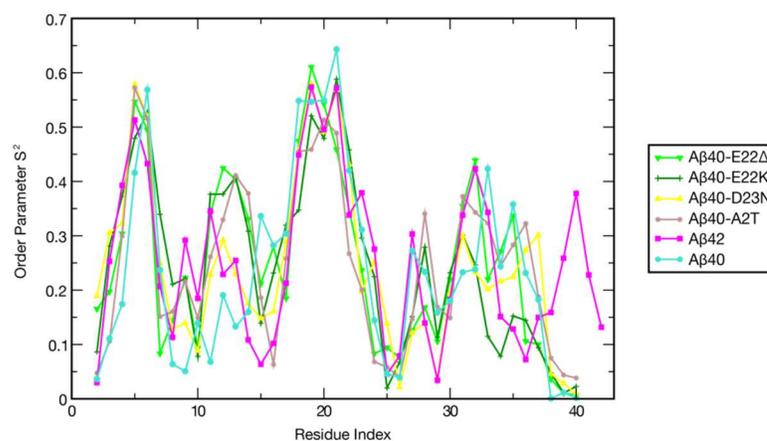


Figure 3.8. Solvent accessible surface (SAS) per residue(354) over the production ensemble of ILDN A β simulations, for the (A) backbone atoms and (B) side chain atoms. The latter is normalized by the standard side chain solvent accessibility for each residue defined by G-X-G tripeptides(365). Error bars are not visible here because error values calculated for these graphs for all points are all very close to zero.

G-X-G tripeptides(365). The pattern of solvent accessibility of the ILDN set is very different from that of OPLS (see Figure 2.10). Notably, the apparent “zig-zag” pattern in this profile that is characteristic of OPLS simulations’ crown motif (see section 2.4.7) is not present; instead the central region (A21-K28) is generally characterized across the ILDN simulations by higher backbone and side chain accessibility compared to the rest of the peptide, along with other turn dominated areas such as D7-S8. In general, SAS profiles are very close to each other across species, even more than observed in the OPLS set. The greatest variation among the FAD mutants is unsurprisingly at site of mutation and the nearby CHC. A β 40-E22 Δ shows the largest deviation, possessing increased backbone and side chain SAS at V18-A21, which may be a consequence of reduced β sheet participation of these CHC residues. A β 40-D23N and A β 40-E22K possess more subtle changes; both increase the backbone exposure of the CHC particularly at V18 as well as residue A30, E22K buries the side chain of the 22nd and 23rd residue, while D23N strongly buries the backbone and side chain of the 23rd residue. The profile of A β 42 is similar to that of the other species, other than a highly increased burial of the backbones and side chains of residue V39 and V40 compared to the 40 residue length species, as well as a more prominent burial of residue V18’s side chain.

3.4.8 Order Parameters

Similar to what was done for OPLS simulations in Section 2.4.8, the isotropic reorientational eigenmode dynamics (iRED) method(353) was also applied to ILDN simulations to measure order parameters (S^2) indicating the angular flexibility of the internuclear vector for each backbone amide group in the system (Figure 3.9). Like OPLS, separability indices are to be 1.3-1.4 for these ensembles, which is at least four fold smaller than that calculated for other simulations of globular proteins(366). Values also range from 0-0.7 like OPLS, further indicating a disordered ensemble. Regions of high S^2 are similar to the OPLS set, but different in terms of intensity. Residues A/T2-D7 and the CHC are now the regions with the highest S^2 ; while A31-M35 also remains a region with large order parameters for most of the simulations, it is less than these other



Separability Index:
 Aβ42 = 1.33; Aβ40 = 1.31; Aβ40-A2T = 1.33;
 Aβ40-E22Δ = 1.33; Aβ40-D23N = 1.32; Aβ40-E22K = 1.33

Figure 3.9. S^2 order parameters for all backbone N-H internuclear vectors (excluding the N-terminus), calculated with the iRED technique(353) over the production period for ILDN simulations. The Prompers-Brüschweiler separability index between the internal modes and global modes is also provided for each species; compare to an index of 5.39 for ubiquitin in bulk water(366).

two regions. N-terminal rigidity is more pronounced in the ILDN set, with residue E11-H14 forming a second hotspot seen only in ILDN simulations (particularly A β 40-A2T and the two E22 mutants). This hotspot is also not seen in ILDN's A β 40 simulation, which, as we have seen through other analysis (Figure 3.2 and 3.3), does not form local β hairpins in this region. An A β 42-specific hotspot is also formed at the C-terminus (V39-I41), suggesting this C-terminal rigidity is a generalizable feature of this alloform. Like OPLS, the central region and the extreme termini form the nadirs of the graph. There is some variation in the central region (particularly the roles of N27 and K28), but the large peaks that were implicated in "crown" motif formation are conspicuously absent.

3.4.9 High Pressure Simulation of A β 40

A β 40 was also simulated under identical REMD conditions, but where the box was equilibrated to 2500 bar (and 300 K) as opposed to 1 bar (and 300 K) before switching to the NVT ensemble. This investigation complements the experimental work described in Chapter 4, where high pressure NMR spectroscopy was used to investigate which regions of the intrinsically disordered A β monomer ensemble are pressure-sensitive and thus may be biased toward structure in ambient conditions. A summary figure of the effects of high pressure in this simulation, using different types of analysis we have discussed previously in this chapter, is illustrated in Figure 3.10. The high pressure simulation is characterized by reduced β sheet interactions between the CHC and the C-terminus, particularly at residue L17-E22, N27-K28, and I31-G33, as evidenced by reduced contacts on the main diagonal in panel Figure 3.10A and reduced DSSP annotation in panel Figure 3.10D, and a concurrent increase in coil/irregular structure. Backbone and side chain solvent accessible surface area of these residues is increased (*data not shown*), while electrostatic interactions involving residues E22 and D23's side chains within the central region are uniformly reduced (for example, the salt bridges between either E22 or D23 and K28 are both populated 2% in the high pressure condition, but are both populated 5% in the ambient pressure condition). Turn and 3-helix character in the central region is also exchanged for coil structure, particularly at

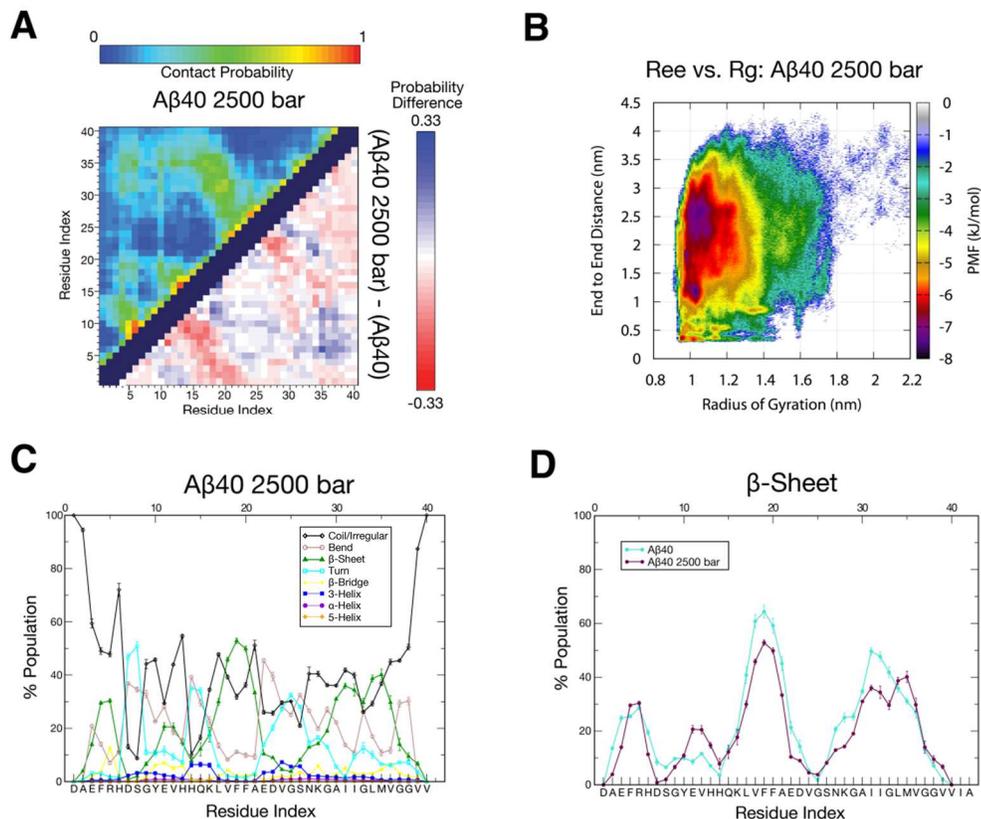


Figure 3.10. Analysis of Aβ40 at 2500 bar simulated with ILDN, using methods covered previously in the chapter. (A, upper corner) Map illustrating the probability of contact over the production ensemble for Aβ40 at 2500 bar. (A, lower corner) Difference between this contact probability map for Aβ40 at 2500 bar compared with Aβ40 at 1 bar (see Figure 3.3). (B) Free energy surface along the coordinates of R_g and R_{ee} in units of kJ/mol for the production ensemble of Aβ40 at 2500 bar, using 0.02 nm binning in both dimensions, compare with Aβ40 at 1 bar shown in Figure 3.7C. (C) Percent composition of secondary structure as determined by DSSP(347, 348) for Aβ40 at 2500 bar, averaged over the production ensemble. Compare to Figure 3.2B. (D) Percent composition of β-sheet alone for Aβ40 at 2500 bar compared to the 1 bar condition. The sequence of wild type Aβ42 is presented in the lower x-axis.

V24-S26, consistent with a loss of central hairpin character. Conversely, the high pressure simulation is also characterized by increased β content in N-terminal residues E11-H13 and extreme C-terminal residues. This aligns with decreased contact between D1-Y10 and the CHC in favor of more global contacts, such as those between F4-E11 and M35-V39, similar to the effects of certain FAD mutations like A β 40-E22 Δ . In general, high pressure does reduce overall β content of the ensemble and replace it with coil/irregular, but only marginally ($22.3\pm 0.8\%$ β content at 1 bar, $19.2\pm 0.8\%$ at 2500 bar). Interestingly, the residues where β character is reduced the most do qualitatively align with the regions characterized by high pressure sensitivity, as we will see in the next chapter.

High pressure also promoted a more diverse free energy landscape, as seen in Panel 3.10B compared to the 1 bar condition depicted in Figure 3.10C. In general, the 2500 bar state favored more states with more extreme lower and higher radii of gyration (R_g), that is, an ensemble with a broader distribution of R_g values. The observations that single linkage clustering performed over the high pressure ensemble, using the same parameters and number and type of input frames as the 1 bar condition, resulted in many more clusters (2029 in the high pressure case, 1366 in the ambient pressure case) and had the top 10 centroids representing a smaller percentage of the ensemble (27.73% at 2500 bar vs. 32.41% at 1 bar) both corroborate this assertion.

3.5 Discussion and Conclusions

In this chapter, we present an alternative characterization of A β simulations to that made in the previous chapter, using a different all atom force field and water model, while keeping all other parameters constant. These characterizations are important because, while many computational simulations have been performed on this system (reviewed in Chapter 1), there are many inconsistencies in the observations made through these studies. We implicate the choice of force field as one of the primary (but not the only) source of divergence between these investigations (other sources are discussed in Chapter 1 and 2). While simulation with state of the art force fields have shown significant reproducibility in determining the folded state of ordered proteins, large divergence has been observed in simulations of intrinsically disordered proteins

and even the unfolded states of globular proteins(28, 46). These cases lack unambiguous energy minima, so even subtle changes in the energy function can have profound effects on the final thermodynamic distribution of states. With this in mind, we suggest that the validation of results through cross-force field investigations be an important consideration before making conclusions about intrinsically disordered systems through *in silico* simulation, along with the factors stressed in the previous chapter: simulation on long timescales using an explicit representation of the chemical groups, and validation with experimental data.

Interestingly, the simulations conducted with the AMBER99sb-ILDN force field and TIP4P-Ew (“ILDN”) combination here are characterized by very similar secondary structure profiles and C α contact maps to the simulations described in the previous chapter using OPLS-AA and TIP3P (“OPLS”) parameters, suggesting a generalizable characterization of A β monomers. In particular, β -hairpin topologies sampled between the CHC and the A30-G33 regions are prevalent in the ensemble, while the A21-G29 region is characterized by an intervening loop structure that rarely contacts any of the residues in the protein except those at the extreme N-terminus. These are prevalent interactions in all ensembles, with the notable exception of A β 40-E22 Δ . Further, A β 42’s extra two residues induce the formation of a second hairpin involving V39-I41, with a loop at G37-G38, and characterized by a prevalent L34 to I41 contact. As discussed in Chapter 2, these secondary structural biases are consistent with experimental NMR chemical shift indices(4) and J-couplings, while these sampled topologies are reminiscent of the intramolecular models of higher order aggregates including fibrils (23, 24, 27), and in the case of A β 42, a toxic, pentameric oligomer believed to be a paranucleus specific to A β 42 aggregation (see Chapter 1 for more information about these aggregates)(26). Further, both A β 40 simulations sample a high percentage of configurations with very low RMSD to an Affibody-trapped A β 40 monomer structure solved by solution NMR(355) (see Figure 2.7 and Figure 3.5); and indeed, both simulations prevalently populate backbone hydrogen bonds in the same V18/M35, F20/G33 register observed in this experimental structure (Table 2.1 and Table 3.1). This provides additional evidence that the hairpin topology represents a transiently occupied structure that unbound A β is prone to sample, and that can interact with other peptides.

Like the OPLS set, this β character in ILDN simulations does not equilibrate until several 100 ns/replica, if at all (Figure 3.1). Similar β topologies have been observed through some of the more recent, extensive, and experimentally refined or validated simulations of A β 40 and A β 42 produced by other groups (9, 11, 17, 18), but we are the first to point out these consistencies among simulations of different force fields. Most importantly, these observations are not only commonalities between simulation sets, but are also backed by experimental data; we therefore have some confidence in these biases as general descriptions of the A β system.

This said, there are key differences between the OPLS and ILDN sets; causing us to view these particular details with greater skepticism. OPLS simulations significantly favored more collapsed states, and intramolecular contacts are populated with higher probability. This said, the global secondary and tertiary biases of A β are quite similar between the sets. Rather, it is the local interactions that prevail between each set that are considerably different. First, while both simulations form loop interactions in key regions (D7-G9, A21-G29, and G37-G38 in A β 42), OPLS is prone to form very rigid bend structure at these positions, stabilized by strongly populated charge-charge interactions. Other types of loop structure, like turn and coil, are rare in wild type A β . Meanwhile, ILDN simulations demonstrate much more flexibility in these regions, with a greater tendency to sample a mixture of bend and turn secondary structures, the latter of which is characterized by medium range, helical backbone hydrogen bonds. The most prominent example of this disparity is the central region of A β . As we saw in the previous chapter, this region was dominated by structures that established a “crown” motif – a bulge of outwardly pointing carbonyl groups stabilized by hydrogen bonds formed by D23’s side chain with the inwardly-drawn backbone amides of the downstream 4 residues, and a salt bridge between E22 and K28. In contrast, these interactions are not strongly populated in ILDN simulations; instead, E22 and D23’s side chains interact with solvent much more frequently. Consequentially, the backbone amides of the central region are not locked into interactions with the D23 side chain carboxyl, while the backbone carbonyls are not forced to point outward to accommodate the “crown” structure. Thus, these central region amides and carbonyls are freer to form helical hydrogen bonds, which are especially well populated between D23/S26 and

V24/N27. Experimentally, it has been previously found that there is a i to $i+3$ α N nuclear Overhauser effect (NOE) constraint between D23 and S26 in A β 40, the longest range unambiguous backbone NOE observed in any monomeric A β (4). Additionally, the D7-Y10 and F20-A30 regions are also the only regions observed to have i to $i+2$ backbone NOEs(4). The presence of helical hydrogen bonds characterizing these regions in ILDN, rather than the pure bend regions of OPLS, may be more consistent with these findings.

Another major difference between the two force fields is the role the N-terminus plays in A β structure. The N-terminus is conspicuously absent or poorly resolved in both the A β 40 trapped monomer solution NMR structure and solid state NMR structures of A β fibrils (23, 24, 27, 176, 251, 355), suggesting that it remains disordered even when the rest of the peptide is stabilized. Further, the N-terminal sequence, like the central region but unlike the rest of the peptide, contains many charges, and is thus most IDP-like in character. It is unsurprising that the two force fields, which emphasize electrostatics in fundamentally different ways (see the Introduction, section 3.1), would also disagree on the structure of this region. OPLS characterizes the N-terminus with a mixture of structures, but primarily coil and bend character, in contrast with the highly structured C-terminus. On the other hand, coil, bend, and turn are prevalent in ILDN simulations, but β character is also highly populated, particularly at E3-R5 and Y10-H13, and particularly in all ensembles except wild type A β 40. Very often, these regions directly form local β sheets, but more global β pairing does occur with other regions of the protein. The N-terminus is also particularly prone to contacting the rest of the protein in OPLS simulations, even at the central region, but it rarely does this through β interactions. The involvement of these regions in β pairing is evidenced by the presence of topologically diverse β structures seen in the top centroid structures of ILDN simulations, but not seen in OPLS simulations where the central hairpin, if anything, is sampled by these structures.

OPLS featured strongly populated electrostatic roles for E22 and D23 that stabilize the “crown” motif structure. Unfortunately, the physicochemical influences of E22 and D23 are not as clear in the ILDN set. These residues do form some of the same salt bridges and hydrogen bonds, but the populations of these interactions are greatly reduced

in population compared to OPLS, in exchange for more frequent interactions with solvent. Consequentially, the FAD mutants of A β that we studied, which target these two residues, have considerably different effects on the A β ensemble for each force field. OPLS saw well defined changes in the central region bend structure associated with key changes in the population of electrostatic interactions, and these biases in the structure changed the binding patterns of hairpins sampled in the ensemble. The effects of these substitutions are not so prevalent among the ILDN set. It is still true that FAD mutations change the frequencies of interactions in the central region, and the distribution of turn region is narrowed while residue 22 and 23 in these systems are more prone to adopting β structure, but these changes are relatively subtle.

More obvious among the differences common to the point mutants of A β studied here and wild type A β 40 unique to the ILDN set is that the N-terminus is more active in forming both local and global β hairpins in the former, as opposed to more localized turn and bend structure in the latter. This behavior was expected for the A2T variant as threonine is a residue far more inclined to adopt β structure than alanine(387, 388), but the mechanism behind this change for FAD mutants of the central region is not as clear. Consider, however, that both the N-terminal region and the central region in question are net negatively charged. These FAD mutations share in common that they reduce the amount of net negative charges in the central region (or in certain mutants like H6R or D7N, the N-terminus itself), so one could speculate that by mutating these residues, we are removing repulsive interactions from the system that would otherwise force the N-terminus to avoid interactions with these regions. This causal hypothesis is challenging to test, however. Interestingly, A β 40-E22 Δ has the most profound effect of the FAD mutants, in that it is the only variant to greatly reduce the β content at the CHC and the C-terminus. This was seen in OPLS simulations as well, although there, the presence of more C-terminal character (M35 and V36) made up for the loss of β character in residues like I31 and I32, which is not true for the ILDN variant. Instead, the heightened N-terminal β character (common to all ILDN FAD mutants, but not their OPLS equivalents) serves as the primary source of β interactions in this peptide.

As discussed above, the increased interactivity of the N-terminus in β pairing seen in these mutants leads to the possibility of more diverse β topologies. One could further

speculate (but, again, find a great challenge in testing) that the presence of these newly allowed β combinations in the monomer ensemble may allow for the seeding of different aggregates that could either account for the novel higher order forms and/or faster aggregation pathways that define the phenotypes of these mutants. This hypothesis is a different alternative to the one proposed in the previous chapter, which suggested the phenotype of FAD mutants may arise from the broader range of registers adopted by the central hairpin seen in OPLS simulations. Since neither of these hypotheses are based on observations that are common to both force field combinations, we cannot endorse one over the other at this time. Further, since A β 40-A2T contributes a similar effect as the FAD mutants on the involvement of the N-terminus, it is also unclear how (or even if) this mutant may confer changes in monomeric structure that lead to its neuroprotective phenotype (though the similarity in behavior would be in line with the finding that this mutant actually enhances A β 40 aggregation(379)).

This said, recently, REMD simulations up to 175 ns/replica with the OPLS-AA/TIP3P force field comparing this A2T variant with wild type as well as the A2V FAD mutant in the context of A β 42 were published(389) that may clarify the involvement of the N-terminus in A β aggregation. Like the findings described here for our A β 40-A2T simulation, it was observed that this point mutation encourages the N-terminus to engage distant regions of the peptide. However, in the context of A β 42 in the Das *et al.* study, it was observed that this behavior by A2T discouraged the formation of double β hairpin conformations involving the CHC and C-terminus (matching those prone to be adopted by wild type A β 42 in the OPLS and ILDN simulations described in the previous and current chapter) in favor of single hairpins at the C-terminus, while the A2V mutant encouraged sampling of these meander structures(389). As described above, these double hairpins are unique to A β 42 simulations and bear strong resemblance to a solid state NMR model of neurotoxic A β 42 oligomers(26).

As we will see in the next chapter, the overall similarities between the ILDN and OPLS simulations make it very challenging to find direct evidence to discount one model over the other. Both simulations yield a good comparison to NMR data, which almost exclusively report on secondary but not tertiary structure biases. This said, while still imperfect, the ILDN model seems to be the better representative of experiment

because the profound changes in structure predicted by the OPLS simulation caused by FAD mutants (particularly A β 40-D23N) are not observed experimentally. As neither model is conclusively correct in explaining the experimental data and there are large gaps in our understanding of these systems based on experiment alone, we value those properties that are consistent among these different investigations. Fortunately, the descriptions that fit this criterion are sufficient to develop a generalized depiction of wild type A β monomers. The agreement between these different techniques is an important and non-trivial finding: we will see in Chapter 5, for example, that REMD simulation of a small peptide from α -synuclein using the very same force fields employed in our study of A β here nevertheless produce very different ensembles for that system, and that neither force field is able to accurately reproduce $^3J_{\text{HNHA}}$ couplings for this system. We describe A β peptides here, in summary, as disordered proteins that are nonetheless biased to forming β topologies that bear resemblance to the intrapeptide models of higher order aggregates. As such, we suspect that the shared conformations presented here for A β may act as the seeds for the formation of these aggregates, and thus, potential targets for drug discovery in Alzheimer's disease.

4. NMR Experiments on A β and Experimental Validation of REMD Data

4.1 Abstract

Amyloid β (A β) is a disordered peptide whose aggregates are implicated in the pathogenesis of Alzheimer's disease; however, the mechanism behind the self-association of these peptides is poorly understood. The simulations in Chapter 2 and Chapter 3 have yielded plausible conformations sampled by A β monomers that could be potential seeds for the formation of similarly structured, higher order aggregates. In this chapter, we assess if these simulation results are experimentally relevant through qualitative and quantitative comparison to observables for each residue collected through solution nuclear magnetic resonance (NMR) spectroscopy for different A β alloforms (A β 40 and A β 42) and mutants (E22 Δ , E22G, D23N, and A2T mutants of A β 40). Chemical shifts, J-couplings, solvent exchange rates, and pressure coefficients (measures of the sensitivity of chemical shifts to pressure) reveal a consistent model for A β that agrees with the β hairpin ensemble biases observed through simulation. Further, good quantitative agreement is observed through direct comparison of experimental chemical shifts, J-couplings, and residual dipolar couplings to equivalent values back-calculated over the simulated ensembles. Nevertheless, the long range changes in the A β ensemble seen for certain mutations through simulation do not agree with the primarily local changes observed experimental NMR values. Further, for A β 40-D23N compared to wild type, the OPLS-AA/TIP3P model consistently models large changes for several different observables in the region near the mutation site that anticorrelate with the experimental trends, suggesting that the local electrostatics at this region may be an artefact of the force field. These inconsistencies aside, the solution NMR data presented in this chapter provide an atomic resolution characterization of different A β monomer forms, revealing secondary structure biases in the ensembles that are consistent with the models produced through simulation.

Portions of this chapter previously appeared as D. J. Rosenman, C. R. Connors, W. Chen, C. Wang, A. E. Garcia, A β monomers transiently sample oligomer and fibril-like configurations: ensemble characterization using a combined MD/NMR approach. *J. Mol. Biol.* **425**, 3338 (2013).

4.2 Introduction

The previous two chapters describe simulations that have revealed important potential biases and metastable states in the ensembles of monomeric amyloid β ($A\beta$) variants that may underlie the differential aggregation of these peptides. However, it is important to run calculations in tandem with experiments that can yield comparable data, for two reasons: to test the validity of the computations at hand and to facilitate the improvement of the theory underlying the simulations for future calculations. As discussed in more detail in the first chapter, intrinsically disordered peptides (IDPs) like $A\beta$ represent an extreme test case for the robustness of our simulation methodology due to their inherent lack of a definitive global energy minimum. As such, we suggest that validation through experiment is of paramount importance in dealing with these IDPs.

4.2.1 Use of Solution NMR in the Study of Intrinsic Disordered Peptides

Solution nuclear magnetic resonance (NMR) spectroscopy is one of the few experimental techniques that are well suited for the study of the structural biases of soluble IDPs at atomic resolution. NMR yields information about protein structure and dynamics at the atomic level. X-ray crystallography also could provide atomistic structural information, but this method often fails for IDPs as their dynamics typically preclude the packing of the protein into a well-ordered, isomorphous lattice. Furthermore, solution NMR can be used to investigate proteins in native buffer conditions. This is important for IDPs because we would expect their ensembles to be particularly sensitive to changes in solution chemistry due to their flexible nature. NMR yields several observables that can be compared to simulation. Chemical shifts are sensitive to the surrounding electronic environment of nuclei, while through-bond J-coupling constants report information on the connectivity of the chemical moieties and constraints on torsion angles; both of these observables are often used as reliable reporters of secondary structure^(390, 391). Reporters that are more global in nature exist and are instrumental in the calculation of structures through NMR, such as through-space correlations produced by the nuclear Overhauser effect (NOE) and splittings through residual dipolar coupling (RDC) which yield information about the angular orientation of bonds relative to the overall alignment of the peptide^(392, 393), though

these experiments are typically more challenging to interpret and analyze. Further, NMR can directly probe the dynamics of nuclei on the picosecond to nanosecond timescale because these motions affect relaxation parameters, such as R_1 , R_2 , and steady-state heteronuclear nuclear Overhauser effect values (ssNOE)(394, 395). This is particularly helpful for studying intrinsically disordered proteins like A β , which are known to substantially sample motions on the nanosecond timescale (and faster) that influence these relaxation parameters(375). Longer timescales can be probed with other observables: $R_{1\rho}$, CPMG, and ZZ-exchange spin relaxation measurements are sensitive to molecular motions and chemical exchange processes that occur on the microsecond to millisecond timescale(396), while amide proton exchange experiments (for example, H/D exchange and CLEANEX-PM(223, 224)) and saturation transfer experiments measure phenomena that occur on the millisecond and greater timescale(395, 397). Finally, solution NMR is also suited to studying the low molecular weight forms of small amyloidogenic peptides like A β because higher order aggregates may be NMR-invisible due to the slower tumbling time of these forms. Indeed, disaggregated A β peptides (see Methods section for more details) show a single set of resonances that almost entirely (>90%) represent the monomeric state in solution(337). No other experimental technique can selectively detect monomer-specific signals under native conditions with such fidelity, which is extremely important because most forms of A β start aggregating immediately under physiological buffer conditions(398).

On the other hand, there are some limitations in the use of NMR to study IDPs. First, the timescale for NMR acquisition is fairly slow (for a 10 kDa protein, it would be typically be at minimum 250-500 ms to minimize artefacts (394)), and the fast interconversion between states of an IDP typically put it in the “fast exchange dynamics” regime(399, 400). As such, the NMR observables described above typically represent a single ensemble-averaged value for each resonance(394); this is a benefit because spectra become less overlapped and easier to assign, but also a weakness because individual states become difficult to study. Secondly, due to the weak structural typing of an IDP, the observed time-averaged chemical shifts are less disperse and are therefore highly prone to spectral overlap, which complicates assignment and analysis of the data(401, 402). Additionally, amide hydrogens – the most common detection nucleus in

multidimensional NMR – are particularly prone to solvent exchange with bulk water in IDP peptides that lack a hydrophobic core, leading to low signal to noise ratios of spectra that use these nuclei(401, 402). Lastly, despite an ongoing effort to design techniques to surpass these limits, most solution NMR methods are only suitable for studying globular proteins up to 40 kDa due to slow tumbling leading to signal broadening(403, 404). IDPs usually tumble faster due to a lack of structure, so the typical monomeric size limit can be surpassed (400, 405), but the size limit still makes investigation of higher order aggregates, like large oligomers and fibrils, impractical or impossible with solution NMR.

In this chapter, we have employed NMR to investigate the biases in structure and dynamics of different monomeric A β peptides and validate the simulations described in the previous two chapters. Our lab has extensive experience overexpressing and purifying recombinant, monomeric A β ; we have expanded upon previously published protocols(406) to improve product yield. In addition to the previously studied A β 40, A β 42, and A β 42-M35ox forms (for example, in (398, 407-410)), we have created samples and acquired experimental observables for variants of A β not previously studied through solution NMR, including the E22 Δ , E22G, and D23N familial AD (FAD) mutants of A β 40 (more information on these species can be found in Chapter 1) and the neuroprotective point mutation A2T (discussed in Chapter 3). For these variants, we present a collection of chemical shift data, residual dipolar couplings, $^3J_{\text{NHb}}$ couplings, and $^3J_{\text{HNHA}}$ couplings, the latter measured through the J-resolved SOFAST-HMQC(411) pulse sequence. These data are then compared to their counterparts derived from the REMD simulations. We find that there is good agreement between experiment and simulation overall, with both force field combinations (OPLS-AA(19)/TIP3P(20) and AMBER99sb-ILDN(21) /TIP4P-Ew(22), called "OPLS" and "ILDN" here, as described in Chapters 2 and 3 respectively) generating fairly similar correlation values in general.

4.2.2 Perturbations and High Pressure NMR

Overall success at matching chemical shifts and J-couplings are an important first step, however, this good agreement may purely stem from the simulation recapitulating the correct local chemistry of the residues, rather than agreement in higher order (ie,

tertiary) structure. As such, we go beyond direct comparison and look at how these observables change through different perturbations of the A β ensemble. One such change would be that of mutation; i.e., how these observables change across different A β species. Since the sequences are highly similar and data was collected under identical conditions, this minimizes the sequence-dependence and impact of systematic errors of the comparison. We find that, particularly between the D23N mutant and wild type A β 40, the OPLS model predicts strong changes that are not present in the experimental data. ILDN performs significantly better than OPLS at this more stringent comparison, but even this model is fairly weak at picking out these subtle species-specific differences.

Other perturbations examined in this chapter are the effects of pressure and the addition of urea on A β resonances. Both of these changes are well known to denature globular proteins, though in fundamentally different ways: urea typically biases the ensemble to states with larger exposed solvent accessible surface area(412-414), while pressure often favors more collapsed states that occupy lower volumes(415). However, these perturbations have also been used to manipulate the metastable states of IDPs and control the aggregation of these peptides(54, 410, 416-420).

Indeed, high pressure has previously been used to investigate structural biases and define distinct states of the intrinsically disordered A β 40 ensemble(410). Here, we supplement this work with an investigation into the effects of pressure on five different A β variants (A β 40, A β 42, A β 40-E22G, A β 40-D23N, and A β 40-A2T). We find that the sensitivity of A β residues to pressure correlates to the secondary structure proclivities seen in simulation, and these map to changes seen explicitly in high pressure simulations (see Chapter 3). Interestingly, A β 42 changes the movement of chemical shifts with pressure for C-terminal residues in a way that is consistent with a bias toward sampling a new C-terminal turn at ambient pressure. Further, the two FAD mutants investigated here change the sensitivity of chemical shifts to pressure in the central region, but in a fairly subtle way that is inconsistent with the profound changes predicted by the OPLS model discussed in Chapter 2.

4.3 Methods

4.3.1 Overexpression and Purification of A β Peptides

Isotopically labeled samples of A β are required for multidimensional solution NMR experiments. ^{15}N labeled recombinant HFIP-treated A β 40 and A β 42 peptides were ordered from rPeptide LLC (Bogart, GA, USA), but other variants of A β not provided by this company at the time of this work, as well as $^{13}\text{C}/^{15}\text{N}$ labeled peptides, were produced in lab. The development and execution of the following protocol for overexpression and purification of recombinant A β monomers was greatly assisted by Nicolina Clemente and Dr. Christopher Connors; a similar protocol is outlined in Dr. Connors's Ph.D. dissertation(54), but it is also documented here for the reader's convenience. A diagram of this procedure is also illustrated in Figure 4.1.

This methodology is adapted from Glockshuber *et al.*(406); deviations from this protocol were implemented to improve peptide yield. The wild type A β 40 plasmid used for overexpression, originally provided as a gift and described by Glockshuber *et al.*(406), encodes an N-terminal 6xHis tag, a (NANP)₁₉ repeat sequence included to improve solubility, a site for TEV protease digestion (RSENL YFQ), and finally, a codon-optimized wild type A β 40 sequence. Vectors encoding D23N, E22 Δ , E22G, and A2T mutants of A β 40 were generated through site-directed mutagenesis by GenScript USA Inc. (Piscataway, NJ, USA). All plasmids were transformed into BL21(*DE3*) *Escherichia coli* cells. The culture was used to make cell stocks in 15% glycerol, which were then stored at -80°C. Success of the mutagenesis and transformation procedures were confirmed by purifying plasmid from cultured cells derived from the stock through Miniprep and then sequencing the product, a service performed by Eurofins MWG Operon (Huntsville, AL, USA).

The medium used for overexpression of labeled peptides is OD2 growth medium with the desired labeling (either ^{15}N or $^{13}\text{C}/^{15}\text{N}$), provided by Silantes GmbH (Munich, Germany). This medium was chosen to avoid reductions in yield observed with standard minimal media such as M9. Unlabeled peptides were instead prepared using Luria broth (LB) medium. Overnight cultures were prepared from the cell stocks using 50 mL of the appropriate medium with 50 $\mu\text{g}/\text{mL}$ fresh ampicillin and incubating at 37°C and 225

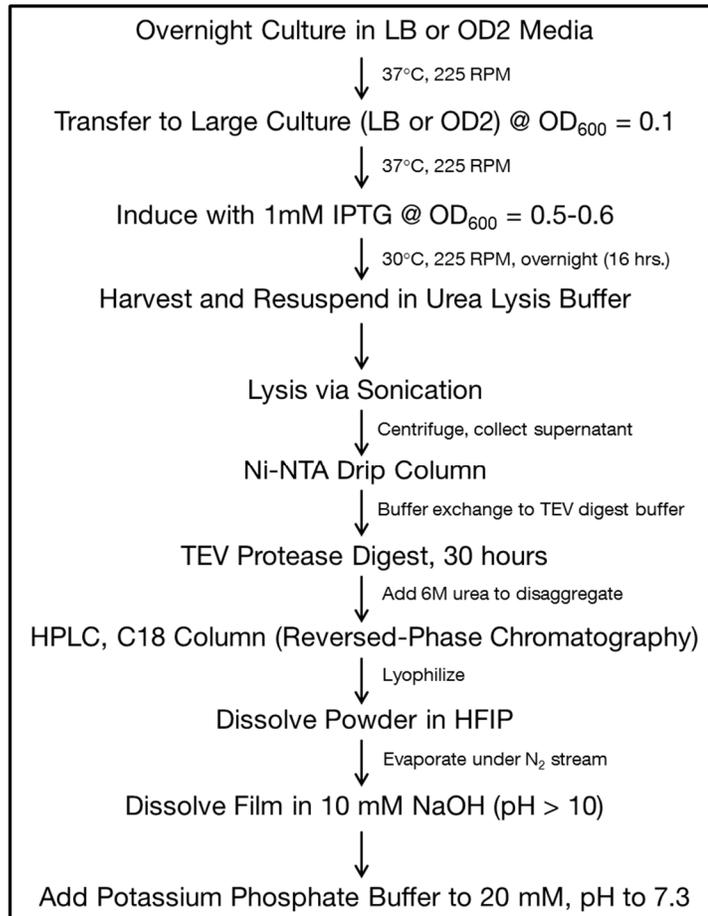


Figure 4.1. A simplified flowchart for the A β monomer sample preparation protocol presented in this chapter. More detail can be found in the text.

RPM. Cells were transferred directly into a larger volume of the same medium (also containing 50 µg/mL fresh ampicillin); the volume of this transfer was selected to start this overexpression culture with an OD₆₀₀ of 0.1 measured through absorbance spectrophotometry. This culture was again incubated at 37°C and 225 RPM until an OD₆₀₀ of 0.5-0.6 was measured, then isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a total concentration of 1 mM to induce expression of the protein. Following this step, the culture was incubated at 30°C and 225 RPM for 16 hours. Cells were then harvested from the medium by centrifugation at 4000 RPM for 30 minutes, and either frozen at -80°C or used immediately for subsequent purification.

Cells were thawed in a water bath (if need be) and resuspended in a “urea lysis buffer” (8 M urea, 100 mM potassium phosphate buffer (using a dibasic K₂HPO₄ stock), pH 8.0, hereafter called ULB); 5 mL ULB was used for each 1 g of cells. Lysis through sonication was performed by using a sonicator with a microtip at 35% power with 30 seconds on and 1 minute off for a combined total of 4 minutes “on” time, while keeping the sample on ice. This sonication step was repeated twice, and fresh ice was added in between steps. The soluble components of the lysate were then removed from the insoluble pellet through centrifugation at 20000 RPM for 45 minutes. Nickel sepharose beads obtained from GE Healthcare Biosciences (Pittsburgh, PA, USA) were washed three times in ULB, and then nutated with the supernatant of the lysate obtained above for at least one hour. The bead mixture was then added to a drip column and the flowthrough was discarded. The column was further washed two independent times with 3 column volumes of ULB each time. Finally, the column was eluted with 3 column volumes of “urea elution buffer” (8 M urea, 100 mM potassium phosphate buffer (using a dibasic K₂HPO₄ stock), pH 8.0, 500 mM imidazole), which was carefully collected for the future steps.

After confirming the NANP-Aβ fusion protein was present by SDS-PAGE, the elution fraction was dialyzed overnight with “TEV digest buffer” (50 mM Tris-HCl, 0.5 mM EDTA, 30 mM NaCl, 1 mM DTT, pH 8.0). After assessing the appropriate activity concentration ratio through test cleavage of a NANP-MBP fusion protein, TEV protease purified in lab was added to the buffer exchanged sample. The sample was nutated at

4°C, allowing the cleavage reaction to occur for ~30 hours. Any precipitate formed by the digest was then dissolved through addition of 6 M urea.

Confirming success of the digest reaction through SDS-PAGE, reversed-phase chromatography was used to separate cleaved A β from uncleaved fusion protein, NANP tag, TEV protease, and other contaminants. The Shimadzu HPLC (Kyoto, Japan) was equipped with a Zorbax SB-C18 semi-preparative RP-HPLC column, obtained from Agilent Technologies (Santa Clara, CA, USA). The column was heated to 80°C, washed with a buffer containing 10% acetonitrile and 0.1% trifluoroacetic acid (TFA), and loaded with the sample containing A β . The second line was connected with a buffer of 90% acetonitrile and 0.1% TFA, and the two buffers were combined with a stepwise gradient to encompass the range 20-30% acetonitrile that was used to elute the A β peptide. A flow rate of 3 mL/min was used throughout the procedure. Fractions with pure A β were determined and confirmed through MALDI-TOF mass spectrometry, then lyophilized and frozen at -80°C. Lyophilization removes water, acetonitrile, and other volatile organic solvents.

4.3.2 NMR Sample Preparation of A β Peptides

Powdered A β samples (including both in lab and rPeptide samples, see previous section) were disaggregated using a HFIP-NaOH treatment, demonstrated to produce A β in primarily monomeric form(337, 398, 407). This entails fully dissolving the powder in hexafluoro-2-propanol (HFIP), then allowing the sample to fully evaporate under a stream of N₂, leaving a film of disaggregated peptide. By fully evaporating the HFIP, we ensure that there is no contamination by organic solvents. The film was re-dissolved in 10 mM NaOH with enough volume to produce a storage concentration of 1 mg/mL, then either used directly for NMR or stored at -80°C to be used in future experiments. NMR samples were produced by dissolving this solution in other buffers to produce the following final conditions: 100 μ M A β , 20 mM potassium phosphate buffer (using a dibasic K₂HPO₄ stock), 2.5 mM NaOH, 10% D₂O, pH 7.3. Experiments were done at a total volume of 400 μ L with a standard NMR tube, except those done with A β 40-E22 Δ and A β 40-A2T, which used 260 μ L volumes and a Shigemi tube (used to conserve sample while maximizing sample concentration, increasing signal to noise ratio).

A few caveats to the sample preparation protocol above apply to certain peptides. M35-oxidized A β 42 was generated through oxidation by H₂O₂ following the same steps as previous studies(202, 398, 409, 421). This entails dissolving HFIP-treated A β 42 from rPeptide in 10 mM NaOH as above to disaggregate the sample, then diluting the sample to 50 μ M in 10 mM potassium phosphate buffer (using a dibasic K₂HPO₄ stock) at pH 7.3. For each 1 mL of volume, 5 μ L of a dilute stock of 7.5% H₂O₂ by weight was added. This sample was incubated at 293 K for four hours while being stirred. The sample was then lyophilized again to remove hydrogen peroxide, and redissolved in water to produce a sample with the same final solution conditions for NMR as listed above. A β 40-E22 Δ is characterized by an extremely fast aggregation rate that would normally impede the ability to collect certain multidimensional NMR spectra within the necessary acquisition time. Further, it has previously been observed that wild type A β 40 acts as an inhibitor for A β 40-E22 Δ aggregation(250). As such, certain experiments (delineated below) required the use of a 1:1 mixture of isotopically labeled A β 40-E22 Δ and unlabeled A β 40; this condition was found to sustain the proportion of A β 40-E22 Δ in NMR active monomer form for sufficient enough durations needed to collect NMR data. All other NMR sample preparation steps for these experiments were identical.

4.3.3 NMR Data Acquisition

Coupled with the monomer-driven sample preparation above, NMR signals of A β have been demonstrated to be primarily (>90%) from the monomeric state, as demonstrated by analysis of NMR relaxation and pulse field experiments(337, 407, 422). The NMR experiments discussed here were run on either a 600 or 800 MHz Bruker NMR spectrometer equipped with a cryoprobe. The high pressure NMR experiments (described in section 4.3.3.6) are an exception - these were collected on a 600 MHz spectrometer without a cryoprobe. These spectrometers are located at the NMR core facility in the Center for Biotechnology and Interdisciplinary Studies at Rensselaer. All experiments were run at 277 K to minimize signal loss due to peptide aggregation and the effects of solvent exchange. All spectra in this thesis were processed using nmrPipe(423) and analyzed with Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco, CA).

4.3.3.1 NMR Assignment

Full assignments of all A β 40, A β 42, and A β 42-M35ox resonances have been previously published(424), while backbone assignments of A β 40-E22 Δ and A β 40-E22G were performed by our lab and published in Dr. Christopher Connors's thesis document(54). Briefly, this involved running the BEST implementations of HNCA and HN(CO)CA (425, 426) with samples prepared identically as above but with the addition of 3 M urea. This was done to prevent signal loss of the monomer through aggregation on the timescale needed to run the experiment, which is a problem for the fast aggregating FAD mutants being studied. ^{15}N - ^1H HSQCs at 0 M urea were then assigned by collecting several HSQCs at varying urea concentrations including 3M, then using these spectra to track back the assignment to the 0M condition.

Newly published in this dissertation are backbone assignments for A β 40-D23N and A β 40-A2T, as well as a partial side chain assignment for A β 40-A2T. All of these assignments were done under the sample conditions listed above and directly at 0 M urea. Backbone assignments were done with the BEST version of triple resonance experiments (HNCA, HNCO, HN(CA)CO, HNCACB, HN(CO)CA, HN(CO)CACB) (425, 426) and using non-uniform sampling (NUS), with a sampling rate of 25%. Both the use of BEST experiments and NUS sampling were done to minimize the amount of NMR acquisition time necessary to collect spectra and avoid signal loss due to aggregation. Reconstructions of NUS spectra were done through the multi-dimensional decomposition method (MDD)(427) using the qMDD software package(428). A control BEST HNCO without NUS was run on the same A β 40-A2T sample as the NUS equivalent; the resulting assignment was fully reproducible. Partial side chain assignments (only including aliphatic carbon and hydrogen shifts) for A β 40-A2T were done through the (H)C(CCO)NH-TOCSY and H(CCCO)NH-TOCSY experiments(429, 430), also collected with NUS with a sampling rate of 25%. Resonance lists containing these assignments are published in Table 7.2 in the Appendix, augmented with resonances from other non-assignment spectra (such as ^1H - ^{15}N HSQC and ^1H - ^{13}C HSQCs). Here, these chemical shifts are used for comparison to those predicted from REMD simulation.

4.3.3.2 NMR Relaxation Measurements

This thesis presents new measurements of ^{15}N R_1 , ^{15}N R_2 , and heteronuclear, steady state ^1H - ^{15}N nuclear Overhauser effect values (ssNOE) for A β 40-A2T, and compares them to previously collected relaxation measurements for A β 40 made by Yan *et al.*(407). Spectra to measure these data for A β 40-A2T were collected in an interleaved manner, to minimize the confounding effect of signal intensity loss over time due to sample aggregation. These spectra were collected using very similar acquisition parameters to the experiments collected by Yan *et al.*(407). For measurement of R_1 , the relaxation delays that spectra were collected for were: {10, 100, 200, 300, 400, 500, 600, 700} ms, with two replicates collected for the 10 and 700 ms points. Meanwhile, R_2 was fit using spectra collected with relaxation delays of {5, 250, 375, 500} ms, with two replicates performed for the 5 and 500 ms points. ssNOE values were obtained by collecting a spectra that interleaves the proton saturated data and the non-proton saturated data at each t_1 point, again, analogous to what was done by Yan *et al.*(407) for wild type A β . The ssNOE data for A β 40-A2T presented here is an average of two separate measurements.

4.3.3.3 Measurement of J Coupling and Residual Dipolar Coupling Data

$^3\text{J}_{\text{HNHA}}$ couplings for A β 40, A β 42, A β 42-M35ox, A β 40-D23N, A β 40-E22 Δ , and A β 40-A2T were collected using the J-resolved SOFAST-HMQC pulse sequence(411); these coupling sets are included in Table 7.3 in the Appendix. Data sets are based on splitting in the indirect J (ω_1) dimension only. Both glycine couplings and couplings of ambiguous assignment (due to spectral overlap) were ignored in all comparisons with simulation data. The experimental residual dipolar couplings (RDCs) studied here for amides of A β 42, A β 40, and A β 42-M35ox were previously determined by Yan *et al.*(218); these were measured in 10% polyacrylamide gel using an IPAP-HSQC technique(431). The experimental $^3\text{J}_{\text{NHB}}$ couplings for A β 40 compared to simulation here were published by Rosenman *et al.*(212) using the HNHB pulse sequence(432).

4.3.3.4 Prediction of NMR Observables from Simulation

Quantitative comparison of chemical shifts, J-coupling (${}^3J_{\text{HNHA}}$ and ${}^3J_{\text{NHB}}$), and RDC data back-calculated from the REMD trajectories described in Chapters 2 and 3, with their corresponding experimental values are used to both monitor equilibration over time and assess the validity of the MD ensemble. Chemical shifts were calculated from simulated coordinates using both SPARTA+(433) and SHIFTX+(434) calculators. J couplings were calculated from the REMD coordinates for each system using the Karplus equation(435):

$${}^3J(\theta) = A \cos^2(\theta) + B \cos(\theta) + C \quad (4.1)$$

where A , B , and C are semi-empirically determined coefficients (Karplus coefficients) and $\theta = (\phi - 60^\circ)$ or $\theta = (\chi_1 - 60^\circ)$ for ${}^3J_{\text{HNHA}}$ or ${}^3J_{\text{NHB}}$ respectively, where ϕ and χ_1 are peptide dihedral angles(394). The Karplus coefficients used for comparison here are those originally published by Vuister and Bax(362) for ${}^3J_{\text{HNHA}}$ couplings, as well as the consensus coefficients published by Pérez *et al.*(436) for ${}^3J_{\text{NHB}}$ couplings.

RDCs were calculated from the REMD trajectories using the PALES program(437, 438). As the alignment tensor magnitude is a function of the extent of alignment of the particles of the media which is dependent on experimental conditions, the reported computational ensemble values were uniformly scaled to minimize the root mean square deviation (RMSD) compared to experimental data. The resulting scaling factors were 0.35 for A β 42 and 0.26 for A β 40 for ILDN simulations and 0.37 for A β 42, 0.30 for A β 40, and 0.42 for A β 42-M35ox(S) for OPLS simulations.

Correlations of these observables between experiment and the REMD ensemble average are examined similarly to previous studies(9, 11) by use of the Pearson's correlation coefficient (PCC):

$$\text{PCC} = \frac{\sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}))}{(n - 1)\sigma_x\sigma_y} \quad (4.2)$$

for data sets x and y of size n with standard deviation σ_x and σ_y , respectively. PCCs before and after random coil chemical shifts have been subtracted from both experimental and computational data sets are presented here.

4.3.3.5 Urea Titrations

The urea titration data for A β 40-E22 Δ discussed in this chapter was first published in Dr. Christopher Connors's Ph.D. dissertation(54). These titrations were performed by collecting ^{15}N - ^1H HSQC after adding various quantities of a urea stock solution (8 M urea, 100 mM potassium phosphate buffer (using a dibasic K_2HPO_4 stock), pH 7.3) to the A β NMR sample. We define chemical shift perturbations (CSP) between two ^1H - ^{15}N HSQCs to be:

$$CSP = \sqrt{(10\Delta\delta_H)^2 + (\Delta\delta_N)^2} \quad (4.3)$$

where $\Delta\delta_H$ and $\Delta\delta_N$ are the change in chemical shift of a given resonance in the hydrogen and nitrogen dimensions, respectively. Changes in the CSPs between conditions where [urea] is 2 M and 0 M in the NMR sample for different A β species were calculated and compared to changes in hydrogen bond populations of the simulations described in Chapters 2 and 3 of this thesis.

4.3.3.6 High Pressure NMR

High pressure NMR of A β was conducted by our group in collaboration with the lab of Prof. Catherine Royer, with the assistance of Nicolina Clemente and Martin Fossat. Samples were prepared identically as above (see "NMR Sample Preparation of A β Peptides" section) but dissolved in a pressure compatible 10 mM deuterated Tris-HCl buffer in place of the more pressure sensitive phosphate buffer used for the standard experiments (all other sample components are present at the same concentrations)(439-441). In addition, 450 or 500 μL sample volumes were used. All high pressure experiments were conducted at 277 K on a 600 MHz Bruker spectrometer equipped with a TXI probe, without a cryoprobe. Samples were loaded in a 2500 bar standard ceramic pressure cell connected to an Xtreme-60 syringe pump system provided by Daedalus Innovations (Aston, PA, USA). This is an in-line system, allowing us to change the pressure without removing the sample from the NMR probe. Temperature calibration in the high pressure cell at ambient pressure was specifically performed by measuring the ^1H chemical shift difference between the methyl and hydroxyl groups of 100% methanol(442) to gauge the actual temperature in the sample. This was done 7 times,

yielding sample temperatures that encompassed the range 273-307 K in a roughly uniform manner; these data collectively yielded a linear fit between sample and sensor temperatures with an R^2 of 0.999 (Figure 7.3 in the Appendix). This fit allowed the interpolation of appropriate sensor temperature to reach a 277 K sample temperature. ^{15}N - ^1H HSQC and ^1H NMR spectra were measured for multiple samples (A β 40, A β 42, A β 40-D23N, A β 40-E22G, A β 40-A2T) at 250 bar increments, from ambient pressure (1 bar) to 2500 bar. ^1H chemical shifts were referenced to the remnant methylene signal from the 10 mM deuterated Tris-HCl buffer across the different pressures. This Tris signal itself was, in turn, referenced to DSS through a second set of ^1H spectra collected across the different pressures, with a sample of 0.1 mM DSS added to the same buffer at the same conditions (10 mM deuterated Tris-HCl buffer, pH 7.3) but without A β peptide. This 2-step referencing was performed because DSS has been observed to interact with A β and affect its aggregation(443); this method avoids using DSS as a buffer component. (Tris chemical shifts, meanwhile, are not observed to not change upon addition of A β .) All peaks involved in the 2-step referencing were observed to be non-overlapped. ^{15}N chemical shifts were then also referenced based on the extrapolated DSS shift using the indirect method proposed by Wishart and colleagues(444). HSQCs were collected at 1 bar, and then every 250 bar increment up to 2500 bar. Each HSQC was collected successively, using the same sample in an in-line fashion. 16 scans were employed for each HSQC, requiring a full acquisition time of about 1 hour, except for A β 40-E22G, which used 32 scans, employed to mitigate losses due to low signal to noise ratio, and thus required ~2 hours for NMR data acquisition per pressure point. A pressure pump failure occurred in the successive collection of the final 2500 bar point for A β 40-E22G; this point had to be re-collected again after resealing the pump at ambient pressure. This significantly changed peak height due to aggregation (but not peak positions), so metrics relating to data intensity values at 2500 bar for this species are not reported here.

Pressure coefficients from the positions of ^{15}N - ^1H HSQC peaks were calculated through a similar method as the one documented by Munte *et al.*(410). This entails subtracting random coil chemical shifts at each given pressure by extrapolating the random coil amide nitrogen and hydrogen chemical shifts at ambient pressure with the

known pressure dependence of these nuclei for each amino acid X in the model peptide Ac-Gly-Gly-X-Ala-NH₂, as determined by Koehler *et al.*(445). The set of these corrected shifts were used as the objective variable δ^* to fit the pressure coefficients B_1^* and B_2^* in the following second order Taylor expansion:

$$\delta^*(p, T_o) = \delta_o^*(p_o, T_o) + B_1^*(p - p_o) + \frac{1}{2} B_2^*(p - p_o)^2 \quad (4.4)$$

where p_o is the ambient pressure and δ_o^* is the random chemical shift at this pressure. This second order model was previously observed to best represent the movement of ¹H chemical shifts with pressure(446). Least squares fitting over all pressures was used to obtain pressure coefficients for both nitrogen and hydrogen nuclei separately.

Solvent exchange measurements were made using the CLEANEX-PM pulse sequence with fHSQC detection scheme(223, 224), using mixing times (τ_m) of {5, 10, 15, 20, 30, 40, 50} ms. Data at longer mixing times were collected, but not used for fitting because the equation below is only valid at short τ_m (224). As per the protocol in the original publication of the sequence(223, 224), data for each τ_m were corrected by the intensity values of an initial HSQC, and two different control experiments were run to correct for losses due to recycle delay and to obtain the R_{1B} value, respectively. Then, fitting to obtain k_{ex} was done according to the following equation:

$$\frac{V}{V_o} = \frac{k_{ex}}{(R_{1A} + k_{ex} - R_{1B})} * \{ \exp(-R_{1B} * \tau_m) - \exp[-(R_{1A} + k_{ex}) * \tau_m] \} \quad (4.5)$$

where V is the series of intensity values of each amide group as a function of mixing time τ_m , V_o is the intensity of each amide in the initial HSQC, k_{ex} is the rate constant of solvent exchange for each amide, and R_{1A} and R_{1B} are the rates for the apparent longitudinal relaxation of protein and water respectively(223, 224).

A series of three CLEANEX-PM experiments were run over the timescale of the high pressure solvent exchange experiments (~26 hours) at ambient pressure for A β 40 in the NMR buffer described above at 277 K, revealing no directional change in the measured k_{ex} values as a function of time (or aggregation), serving as an important control for these possible confounding variables (Figure 7.7 in the Appendix). Though

almost ~50% signal had been lost due to aggregation during the control experiment (see Figure 7.7's caption for specific losses over the experiment), correcting CLEANEX-PM data heights by the intensities of the initial HSQC essentially removed this effect from the fit. A similar control was run for A β 42 solvent exchange, this time as a final experiment collected at the same pressure condition of the first experiment (1000 bar) after all of the main pressure experiments were run; there were no directional changes in k_{ex} values observed between the initial and final run (*data not shown*). For the main experiments, solvent exchange measurements were collected at {1, 1000, 2000} bar for both species.

As mentioned, we used Tris buffer for these experiments, and while it is both one of the most commonly used and one of the most baroresistant buffers for high pressure studies(447, 448), the pH of this buffer has still been observed to increase on the order of 0.1 pH unit/kbar(440). Even though this would be a relatively subtle change for our experiments, it could dramatically affect the solvent exchange rates, which are highly sensitive to pH conditions(449). Fortunately, the effect of pressure on Tris NMR signals has been previously characterized, and can be used as an indicator of the pH state of the sample(450). pH was thus calculated at each pressure covered in our experiments using the reported ^1H shifts of the protonated and deprotonated forms of the Tris methylene signal and the pK_a (450), as well as the change in the true remnant Tris methylene shift referenced to DSS as a function of pressure (collected during the indirect referencing method discussed above). We calculate that the maximum pH increase over pressures occurred from 7.30 to 7.44, occurring from 1 bar to 2500 bar. Using these pH values, we calculated theoretical solvent exchange measurements for A β 40 using SPHERE(451, 452). These revealed that the k_{ex} increased uniformly for most residues at pH 7.44, but only by ~27-28% their original values at pH 7.30. Importantly, no residue pair exceeded this 28% gain. The increases in k_{ex} observed for A β peptides with increasing pressure far surpass this nominal gain (see Figure 4.9), suggesting that other factors beyond pressure-induced changes to pH of the buffer are responsible for these observations.

4.4 Results

4.4.1 Solution NMR of A β Monomers at Ambient Pressure

New solution NMR data are presented in this thesis for two mutants of A β 40: D23N and A2T. Table 7.2 in the Appendix presents partial assignments of both of these mutants based on the conducted ^{15}N and ^{13}C , ^{15}N experiments, as described in the Methods section (section 4.3.3.1). Figure 4.2 is a ^1H - ^{15}N HSQC overlay of A β 40-D23N with A β 40. The presence of new resonances for N23's side chain amide group, and the dramatic perturbation of the D23 to N23 backbone amide position clearly confirm that we are indeed studying the correct mutant form of this A β . Overall, A β 40-D23N produces a similar HSQC profile to wild type, except in residues close to the site of the mutation. Several of the residues preceding the mutation (L17, F20, A21) are shifted upfield, while some residues downstream of the mutation (S26 and N27) are instead shifted downfield. However, residues of the N and C-terminus are mostly unaffected by the mutation, in sharp contrast to the large changes in the C-terminal conformations predicted by the OPLS model in Chapter 2.

To facilitate further comparison between these ^1H - ^{15}N HSQCs, Figure 4.3 plots chemical shift perturbations of the amides of all of the A β mutants studied by our lab thus far with respect to their wild type A β 40 original form, as well as the perturbations between A β 42 and A β 40 (Figure 4.3E) and between methionine 35 oxidized A β 42 and its normal reduced form (Figure 4.3F). Many of the mutations are more conservative in their changes than one might expect based on the simulation data. As we noted in the previous paragraph, A β 40-D23N is one of the more conservative mutants, with only residue 23 at the site of the mutation having a very large perturbation. The first two residues of any A β are unobservable with ^1H - ^{15}N techniques due to solvent exchange, but interestingly, A β 40-A2T features a large change in the NMR-visible, non-mutated, adjacent residue E3. Besides this and the subsequent residue F4, however, the rest of the peptide remains largely unaffected by this mutation, in terms of chemical shift. Next, A β 40-E22G is one of the most perturbing of the FAD mutations examined here (note the y -axis of Figure 4.3C compared to the other panels), with not only residue 22 having an enormous perturbation of ~ 15.7 ppm, but adjacent residues A21 and D23 also

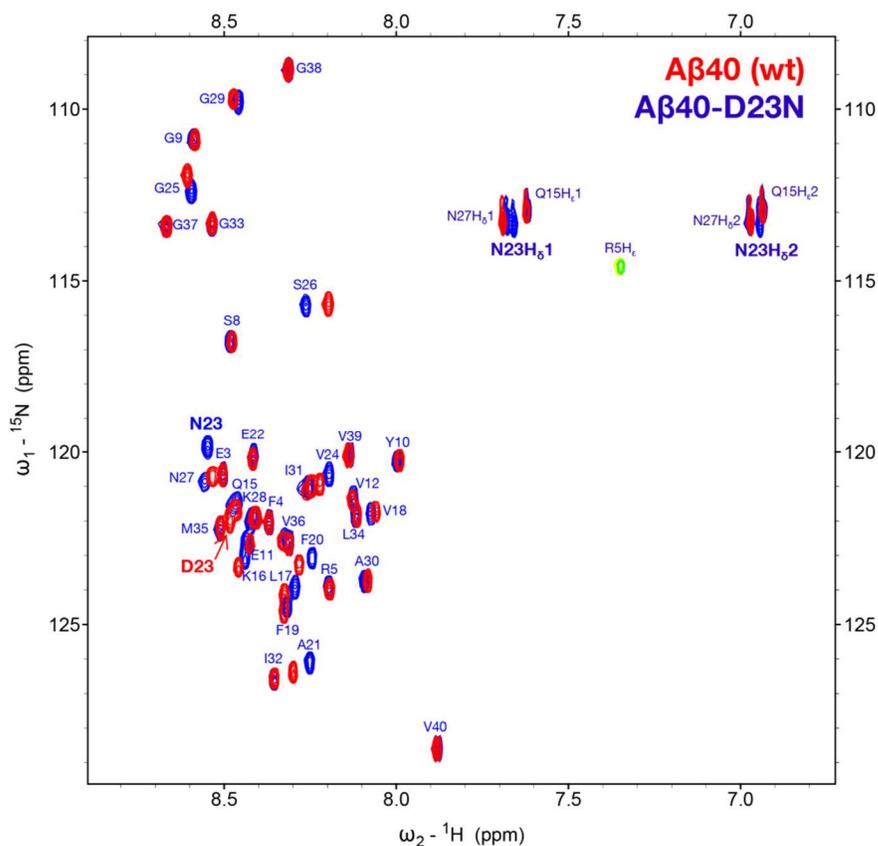


Figure 4.2. ^1H - ^{15}N HSQC spectrum for $\text{A}\beta 40$ -D23N (in blue/yellow), overlaid with the same spectrum for wild type $\text{A}\beta 40$ (in red/green). The assignment for $\text{A}\beta 40$ -D23N, in blue text, is included. Assignments for new resonances caused by the mutation are bolded, while the assignment of the wild type D23 amide is in red text. Data collected on 800 MHz spectrometer at 277 K, at identical buffer conditions as described in the text.

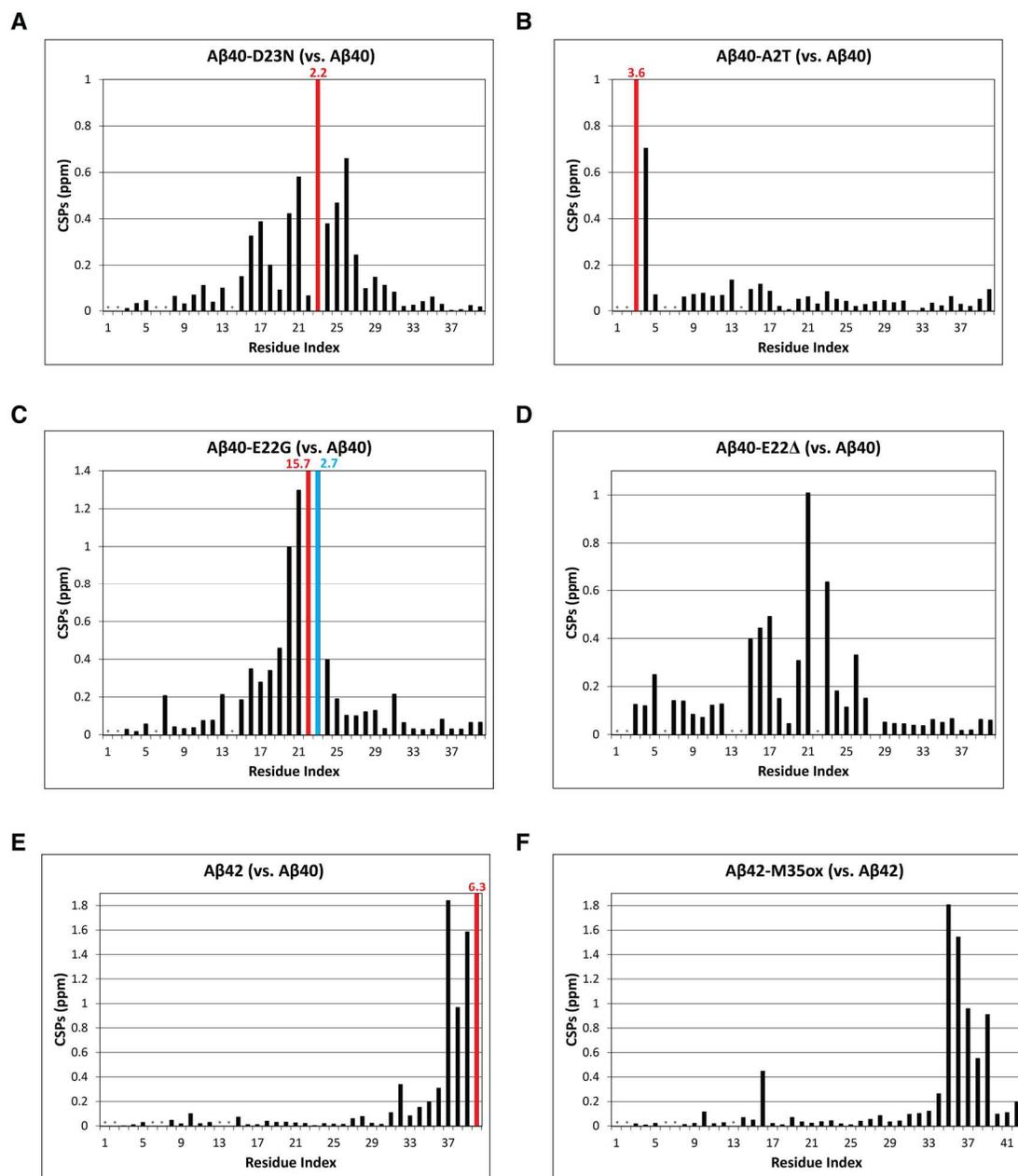


Figure 4.3. (A-F) Chemical shift perturbation (CSP) in ^1H - ^{15}N HSQC spectra between different $\text{A}\beta$ mutants/isoforms, as listed in the title of each graph. CSP was calculated with Equation 4.3, listed in the Methods. Values that exceeded the y-axis of the graph have the bar of the residue in question colored in a color other than black, and have their CSP explicitly printed above their bar. Asterisks were placed over residues which were unassigned in at least one of the spectra being compared, due to solvent/chemical exchange or spectral overlap. Data collected on 800 MHz spectrometer at 277 K, at identical buffer conditions as described in the text. Data generated with the assistance of Dr. Christopher Connors and Nicolina Clemente.

experiencing large shifts. An example of A β 40-E22G's HSQC can be seen in Figure 7.4 in the Appendix; curiously, the new G22 resonance is unusually upfield in both nitrogen and hydrogen axes in this HSQC compared to other glycines in the sequence, suggesting that this mutation samples conformations that shield both the backbone amide nitrogen and hydrogen of this residue. Finally, A β 40-E22 Δ shows rather low perturbations overall, which is surprising considering the profound effect this mutation has on A β aggregation discussed in Chapter 1, and the ability of this mutation to disrupt the CHC to C-terminal hairpin observed through both OPLS and ILDN simulations in Chapters 2 and 3. This said, while there is no 22nd residue to compare to for this mutant, the adjacent residue A21 is substantially perturbed by the change. It is important to stress that none of these point mutations of A β 40 show substantial changes distal to the site of the mutation (i.e., for FAD mutants, few changes are observed at either of the termini of the peptide), in contrast to many of the long range changes predicted by the simulations in Chapters 2 and 3. However, we note that all of the FAD-linked mutants (A β 40-D23N, A β 40-E22G, A β 40-E22 Δ) noticeably perturb at least some of the amides of the CHC (L17-A21), representing short- and medium-range changes to the chemistry of the peptide relative to the site of mutation. Figure 4.3E illustrates that the changes in amide chemical shift caused by the addition of the two residues at the C-terminus for A β 42 are also chiefly local, with only residue G37-V40 experiencing large perturbations. Methionine 35 oxidation is also generally affiliated with local changes, primarily affecting residues M/MTO35-G37 and V39 (MTO = methionine sulfoxide). Curiously, however, the amide of K16 is also perturbed. This may be caused by the formation of new interactions involving K16 and the extreme C-terminus in this peptide, predicted by the OPLS model in Chapter 2.

In addition to chemical shift measurements, Figure 4.4 depicts a comparison of the backbone dynamics of A β 40-A2T and A β 40, by means of experimentally-determined ¹⁵N R_1 , ¹⁵N R_2 , and heteronuclear, steady state ¹H-¹⁵N nuclear Overhauser effect values (ssNOE) for each backbone amide of the peptide. R_1 is a constant that describes longitudinal (spin-lattice) relaxation, that is, the rate constant of recovery of the z -component of magnetization to equilibrium for a certain nucleus after perturbation by a

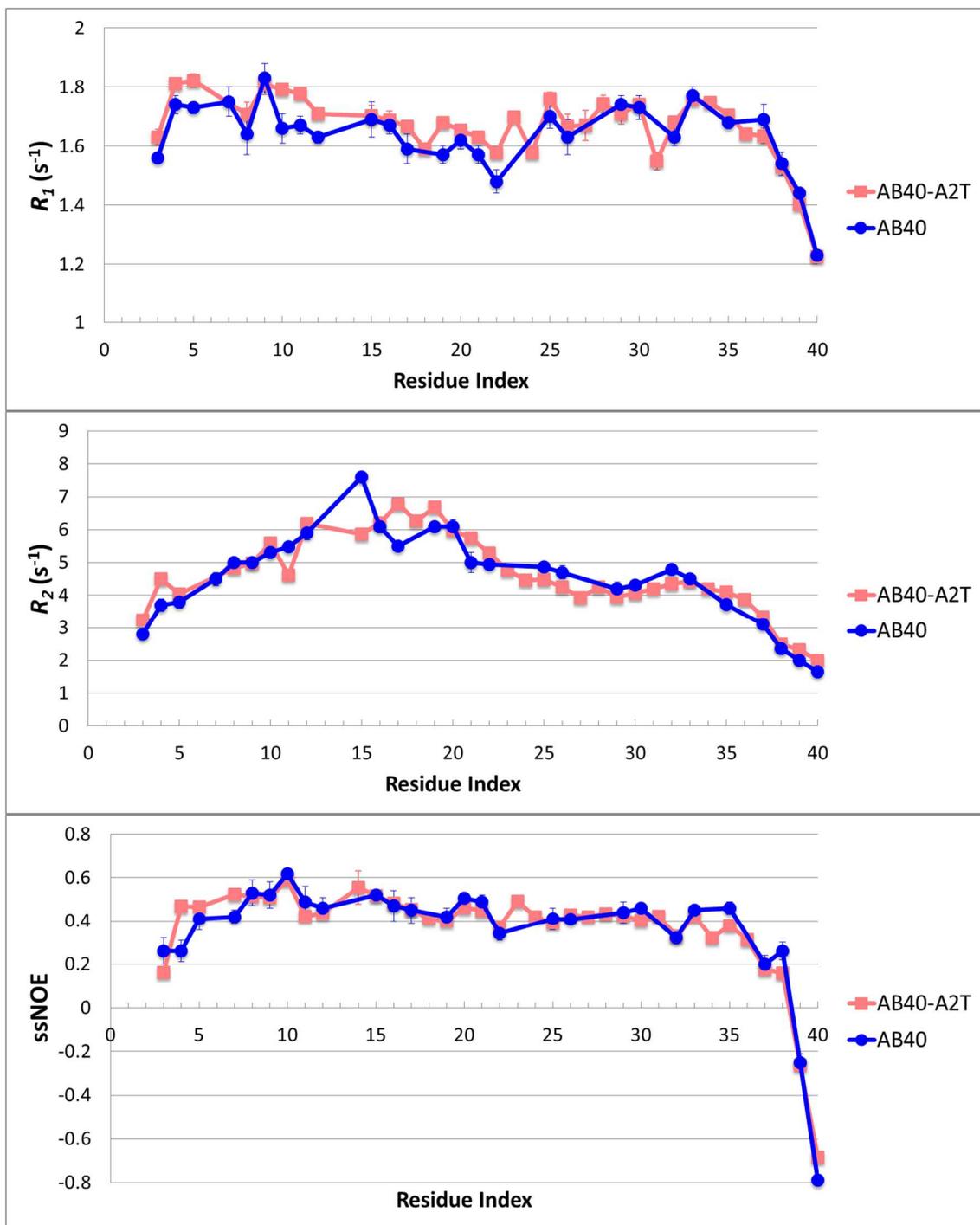


Figure 4.4. Comparison of backbone dynamics of Aβ40-A2T versus Aβ40, as determined by ^{15}N R_1 , ^{15}N R_2 , and heteronuclear, steady state 1H - ^{15}N nuclear Overhauser effect (ssNOE), plotted versus residue index. Values for Aβ40 were previously published by Yan *et al.*(407). New data for Aβ40-A2T generated with the assistance of Prof. Chunyu Wang and Nicolina Clemente. The ssNOE data for this species is an average of two replicates.

resonant radio frequency pulse(394). Meanwhile, R_2 is constant of transverse (spin-spin) relaxation, describing the rate of decay of the magnetization in the xy -plane for a resonance after excitement by the same type of pulse(394). The third parameter, ssNOE, measures the extent to which the saturation of the ^1H spins also changes the magnetization of ^{15}N spins through the nuclear Overhauser effect (NOE), which is mediated by cross-relaxation induced by dipolar coupling(394). All three of these values are sensitive to movement on the picosecond to nanosecond timescale(394). For intrinsically disordered proteins, such as $\text{A}\beta$, that are known to sample faster than nanosecond timescale motions that affect relaxation, it has been suggested that amides with higher ^{15}N R_1 , ^{15}N R_2 , and ssNOE are directly affiliated with slower internal motions, and are thus characterized by greater backbone rigidity(375, 453). However, it is also important to note that both R_1 and R_2 are affected by other factors, complicating their interpretation: both R_1 and R_2 are sensitive to anisotropic global tumbling, while R_2 is also sensitive to the microsecond to millisecond timescale due to the effects of chemical exchange(394, 454).

With this established, we can discuss the salient changes caused by the A2T mutation on backbone relaxation, as seen in Figure 4.4. The mutation, in general, significantly increases R_1 for many residues in the first half of the peptide: E3, F4, R5, Y10, E11, V12, L17, F19, A21, E22, and G25, while the effect on R_1 is mostly insignificant for the rest of the peptide. For R_2 , many residues are affected: residues with faster rates include F4, L17, F19, A21, L34, V39, V40, while amides with significantly decreased R_2 are E11, Q15, V24, G25, G29. However, while the R_2 shows interesting and consistent regional trends, such as faster relaxation rate in the CHC and C-terminus, and slower relaxation in the central region, these differences are fairly subtle in magnitude. The salient changes in ssNOE, meanwhile, include an increase for most amides of the N-terminus (F4 and D7), and a decrease in the C-terminus (M35 and G38). In general, the changes in these three parameters caused by A2T are rather disparate, though the consensus between these measures would indicate a more rigid N-terminus and CHC. Recall from the REMD simulations in Chapter 3 that the A2T mutation primarily promoted β character and higher order parameters (S^2) in E3-R5 and E11-H14, while decreasing the magnitude of these values in the CHC (V18-D23) (Figure 3.2B and

Figure 3.8). Therefore, qualitatively, the effects of this mutation as gauged by simulation are only partially sustained by the experimental data presented here (namely, increases in rigidity near the mutation site, but not the changes predicted at the CHC).

On this note, in normal circumstances, standard model-free analysis could be used to derive order parameters S^2 from the experimentally measured R_1 and R_2 measurements(367), which could provide further insight into the rigidity of each amide bond vector, as well as providing another means of directly validating simulation data. However, when we calculated S^2 from REMD simulation data (see section 2.4.8 and 3.4.8), we observed that the separability between global and internal motions was very low for A β monomers in the simulations. This is likely also true for the NMR samples, which would also violate the basic assumptions of the standard model-free formalism needed to determine the S^2 from experiment. Therefore, order parameters derived from experiment are not presented here.

Lastly, Table 7.3 in the Appendix contains one more category of NMR data measured at ambient pressure: $^3J_{\text{HNHA}}$ couplings for some of the A β monomer species described above. These couplings are described in more detail in section 4.4.2.2.

4.4.2 NMR Observables and Simulation Comparison

4.4.2.1 Chemical Shifts

Chemical shifts are the easiest and potentially the most information-rich NMR observable that can be measured, however, the physical relationship between these values and the conformational state(s) that produce them are not well understood(455). As such, most computational predictors of chemical shifts from three-dimensional structures rely on empirical or semi-empirical approaches, based on a training set of reference proteins that are usually well-ordered(455). Though these calculators may not be well parameterized for the study of intrinsically disordered proteins, nevertheless, we have used two of the best cited chemical shift predictors in the literature, SPARTA+(433) and SHIFTX+(434), to obtain ensemble averaged chemical shift data from the ILDN and OPLS production ensembles of A β (see Chapters 2 and 3) and compared them to our collection of experimentally determined chemical shifts, including

multiple nuclei for all of the A β peptides that have been investigated by us with NMR (Figure 4.5).

SPARTA+ works through a neural network algorithm trained on a database of high resolution X-ray structures that also have mostly complete backbone and C β chemical shift information(433). While this database is unlikely to represent the case of intrinsically disordered sequences well, SPARTA+ has one of the best performances of the calculators in the literature, making predictions on the test set provided by the investigators with a RMSD of {1.14 1.09, 0.94, 2.45, 0.25, and 0.49} ppm for {C α , C β , C', N, H α , H β } shifts respectively(433). Meanwhile, SHIFTX+ is a subroutine of the SHIFTX2 program(434) representing an improvement on the SHIFTX algorithm(456), and, like SPARTA+, yields purely structure-based predictions of chemical shift based on high resolution, well-folded protein structures with >90% resonance assignment. SHIFTX2, by default, would normally combine predictions from SHIFTX+ with sequence based predictions yielded by the SHIFTY+ algorithm; we choose to ignore the latter because SHIFTY+ employs the bioinformatics of previously assigned proteins; as we discussed, the vast majority of these are globular proteins that differ in nature from the IDPs being investigated here(434, 455). The RMSDs for SHIFTX+ on the original author's test set are comparable to those cited above for SPARTA+ for each of the atom types tested(433, 434).

Panel A of Figure 4.5 presents a direct comparison of the chemical shift sets in terms of PCC and RMSD. A better fit will improve correlation and reduce RMSD. In an effort to minimize the trivial correlation due to agreement in identity between the sequences being compared, secondary chemical shifts – that is, the deviation of each shift from random coil chemical shift of the given residue – were also compared. In general, both chemical shift predictors yield good agreement with the simulated ensembles. The predicted chemical shifts for all nuclei have RMSDs to experimental chemical shifts on the order or lower than the typical deviations cited in the previous paragraph(433, 434), with the exception of the nitrogen chemical shifts predicted by SHIFTX+. Meanwhile, the primary PCCs for N and C α shifts are higher than values cited in the literature for other simulations of A β , while H α values are of a similar

A

Aβ40									
Atom Type	ILDN Primary / Secondary PCC		ILDN RMSD (ppm)		OPLS Primary / Secondary PCC		OPLS RMSD (ppm)		
	<i>SPARTA+</i>	<i>SHIFTX+</i>			<i>SPARTA+</i>	<i>SHIFTX+</i>			
C α	0.99 / 0.28	0.99 / 0.15	0.53	0.78	0.99 / 0.24	0.99 / 0.19	0.59	0.72	
C β	1.00 / 0.51	1.00 / 0.52	0.70	1.00	1.00 / 0.08	1.00 / 0.07	0.81	1.04	
N	0.97 / 0.71	0.94 / 0.64	1.99	2.99	0.97 / 0.70	0.96 / 0.68	1.85	2.74	
H α	0.83 / 0.15	0.88 / 0.16	0.18	0.16	0.84 / 0.26	0.91 / 0.36	0.22	0.17	
H _N	0.25 / 0.06	0.46 / 0.28	0.28	0.25	0.26 / 0.11	0.56 / 0.44	0.21	0.16	
Aβ40-D23N									
C α	1.00 / 0.44	0.99 / 0.19	0.62	0.84	1.00 / 0.45	1.00 / 0.35	0.67	0.72	
C β	1.00 / 0.52	1.00 / 0.57	0.93	1.24	1.00 / 0.25	1.00 / 0.42	0.88	1.02	
C'	0.96 / 0.80	0.75 / 0.50	0.58	1.05	0.93 / 0.71	0.69 / 0.26	0.73	1.17	
N	0.97 / 0.73	0.95 / 0.68	1.81	2.86	0.97 / 0.71	0.96 / 0.66	1.59	2.30	
H α	0.81 / 0.12	0.87 / 0.10	0.21	0.16	0.85 / 0.27	0.91 / 0.37	0.21	0.16	
H _N	0.20 / 0.00	0.57 / 0.40	0.26	0.21	0.32 / 0.12	0.53 / 0.40	0.19	0.17	
Aβ40-E22Δ									
N	0.97 / 0.72	0.96 / 0.74	2.13	3.14	0.98 / 0.76	0.97 / 0.72	1.87	2.71	
H _N	0.39 / 0.19	0.69 / 0.55	0.27	0.25	0.12 / 0.01	0.41 / 0.26	0.23	0.16	
Aβ40-A2T									
C α	1.00 / 0.49	1.00 / 0.34	0.55	0.82					
C β	1.00 / 0.53	1.00 / 0.58	0.84	1.11					
C'	0.96 / 0.82	0.86 / 0.58	0.57	0.89					
N	0.97 / 0.73	0.96 / 0.74	1.99	2.93					
H α	0.82 / 0.21	0.88 / 0.09	0.19	0.14					
H _N	0.32 / 0.21	0.64 / 0.49	0.26	0.25					

B

(Aβ40-D23N)—(Aβ40)									
Atom Type	ILDN PCC		ILDN RMSD (ppm)		OPLS PCC		OPLS RMSD (ppm)		
C α	0.25	0.23	0.30	0.32	-0.30	-0.03	0.41	0.39	
N	0.33	0.53	0.54	0.47	-0.26	-0.19	0.94	1.19	

C

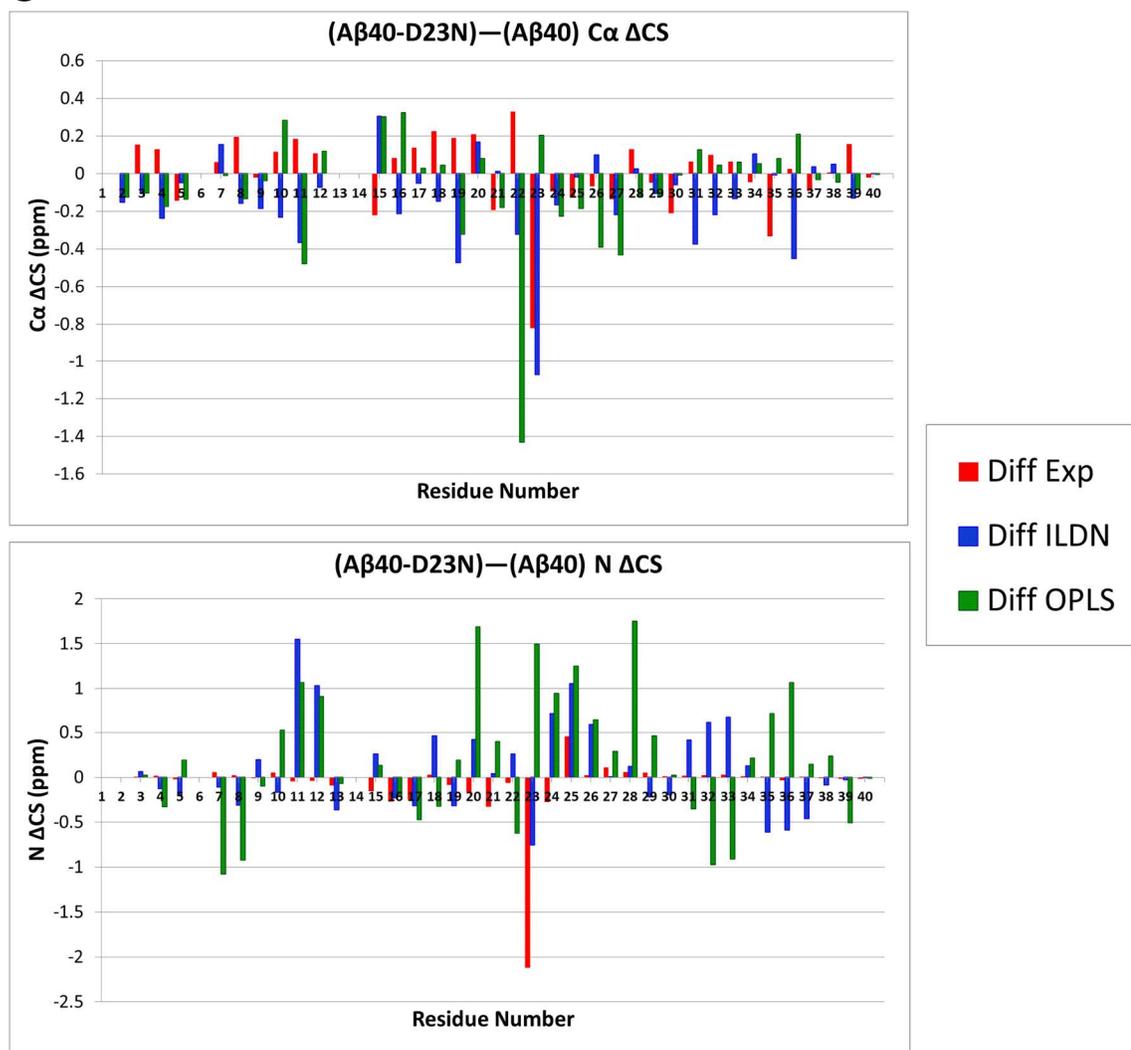


Figure 4.5. Correlation and RMSD of predicted chemical shifts with experimentally measured values, with predictions based on frames from the ILDN and OPLS Aβ ensembles. (A) Direct comparison of chemical shifts. In red are the comparisons using shifts predicted by SPARTA+(433), while light blue use values predicted by SHIFTX+(434). “Primary” refers to the comparison of the chemical shifts as they are, while “secondary” correlation is measured once both the predicted and experimental chemical shifts are corrected for their random coil values. (B) Comparison for the difference in chemical shift (Δ CS) between Aβ40-D23N and wild type Aβ40 between experimental and simulated shifts for two well-predicted atom types, Cα and N. (C) Plot as a function of residue number for these Δ CS for experimental data and predicted from the two different sets of simulated ensembles using the SPARTA+ predictor. The final residue, V40, is unreliably predicted by both calculators and is ignored in all correlation calculations. Stereospecific H α shifts are provided by SPARTA+ for glycine; these were averaged for H α comparisons.

magnitude(18). The set of PCCs and RMSDs produced are comparable for both the OPLS and ILDN runs for most of the atom types. According to the secondary PCCs, hydrogen nuclei are not predicted well by the chemical shift calculators in general, though OPLS seems to do a somewhat better job. In contrast, for C β chemical shifts of A β 40 and A β 40-D23N, ILDN produces a reliable prediction while the secondary PCC associated with OPLS is much poorer.

Beyond looking at these direct comparisons, we are interested in considering how each force field is able to reproduce the changes in NMR observables when a perturbation to the ensemble is considered, such as a point mutation. This analysis would help to minimize the effect of systematic errors common to the multiple measurements involved and sequence dependence of the comparison (even in the case of a point mutation, 39/40 of the residues are identical). This said, we must also consider the results with some skepticism because it is unclear whether the sensitivity in the difference of these observables between species outweighs the potentially sizable error inherent in the prediction of these observables from simulation using chemical shift predictors that are calibrated for globular proteins. Nevertheless, we proceed with this analysis because we have full sets of simulation data and experimental chemical shifts for A β 40 and A β 40-D23N, and because the structural effect of the D23N mutation in the OPLS ensemble was so radically different from that of ILDN (see Chapters 2 and 3). Figure 4.5B and Figure 4.5C compare the experimental and computational differences in chemical shifts between these species (Δ CS) for two of the resonances that had high correlation in the direct comparisons for these ensembles, N and C α . Both force fields strongly over-predict the effect of the mutation on changes in chemical shifts, particularly at the N- and C-terminus. However, unlike the ILDN force field, the OPLS force field is associated with an overall negative correlation and increased RMSD for these comparisons with both SPARTA+ and SHIFTX+ predictors, driven by strong anticorrelation on the effect of this mutation on the Δ CSs of residue E22-D/N23 for C α and residue D/N23-V24 for N shifts in the case of SPARTA+. These strong anticorrelations for Δ CS of residue D/N23 are not present for the ILDN force field. While this may be a weak argument given the inherent error that may be involved in measurement or prediction of chemical shifts in the first place, the differences we

examine for $^3J_{\text{HNHA}}$ couplings in the next section also support the finding that the D23N mutation causes local structural changes compared to wild type when simulated with the OPLS force field that are not supported by the experimental data.

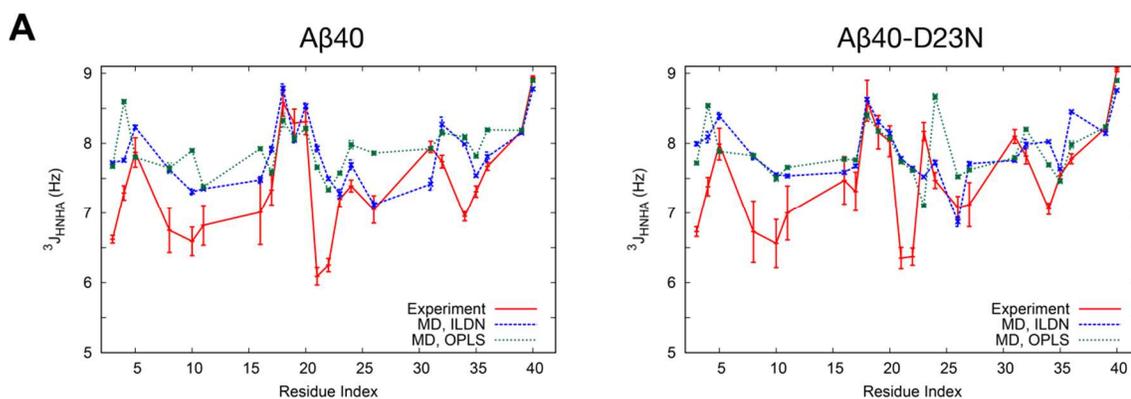
4.4.2.2 J-couplings and Residual Dipolar Couplings

J-couplings are observable scalar splittings mediated by the intervening electrons between two correlated nuclei; as such, they are dependent on the geometry and environment of the chemical bonds connecting the coupled partners. 3-bond J-couplings, such as $^3J_{\text{HNHA}}$ and $^3J_{\text{NHB}}$, are directly related to the dihedral angle between the correlated atoms (backbone angle ϕ and side chain angle χ_1 for $^3J_{\text{HNHA}}$ and $^3J_{\text{NHB}}$ respectively) by an empirically-derived relationship known as the Karplus equation(435)(see Methods). Thus, we can directly back-calculate these values from simulation based on the ensemble average of the dihedral angle in question. In general, $^3J_{\text{HNHA}}$ is sensitive to secondary structure; high values (> 8.0 Hz) are associated with β sheet, low $^3J_{\text{HNHA}}$ (< 5.5 Hz) with α -helix, and the interim range is associated with random coil or a mixed ensemble of states(457). Meanwhile, RDCs relate to the orientation of nitrogen-hydrogen bond vectors for each amide group relative to the steric alignment frame of each conformation (in the computational case, as determined by the PALES program(437, 438)), and thus serve as reporters of global angular orientation. We present a comprehensive set of $^3J_{\text{HNHA}}$ values for A β monomers in Table 7.3 of the Appendix, acquired using the J-resolved SOFAST-HMQC(411) pulse sequence, designed to mitigate spectral overlap, a problem particularly inherent to measurement of these values in intrinsically disordered peptides through other means. In general, the trend of the experimental $^3J_{\text{HNHA}}$ couplings is very similar for all A β species. The regions with the highest $^3J_{\text{HNHA}}$ coupling are L17-F20, I31-I32, V39-V40 for A β 40, and I41 for A β 42, all of which are hydrophobic residues. Interestingly, with the exception of V39 and V40 in A β 40, all of these residues also match regions of high β propensity as observed through simulation (Figure 2.2 and Figure 3.2).

These $^3J_{\text{HNHA}}$ coupling values, in addition to previously published RDCs and $^3J_{\text{NHB}}$ couplings(212, 218) for A β , are directly compared to their computationally back-calculated values in Figure 4.6A and Figure 4.6B over the production period for both

ILDN and OPLS simulations. As observed in Panel A, when using Vuister and Bax's Karplus equation parameter set(362) for the determination of computational $^3J_{\text{HNHA}}$ as we have done here, these values are too high compared to experiment overall, except in certain regions (ie, V18-F20, V40) where the experimental $^3J_{\text{HNHA}}$ is the highest. This also means that the range of computational $^3J_{\text{HNHA}}$ values is less than their experimental counterparts. This bias toward higher values may suggest oversampling of β -conformations in our simulated ensembles, but it could also indicate that the Vuister and Bax parameters of the Karplus equation are inappropriate for this system. Indeed, in the past, we have also fitted these Karplus parameters using least square minimization of the RMSD between the simulated and experimental data, to account for motional averaging effects within our trajectories that may not be represented by parameter sets determined by fitting to well-ordered X-ray structures(9, 11, 212). This produced curves where the average values and range of the J-couplings match for both ILDN and OPLS, reducing RMSD while producing similar PCC values. However, because this fitting was not necessarily producing parameters that were physically relevant (ie, we saw that fitting could greatly change the Karplus curve for positive ϕ (212)), for the sake of this thesis, we decided to only present comparisons using the unaltered Vuister and Bax parameter set(362).

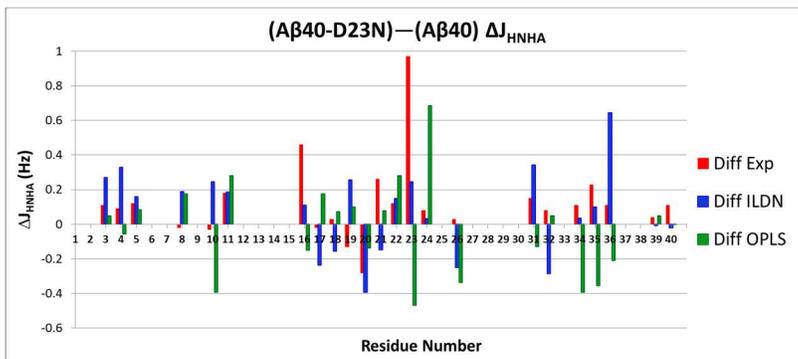
Even with these Karplus parameters, comparison of $^3J_{\text{HNHA}}$ yields variations along the sequence that agree well for both force fields (Figure 4.6A) yielding strong PCC values (Figure 4.6B). Disagreement occurs at regions where the measured $^3J_{\text{HNHA}}$ is lowest, especially D7-E11, A21-E22, and L34. Both ILDN and OPLS produce very similar back-calculated $^3J_{\text{HNHA}}$ curves for wild type A β 40. Differences between the simulated data lie primarily in the N-terminus and the central region (ie, F4, Y10, K16, V24, S26, A30, V36), where OPLS favors somewhat higher values. In general, among these differences, ILDN better represents the trends of the experimental data, except at A30. For A β 42, however, OPLS does a better job of representing the C-terminus; experimentally, V40 and A42 have a lower $^3J_{\text{HNHA}}$ than I41, a trend that OPLS captures well but ILDN anticorrelates(*data not shown, presented for OPLS simulations by Rosenman et al.(212)*). Nevertheless, in general, the direct comparison of $^3J_{\text{HNHA}}$ results



B

$^3J_{\text{HNHA}}$					
Species	ILDN PCC	ILDN RMSD (Hz)		OPLS PCC	OPLS RMSD (Hz)
A β 40	0.71	0.68		0.72	0.77
A β 42	0.42	0.65		0.47	0.64
A β 42-M35ox	<i>N/A</i>	<i>N/A</i>		0.48	1.08
A β 40-D23N	0.61	0.68		0.53	0.72
A β 40-E22 Δ	0.61	1.06		0.65	1.21
A β 40-A2T	0.65	0.61		<i>N/A</i>	<i>N/A</i>
RDC					
A β 40	0.30	1.89		0.50	1.69
A β 42	0.33	1.39		0.39	1.66
A β 42-M35ox	<i>N/A</i>	<i>N/A</i>		0.44	1.45
$^3J_{\text{NHB}}$					
A β 40	0.32	1.08		0.56	0.97

C



ILDN PCC	ILDN RMSD (Hz)
0.28	0.28
OPLS PCC	OPLS RMSD (Hz)
-0.31	0.41

Figure 4.6. Correlation and RMSD of calculated $^3J_{\text{HNHA}}$, $^3J_{\text{NHB}}$, and residual dipolar couplings from ILDN and OPLS A β production ensembles compared to experimentally determined values. (A) Plot of experimentally and computationally derived $^3J_{\text{HNHA}}$ couplings as a function of residue index for A β 40 and A β 40-D23N. (B) Direct comparison of experimental NMR observables with computationally back-calculated values. (C) Plot as a function of residue number for the difference in $^3J_{\text{HNHA}}$ between A β 40-D23N and wild type A β 40. PCC and RMSD between the computational and experimentally derived differences are also provided. When multiple experimental $^3J_{\text{HNHA}}$ were available for a given species, comparisons were made to the averaged values. Computational $^3J_{\text{HNHA}}$ couplings were back-calculated with Karplus parameters published by Vuister and Bax(362), while $^3J_{\text{NHB}}$ values use Pérez *et al.*(436) values. Glycines are not included in comparisons of $^3J_{\text{HNHA}}$ due to experimental ambiguity.

in a similar set of PCCs between the two simulation schemes, while OPLS has slightly higher RMSDs to experiment. Lastly, it should also be mentioned that the PCC of ${}^3J_{\text{HNHA}}$ from simulation to experimental values was also monitored as a function of simulation time over 20 ns/replica windows at the production temperatures (*data not shown, presented for OPLS simulations by Rosenman et al.(212)*); as a general rule, regardless of force field or species, PCC steadily improved with simulation time on the 100 ns/replica timescale. This suggests that long timescale simulations are needed to better match the properties of the experimental ensemble.

The A β 40-D23N mutation, which resulted in very different effects on the ensemble for OPLS and ILDN (see Chapters 2 and 3), represents an interesting test case, producing an anticorrelation between experiment and simulation in the trend of residue N23 and V24 that is particularly egregious in the case of OPLS (Figure 4.6B). We will examine this more carefully by looking at Δ^3J_{HNHA} between A β 40-D23N and A β 40. Like comparing ΔCS values, we proceed with this analysis with the caveat that the inherent error associated with measurement and other systematic errors (for example, by using inappropriate parameters for the Karplus equation) may outweigh the subtle changes in ${}^3J_{\text{HNHA}}$ between the different A β species. For the case of the measurement error associated with Δ^3J_{HNHA} , however, we were able to get some assessment of the significance of PCC comparisons for these two species by running Monte Carlo simulations (Figure 7.2 in the Appendix; the details are more explicitly explained in the figure caption) by choosing randomized values within the experimental error bars of the data compared to the actual measurement and calculating PCC for each. The average of the PCC values over 100000 runs when using the explicit experimental values and error bars for the simulation remained a very high value of 0.82. We conclude that experimental error has a non-negligible effect, but not substantial enough to render the comparison meaningless. With that in mind, we proceed to examine Δ^3J_{HNHA} between A β 40-D23N and A β 40 in Figure 4.6C. Like with the ΔCS s discussed in the previous section, OPLS produces an overall negative PCC of -0.31 for this comparison, while ILDN produces a weak positive correlation of 0.28; OPLS also produces a larger RMSD than ILDN (0.41 Hz vs. 0.28 Hz). Negative correlation for Δ^3J_{HNHA} occurs for D/N23 in OPLS but not ILDN, which reproduces the trends we saw for ΔCS for N and C α nuclei.

This is the largest experimental Δ^3J_{HNHA} , and the most egregious disagreement with OPLS. The effect of the mutation on $^3J_{\text{HNHA}}$ for adjacent residue V24, meanwhile, is increased for both simulations and experiment; however, the magnitude of this change is overpredicted for OPLS. Further, OPLS is also the only force field to anticorrelate Δ^3J_{HNHA} for residues K16 and L34-V36. The disagreement in the latter region is interesting, because one of the conclusions we made with OPLS simulations in Chapter 2 is that the D23N mutation narrows the extent of the central bend and promotes earlier registers for β interactions, shifting the distribution of β content in the C-terminal hotspot toward residue N27-G29 and reducing content at G33-V36, which may account for the reduction in Δ^3J_{HNHA} for the OPLS set. That the experimental data disagree with this trend consistently among these C-terminal residues may cast doubt on this model of hydrogen bonding. ILDN does not make such a dramatic change in β character of this region, and it performs better at reproducing the experimental Δ^3J_{HNHA} . This said, the ILDN model is not perfect either, overpredicting the changes in several values, and exclusively anticorrelating Δ^3J_{HNHA} of certain residues, even ones close to the mutation such as V18 and A21.

While OPLS has produced rather serious flaws in the comparison of Δ^3J_{HNHA} , other couplings such as RDCs and $^3J_{\text{NHB}}$ both actually support this as the superior model compared to ILDN. Positive correlations are seen for all comparisons, but in general, PCCs are somewhat higher for OPLS simulations compared to ILDN. RMSDs are also lower for OPLS in comparisons involving A β 40. In general, this may suggest that OPLS may better model the distributions of side chain angles, as well as the overall global orientation of the molecule. Nevertheless, recall that OPLS was also substantially worse at modelling the secondary chemical shifts of C β atoms (see previous section) for both A β 40 and A β 40-D23N, lending scepticism to the former conclusion. Also, experimental RDCs and $^3J_{\text{NHB}}$ were only collected for wild type A β , so it is unclear if these observations would hold in the case of FAD mutations.

4.4.2.3 Hydrogen Bonding and Urea Titrations

Cosolvents can be used to perturb the ensemble of A β ; in particular, our lab has investigated the effect of urea. Urea is a commonly used, potent denaturant of globular

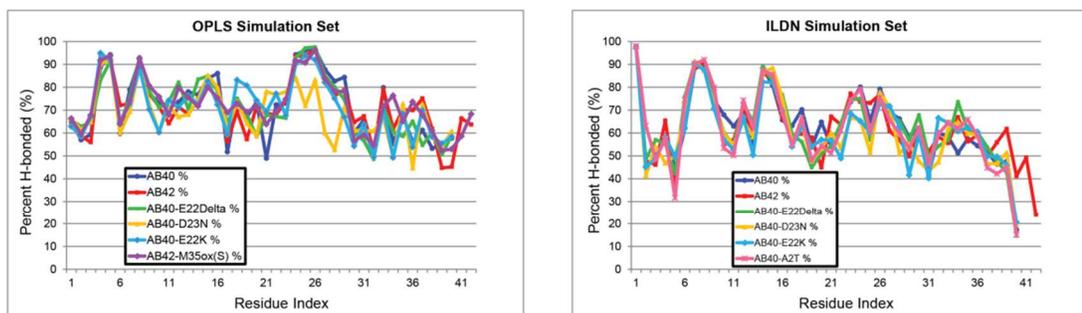
proteins, but its mechanism of action is just beginning to become understood. Experimental evidence supports a mechanism where urea interacts directly with the protein, but simulations reveal that both dispersion and electrostatic interactions with urea are involved(413, 458). Furthermore, both direct backbone and side chain interactions contribute to the interaction coefficient of the molecule(414). It is clear from these studies that urea ultimately works to disrupt the pre-existing electrostatic and hydrophobic interactions of the system.

Previously, our lab has calculated the effects of urea on the chemical shift perturbation (CSP, see Methods for definition) of amide groups of A β 40 and other FAD mutants, including A β 40-E22 Δ (54). Since A β is a mostly solvent exposed IDP with only transient structure (see Chapter 1, Section 1.2.7), we surmised that changes in the CSP of these backbone amides would not be impacted strongly by loss of a global fold or changes in solvent exposure, but rather, would be dominated by the loss or gain of hydrogen bonds (ie, loss of intrapeptide bonds and gain of peptide-urea bonds) involving those amide groups(54). On the other hand, changes in certain side chain CSPs may better reflect the changes in nonpolar interactions induced by urea(54). In the future, we may acquire ¹H-¹³C HSQCs or other experiments to examine these effects in the side chains, but for now, these data are currently unavailable.

First, to study the nature of hydrogen bonding involving these amides, we calculated the percent of frames in which each backbone amide of A β is involved in an intrapeptide hydrogen bond, illustrated in Figure 4.7A (i.e., hydrogen bond occupancy). OPLS and ILDN simulations exhibit rather different profiles for hydrogen bonding; given the propensity of the two simulation conditions to sample very different local electrostatic interactions as described in Chapters 2 and 3 in spite of the shared structural similarity, this came as no surprise. Therefore, comparison with a reliable experimental reporter on this attribute may be able to distinguish between the two force fields.

With that in mind, we measured the CSP caused by urea at 2 M compared to 0 M for both A β 40 and A β 40-E22 Δ , then compared the difference between species in the perturbations caused by urea (Δ CSP). As we have not performed any simulations in urea, this Δ CSP was instead just compared to the changes in hydrogen bond occupancy

A



B

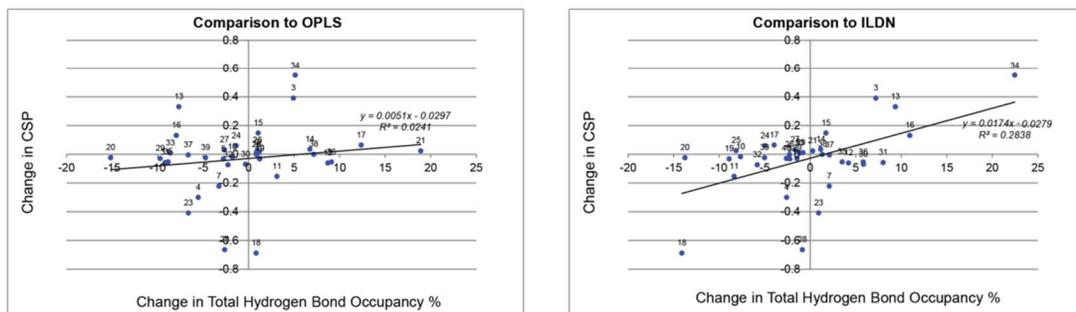


Figure 4.7. Comparison of the perturbations on chemical shift caused by urea versus hydrogen bond occupancy in simulations between different A β species. (A) Percent of frames in which the backbone amide of the current residue is involved in at least one intramolecular hydrogen bond, for the OPLS and ILDN simulation. (B) Scatter plots comparing the difference in this percent of frames where amides are involved in hydrogen bonding between A β 40-E22 Δ and A β 40 with the difference in chemical shift perturbation caused by 2M urea between these two species (ie, the Δ CSP). PCC of this comparison for OPLS is 0.16 and for ILDN is 0.53. Experimental data provided by Dr. Christopher Connors.

of simulations of the two different species. Unlike the chemical shifts and $^3J_{\text{HNHA}}$ comparisons above, this is not a direct comparison of equivalent phenomena on the computation and experimental side. The comparison is also imperfect, not only because of the associated errors with calculating a ΔCSP , but also because, on the simulation side, we are discounting possible hydrophobic and conformational-related interactions on amides that may affect CSP and we are also ignoring the direct interactions of urea on the peptide (since it has not been modelled). As this is the best we can do with our current data, we proceed with this analysis, depicted in Figure 4.7B. Despite the shortcomings, the change in hydrogen bonding between species is able to capture some linear dependence with the ΔCSPs , and it distinguishes ILDN as the better model, with a PCC of 0.53 vs a weaker PCC of 0.16 for the OPLS set. In particular, where experimental ΔCSP is highest (E3, H13, K16, V18, L34), the trend is well captured by the change in hydrogen bond occupancy in ILDN. Outliers in this regard are D7, D23, and K28. Like with our other metrics, much of the deviation in correlation lies in over-prediction on the side of the simulations. Between the species, many residues show changes in hydrogen bonding propensity in simulation, but they do not have a strong change in CSP caused by urea; this is particularly true for the OPLS model. In general, this may serve as more evidence that ILDN models the local electrostatics of the system better than OPLS, though other factors (i.e., overall hydrophobicity of the peptide) could also be affecting these profiles.

4.4.3 High Pressure NMR of A β Monomers

High pressure also generally denatures proteins, but in a different manner, usually favoring lower volume states(415). High pressure NMR has previously found extensive utility in the study of globular proteins, including measurement of thermodynamic parameters, volume changes, and structural data about the intermediates of folding(415, 441, 459-463). If pressure also “unfolds” the partially structured ensembles of IDPs (as is suggested through previous studies (410, 420)), the resulting effects may allow us to identify residual structure and interactions in the ensemble and check the conclusions we made through simulation. High pressure can also be used to control and study the aggregation and dissolution of amyloidogenic proteins; we refer the reader to the book

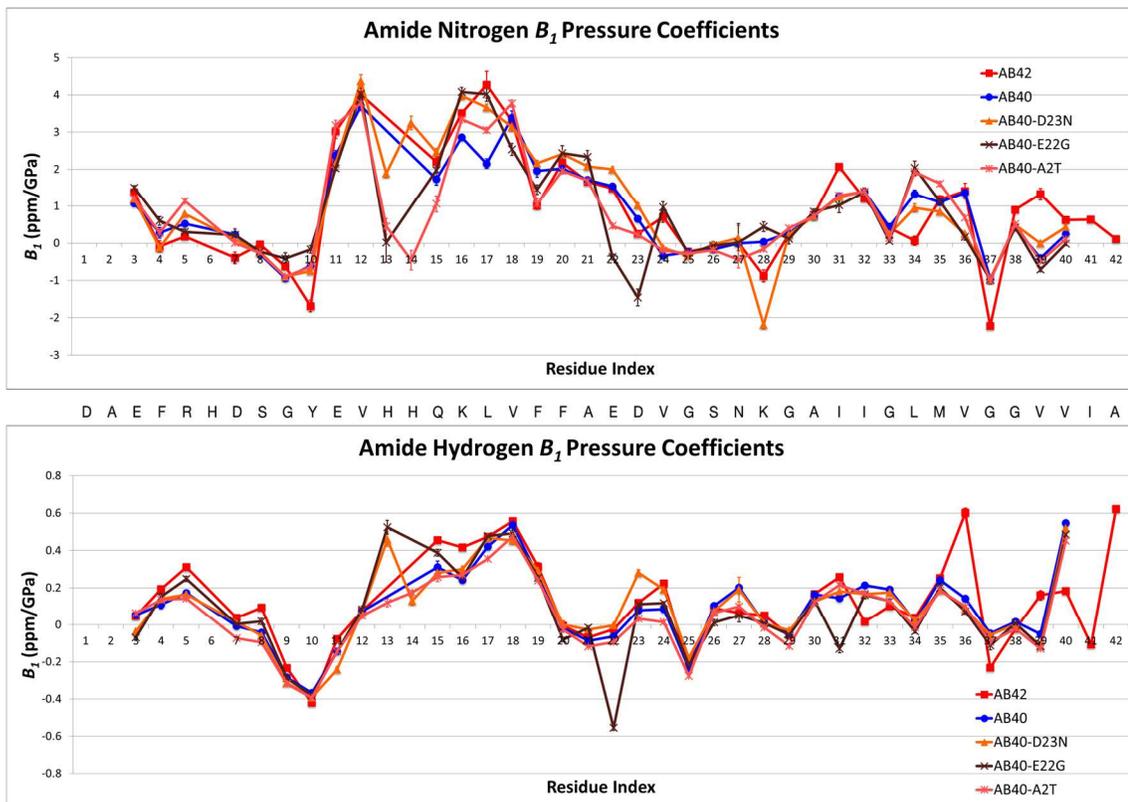
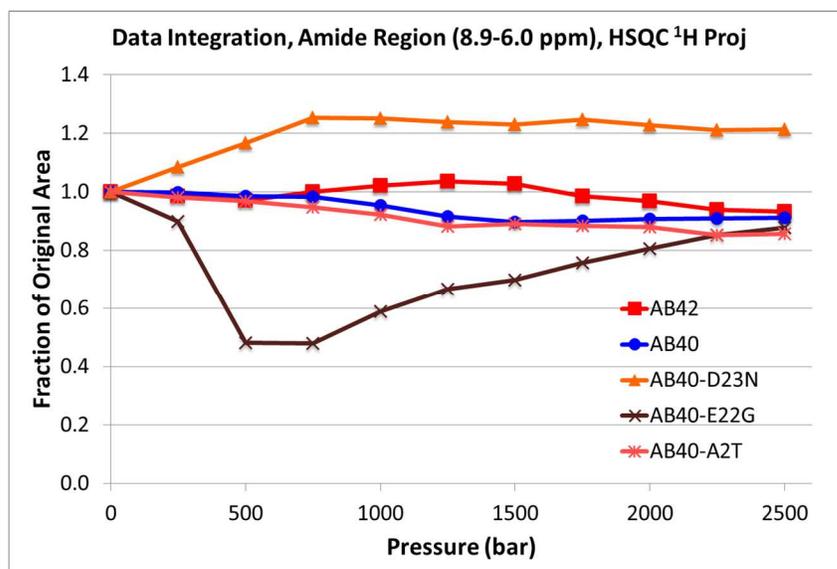
chapter written by Kim *et al.*(417) and the reviews by Meersman *et al.*(418, 419) for a comprehensive examination of this topic. Among the techniques that can be employed to make these observations, changes in aggregation can be monitored through loss of signal in solution NMR because only low molecular weight signals are typically observable (for example, as we discussed, signals in solution NMR are almost exclusively represented by the monomer form of A β (337)). This has been done, for example, by Kamatari *et al.*(416), who used pressure-jump NMR to study aggregation of an intrinsically disordered variant of hen lysozyme.

Here, we expand on the previous work of Munte *et al.*(410) to define the biases in the pressure sensitivity of chemical shifts and solvent exchange rates over the residues of different monomeric A β species. Chemical shifts of A β 40, A β 42, A β 40-E22G, A β 40-D23N, and A β 40-A2T were measured using ^1H - ^{15}N HSQC spectra ranging from ambient pressure to 2500 bar at 250 bar increments; they are illustrated in Figure 7.4 in the Appendix. Both the changes in the positions and intensities of the HSQC resonances as a function of pressure are useful sets of information. The differential changes in peak position with pressure indicate the relative chemical sensitivity of amide groups to the effects of pressure. Meanwhile, changes in peak intensity with pressure can reflect an amalgam of different phenomena, including chemical exchange, solvent exchange, and aggregation (the latter being particularly important for A β). In addition to the multifactorial influence underlying peak intensities, these values are particularly prone to being affected by spectral overlap and may also be decreased with experiment time for these A β systems (unlike peak positions) due to aggregation, making them very difficult to interpret. Nevertheless, both types of data will be explained here in order to present the most comprehensive analysis possible from the measurements.

Focusing on the effects of pressure on chemical shift position first, we begin by noting that in all A β ^1H - ^{15}N HSQC (Figure 7.4 in the Appendix), only one set of resonances is observed at any given pressure. This implies that all peaks of A β undergo fast exchange with the pressure perturbation. Slow exchange is observed for proteins that undergo large conformational changes with pressure(463, 464), however, this phenomenon is unlikely occurring in any A β system. Next, we observe a general trend in pressure where almost all resonances progressively move downfield in both nitrogen and

hydrogen chemical shifts, indicative of increased deshielding. The shortening and straightening of hydrogen bonds with pressure has been implicated in this trend(463, 465). The few exceptions are the hydrogen shifts of the side chain of R5, and the G22 mutation in A β 40-E22G (which, at ambient pressure, is already unusually upfield in both nuclei). Our HSQC data for A β 40 also qualitatively reproduce the spectra collected by Munte *et al.*(410) very well. While the analysis of peak positions in the next paragraph will be based on HSQCs collected with ascending pressure, it is also important to note that the positions of peaks for all A β are identical when descending in pressure; that is, the changes in chemical shift induced by pressure are completely reversible. We also observe that equilibration of A β chemical shifts with pressure is very fast (at least as fast as 1 minute), and also do not change on the hour-long timescales employed in this experiment. Therefore, unlike peak intensities, analysis of resonance positions allows us to consider the effect of pressure on chemical shifts while disregarding other variables, such as the effect of experiment time and aggregation.

Having established these caveats, the effect of chemical shift positions were analyzed more carefully by removing the pressure dependent shift changes for both nuclei observed in random coil model peptides(445) and fitting to a second order Taylor expansion, as described in the Methods (Equation 4.4). The linear B_1 pressure coefficient for each species and nucleus is illustrated in Figure 4.8A, while the nonlinear B_2 coefficients are printed in Figure 7.5A in the Appendix. We observe that B_2 , for each value, almost always anticorrelates to their respective the B_1 value with similar relative magnitudes, and thus contribute little extra to the discussion; further, unlike B_1 , the B_2 term is observed to vary widely only in the context of globular proteins and not for small peptides(463). Lastly, consider that because we are removing random coil shifts at each pressure, negative B_1 pressure coefficients are possible, and represent a greater resistance to downfield changes in pressure relative to random coil behavior. Focusing on B_1 values (Figure 4.8A) and looking at just the values of wild type A β 40 first, we see clusters in the pattern of B_1 coefficients that match the secondary structure characterizations of the peptide made in Chapters 2 and 3 through simulation (see Figure 2.2 and Figure 3.2). For example, the region V24-G29 is generally characterized by small nitrogen B_1 values,

A**B**

<Figure continued on previous page>

C

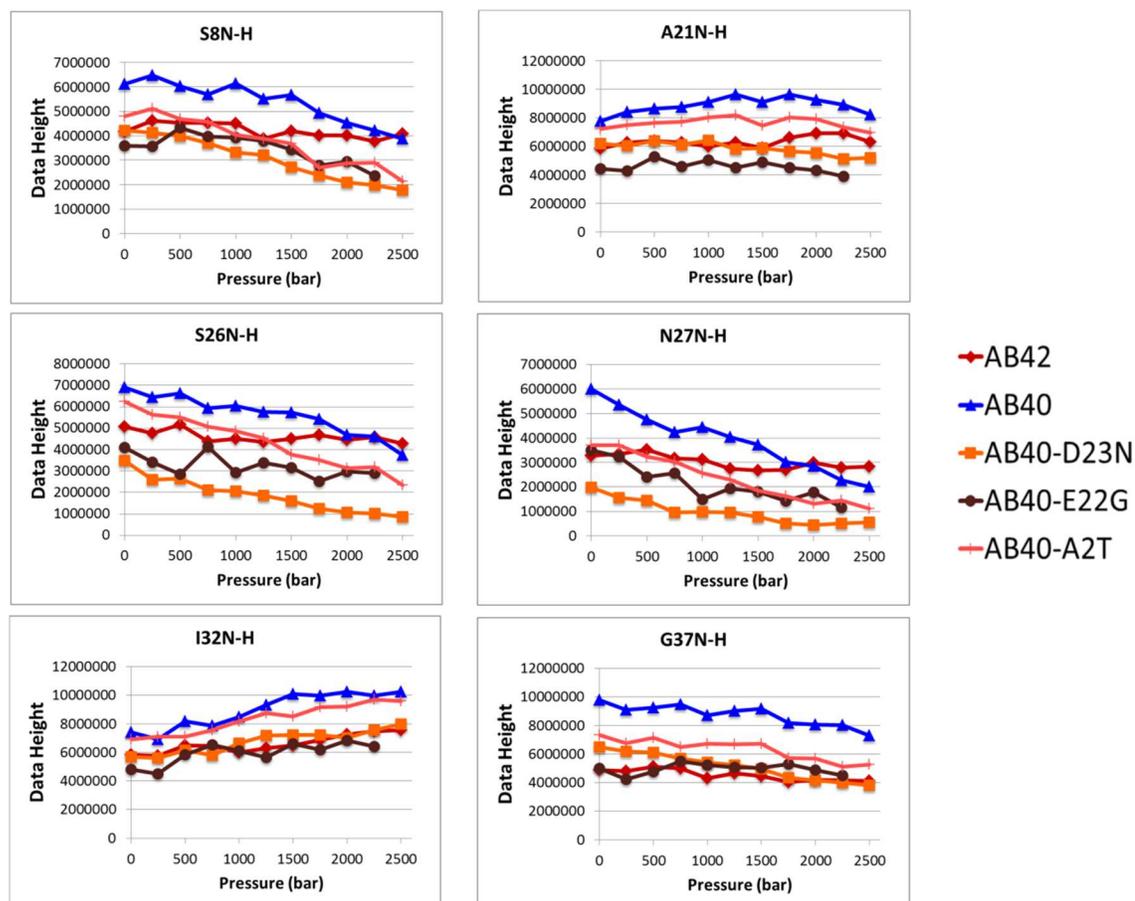


Figure 4.8. Effect of pressure on the chemical shifts and peak intensities of monomeric A β . (A) Plot of the first order (B_1) pressure coefficients for the amide nitrogen and hydrogen nuclei for different species of monomeric A β . B_2 coefficients of the second order Taylor polynomial (Equation 4.4) are presented in Figure 7.4 of the Appendix. The sequence of wild type A β 42 is presented in between the two graphs. (B) Effect of pressure (and experiment time, see the main text for explanation) on the overall monomer signal, as determined by integrating over a 1D projection of the entire amide region (8.9-6.0 ppm) of each ^1H - ^{15}N HSQC spectrum. (C) Plot of the intensity of select residues, as a function of pressure (and time). All residues chosen for this panel do not experience spectral overlap. Data heights were corrected by the overall loss of signal to better observe residue-specific loss (see Panel B). The point at 2500 bar is not presented for A β 40-E22G because the HSQC collected at this final point was interrupted by a pump failure (see Methods). Coefficients and data heights are based on the position and magnitude of peaks, respectively, on ^1H - ^{15}N HSQC spectra collected with high pressure NMR, ranging from ambient pressure to 2500 bar at 250 bar increments. Data collected on a 600 MHz spectrometer at 277 K, using the high pressure pump and buffer conditions as described in the Methods. Peak positions were corrected with random coil shifts as a function of pressure, as described in the text.

while for hydrogen, the region of low B_I encompasses F20-G29 except N27. These pressure-insensitive regions are annotated with turn/bend structure in the OPLS and ILDN models. Meanwhile, the central hydrophobic cluster, denoted as strongly β sheet prone by both OPLS and ILDN, is also characterized by high sensitivity to pressure in Q15-A21 for nitrogen and in Q15-F19 for hydrogen. Meanwhile, the residues of A30-V36 are also characterized by high B_I values (with the exception of G33 for nitrogen and L34 for hydrogen), albeit significantly less than the CHC; this is the second strongly β sheet-prone region shared by both force fields. Finally, the C-terminus of A β 40 is largely resistant to pressure changes, except the final residue's hydrogen atom, matching a shared characterization of the region as largely disordered. It has been suggested that the final C-terminal residue may have a high hydrogen B_I pressure coefficient due to changes of the protonation of the carboxyl group with pressure(410). Thus, the pressure coefficients are consistent with the generalized model of A β we derived through simulations with multiple force fields. However, the coefficients offer insight into the contentious identity of the N-terminus. The most distinctive feature in this region are the negative B_I hydrogen coefficients at G9-E11, indicative of resistance to deshielding with pressure. H6-G9 is strongly prone to turn/bend character in ILDN and OPLS, which may explain these values. E11/V12 also have very high nitrogen B_I but low hydrogen chemical shift change; these are weakly characterized by β structure in the simulations. The intervening H13/H14 show poor signal to noise in the HSQCs (likely due to solvent/chemical exchange), so their trends in pressure coefficient are unclear. The extreme N-terminus (E3-D7) is characterized by moderate pressure sensitivity (higher for the nitrogen at E3 and for both nuclei at R5), however, it is difficult to conclude if these data distinguish between the β -structured N-terminus of ILDN and the weakly structured characterization made by OPLS.

In general, despite it being a simplified representation of the ensemble, we can successfully use the overall ensemble β sheet content of A β 40 in REMD simulations to explain much of the variation in the magnitude B_I coefficients. Figure 4.10 provides a quantitative comparison of these variables, resulting in PCCs of 0.5-0.6 for both force fields and for both nitrogen and hydrogen NMR data. We can also link some of the trends in B_I with pressure-related changes to the REMD ensemble of A β 40 under high

pressure, as examined in Chapter 3. Figure 3.10D in that chapter reveals that L17-D23 and I31-G33 are regions particularly prone to lose β character when going from the high pressure to the 1 bar REMD ensemble, primarily in favor of coil/irregular structure. Save for G33 for nitrogen and F20-D23 for hydrogen, these regions are also affiliated with high B_I values. This said, residues N27-G29 are also associated with substantial β character loss, but have low pressure sensitivity. Meanwhile, E11-V12 and V36 are also regions of high nitrogen B_I , but they are affiliated with the opposite trend – a gain of β character in the high pressure simulation. Due to these anticorrelations, an overall comparison of $\Delta\beta$ content (content in the 1 bar case subtracted by that of the 2500 bar case) actually yield overall weak PCCs of 0.12 and 0.35 for nitrogen and hydrogen B_I , respectively. However, the regions that anticorrelate (E11-V12, N27-G29, V36) are also more strongly typed by coil/irregular structure in the 2500 bar ensemble, and exchange these structures for more turn/bend character in the ambient pressure case, rather than β content. Loss of high curvature structure may thus still account for the heightened pressure sensitivity of these residues.

The four other species of A β studied here generally show very similar profiles of chemical shift variation with pressures as A β 40; this said, we will now discuss the most salient differences in B_I between these species. We begin by noting that, although we have included data for H13/H14 for certain A β species where available, the signal to noise ratio for these residues is characteristically very low for A β peptides for the buffer conditions used in this experiment, presumably due to heightened solvent and/or chemical exchange. Further, although we have corrected for random coil effects in our method, the protonation state of these histidine side chains may also change due to pressure, which may affect the coefficients observed here. Due to these complications, we will assign no meaning to the large variation seen among the different species at this spot in the remainder of this section.

Comparing A β 42 with A β 40 at other regions reveals significant changes in pressure coefficients, particularly at the C-terminus. Ignoring comparisons involving the very terminal residue (V40 or A42) for either peptide (which, as we covered above, is consistently characterized with high hydrogen B_I pressure coefficient that may be induced by protonation changes to the C-terminal carboxyl group(410)), we observe

that: V36 is characterized by a large increase in hydrogen pressure coefficient (and a subtle decrease in nitrogen coefficient), G37 becomes less susceptible to nitrogen and hydrogen movement, while V39 is marked by higher sensitivity to pressure for both nuclei. We saw previously that other turn-implicated regions (as defined by our simulations) have low pressure coefficients, while more structured regions, particularly those prone to intrapeptide interaction, are associated with high pressure sensitivity. One model, then, that may be consistent with the observed trends in pressure coefficient of A β 42 versus A β 40 is the sampling of a new C-terminal β hairpin at ambient pressure that imposes turn character on G37 while promoting the hydrogen bonding of V36 and V39, a model consistent with our characterization of this alloform through simulations in Chapters 2 and 3. Indeed, direct backbone hydrogen bonds between V36 and V39 are observed in both OPLS and ILDN models of A β 42 but not A β 40. Specifically, using the donor-acceptor/acceptor-donor nomenclature from Chapters 2 and 3, we see populations of 15.6/0.1% for OPLS A β 42 and 4.9/9.0% for ILDN A β 42, but only 0.5/0.0% for OPLS A β 40 and 0.2/1.3% for ILDN A β 40. However, this does not necessarily vindicate the REMD simulations, since both force fields predict that the in-register bonds between L34 and I41 are also well populated (15.9/16.7% in OPLS A β 42, 4.9/9.0% in ILDN A β 42), but both of these amides are mostly pressure-insensitive in A β 42. Nevertheless, these C-terminal changes in pressure coefficients of A β 42 suggest a new topology at the C-terminus in the monomer form. Lastly, residue Q15 and K16 also show a higher hydrogen coefficient for A β 42 compared to other forms; these residues may also be more involved in hydrogen bonds in this form.

Other changes in B_I among the different A β species are also observed. We note that A β 40 shows reduced susceptibility to pressure in the nitrogen dimension compared to all of the other A β forms at residues K16 and L17. Most of the other forms studied here are known to promote A β aggregation (A β 42, A β 40-D23N, and A β 40-E22G), and increases in the structural order K16 and L17 in the CHC, as indicated by these coefficient changes, may be involved in that observation. Besides A β 42, A β 40-E22G also produces some distinctive changes. For one, residue 22 is characterized with very low B_I for both nitrogen and hydrogen in this mutant, even when compared to other glycines in the sequence. The adjacent D23 is also characterized with very low nitrogen chemical shift

movement, while V24 is more sensitive than wild type A β 40. I31 also shows a substantially reduced hydrogen B_1 coefficient in this mutant; the reason why is currently unknown. G22 in A β 40-E22G, an already unusually upfield residue in the HSQC at ambient pressure, is affiliated with the only backbone amide signal of all the A β forms studied here to ever move upfield in the hydrogen dimension with pressure (Figure 7.4 in the Appendix). In REMD simulations of A β 40-E22G, described in Chapter 6, we observe that E22G both exchanges β character for coil/irregular annotations and promotes disorder of backbone amide groups (as gauged by S^2 order parameters) of residues local to the mutation site, relative to wild type A β 40 (see Figure 7.9 in the Appendix). That the E22G substitution promotes disordered coil is experimentally supported by the lowered B_1 coefficients of G22 and D23 relative to wild type (Figure 4.8A), which suggest the region is less prone to the denaturing effect of high pressure. The significantly lowered R_2 relaxation values near the site of the mutation in this variant compared to wild type(54) also indicate that this mutant promotes greater local flexibility in the NMR tube. Other changes in B_1 caused by the mutation, such as the gains at K16-L17, may also be in line with the shifted distributions in β character caused by the E22G mutation in simulation (Figure 7.9D in the Appendix).

Meanwhile, D23N has a more subtle effect on pressure sensitivity, primarily increasing B_1 near the site of the substitution, primarily for residue 23 in the hydrogen dimension (Figure 4.8A). This increase in pressure susceptibility near the mutation site could be consistent with an observed increase in β structure near the mutation seen in both OPLS and ILDN simulations. Meanwhile, the nitrogen pressure sensitivity is substantially reduced for residue K28 in the A β 40-D23N variant. Recall that the D23N mutation disrupts a D23-K28 salt bridge that normally exists in wild type, striated ribbon fibrils(23, 251). That D23N induces a change in the pressure coefficient of residue K28 in the monomer form may be indicative of a loss of this interaction caused by the mutant. This mostly disagrees with REMD simulations, where OPLS primarily implicates E22 as the partner to K28 in salt bridge for A β 40, while both E22-K28 and D23-K28 salt bridges are only modestly populated in the ILDN model (Table 2.2 and Table 3.2). Note that this decrease in K28 pressure sensitivity is not seen for A β 40-E22G, which might mean that E22G does not disrupt salt bridge formation (implying

that either E22 is not an ionic partner for K28 in wild type monomers at ambient pressure, or D23 takes over E22's role in the mutant). All said, this is a change in pressure coefficient seen in only one residue (K28), and we caution that this K28 peak often overlaps with D7 in both wild type and D23N HSQCs; while site-specific assignment is possible at ambient pressure and we did our best to track the movement of this peak with pressure, assignment error is still possible. We also point out that besides these changes, A β 40-D23N shares a fairly similar profile to that of wild type A β 40, inconsistent with dramatic changes in central region electrostatics and cross hairpin hydrogen bonding seen especially in the OPLS model. The last variant, A β 40-A2T, has a very similar profile of pressure coefficients to the wild type, except at the CHC, as previously mentioned. Surprisingly, though residue 2 is not observable in the HSQCs (presumably due to solvent exchange), nearby downstream residues of the N-terminus that are visible (E3 and onward) are similar in their response to pressure.

Panel B and C of Figure 4.8 contain information about the peak intensities from the HSQCs. As mentioned, analyses of these data are challenging due to the effects of spectral overlap and external factors. Panel B attempts to study the latter by plotting the overall changes in integration of signal of the entire HSQC ^1H projection over the amide region (8.9-6.0 ppm). First, we caution that, although the plot maps data height to pressure, it is implied here that the x -axis is actually a convolution of both pressure and experiment time. As discussed in the Methods, each pressure data point was collected as a successive HSQC in ~ 1 hour intervals, except for A β 40-E22G, which required ~ 2 hour HSQCs to acquire (done to mitigate insufficient signal to noise ratio). Next, although other effects could change the signal intensities of this region (for example, solvent exchange), we suspect that by averaging over many peaks in the amide region, this metric is more reflective of large overall changes in the data intensity, which could largely be driven by peptide aggregation. The effect of solvent exchange on each residue is examined in more detail later in this discussion.

With these considerations in mind, we observe in Figure 4.8B that A β 40-E22G undergoes an interesting change wherein signal is rapidly lost until ~ 500 bar, then intensity progressively begins to recover, reaching values approaching the starting intensity at the end of the experiments. We hypothesize that the first phase's loss of

signal is due to aggregation of this species over time, while the recovery in the second phase is due to a disaggregation effect caused by high pressure. High pressure has been previously observed to disassemble other aggregates of intrinsically disordered proteins, thereby recovering intensity of monomer-associated peaks in solution NMR(416). It was further observed that aggregation occurred very rapidly when the A β 40-E22G sample was equilibrated to 1 bar after collecting the data for Figure 4.8B, with only 2% of the starting monomer signal remaining after six hours at ambient pressure and 277 K (*data not shown*). The resulting sample possessed both soluble and insoluble aggregates, and atomic force microscopy (AFM) of this aggregated A β 40-E22G sample confirmed small, narrow, straight structures, matching characteristic protofibrils and fibrils of A β 40-E22G reported in the literature(270, 277)(Figure 7.6 in the Appendix). The disaggregation effect caused by pressure may also be implicated in A β 40-D23N's recovery of signal seen in Figure 4.8B, where intensity is progressively gained until ~120% of the original value at 750 bar, then maintained throughout the experiment. A β 40-D23N can associate into novel fibril forms(247, 251, 282); it is possible the disaggregation protocol employed in the Methods is imperfect for these structures, and increased pressure has disrupted pre-existing aggregates in the sample. Finally, A β 40 and A β 40-A2T demonstrate similar profiles marked by a very subtle and progressive loss of signal, while A β 42 marginally recovers some overall intensity at ~1250 bar, which is then lost at later stages of the experiment. Clearly, more controlled experiments to deconvolute the effect of pressure and the effect of experiment time/aggregation are necessary to properly study the aggregation and disaggregation of A β in this context. As described in section 6.2.2, our lab is planning on running HSQC series using multiple HFIP-NaOH disaggregated samples (see Methods) starting at different pressures to study aggregation at different pressures without starting time as a confounding variable. Further, we plan to run equilibrium pressure-jump NMR experiments akin to Kamatari *et al.*(416), who reliably used both jumps to high pressure to disassociate aggregates into monomers and jumps back to low pressure to reassociate these peptides, allowing them to study the kinetics and mechanisms underlying these processes. This said, the findings presented here serve as a prelude to these more rigorous investigations, demonstrating that A β aggregation can be monitored and manipulated through high pressure NMR.

While Figure 4.8B was an analysis of the overall changes in signal with pressure and time, Figure 4.8C plots the effect of pressure on certain residues while discounting these global changes. Again, we caution that the *x*-axis is actually a convolution of pressure and experiment time, as described in the previous paragraph; however, by dividing the data heights by the integrations in Panel B, we are removing overall effects such as aggregation, which is the primary way by which experiment time is expected to change the data heights. Panel C shows changes in data height for a select few residues, chosen because they avoid spectral overlap in the HSQCs for all species over all pressures (and also because they are not directly mutated in any of the peptides studied here). Changes in data height after dividing by changes in overall signal are summarized for all residues in Figure 7.5B in the Appendix as the slope of a linear fit over all pressures, however, some of the data height trends are compromised by spectral overlap and ambiguous assignment in at least one of the HSQCs, and there is large variation amongst the different A β species (unlike the position-related pressure coefficients discussed above). Nonetheless, examination of these slopes leads to an interesting observation for wild type A β 40: turn characterized regions of the peptide as described by simulation (see earlier in this section) are affiliated with decreases in data height (for example, D7-G9, S26-G29, V36-G38), while other parts of the peptide that were characterized with structure above actually gain signal (ie, Y10-V12, L17-D23, I32). Many of the residues affiliated with the former loss of signal, in particular, are unoverlapped and have a low error affiliated with a linear fit; some of these resonances (S8, S26, N27, G37) are shown as examples in Figure 4.8C. Meanwhile, other unambiguous resonances that are not associated with these regions are depicted for contrast (A21 and I32). Some interesting observations follow from these graphs. Among the turn-associated residues that decrease in data intensity with pressure, A β 42 is notably the only species to be mostly protected from this loss. Further, we see that, even though progressive losses in data height with pressure are observed for all of the 40-residue A β species studied here, not all of the species start with the same data height. For example, A β 40-D23N starts with a uniquely reduced data height compared to wild type, particularly for S26 and N27, which are then further reduced with pressure. A reduced apparent content of A β monomers might explain some of the ambient pressure signal

reduction affiliated with this mutant, however, the starting data height of this mutant for other residues like A21 and I32 is far less reduced compared to wild type.

Previously, the CLEANEX-PM-fHSQC experiment(223, 224) was run at ambient pressure on A β monomers(225, 226). Examining these data, we note that many of the same residues we implicate as turn susceptible above were also characterized with substantially increased solvent exchange relative to the rest of the peptide in A β 40 monomers. Furthermore, it was also observed that A β 42 demonstrated dramatically reduced exchange rates compared to A β 40, particularly in these turn regions(225). Based on this information, we suspected that all of the residue-specific changes in data height with pressure discussed in the previous paragraph were related to solvent exchange, with the following hypotheses that work together to explain the data: 1) solvent exchange rate of each residue increases proportionally with pressure, leading to the progressively reducing data heights with pressure – particularly in loop regions that already have high solvent exchange at 1 bar, and 2) unlike A β 40, A β 42 shows a resistance to decreasing data height with pressure, along with substantially reduced solvent exchange for almost all residues at 1 bar(225), so we hypothesized that this species forms unique, NMR-invisible (ie, “dark state”) aggregates that monomers can exchange with, and that this dark state protects against signal losses in the monomer due to solvent exchange, particularly as pressure increases. Exchange of A β monomers with the NMR dark state has been investigated using ¹⁵N dark-state exchange saturation transfer (DEST)(397). For A β concentrations on the 100 μ M scale (similar to the concentrations used in the experiments discussed here), it was observed that the apparent rate constants for exchange of both A β 40 and A β 42 monomers onto aggregates of the dark state is on the seconds⁻¹ scale while the reverse “off” rate occurs on the magnitude of the (tens of seconds)⁻¹(397). The rates of solvent exchange for A β of similar concentration occur exactly on this same seconds⁻¹ to (tens of seconds)⁻¹ scale(225), so it is possible that exchange with the dark state could also affect the apparent solvent exchange rates.

To directly test the two hypotheses above, we used CLEANEX-PM-fHSQC to observe solvent exchange rates for each residue of A β 40 and A β 42 as a function of pressure (Figure 4.9). This is the first time, in our knowledge, that this pulse sequence

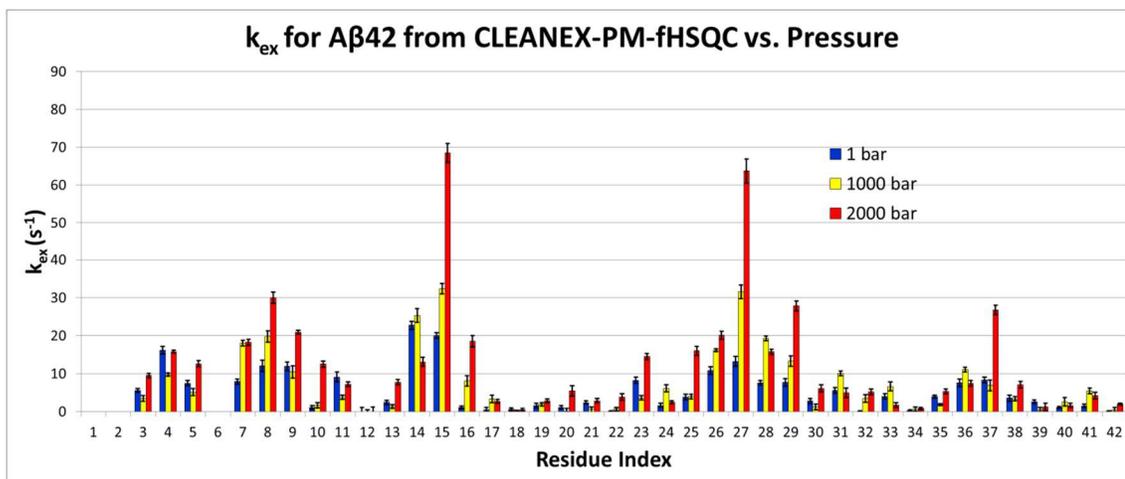
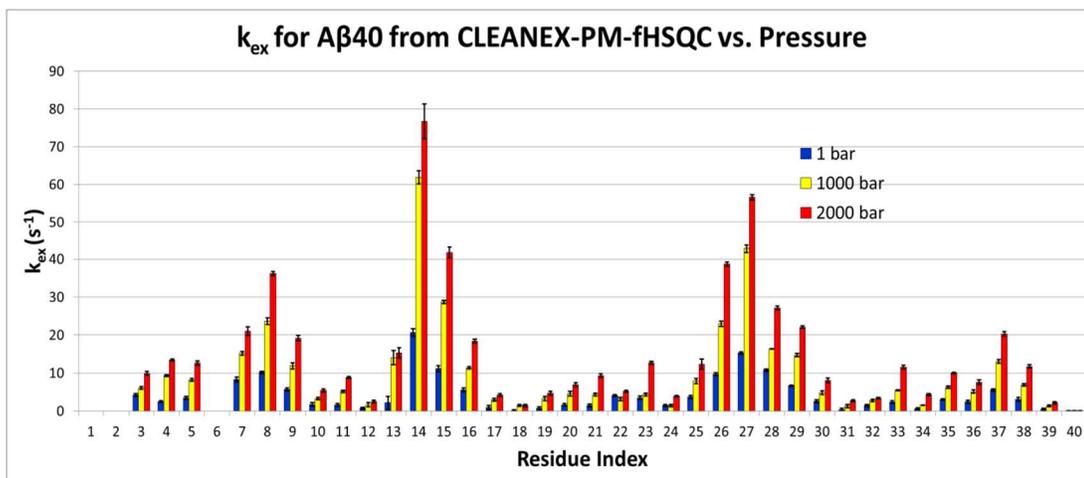


Figure 4.9. Solvent exchange rates (k_{ex}) for each residue as a function of pressure for monomeric A β 40 and A β 42, measured through CLEANEX-PM-fHSQC(223, 224). Data was collected at 277 K on a 600 MHz spectrometer, using buffer conditions as described in the Methods.

been applied to measure solvent exchange rates for peptides as a function of pressure, though amide H/D exchange has been used for this purpose in the past(466-468). It has been noted in the Methods section (Section 4.3.3.6) and in Figure 7.7 of the Appendix that the experimental timescale involved in the experiment did not cause any substantial directional changes in measured k_{ex} of the monomer form at ambient pressure, and that, although going up to 2500 bar does increase the pH of the sample an estimated 0.14 units compared to 1 bar, this change would only account for a small proportion of the increases in k_{ex} observed in Figure 4.9. The solvent exchange measurements in Figure 4.9 are consistent with previous studies that employed the same pulse sequence(225, 226), especially in the observation that turn associated regions of A β are affiliated with the highest solvent exchange rates. The magnitudes of the exchange rates at ambient pressure also match these other studies; however, there are some subtle differences between these data that are likely due to inconsistencies in assignment. In particular, for A β 40, we characterize D7-G9, H14-K16, S26-G29, and G37-G38 as the regions affiliated with the highest recovery of CLEANEX-PM-fHSQC signal at all pressures (we observe that A β 42 has a roughly similar profile). In contrast, Rezaei-Galeh *et al.*(225) observe that A β 40's H14 and Q15 possess lower exchange rates while D23 has a very high relative recovery compared to the other residues. This characterization is unsurprising because H14's assignment is often ambiguous, while Q15 and D23 are typically overlapped, but we are more confident in our assignment because of the higher sample concentration used here (100 μ M here vs. 30 μ M in the Rezaei-Galeh *et al.* study). We also point out that the locations of highest solvent exchange rates for A β monomers observed here also match the locations of high solvent exchange (H13-Q15, G25-G29, and G37-G38) made by Ahmed *et al.* for a neurotoxic oligomer of A β 42 through amide H/D exchange, leading them to classify these regions as turn prone in their molecular model(26). Next, we observe in Figure 4.9 for both species that pressure increases the solvent exchange rate overall, in a manner that is roughly proportional to the original exchange rate at 1 bar (that is, residues with larger k_{ex} at 1 bar also have larger rates of change in k_{ex} with pressure). Previous studies have observed through amide H/D exchange experiments that solvent exchange rates increase and protection factors decrease overall with pressure(466-468). For example, it has been previously

observed for H/D exchange of T4 lysozyme for nine different points from 1 to 2000 bar that k_{ex} increased for all but 1 of the 44 studied amide groups(466). However, Hitchens and Bryant also observed that k_{ex} related exponentially with pressure at these sites(466), whereas the data in Figure 4.9 seem to suggest a linear relationship between these variables for most amides (though it is difficult to say since we only have three points for the pressure variable). Either way, our findings support “hypothesis 1” made in the previous paragraph: solvent exchange rates increase both substantially and proportionally with pressure, which can account for both the progressive loss of data height with pressure seen for A β 40 and the accelerated loss observed in turn affiliated residues in particular. However, unlike Rezaei-Galeh *et al.*(225) who found that A β 42 was largely protected from solvent exchange compared to A β 40, we observe that monomeric A β 42’s solvent exchange rates are actually comparable to A β 40 in magnitude and trend at 1 bar. Why this disparity exists is a mystery, though it is possible that the A β 42 used in the former study could have more advanced aggregates that do in fact protect against solvent exchange. We also observe that, like A β 40, A β 42’s rates also increase in a roughly proportional way with pressure (with a few exceptions, like at H14). If this is a true representation of the solvent exchange of this peptide in the NMR tube with pressure, then this invalidates “hypothesis 2” made above, and we are left with no explanation for why A β 42’s data height does not decrease in turn regions like the other peptides.

4.5 Discussion

In order to assess the validity of the models produced by long timescale, enhanced sampling simulation as discussed in the previous 2 chapters, we studied the biophysical properties of A β monomers experimentally using solution NMR. This thesis presents a number of new observables for each of the residues of free A β monomers, including chemical shifts, J-couplings, relaxation rates, pressure coefficients, and solvent exchange rates. It is important to point out that these particular measurements almost exclusively pertain to the backbone amide groups of these peptides, provide little information about tertiary structure, and do not inform us about individual structures sampled by A β .

Nevertheless, these measurements, in general, provide a consistent, ensemble-wide characterization of the secondary structure biases in sampling of this intrinsically disordered protein. In A β 40, regions such as the central hydrophobic cluster (CHC, L17-F20) and I31-I32 are simultaneously characterized with high $^3J_{\text{HNHA}}$ coupling, high nitrogen B_I pressure coefficients, and low solvent exchange rates, all in line with a relatively packed β sheet secondary structure (see Figure 4.6A, Figure 4.8A, and Figure 4.9A). In contrast, regions such as S8-G9, S26-G29, and G37-G38 are associated with low nitrogen pressure sensitivity and high solvent exchange rates, telltale signs of turn character (see Figure 4.8A and Figure 4.9A). The characterization of A β 42 is also mostly consistent among the NMR data sets: V36 sees a heightened hydrogen B_I and $^3J_{\text{HNHA}}$ coupling compared to A β 40, G37 has a reduced B_I and increased k_{ex} compared to A β 40, V39 and V40 are characterized by higher than average sensitivity to pressure, and V40 and I41 are distinguished with higher than average $^3J_{\text{HNHA}}$ coupling (see Figure 4.8A, as well as Table 7.3 in the Appendix). These data are consistent with a characterization of β character at V36 and V39-I41, and turn character at G37. While the structural trends of other regions of A β are less clear (ie, E11-V12 are highly sensitive to pressure but have average $^3J_{\text{HNHA}}$ s, while H14-K16 are highly solvent accessible but the pressure sensitivity is unclear from the data), the NMR data in this chapter form a coherent, if coarse, depiction of the A β ensemble.

Further, the experiment-based trends outlined in the previous paragraph are both qualitatively and quantitatively consistent with the β -hairpin dominated models for A β monomers produced by simulation with either of the two different force fields used in this thesis, OPLS and ILDN. This agreement is seen in the matching of secondary structure biases for either force field with the experimental characterizations we have made in the previous paragraph (Figure 2.2 and Figure 3.2). Figure 4.10 quantifies this apparent alignment through comparisons of the different NMR observables described in the previous paragraph, as well as the percent β sheet content of the REMD ensembles for A β 40. Not only does the β sheet profile agree very well between simulations with the two different force fields (OPLS and ILDN, as emphasized in Chapter 3), but substantial correlations with $^3J_{\text{HNHA}}$ and pressure coefficients and an anticorrelation with solvent

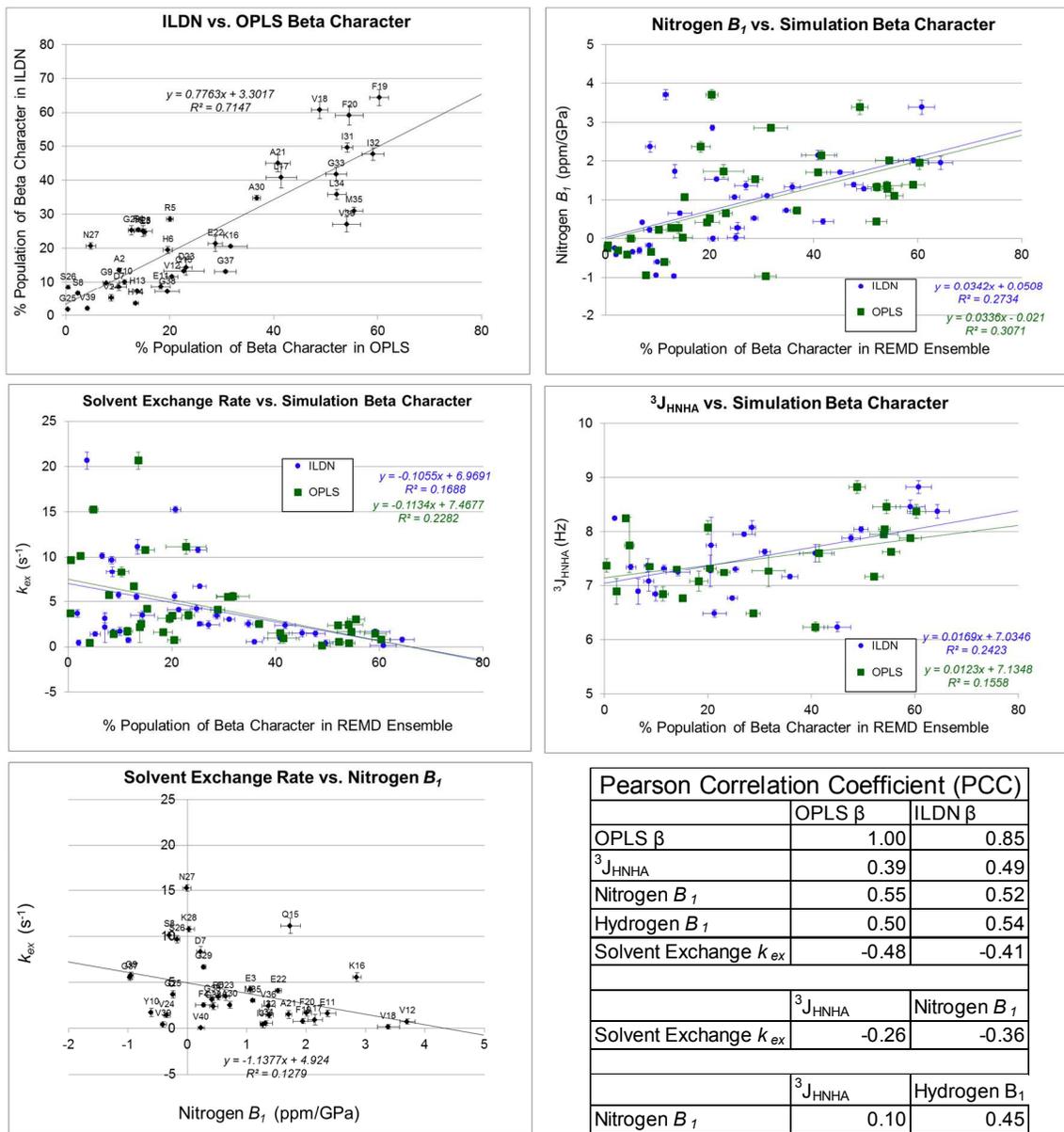


Figure 4.10. Correlation analysis for A β 40 relating β sheet character annotated through REMD simulations using DSSP(347, 348) with various NMR observables, including first order (B_1) pressure coefficients, experimental $^3J_{\text{HNHA}}$ couplings, and solvent exchange rates (k_{ex}). The first and last residue (D1 and V40) are outliers that are omitted from all comparisons.

exchange rates are also observed with either set of simulation data. That such a minimalist interpretation of the A β ensemble as the ensemble population of β sheet content per residue alone could capture so much of the variation seen in these NMR data is interesting, since many other factors could contribute to the deviations in the correlation. On this note, the scatter plots involving NMR data in Figure 4.10 (i.e., nitrogen $B_1/k_{ex}/^3J_{HNHA}$ vs. simulation β character) demonstrate apparent heteroscedasticity: in general, regions of high β simulation character have a narrower distribution for the NMR observable in question, but regions of low and medium β content can adopt a broader set of values that allow for more outliers to the linear correlation. Deviations in these regions could be explained by the sampling of other types of structure besides β sheet which might also contribute to heightened pressure sensitivity, protection from solvent exchange, etc.

A more direct quantitative agreement was observed by calculating the correlation and deviation of experimentally measured chemical shifts, J-couplings, and RDCs to the equivalents as calculated as ensemble-averaged values from frames of the REMD simulations (Figure 4.6). The overall comparison showed good agreement for either of the two force fields, with ILDN yielding very similar PCCs and slightly lower RMSDs to experiment compared to OPLS for chemical shifts and $^3J_{HNHAS}$, but with OPLS performing better at reproducing the trends observed amongst $^3J_{NHBS}$ and RDCs. The factors that largely contribute to good agreement with experiment are also the attributes that these A β REMD ensembles share with each other across the two force fields. For example, in Figure 4.6A, simulations with both force fields and experiment largely converge upon a similar, higher than average $^3J_{HNHA}$ for the central hydrophobic cluster, and both OPLS and ILDN characterized this region as β -rich (Figure 2.2 and Figure 3.2). We conclude that the commonalities between the two force fields and the experimental data contribute to a model for the ensembles of wild type A β monomers that is supported by multiple different methods. This concurrence between complementary methods lends much needed confidence to the descriptions of A β made throughout this thesis: our NMR data provide only ensemble-averaged, backbone-specific insight into A β structure, while our simulations are prone to the parameter-dependence described in detail in Chapter 1.

Though it is consistent amongst multiple characterizations, however, the depth of this model is limited: simulations only share this agreement primarily on backbone conformational biases (as seen through secondary structure annotation and contact maps (Figure 2.2, Figure 2.3, Figure 3.2, and Figure 3.3)), while the vast majority of our collected NMR data only report on local structural information at the backbone-level. We saw in Chapter 3 that OPLS and ILDN are characterized by very different electrostatic interactions. The point mutations of A β 40 studied in this thesis are not only relevant to Alzheimer's disease, but their characterization allows us to study the role of that residue (both in wild type and in the mutant) in A β structure, even if our NMR data are largely specific to backbone conformations.

To compare this effect of mutation with simulation, we used comparisons between wild type and D23N A β 40 as a test case. D23N is a conservative mutation (i.e., as defined by PAM 250 substitution matrix-based⁽⁴⁶⁹⁾ criteria used in the ClustalW⁽⁴⁷⁰⁾ alignment tool) that supplants the negatively charged aspartate with a sterically similar, neutrally charged asparagine. Further, we observed that the NMR data predict that the effect for all of the studied point mutations on the A β monomer backbone is almost exclusively local, as seen, for example, through sparse CSPs in residues distal to the site of each point mutation in Figure 4.3; Panel A of that figure reveals that A β 40-D23N is no exception to this trend. Nevertheless, this mutation resulted in very different characterizations produced by the OPLS and ILDN force fields, as discussed in Chapter 3. To recap, in OPLS, wild type A β featured a "crown motif" where D23's carboxyl interacted with the backbone amides of the subsequent four residues to produce a broad bend. D23N abolishes these interactions, reducing the breadth of the bend and changing the nature of hydrogen bonding between the central hydrophobic cluster and the C-terminus. ILDN did not feature this crown motif, instead characterizing the region with a more diverse set of turn-like annotations that frequently formed helical hydrogen bonds. The primary effect of D23N among the ILDN set was to promote N-terminal β character and increase global contact of the region with the rest of the peptide; it also reduced the extent of turn character in the central region, but to a much more limited extent than the effect of OPLS.

In this chapter, we made direct comparisons in the change of chemical shifts and $^3J_{\text{HNHAs}}$ between A β 40-D23N and A β 40, both observed through simulation and experiment. While there is reason to be skeptical of the significance of these analyses (see section 4.4.2.1 and 4.4.2.2), we propose the two results in tandem independently lead to the same conclusions. Both suggest that ILDN and OPLS overestimate changes to distal terminal regions, and they both demonstrate that OPLS, in particular, poorly simulates even local changes caused by the mutation. For N and C α Δ CSs and Δ^3J_{HNHA} , OPLS has its largest anticorrelation at the D23N mutation site itself, and an overall negative PCC for the whole sequence that is primarily driven by this change (Figure 4.5C and Figure 4.6C). It is possible that these inconsistencies are caused by insufficient sampling, but the observation that the crown motif is such a prevalent and enduring feature of the long timescale REMD ensemble produced by REMD over many different A β simulations suggests that this is an artefact produced by the force field that is not consistent with NMR data. We further hypothesize that OPLS overemphasizes protein-protein electrostatic interactions; if true, this may explain why OPLS adopts a far more collapsed and bend-prone ensemble than ILDN for a smaller IDP (an 18 residue fragment of α -synuclein) studied in Chapter 5. Either way, it is clear that simulations and experiments are not in accord on the nature of electrostatics of the central region of A β and the effect of point mutations, so we declare that conclusions of these studies are not generalizable between methods and merit further study. (One exception to this may be the A β 40-E22G variant, where lowered B_1 pressure coefficients (Figure 4.8A) and R_2 relaxation rates⁽⁵⁴⁾ support the observation of lost structure and greater disorder local to the mutation site observed in REMD simulations, as described in Chapter 6.) Plans for future work aimed at correcting force field issues and acquiring new NMR data that report on more global properties of the ensemble, as well as those that report on the side chains, are also proposed in Chapter 6.

Having established the inconsistencies between simulations with multiple force fields and the experimental data, we propose that the conclusions discussed above that are shared among multiple characterizations should be emphasized over those that are not. These conclusions, in addition, have important implications for the self-association of the peptide; the similarity of these models to higher order aggregates has been

covered at length in the Discussion sections of Chapters 2 and 3. The NMR data presented in this chapter lend credence to the hypotheses posited in those sections, particularly that the β hairpin conformations observed through simulation are indeed sampled in the ensemble of free monomers and could act as seeds of aggregation.

5. MD/REMD Simulations of Other Disease-Relevant IDPs: Fragments of the SOD1 Protein and α -synuclein 1-18 A18D

5.1 Abstract

This chapter presents two projects that test the robustness of the methodology used to simulate A β in previous chapters. The first involves long timescale molecular dynamics (MD) and replica exchange MD (REMD) of peptides, each of ~ 30 residues in length, encompassing the sequence of human cytoplasmic copper/zinc superoxide (SOD1), in order to identify the region and possible structural states involved in SOD1 aggregation in amyotrophic lateral sclerosis (ALS). End to end distances and radii of gyration in the simulations were compared to experimental data collected with Förster resonance energy transfer and small-angle X-ray scattering. Inconsistencies between theoretical predictions and experimental observations suggest that our simulation methodology is biased toward sampling collapsed states. The second project involves α -synuclein, which forms aggregates that are implicated in Parkinson's disease. The system studied is a small 18-residue N-terminal fragment of this peptide that is implicated in membrane binding and cytotoxicity of the protein. REMD simulations were conducted on the fragment using both the OPLS-AA/TIP3P and AMBER99sb-ILDN/TIP4P-Ew force field combinations examined for A β in Chapter 2 and 3. (The primary differences and similarities between these force fields and water models are discussed in the Introduction of Chapter 3.) Further, both NMR and simulations were done on this system at multiple pressures. As a smaller disordered peptide than A β , the potential for structural biases are reduced, and unlike A β , the OPLS and ILDN runs yielded very different ensembles in terms of structure and shape. Further, neither simulation matched the observed $^3J_{\text{HNHA}}$ couplings for the system. This shows that the concurrence of these characterizations for A β made in Chapter 2-4 is a non-trivial result, but also illustrates that our simulation parameters may be fundamentally flawed in reproducing the ensembles of highly disordered proteins.

5.2 Introduction

An important case for assessing the validity of the intrinsically disordered peptide (IDP) simulation methodology and parameters we are employing in this thesis is to

observe if the simulated ensembles for a variety of different IDP sequences recapture their respective experimental properties. In this chapter, we present two ongoing projects that apply the methodology described in Chapter 2-4 for A β to other small peptides of comparable length (18-36 residues long) that are subsequences of larger aggregation-prone proteins with relevance to neurological disease. Some of these peptides are known to be IDPs themselves, and are being investigated for their role in driving the aggregation of the respective full length protein.

5.2.1 SOD1 Fragments

The first project focuses on the human cytoplasmic copper/zinc superoxide dismutase, also known as SOD1, a protein whose aggregates are implicated in amyotrophic lateral sclerosis (ALS). Chapter 1 (section 1.3.1) reviews the relationship of SOD1 in this disease, along with important features of the known SOD1 structure.

Researchers in the lab of Prof. C. Robert Matthews at the University of Massachusetts have observed friction-limited folding (FLF) and unfolding of full length SOD1 upon reduction of its single, highly conserved intramolecular disulfide bond C57-C146, with changes in circular dichroism signal through urea titrations showing a dependence on viscosity of the surrounding media (*unpublished data presented at a conference(471); noted here with permission*). Aggregation could fundamentally be behind these observations, as the different solution viscosities could affect the kinetics and thermodynamics of self-association. To further identify the regions that may be responsible for this phenomenon, the same group acquired samples for five different subpeptides of around 30 residues each, in tandem entirely encompassing the sequence of SOD1; see the Methods section for more details about these peptides. In this chapter, these are named N-terminal peptide (NTP), 32-57, 58-86, 87-120, and C-terminal peptide (CTP), in sequential order. Experiments were done in a reducing environment and in the absence of metal ions and the presence of chelating agents; at least two of these peptides (58-86 and the C-terminal peptide) are dominated by the intrinsically disordered sequences listed Chapter 1 and are therefore very likely disordered themselves. These peptides were appended with Förster resonance energy transfer (FRET) acceptors and donors at or near the N- and C- termini, and analyzed with FRET,

small-angle X-ray scattering (SAXS), and circular dichroism (CD) to obtain information on the secondary structure, end to end distance, and radius of gyration of the peptides as a function of increasing urea concentration. The experimental data, some of which is covered in more detail in the Results section, indicate that the CTP, containing the electrostatic loop, was the only peptide to collapse and the only peptide to show a dependence for FRET efficiency on urea content (*unpublished data presented at a conference*(471); *noted here with permission*). Since the CTP also contains one of the participants of the SOD1 intramolecular disulfide bond that needs to be broken for FLF, these observations suggest this fragment may be involved in this phenomenon.

To complement this study, our group has simulated each of these fragments of SOD1 with molecular dynamics. For NTP and CTP, we have run replica exchange molecular dynamics (REMD) simulations in a similar vein as we have done in previous chapters with A β , using the enhanced sampling provided by this method to generate an ensemble of structures needed, in particular, to understand the intrinsically disordered C-terminal region. Standard MD simulations were also run for all peptides, complementing the REMD simulations because the former permits the accurate exploration of the kinetics of conformational switching that are typically precluded by discontinuities in the trajectories produced by the replica exchanges of the latter(472). Our simulations have defined a wide variety of states for these peptides, particularly CTP, with the ultimate goal of elucidating the molecular basis for FLF in full length SOD1 and defining the seeds that may drive SOD1 aggregation in ALS. Peptides are shown to sample some conformations with structural resemblance (i.e., characterization of secondary structure) to the same sequence in the full length peptide, but nevertheless, sample very different conformations, with almost no sampling of frames with C α root mean square deviations (RMSDs) < 4.0 Å to full SOD1. Further, comparing physical properties of the MD ensembles (radius of gyration and end to end distance) with their experimental counterparts reveals that, although good agreement is seen for the collapsed CTP, these parameters are not reproduced for the extended NTP. This suggests the simulations may be biasing the structural ensembles to lower radius of gyration states.

5.2.1.1 α -synuclein 1-18 A18D ($N_{\alpha\text{-syn}}$)

The second project described in this chapter centers on α -synuclein, an intrinsically disordered protein in solution(89), whose aggregates are implicated in Parkinson's disease, Lewy body dementia, and Alzheimer's disease(88-90, 289-291). The structural features of α -synuclein and role of this protein in disease are reviewed in Chapter 1 (section 1.3.2).

The laboratory of Prof. Ad Bax has previously analyzed the effect of pressure on the ensemble of full length α -synuclein, acquiring observables such as $^3J_{\text{HNHA}}$ couplings, chemical shifts and pressure coefficients, and relaxation rates, and concluding that the protein undergoes a slight increase in population of polyproline II character with pressure(420). The same lab has created a construct (further described in the Methods section), called $N_{\alpha\text{-syn}}$ in this chapter, of 18 residues length derived from the N-terminus of the α -synuclein protein. This sequence contains both the first 11 residues implicated in membrane binding (as described in Chapter 1)(91), and the first repeat of the consensus sequence of the N-terminal region ($^{10}\text{KAKEG}^{16}\text{V}$) (whose repeats in the full length protein form an α -helix in apolar environments(292-295)). In the absence of nonpolar solvents, this peptide is expected to be intrinsically disordered as it is a subsequence of full α -synuclein that is known itself to be natively unfolded under these conditions(89, 91, 294). To study the balance of secondary structure states in a highly truncated peptide, the Bax lab have again characterized this system in water with $^3J_{\text{HNHA}}$ couplings and chemical shifts as a function of increasing pressure, with the results described in more detail in the Results section. To complement this analysis and define states of interest, we have examined this small peptide with extensive replica exchange molecular dynamics (REMD) simulations using the same force field and water model combinations used to simulate $A\beta$ variants in Chapter 2 and 3: the OPLS-AA/L(19)/TIP3P(20) and AMBER99sb-ILDN(21)/TIP4P-Ew(22) (hereafter using the "OPLS" and "ILDN" nomenclature as defined in those previous chapters). Additional simulations were also conducted at 2500 bar. Unlike $A\beta$, the two force field combinations (at ambient pressure) produce very different ensembles in terms of secondary and tertiary structure, and neither force field is able to accurately reproduce the measured NMR observables. This highlights the significance of the agreement

between multiple simulations and experimental data seen for A β in Chapter 2-4, while pointing to the limitations of the employed simulation methods and parameters for accurately representing IDPs in general.

5.3 Methods

5.3.1 Simulation Details

5.3.1.1 MD and REMD on SOD1 Fragments

The sequences simulated by our group for this project are fragments of human soluble superoxide dismutase, form 1 (SOD1) originally chosen for analysis by the lab of Prof. C. Robert Matthews at the University of Massachusetts for FRET as described in section 5.3.3. The five peptides span the entire sequence of the SOD1 peptide and were chosen to be roughly the same size (26-36 residues), encompass full segments of secondary structure in the full protein, minimize the need for additional tryptophan residues (if possible), and finally minimize redundancy between the peptides. The sequences of the computational study include:

- NTP (residues 1-36) (add 0 to index for full protein):
 $^1(A)TKAVCVLK^{10}GDGPVQGIIN^{20}FEQKESNGPV^{30}KVWGSIK$
- 32-57 Peptide (add 31 to index for full protein):
 $^1WGSIKGLTE^{10}GLHGFHVHEF^{20}GDNTAGC$
- 58-86 Peptide (add 56 to index for full protein):
 $^1WTSAGPHFN^{10}PLSRKHGGPK^{20}DEERHVGDLG^{30}N$
- 87-120 Peptide (add 86 to index for full protein):
 $^1VTADKDGVA^{10}DVSIEDSVIS^{20}LSGDHCIIGR^{30}TLVVHW$
- CTP (residues 121-155) (add 120 to index for full protein):
 $^1EKADDLGKG^{10}GNEESTKTGN^{20}AGSRLACGVI^{30}GIAQW$

Note that numbering of the individual peptides will be used in our discussion of particular residues, then the numbers from the full length sequence will be provided in parentheses. Throughout this chapter, the indexing for full length sequence is based on the indexing of residues in PDB:1SOS(301). Underlined residues in the sequences above are the two cysteines involved in an intramolecular disulfide bond in the full protein. Bolded residues are tryptophan residues not present in the wild type sequence, but were

included in the fragments as an intrinsically fluorescent reporter. There are two variants of the experimental peptides used for FRET: donor-only peptides, whose sequences are exactly as listed above, and donor-acceptor peptides, which also possess glutamine attached to ethyldiaminonaphthalene-1-sulfonic acid (E-EDANS) at one of the termini of each peptide. These tags are located at the N-terminus for NTP, 87-120 peptide, and CTP; while they are present at the C-terminus for the 32-57 and 58-86 peptides. This makes the E-EDANS group occupy the opposite terminus of the tryptophan residue of each donor-acceptor peptide, so that the FRET efficiency acts as a reporter of end to end distance (except NTP, where tryptophan is at position 32 which is close but not exactly the C-terminus, however, E-EDANS is affixed to the N-terminus so this relates closely to the end to end metric regardless). The simulated peptides match the donor-only constructs; that is, they contain the added tryptophan residues, but not the E-EDANS moieties. The N-terminal alanine of the NTP is present in experiments and REMD simulations, but not regular MD of this peptide. All sequences were modelled with zwitterionic ends, and side chains were protonated matching the experimental pH of 7.2.

The force field and water model used for all simulations for this project was AMBER99sb-ILDN(21) and TIP4P-Ew(22), following the relative success with this combination in predicting experimental observables for A β seen in Chapter 3 and 4. Each of the SOD1 peptides was equilibrated from a fully extended conformation using a similar method as that of A β peptides described in Chapters 2 and 3. This entails running a 1 nanosecond *in vacuo* MD simulation at 700 K to collapse the structure, then solvating each cubic box and replacing waters with ions to match 0.15 M excess NaCl, chosen to better matching the experimental conditions while neutralizing the charge of the system. The number of TIP4P-Ew waters, number of ions, and size of the box in each dimension after solvation for each peptide are described in Table 5.1. Then, a 1 ns NPT equilibration simulation was run at 298 K and 1 bar, with an integration step of 1 fs, with temperature and pressure coupled using the standard Berendsen method(342). After this step, it was observed that the box sizes changed, but only marginally, where the side of each box became the value denoted in the final column of Table 5.1.

Table 5.1. Number of waters, Na⁺ ions, and Cl⁻ ions for each SOD1 peptide system, along with the size of the cubic box in each dimension directly after solvation and following NPT equilibration simulations, as described in the text.

Peptide	Waters	Na ⁺ ions	Cl ⁻ ions	Size of box, after solvation (nm)	Size of box, after NPT (nm)
NTP	5158	15	17	5.5	5.4
32-57	3345	12	10	4.7	4.7
58-86	5147	15	15	5.5	5.4
87-120	4961	18	14	5.4	5.4
CTP	4988	15	14	5.4	5.4

Standard NPT MD runs were run in triplicate for each SOD1 peptide for a simulation time of 500 ns each at 298 K, matching the room temperature condition at which the FRET experiments were run. These simulations use a timestep of 2 fs, permitted by use of LINCS(343) and SETTLE(344) algorithms to constrain bond lengths. Lennard-Jones interactions use a distance cutoff of 1 nm, while electrostatics were calculated using the particle mesh Ewald method(345). Temperature coupling to 298 K was achieved through the Nose-Hoover algorithm(346), while pressure was coupled to 1 bar with the Parrinello-Rahman barostat(473). These runs were performed using 39 CPUs with 12 PME ranks on a Linux cluster at Rensselaer Polytechnic Institute.

The NTP and CTP peptides were also run as REMD simulations(35, 36). Standard 10 ns NPT MD simulations with similar parameters as the simulations in the previous paragraph, repeated every 25 K to span the temperatures 275 to 600 K, were used to calculate ensemble-averaged energies. These energies were used to calculate the optimal amount of replicas to span this temperature range with an average exchange ratio of 15% for each peptide with the method originally described by García *et al.*(474) and using a Mathematica script that was also published by our group(475). Both the NTP and CTP were represented by 64 replicas each, covering the range 270.0 K to 588.4 K for the former system and 270.0 K to 599.5 K for the latter; shuffling of replicas between temperatures under these schedules was observed to be sufficient (Figure 7.1 in the Appendix). NVT REMD simulations were run for a total of 500 ns/replica, or 32 μ s cumulative simulation time, for each system, with exchanges being attempted according to the Metropolis Monte Carlo criterion every 4 ps between adjacent temperatures over the course of the simulation. All other simulation parameters, such as bond length

constraints, timestep, cutoff for Lennard-Jones interactions, and temperature coupling (no pressure coupling is present) are identical to the paragraph above. These parameters are also identical to that of the A β simulations of Chapter 2 and 3. SOD1 peptide REMD simulations were performed with 256 CPUs each on the XSEDE Stampede cluster at the Texas Advanced Computing Center at the University of Texas.

5.3.1.2 REMD on α -synuclein 1-18 A18D ($N_{\alpha\text{-syn}}$)

REMD simulations were also performed on a fragment derived from the N-terminus of α -synuclein, matching the experimental construct used by the lab of Prof. Ad Bax in the NMR experiments used for comparison to these simulations (see section 5.3.2). This peptide contains the first 17 residues of the sequence, followed by a non-native C-terminal aspartate instead of the wild type alanine. This residue is a remnant of a TEV protease cleavage site used to isolate this subpeptide experimentally. The sequence of this construct, hereafter termed $N_{\alpha\text{-syn}}$, is:

- $N_{\alpha\text{-syn}}$ (α -synuclein 1-18 A18D):
 ${}^1\text{MDVFMKGLS}^{10}\text{KAKEGVVAD}$

Experiments were done at pH 6.0 (see section 5.3.3); residues were protonated accordingly (though this also matches the protonation state that would occur at pH 7.0). Zwitterionic ends were used for the peptide termini. Two different force field and water model combinations were used for this project matching the two used for A β in Chapter 2 and 3, namely OPLS-AA(19)/TIP3P(20) (referred to as “OPLS” in this chapter) and AMBER99sb-ILDN(21)/TIP4P-Ew(22) (designated “ILDN”). REMD simulations were done for both force fields at 1 bar; in addition, a simulation was conducted using the ILDN combination at 2500 bar.

The $N_{\alpha\text{-syn}}$ sequence was initially modelled as an extended conformation, with charged N- and C- termini. Simulations were equilibrated in a similar manner to that of the SOD1 peptides in the previous section. One difference is that no ions were included in the system; this emulates the low salt conditions used for NMR experiments. Further, the $N_{\alpha\text{-syn}}$ peptide has a net charge of zero, so charge balancing was not necessary. The system was solvated with 2869 waters, producing a 4.5 nm box for both force field conditions. The final box size after the NPT equilibration for the ambient pressure

simulations was changed only marginally, still remaining 4.5 nm for both OPLS and ILDN force field setups. As for the high pressure ILDN run, this NPT MD simulation was set to 2500 bar, still employing the Berendsen method(342), which changed the box size to 4.3 nm.

Like the SOD1 peptides in the previous section, energies were calculated over 10 ns NPT MD simulations from 275-600 K, and an optimized temperature schedule for REMD was derived for ILDN and OPLS simulations(474, 475), this time matching an average exchange ratio of 20%. For ILDN simulations (both ambient and high pressure), 56 replicas were used, spanning 270.0 K to 612.0 K. For the OPLS run, 270.0 K to 609.2 K was covered using 50 replicas. These parameters yielded frequent random walks through temperature space for these simulations (Figure 7.1 in the Appendix). The 1 bar ILDN REMD simulation was run to 1000 ns/replica, while the 2500 bar ILDN and 1 bar OPLS run were run to 500 ns/replica each. All other parameters for the NVT REMD simulations are identical to that of the SOD1 peptides above (and to the A β REMD simulations discussed in Chapter 2 and 3). The N $_{\alpha}$ -syn REMD simulations were also performed on Texas Advanced Computing Center and XSEDE's Stampede cluster, using 224 CPUs (for ILDN runs) or 200 CPUs (for OPLS runs).

5.3.2 Ensemble Analysis

Simulations for both projects were analysed in similar ways. Individual MD simulations of SOD1 peptides were studied from 100-500 ns, while both REMD simulations were analysed from 100-500 ns/replica over temperature space replicas 8 through 15 for both systems (encompassing 289.4-311.0 K for NTP and 289.8-311.9 K for CTP). For N $_{\alpha}$ -syn, the production ensemble for the ILDN 1 bar simulation was 400-1000 ns/replica, while the timescale for analysis was 150-500 ns/replica for both the OPLS 1 bar and ILDN 2500 bar runs. The replicas studied were 6 through 12 for the 2 ILDN runs (covering 289.8-312.1 K) and 5 through 10 for the OPLS run (spanning 289.3-310.6 K). For each simulation, these frames will be referred to as the "production ensemble" for that system in this chapter. Analysis of these simulations was done in a very similar manner to that of the A β simulations in Chapter 2 and 3. Clustering was done with single linkage clustering with an RMSD of 2.0 Å for the REMD simulations

of SOD1 peptides (349), but for $N_{\alpha\text{-syn}}$, Daura clustering with an RMSD of 3.0 Å (351) was employed. In the case of the former, centroids from single linkage clustering were found to be the most representative of individual basins in the ensembles, but in the latter, the resulting cluster did not properly distinguish members of its cluster and members outside its cluster, as ascertained by RMSD histograms to the centroid structures. This is likely due to an undesired “chaining” effect, inherent to linkage clustering, grouping dissimilar structures. This may be promoted by the lower degrees of freedom of the $N_{\alpha\text{-syn}}$ system given its small sequence size compared to A β and the SOD1 peptides studied here. Daura clustering avoids the chaining problem, and was found to produce more discriminating centroid structures. Clustering was still done for every 100 ps-th frame for all simulations, in spite of the different production ensemble size. For $N_{\alpha\text{-syn}}$ ILDN 1 bar, clustering, radius of gyration, and end to end distance were calculated for both the 400-1000 ns/replica window, and for the 150-500 ns/replica window; the latter was done to better compare these statistics to the other simulations.

5.3.3 Experiment Details

All experimental data and constructs for SOD1 peptides were generated by Noah Cohen, Prof. Osman Bilsel, and Prof. C. Robert Matthews at the University of Massachusetts. These data are not published in text at the time of writing this thesis and presented here with the investigators’ explicit permission, though it was presented at the 2015 American Society for Biochemistry and Molecular Biology annual meeting(471). The donor-only and donor-acceptor constructs used in the experiments are described above in section 5.3.1.1. All experiments were done at 20 mM HEPES, 1 mM EDTA, 1 mM TCEP, pH 7.2, 298 K, except those involving zinc binding, which was done at 20 mM HEPES, 1 mM Zn, 1 mM TCEP, pH 7.2, 298 K. Data were collected using a [peptide] range of 2-20 μM ; observed FRET efficiencies were highly reproducible regardless of concentration, serving as an argument against the significant contribution of intermolecular FRET to the observed efficiencies since we would expect this factor to be highly concentration dependent. The data discussed in this thesis is based on 3 repeats using 2 μM of peptide. FRET was calculated based on equilibrium fluorescence lifetime measurements made using a continuous flow time-correlated single photon counting

(CF-TCSPC) fluorescence spectrometer equipped with a home-built autosampler, described in other studies(476, 477). Experiments were run using a customized version of LabVIEW software, specialized for operating the autosampler and processing the data. These software add-ons are available at www.osmanbilsel.net.

Amplitude-averaged fluorescence lifetimes ($\langle \tau \rangle$) were calculated from the data by fitting to the following multi-exponential model:

$$\langle \tau \rangle = \frac{\sum_i \alpha_i \tau_i}{\sum_i \alpha_i} \quad (5.1)$$

where α_i and τ_i are the amplitude and lifetime, respectively, of the i th exponential of the excited state decay. These average lifetimes were calculated for both the donor-only and donor-acceptor peptides, defined as $\langle \tau \rangle_D$ and $\langle \tau \rangle_{DA}$. Then, the FRET efficiency is determined according to the following equation:

$$E_{FRET} = 1 - \frac{\langle \tau \rangle_{DA}}{\langle \tau \rangle_D} \quad (5.2)$$

$\langle \tau \rangle_D$ is present in the equation to account for the possibility that global factors, such as the behavior of the solvent, could affect the FRET efficiency. FRET efficiencies were calculated for the peptides as a function of increasing urea concentration. Additionally, FRET efficiencies at 0M urea were further processed using Laplace inversion via the maximum entropy method (MEM)(478) to calculate the probability of states given the donor-acceptor distance(477, 479, 480).

NMR data for the $N_{\alpha\text{-syn}}$ were provided by Dr. Julien Roche and Prof. Ad Bax at the Laboratory of Chemical Physics of National Institute of Diabetes and Digestive and Kidney Diseases and National Institutes of Health. These data are currently unpublished at the time of writing this document and presented in this thesis with the authors' explicit permission. High pressure NMR for the $N_{\alpha\text{-syn}}$ system was done analogously to that of full length α -synuclein as described by Roche *et al.*(420); these experiments use the same buffer (20 mM Na_2HPO_4 , pH 6.0) and temperature (288 K) and with the same equipment and NMR acquisition parameters. This said, there are two important distinctions compared to the construct used in the Roche *et al.*(420) investigation: the $N_{\alpha\text{-syn}}$ peptide has an A18D mutation (see section 5.3.1.2) and it has a normal cationic N-terminus instead of the N-acetylated terminus used in the full α -synuclein study.

Chemical shift information was collected at the following intervals: {1, 750, 1250, 1800, 2500} bar, while $^3J_{\text{HNHA}}$ was provided for 1 bar and 2500 bar. Prediction of chemical shifts and $^3J_{\text{HNHAS}}$ from simulations and comparison to experimental data, and calculation of pressure coefficients from chemical shifts were done in an identical manner to that of A β , as described in the Methods section of Chapter 4.

5.4 Results

5.4.1 SOD1 Fragments

5.4.1.1 Secondary and Tertiary Structure

Profiles of secondary structure for the simulations of SOD1 subpeptides are depicted in Figure 5.1, while C α contact maps calculated over each ensemble is reported in Figure 5.2. For the REMD ensembles (Panel A of both figures), we first observe that the NTP is prone to forming β structure, particularly at the centers A4-L8, I18-N20, and K30-W32, which corresponds to centers of the three β antiparallel strands in the crystal structure of the full protein. And indeed, the ensemble contact map reveals 2 sets of antiparallel contacts between these regions, compatible with the formation of a meander (that is, two consecutive antiparallel β -hairpins) between these regions. This said, even the areas rich in β character are only annotated between 40-55% with this designation, with the remaining character typically annotated strongly by coil/irregular structure. A mixture of bend, turn, and coil structures represent the interim between the β centers, suggesting that flexible loops intervene between the β hairpin structures. Further, there is a strong basin of contacts where residues of the N-terminal β strand directly contacts the residues of the C-terminal strand; contacts that are likely implicated in the low end to end distance frames of the trajectory. All of this suggests significant conformational diversity in the REMD ensemble of NTP, despite the definitive β typing of the fragment in the full length SOD1 structure.

The CTP REMD ensemble, as expected given its disordered character in metal-free, reduced SOD1(331, 332), is marked with even less structural character. A mixture of bend, coil, and turn structure pervades the peptide. Small amount of β character is observed in the C-terminus, with the highest percentages at R23-L24 (R143-L144) of 20-25%. This β character is at least partially affiliated with the short C-terminal

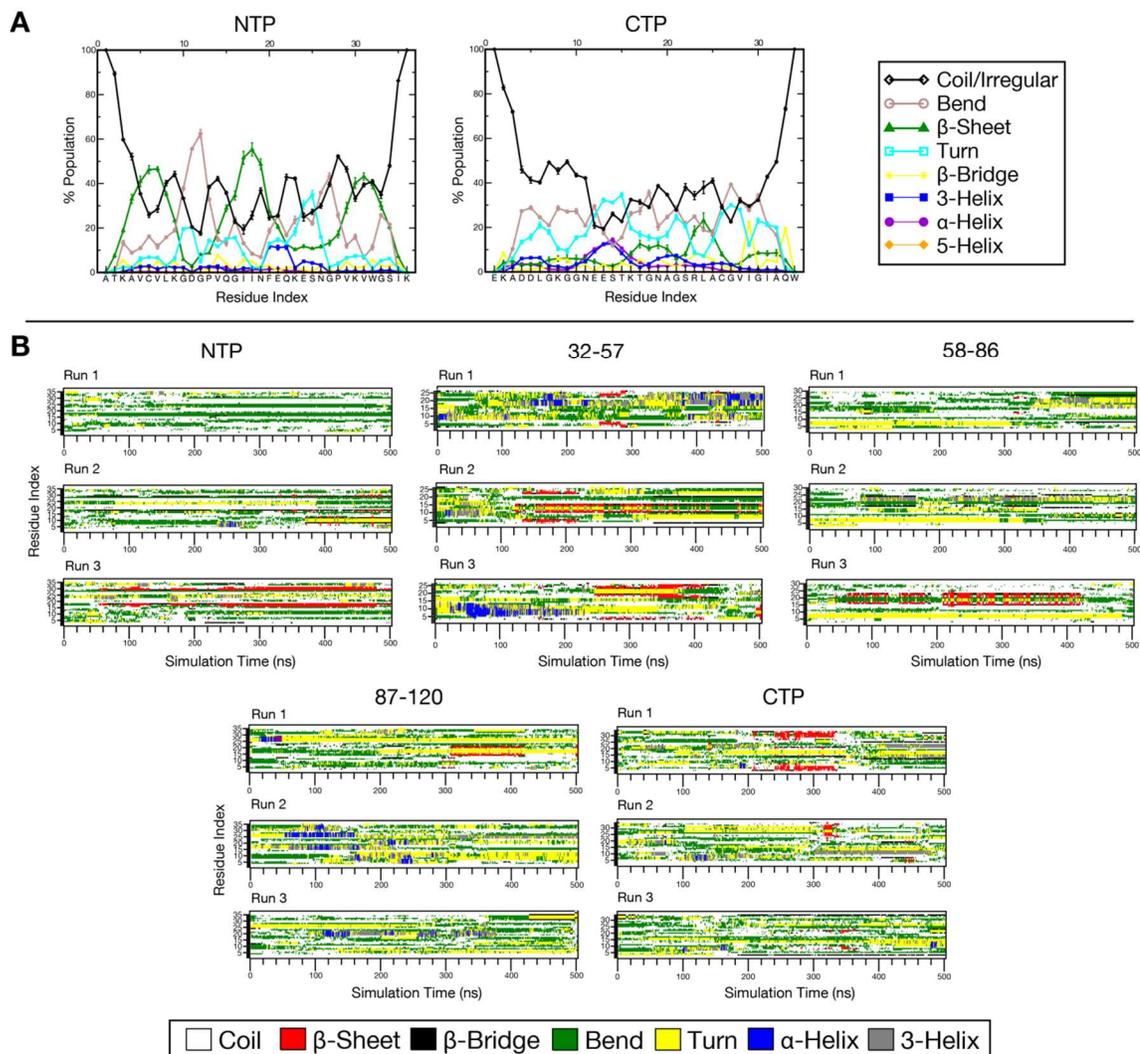


Figure 5.1. Percent composition of secondary structure, as determined by DSSP(347, 348), (A) as a function of residue index for the REMD ensemble of NTP and CTP (values were averaged over the production ensemble for each system), (B) as a function of residue index and simulation time for the MD simulations of the five SOD1 peptides.

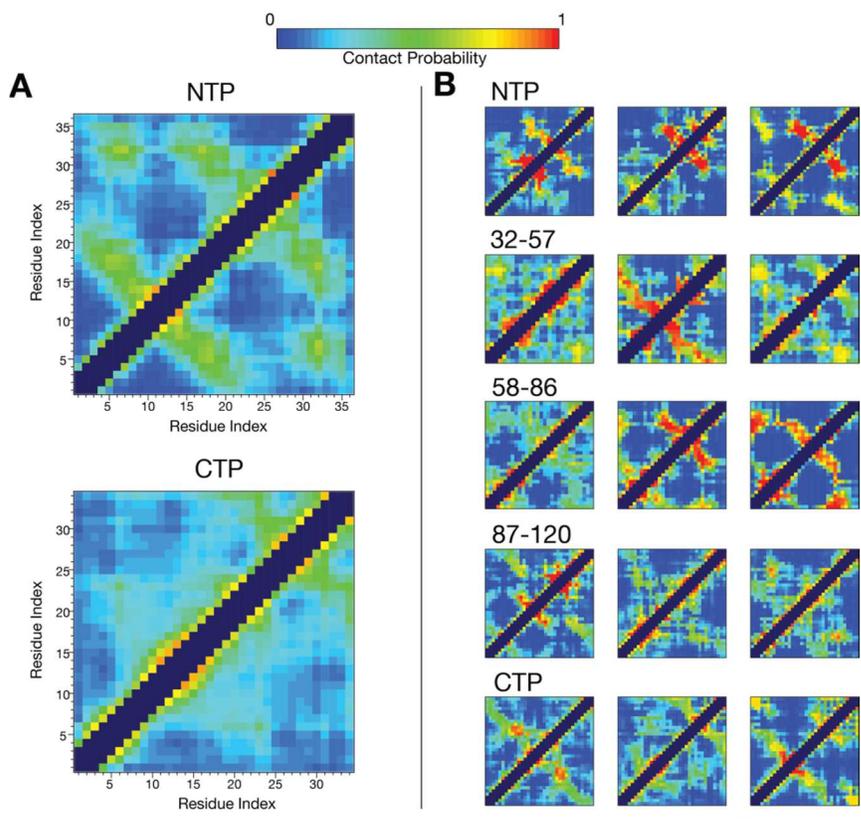


Figure 5.2. Maps illustrating the probability of contact for $C\alpha$ atoms over the production ensemble for different SOD1 peptides derived with (A) REMD and (B) MD. Here, two residues are defined to be in contact if their $C\alpha$ - $C\alpha$ distance is ≤ 8 Å(352). Self or sequential contacts, such as between (i, i) , $(i, i+1)$, and $(i, i+2)$, and invalid squares on the map were not computed and are represented by the dark purple regions.

antidiagonal from S22-L25 to G27-I31 (S142-L145 to G147-I151) in the contact map. These regions also host the only prevalent salt bridge observed in the full REMD ensemble (observed in 7.93% of the frames), between R23 (R143) side chain and the C-terminal anion. In contrast, the N-terminus of the peptide has some α -helical content, with a center at E12-T15 (E132-E135) annotated with a mixture of α -helix, 3-helix, and turn indicative of the sampling of different patterns of helical hydrogen bonds; these also manifest as parallel contacts i to $i+(4,5,6)$ contacts off the main diagonal in the contact map. In general, contacts between residues near each termini exist, but are populated more sparsely. These interactions are fairly different from the role of the fragment in context of the disulfide-linked, *holo*, full protein structure, which forms an extended electrostatic loop and concludes in the eighth β strand (see Chapter 1). However, in the absence of surrounding residues, including the residues of the directly interacting seventh β strand and the disulfide partner C57 (binding to C146 in the fragment), this finding is not too surprising.

Panel B of each figure represents the data for each of the standard MD simulations. In general, reproducibility of these ensemble characteristics between the 3 runs of each peptide in these simulations is poor: only fragment 58-86 shows a similar ensemble over its three runs, characterized by an N-terminal turn and antiparallel β interactions in the center of the peptide between residue S12-H15 and K19-R23 (S68-H71 and K75-R79). That this peptide was the only one to yield consistent structure was actually somewhat surprising because the 58-86 fragment contains a large segment of the intrinsically disordered metal-free zinc binding loop (residues R69-E100)(331, 332). Many of the DSSP plots also reveal frequent interconversions between different minima, some of which are short lived.

The REMD contact map of the NTP bears similarity to the MD contacts for the peptide, with central to C-terminal antiparallel interactions being mostly conserved in all runs. Contacts between the N- and C- termini exist on the map for MD run 3 due to β bridge interactions between the terminal regions at ~ 250 ns, in line with a β rearrangement, which we propose as a possibility in the discussion of the REMD ensemble above. On the other hand, the REMD CTP contact map bears resemblance to run 2 (and to a lesser extent, run 3) of the MD ensemble, however, run 1 has a very

different profile. In this run, a transient extended β hairpin forms between the termini, however, this structure is lost after about 100 ns of sampling. Run 3 also samples low end to end distance in the form of a longer lived β bridge between the termini. Though these conformations are very likely local minima due to their short lifetimes, these could explain N- to C-terminal contacts between the peptide that could possibly account for the urea-dependence of the FRET interactions described in section 5.4.1.4.

5.4.1.2 RMSD Based Clustering and Trajectory Superposition

The NTP and CTP systems were clustered using a similar methodology and cutoff as that used to cluster A β in Chapter 2 and 3. The results of this single linkage clustering are summarized in Figure 5.3A. Despite being clustered over fewer frames (32000 vs 42000) and over a smaller window of simulation time (100-500 ns/replica vs 400-1000 ns/replica), both the NTP and CTP ensembles, interestingly, yield more clusters while reducing the percent of the total ensemble represented by the top 10 clusters, compared to the ILDN simulated A β peptides (see Figure 3.4). This suggests that the SOD1 peptides both represent a more diverse ensemble of structures, a surprising result given the known intrinsic disordered property of A β . Furthermore, inherent differences in sequence length or diversity would not seem to explain this finding on their own: A β 40/A β 42 is only a few residues longer than NTP/CTP (40/42 residues compared to 36/34 residues), and the sequence complexities are similar (the most frequently occurring amino acid composes 15% of the sequence in A β 40 [glycine], 14% in NTP [glycine/leucine/valine], and 21% in CTP [glycine]).

Despite this apparent increase in diversity of centroids in the SOD1 systems, the top 3 NTP centroid structures all bear some commonality in the presence of antiparallel β structure at the N-terminus and center of the peptide and a second hairpin with the C-terminus, matching the ensemble-averaged characterization we made in the previous section through secondary structure and contact maps. Centroid 1 and 3, in particular, both occupy extended N-terminal-central-C-terminal β meanders that each share some structural resemblance to the role of the peptide in the full protein PDB (though there are some differences in the location of bend and β segments, and the C α RMSD is still 6-7Å to either centroid). Centroid 2 is also a double β hairpin with a similar topology, but the

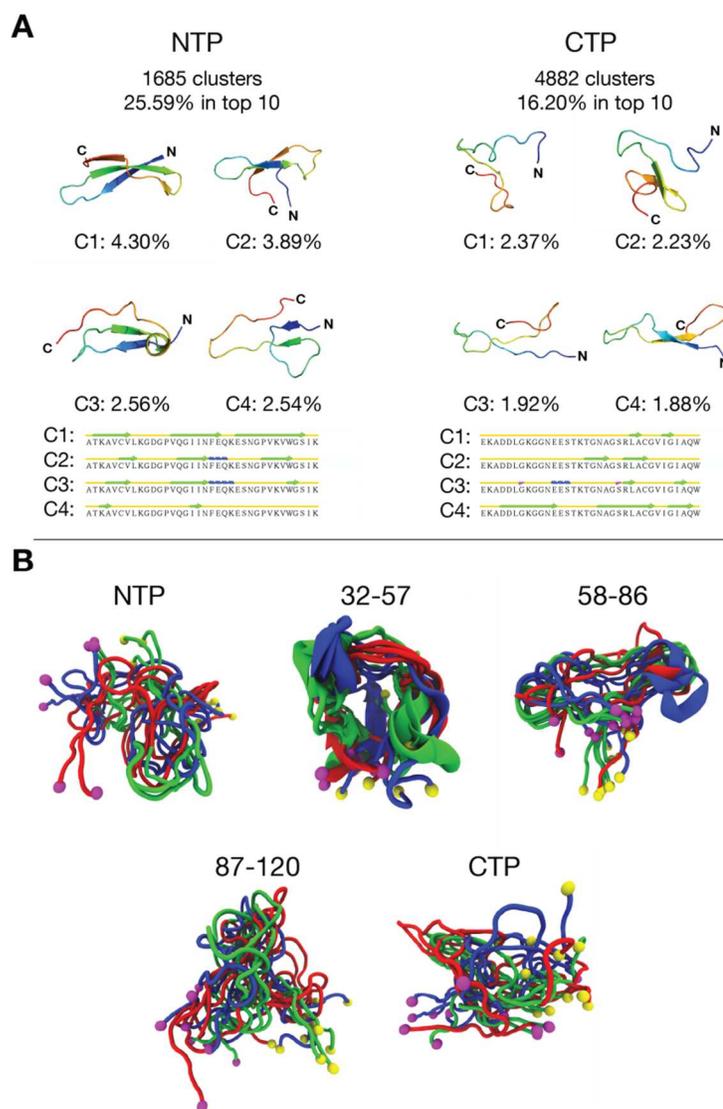


Figure 5.3. Representative conformations for simulations of SOD1 peptides. (A) The four highest-populated centroids from single linkage cluster analysis(349) around $C\alpha$ atoms with an RMSD cutoff of 2.0 Å for the NTP and CTP REMD ensembles over the production period. Ribbons are colored by a blue-to-red rainbow spectrum from the N-to-C terminus. Secondary structure strips for each centroid, as calculated by STRIDE(361), are also shown. Figure rendered in Pymol(360). **(B)** Ribbon diagram of the RMSD-minimized superposition of the 3 standard MD runs for each SOD1 peptide. Each image represents an average of all of the 10 ps frames in a 100 ns interval, within the 100 ns to 500 ns time frame. Run 1 is colored red, Run 2 is colored blue, and Run 3 is colored green. The purple spheres represent the N-termini of the peptide, while the yellow spheres are the C-termini. Figure rendered in VMD(481).

extreme C-terminus folds back toward the N-terminus. Centroid 4 is the only one of the conformations to be weakly structured.

Meanwhile, clustering the CTP run using the same criteria reveals almost 3 times the number of clusters as that of the NTP, suggesting this peptide samples a very diverse and disordered ensemble. Further, of the four most populated centroids, there are only a few commonalities among the structural elements associated with these structures. The N-terminus, for example, remains mostly unstructured in centroid 1-3, but forms an extended β sheet in centroid 4. Residue L24 (L144) and nearby residues are affiliated with β character in all 4 centroids, generally involved with interactions with the central region or C-terminus. As a final note, none of the centroid structures bring the N- and C-termini in proximity of each other.

Panel B of Figure 5.3 illustrates the trajectories of the three standard MD runs of each SOD1 peptide as a RMSD-minimized superposition of the averaged structure over each 100 ns in the run. This representation illustrates the conformational flexibility of each ensemble, as well as the relative separation of the N- and C-termini (colored as purple and yellow spheres, respectively). As discussed in the previous section, the three MD runs were fairly disparate from each other for all of the peptides except the 58-86 runs, so it is difficult to make generalizable conclusions about these ensembles.

As an important final note, histograms of $C\alpha$ RMSD were determined for each frame of the REMD and MD ensembles to the corresponding fragment in the full length 1SOS PDB structure of SOD1 (*data not shown*). There were no significant populations (all MD runs and the CTP REMD had 0%, while the NTP REMD had 0.0004%) with an RMSD less than 4 Å for any of the simulations. This suggests that, though the collective ensembles of these peptides are diverse and the biases in the secondary and tertiary structure may have some functional similarity to the full length protein (see previous section), it is possible that none of these peptides may be able to fully fold properly independently without the context of the full length protein. If so, because the underlying fragments almost never sample native conformers independently, these findings would support the nucleation-condensation model of protein folding for SOD1 where nuclei are dependent on long range interactions to adopt the proper state, so secondary and tertiary

structures are formed concomitantly(482). This said, it is certainly also possible that the simulations have not converged or the simulation parameters are invalid.

5.4.1.3 Radius of Gyration, End to End Distance, and Comparison to Experiment

FRET is a fluorometric technique where a specific chromophore is excited (the donor) and through nonradiative dipole-dipole coupling, energy is transferred to a nearby second chromophore in space (the acceptor)(483, 484). The efficiency of this FRET energy transfer is proportional to $1/R_{DA}^6$ where R_{DA} is the FRET donor-acceptor distance; as such, this technique can be used to gauge how close two chromophores are in a system if they are close enough(483, 484). For the experimental peptides studied here, the two fluorophores employed are tryptophan and a glutamine attached to ethyldiaminonaphthalene-1-sulfonic acid (E-EDANS). The peptides have been designed so these chromophores are located at opposite termini (see Methods, section 5.3.1.1), making the FRET a reliable measurement of intramolecular end to end distance (for certain distances, see below), provided intermolecular FRETs can be discounted (see Methods, section 5.3.3, for why we assume that to be true in our analysis). The exception is NTP where E-EDANS is at the N-terminus and the tryptophan is at the naturally occurring residue 32 instead of the C-terminal K36, designed so the peptide can complete the β strand in the full length *holo* structure without mutation. This residue is close enough to the C-terminus, however, for it to be a reasonable match for the end to end distance metric. FRET using tryptophan and EDANS can be resolved with confidence using maximum entropy method (MEM) (see Methods for more information) when the donor and acceptor are within the distance range of 1.1-3.3 nm (This range is based on personal communication with Noah Cohen. This is corroborated by the literature, which reports the Förster radius (R_o) of this pair as 2.1-2.4 nm(485, 486), this is the distance where transfer efficiency is 50%. 1.1 nm and 3.3 nm are $\sim R_o \pm (1/2)R_o$, and empirically, where FRET efficiency is observed to be 0% and 100%, within error.)

Figure 5.4A depicts the FRET efficiency of each of the SOD1 peptides in this study as a function of urea concentration. Interestingly, it can be observed that only the CTP peptide shows a significant effect with increasing urea concentration; the efficiency

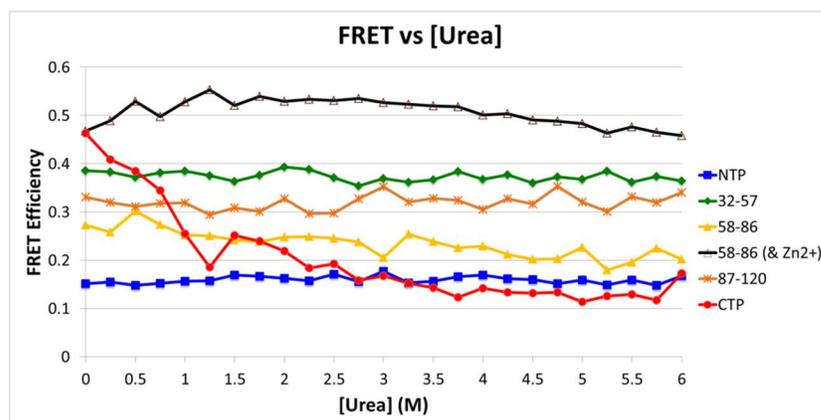
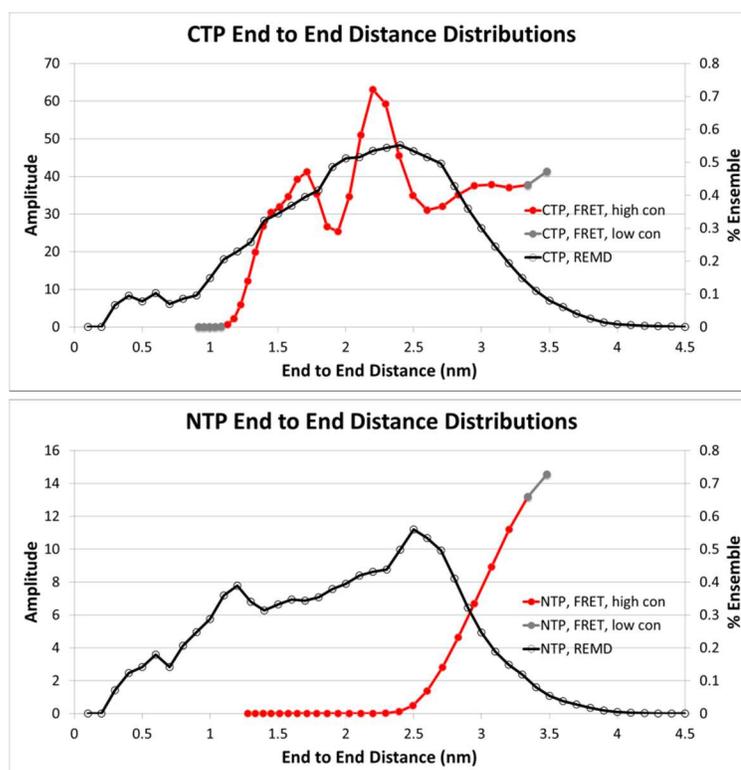
A**B**

Figure 5.4. FRET data for the SOD1 peptides studied in this chapter, obtained and provided by Noah Cohen, Prof. Osman Bilsel, and Prof. C. Robert Matthews at the University of Massachusetts. (A) Equilibrium FRET efficiency for the SOD1 peptides as a function of increasing urea concentration. (B) Probability of states as a function of donor-acceptor distance calculated from FRET data at 0M for the CTP and NTP, using Laplace inversion through the MEM(477-480). Experimental data in red are fall within points where the donor-acceptor distance is resolved with high confidence using MEM and the tryptophan-EDANS system (1.1-3.3 nm, see the main text for the rationale), while points in gray fall outside this interval. Data are compared to histograms for end to end distance of the corresponding REMD ensembles using 0.1 nm bin widths.

progressively decreases with additional urea. If the FRET signal is purely affected by intramolecular distances, this suggests the peptide ensemble transitions from collapsed states to more expanded states, and that this a phenomenon that does not occur with the other peptides. We did not simulate the effect of urea on these peptides; however, we can still use the FRET data to assess the fidelity of our simulations in the absence of urea. If we rank each of the *apo* peptides according to the FRET efficiency at 0M, we have: NTP (15.18%), 58-86 (27.32%), 87-120 (33.12%), 32-57 (38.57%), and finally CTP (46.33%). The end to end distributions of the different MD runs are shown in Figure 5.5B. The distributions of these distances are very different for individual MD runs, consistent with the poor reproducibility of these MD runs in general. That said, if we calculate an expectation value of end to end distance over each of the 3 MD runs, we do observe a trend qualitatively consistent with the trend revealed by FRET data, though the error values are very high: NTP (2.16 ± 0.93 nm), 87-120 (2.02 ± 0.73 nm), 58-86 (1.56 ± 0.57 nm), CTP (1.48 ± 0.92 nm), and 32-57 (1.24 ± 0.60 nm). This trend is actually reversed, however, in the ensemble average of end to end distance in production REMD ensembles of NTP (1.98 ± 0.78 nm) and CTP (2.14 ± 0.72 nm), contrasting with the direction and broad disparity in the FRET efficiency values of these peptides. The REMD distribution also does not agree well with individual simulations (Figure 5.5B).

Panel 5.4B compares the FRET-based distributions using MEM with end to end histograms of the REMD ensemble, revealing that the REMD end to end distribution of the CTP peptide may be consistent with the experimental distribution. REMD may underrepresent the proportion of values with an end to end distance > 3.0 nm, however, the amplitudes are unreliable at these larger distance for tryptophan-EDANS based FRET because energy transfer is insufficient (recall, the effective range was 1.1-3.3 nm as discussed earlier in this section). Likewise, it may be true that the CTP over-represents very low end to end states, but again, the FRET transfer is unreliably measured at this lower boundary. In the interval where the FRET distribution can be accepted with confidence, the agreement with the REMD is qualitatively very good. The REMD distribution is broader and smoother, lacking the different basins seen in the experimental data, nevertheless, the mode of both data sets match (2.0-2.5 nm) and the relative proportions of states along the distribution also match.

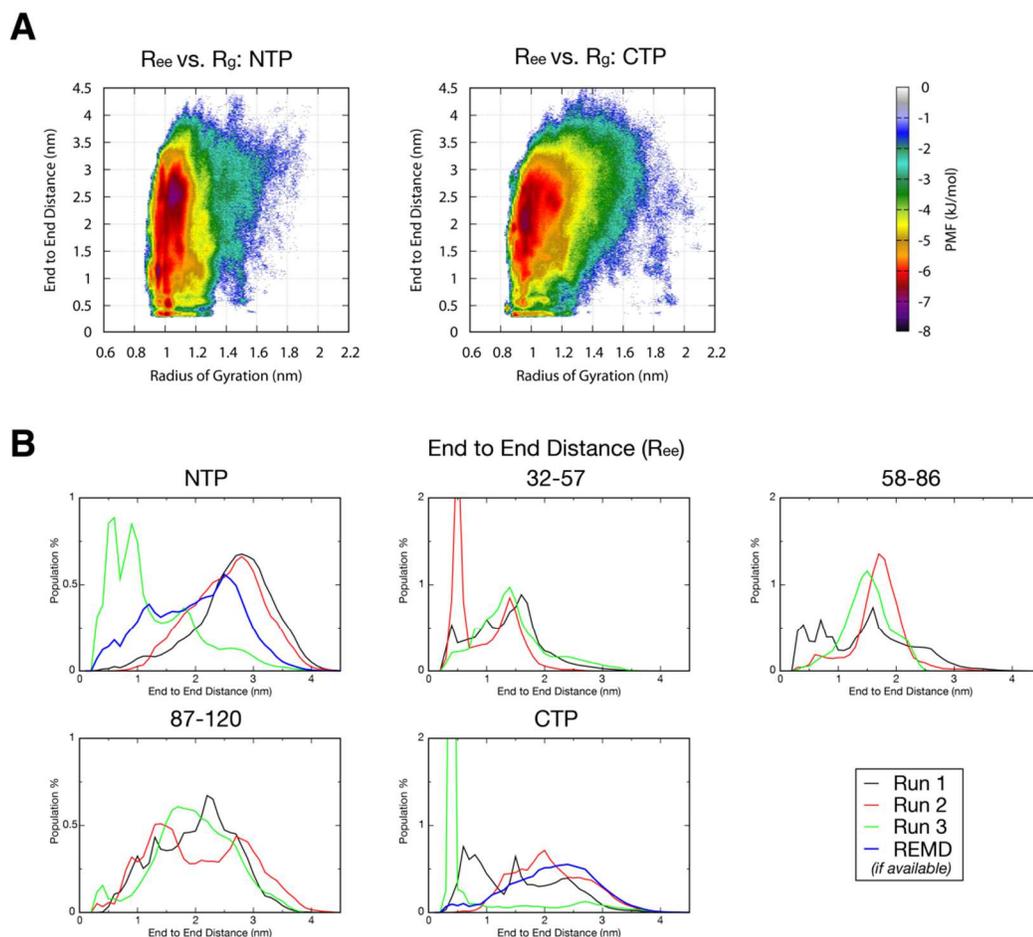


Figure 5.5. Radius of gyration (R_g) and end to end distance (R_{ee}) for simulations of SOD1 peptides. (A) Free energy surface along the coordinates of R_g and R_{ee} in units of kJ/mol for the production ensemble of the SOD1 NTP and CTP REMD simulations using 0.02 nm binning in both dimensions. (B) Histogram of R_{ee} for MD simulations of the 5 SOD1 peptides, binned over 0.1 nm lengths.

On the other hand, the distribution for NTP is vastly different between simulation and experiment. The FRET-based distribution is highly biased to very high end to end distances, but the REMD ensemble features many low end to end distance states. The high terminal distances seen through experiment suggest that the peptide does not collapse, even under physiological conditions. Indeed, SAXS data were also collected for this NTP peptide by the Matthews lab, yielding an experimental expectation of radius of gyration that is invariant with urea: 1.60 nm at 0M urea and 1.64 nm at 6M urea. In contrast, the REMD ensemble features many low radius of gyration populations (Figure 5.5A), and the expectation value is substantially reduced (1.07 ± 0.11 nm). Using the Kolinski-Skolnick equation(363) listed in Chapter 2 (Equation 2.1), a fully collapsed protein matching the length of the NTP would be 0.84 nm, while the parameters for a fully extended protein could be described by modifying the exponential in this equation to account for a denaturing solvent(363, 487):

$$\langle R_g \rangle = 0.22n^{0.60} \quad (5.3)$$

yielding 1.89 nm. Based on these theoretical estimates, we observe that the experimental value is consistent with an almost entirely extended ensemble even at 0M urea, while the REMD-based expectation is almost on the other end of the spectrum, biased to sampling more collapsed states. As a final note, we do report that 2 of the standard MD runs of NTP do consistently sample high end to end distance states that could be more consistent with the FRET distribution, but third MD run shows that even through this method, low end to end states can be sampled, while the long timescale REMD shows that these collapsed states are more frequently sampled at low temperatures, and are presumably more energetically favored by the system (provided convergence has been reached).

5.4.2 α -synuclein 1-18 A18D ($N_{\alpha\text{-syn}}$)

5.4.2.1 Secondary and Tertiary Structure

Figure 5.6 contains the annotated content of all DSSP secondary structural elements for each of the $N_{\alpha\text{-syn}}$ simulations, while Figure 5.7 shows maps of $C\alpha$ probability of contact between residues for these ensembles. First and foremost, these two figures

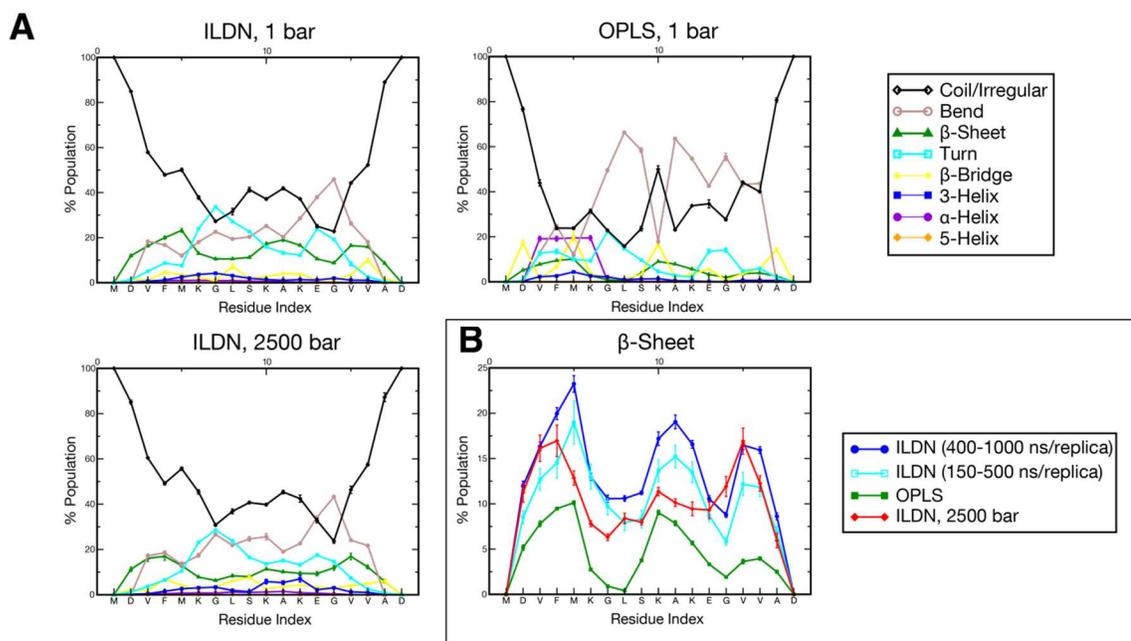


Figure 5.6. Percent composition of secondary structure as a function of residue index, as determined by DSSP(347, 348), for N_{α} -syn. Values were averaged over the production ensemble for all 3 simulations performed. Values presented for (A) all DSSP elements and (B) just β -sheet character (for the sake of comparison).

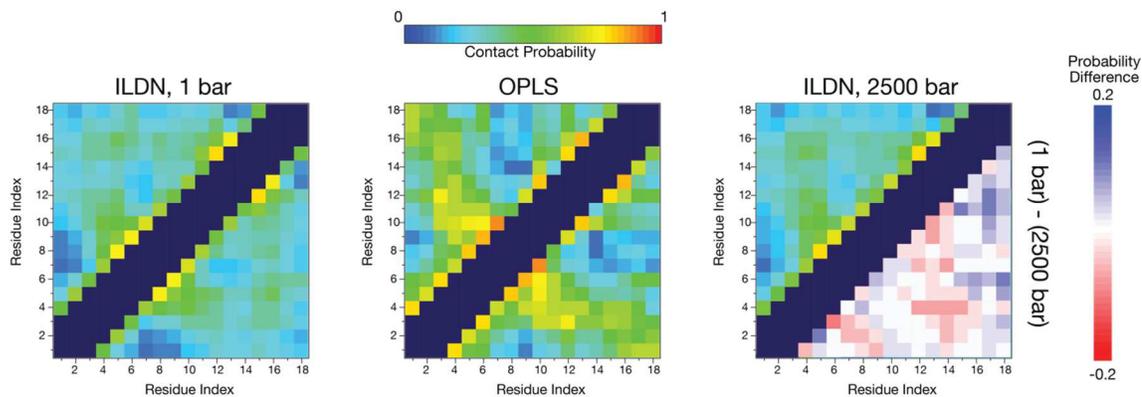


Figure 5.7. (upper corner) Maps illustrating the probability of contact for $C\alpha$ atoms over the production ensemble for the different $N_{\alpha\text{-syn}}$ REMD ensembles. Here, two residues are defined to be in contact if their $C\alpha$ - $C\alpha$ distance is $\leq 8 \text{ \AA}$ (352). (lower corner) Difference between these contact probability maps for the ILDN simulation at 1 bar and the ILDN simulation at 2500 bar. A red color indicates a reduced likelihood of contact while a blue value indicates increased contact probability for the minuend species compared to the subtrahend species. Self or sequential contacts, such as between (i, i) , $(i, i+1)$, and $(i, i+2)$, and invalid squares on the map were not computed and are represented by the dark purple regions.

indicate a profound difference between the ensembles produced by ILDN and that produced by the OPLS force field. In general, the OPLS highly populates bend character at G7-S9 and A11-V17. Besides coil, other annotations are rare, except for an α -helix populated 20% of the time between D2-K6 and some β bridge content at D2, M5, K10, and A17. β sheet content exists, but is sparsely populated, peaking at $\sim 10\%$. ILDN shares some similarity in turn locations, populating a mixture of turn and bend (and other structures) in residues around G7 and around E11-G12. This ensemble is more prone to adopting a more diverse set of secondary structures, especially turn, bend, coil, and β sheet character. Indeed, under this force field β structure is sampled throughout most of the peptide, peaking at populations of $\sim 15\text{-}25\%$ for F4-M5, K10-K12, and V16-V17. α character is poorly sampled in the ILDN ensemble. There is also a significant difference in the contact maps for the two force fields. While the trends in the maps bear some similarity (ie, antiparallel character between F4-M5 and L8-K12), OPLS populates these interresidue contacts with a far higher probability, particularly favoring this antiparallel as well as N- to C- terminal contacts. While the main antiparallel for A β 40 was also characterized by higher contact probabilities for the OPLS force field (see Chapter 2 and 3), the effect for this system is much more prevalent. This suggests the OPLS ensemble is significantly more collapsed than the ILDN run, an observation supported by analysis of radii of gyration later in this chapter.

Simulation at 2500 bar reveals that the primary effect of high pressure on the ILDN ensemble is an exchange of extended β sheet and turn structure for random coil. In particular, β structure near the center of the peptide (K10-K12) is most prone to being lost, followed by structure at the N-terminus (F4-M5). The difference contact map in the lower right reveals a loss of contacts between the C-terminus and the rest of the peptide, while contact within the N-terminal and central regions are promoted. Both the changes in secondary structure (loss of β and turn for coil) and tertiary structure (favoring local contacts and disfavoring contact between termini) are very similar to the effects of high pressure on A β 40 with the same force field, as described in Chapter 3.

5.4.2.2 RMSD Based Clustering and Salt Bridges

To identify relevant conformations of the ensemble, we employed clustering on frames of the $N_{\alpha\text{-syn}}$ simulations; the top centroids are illustrated in Figure 5.8. As discussed in the Methods section, the single linkage methodology employed to study A β and SOD1 peptides in this thesis was found to be ineffective in clustering the $N_{\alpha\text{-syn}}$ ensembles, so instead, Daura cluster analysis(351) was employed with an optimized C α RMSD cutoff of 3.0 Å.

The clustering reveals very different results produced by the OPLS and ILDN configurations. The resulting structures for the ILDN ensemble share some similarity in that they are all relatively extended structures that nonetheless have at least one major turn in the center of the peptide, forming “horseshoe-like” structures. Explicit antiparallel β hairpins are formed in centroid 1-3 between the regions that fold back on each other. On the other hand, clustering the OPLS simulation with the same criteria produced substantially fewer clusters, with the top 10 clusters representing a far greater proportion of the ensemble, even when the equivalent period of simulation time is considered (i.e., using the clustering statistics in parentheses for ILDN in the figure); we also note this was true, but to a substantially lesser extent, for the clustering of A β simulations (see Chapter 2 and 3). This suggests that the OPLS ensemble is far less dynamic than the ILDN equivalent. Further, the representative structures for OPLS are noticeably more collapsed than their ILDN counterparts, in line with a substantial, ensemble-wide bias toward lower R_g states, as we will see in the next section. Unlike ILDN, no extended β structure is formed for any of the top centroids; in fact, the only secondary structure observed is a small α -helix in the N-terminus of centroid 1.

The more ball-like conformations of OPLS are also highly stabilized by salt bridges between distant residues in the peptide sequence; for example, centroid 1 has salt bridges between the zwitterionic ends as well as K6 and D18’s side chains, centroid 2 also has direct contact between the charged termini which also participates in an ionic network with D2 and K12’s side chains, centroid 3 has a M1’s cationic end interact with D18’s side chain, and centroid 4 has a D2 to K12 salt bridge. In contrast, ILDN only has 2 centroids with salt bridges: both centroid 2 and 4 have a single bridge between K6 and

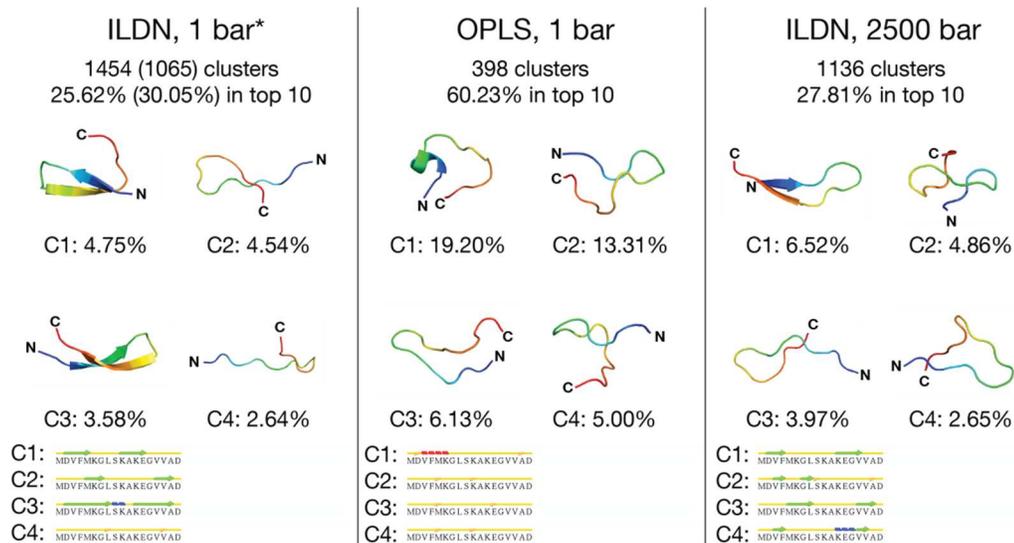


Figure 5.8. The four highest-populated centroids from Daura cluster analysis(351) around $C\alpha$ atoms with an RMSD cutoff of 3.0 \AA for the N_{α} -syn REMD ensembles over the production period. Ribbons are colored by a blue-to-red rainbow spectrum from the N-to-C terminus. Secondary structure strips for each centroid, as calculated by STRIDE(361), are also shown. Figure rendered in Pymol(360). The clusters for ILDN 1 bar's ensemble were calculated over more structures (42000 vs. 24500) due to the greater simulation time (400-1000 ns/replica), this is denoted with an asterisk. In parentheses are the statistics for clustering for that simulation using the same time window as the other simulations (150-500 ns/replica) for the sake of comparison.

D18's side chains. As a general note, over the entire ensemble, ILDN only samples one salt bridge with high propensity, the same one between K6 and D18 (populated 6.4%). Meanwhile, OPLS samples several: K6 and D18's side chains (13.5%), K10 and E13 (40.6%), D2 and K12 (13.8%), and K12 and D18's side chain (8.5%). The higher stability of ionic interactions under the OPLS combination may be a driving force for the more collapsed (see next section) and rigid nature of the resulting ensemble (though it may also be a concurrent symptom; for example, these phenomena could both occur if OPLS were inherently a more hydrophobic peptide model).

A comparison of the clustering results in Figure 5.8 with ILDN at 2500 bar versus 1 bar (over the same window of time, see statistics in parentheses of the figure for ILDN 1 bar) reveals that pressure marginally increases the number of identified clusters and reduces the percent of representation of the top 10 centroids. We identified a similar effect for the application of high pressure in the REMD of A β 40 (see section 3.4.9) and suggested that this implies that pressure increases the diversity of the free energy landscape (an assertion we corroborate in the next section), however, the changes in these values seen here for the much smaller N $_{\alpha}$ -syn peptide are more subtle than the A β 40 case. Observing the top 4 centroid structures in the 2500 bar simulation reveals similar overall properties to that of 1 bar; for example, centroid 1 and 3 of the 2500 bar set also form extended β interactions. Nevertheless, more collapsed structures, like centroid 2 and 4, are also present.

5.4.2.3 Radius of Gyration

Illustrated in Figure 5.9A are histograms for the radii of gyration for the different N $_{\alpha}$ -syn simulations. The Kolinski-Skolnick radius of gyration (363) for a hypothetical, fully collapsed 18 residue “protein” is 0.66 nm. The OPLS histogram has a single peak close to this value, ranging from ~0.62 to 1.00 nm, while the ILDN profiles have a very different, much broader distribution of extended states, nontrivially populating R_g values up to 1.6 nm. In the case of A β simulations, OPLS did populate a narrower distribution of overall smaller R_g s as well compared to ILDN (see Chapter 2 and 3), but the difference there was much more subtle compared the results we see here for the N $_{\alpha}$ -syn

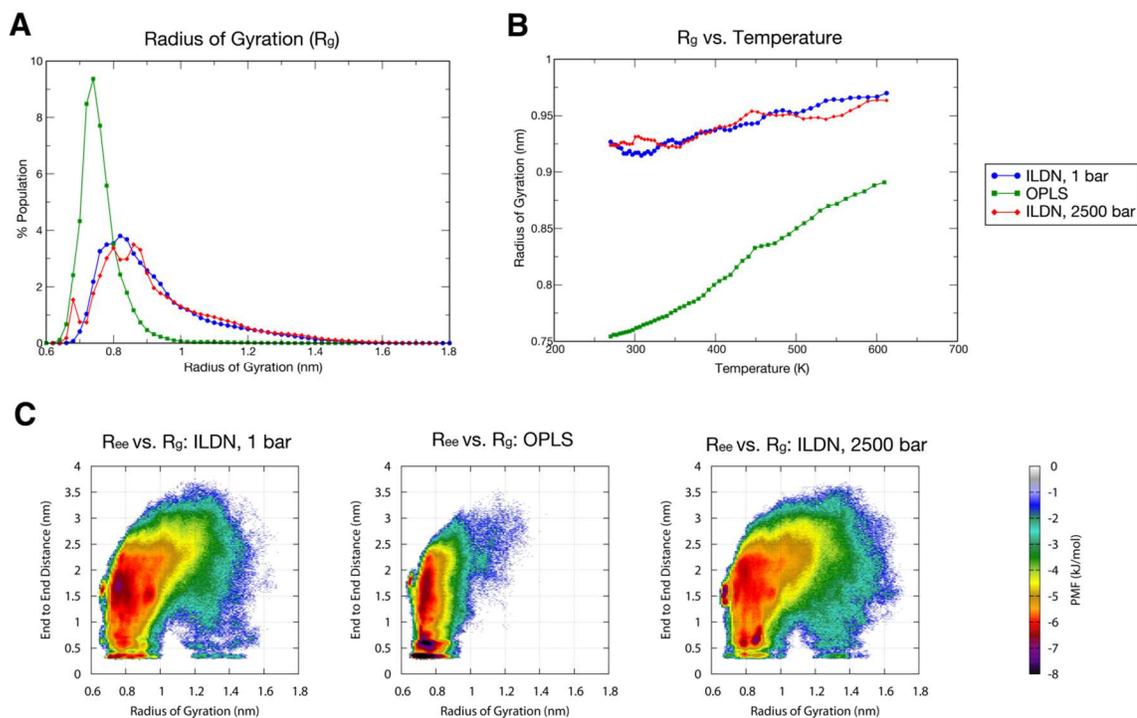


Figure 5.9. Radius of gyration (R_g) and end to end distance (R_{ee}) for simulations of N_{α} -syn. (A) Histogram for R_g over the production ensemble for each of the N_{α} -syn simulations, using 0.02 nm-wide bins. (B) Average R_g as a function of temperature over all frames in the production ensemble. (C) Free energy surface along the coordinates of R_g and R_{ee} in units of kJ/mol for the production ensemble using 0.02 nm binning in both dimensions. The free energy surface of ILDN, 1 bar is calculated over the time window as the other simulations (150-500 ns/replica) for the sake of comparison.

system. Similarly disparate are the plots of radius of gyration as a function of temperature in panel B of the same figure. OPLS shows a strong increase of R_g with temperature, while ILDN starts with a much higher average R_g value at low temperatures to begin with, so the increase with temperature is weak. The dissimilarities between OPLS and ILDN are such that the average R_g s at OPLS's highest REMD temperatures (going up to 612.0 K) are still lower than the same values at ILDN's lowest REMD temperatures (as low as 270.0 K) by 0.03-0.04 nm. This suggests that the forces promoting collapsed states specific to the OPLS configuration (such as an increased stabilization of electrostatic interactions, as discussed in the previous section) are not fully overcome (on an ensemble-averaged basis) even when the system is thermalized to very high temperatures.

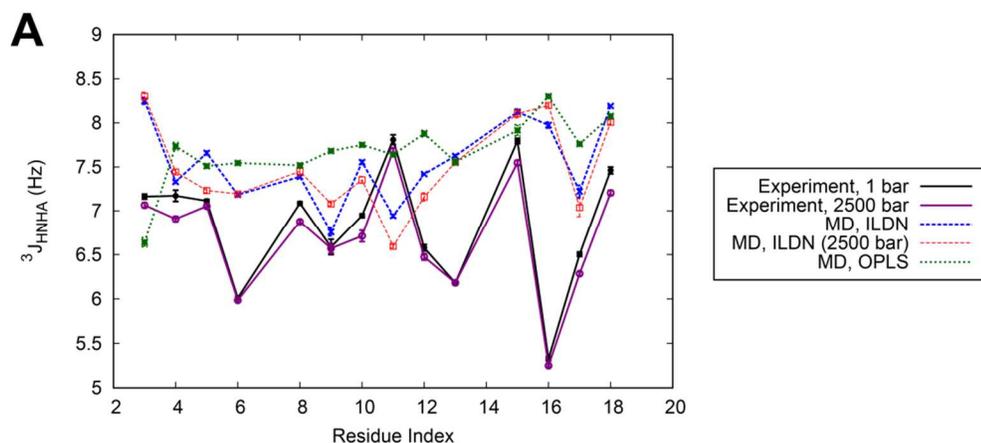
Figure 5.9C illustrates the free energy plots using radius of gyration and end to end distance as order parameters for all of the $N_{\alpha\text{-syn}}$ runs. To make the comparison between runs more straightforward, this surface was calculated over the same time window for all simulations (150-500 ns/replica). As we have covered, OPLS is highly biased to low R_g states compared to ILDN, which reflects as a dramatic change in the surface wherein essentially the entire right side of the landscape (>1.3 nm) is inaccessible to the OPLS ensemble. But interestingly, of those low R_g states, OPLS seems to share some basins in common with ILDN. Though low end to end, low R_g states are accessible to ILDN, there is a strong bias to these states in OPLS. ILDN can also sample low end to end, high R_g states (largely inaccessible to OPLS), which corresponds to the observation seen in the centroids of the peptide folding upon itself but nevertheless sampling extended states. The effect of pressure on the free energy surface for ILDN $N_{\alpha\text{-syn}}$ is more subtle, but is similar to the effect of pressure on A β 40 (see section 3.4.9) – favoring a broader landscape with fewer deep energy minima, and in particular, biasing the ensemble to populations with more extreme R_g values (the population with $R_g < 0.7$ nm as well as the right side of the surface with $R_g > 1.0$ nm).

5.4.2.4 Comparison of Chemical Shifts and $^3J_{\text{HNHA}}$ to Experiment

In Chapter 4, we discussed the utility of NMR in the experimental validation of IDPs, described observables such as chemical shifts and $^3J_{\text{HNHA}}$ couplings typically used

for these purposes, and compared these experimentally determined values to the computational equivalents for different forms of A β . Using NMR data collected as a function of different pressures for N $_{\alpha\text{-syn}}$, in Figure 5.10 we present the results of similar comparisons for this system with our simulation. Comparison of both N and H $_N$ chemical shifts yield very favorable results for the ILDN simulations, while OPLS is affiliated with substantially worse Pearson correlation coefficients (PCCs) and RMSDs. This is unlike the case of A β in Chapter 4, where direct comparison of chemical shifts (or $^3J_{\text{HNHA}}$ couplings) did not differentiate the two force field combinations well. It is also worth noting that, while the comparison of N chemical shifts led to better agreement for the A β systems compared to full N $_{\alpha\text{-syn}}$, the H $_N$ chemical shifts are much better predicted for ILDN N $_{\alpha\text{-syn}}$ than the A β equivalent. For ILDN simulations, among the chemical shifts predictors used here, SPARTA+(433) led to lower RMSDs for predictions compared to experiment versus SHIFTX+(434), and generally yielded better correlations for most data sets too. In general, it was observed that a large source of disagreement with chemical shifts for N $_{\alpha\text{-syn}}$ occurred with the residues forming the termini of the peptide. V16 in particular was the worst predicted residue for all simulations (also true for $^3J_{\text{HNHA}}$ coupling as we will discuss in the next paragraph), and removal of V3 and V16-A17 from the data uniformly and substantially improved the fidelity of the prediction. Lastly, while PCC was substantially unaffected, RMSD of prediction increased substantially when comparing predicted chemical shifts for simulation data at 2500 bar compared to the experimental chemical shifts taken at the same pressure, either suggesting that our simulation of the system at high pressure may not be accurately capturing the effect of this perturbation on the experimental ensemble, or the chemical shift predictors themselves may not be well calibrated for predicting chemical shifts for high pressure NMR.

While validation of the N $_{\alpha\text{-syn}}$ simulated ensembles through chemical shifts yielded promising results, particularly for ILDN simulations, the comparisons between experiment and simulation made using $^3J_{\text{HNHA}}$ couplings were far less encouraging. With A β (Chapter 4), we saw that calculation of J-couplings from simulation using Vuister and Bax's (362) parameters for the Karplus equation led to an overall bias to higher



B

Chemical Shift Index (full peptide)

Atom Type	ILDN Primary / Secondary PCC		ILDN RMSD (ppm)		OPLS Primary / Secondary PCC		OPLS RMSD (ppm)		ILDN, 2500 bar Primary / Secondary PCC		ILDN, 2500 bar RMSD (ppm)	
	N	0.96/ 0.68	0.93/ 0.59	2.11	3.40	0.94/ 0.42	0.91/ 0.47	2.65	3.58	0.96/ 0.67	0.94/ 0.61	2.70
H _N	0.78/ 0.74	0.81/ 0.78	0.14	0.15	0.20/ 0.20	0.53/ 0.49	0.27	0.19	0.63/ 0.60	0.79/ 0.77	0.23	0.23

Chemical Shift Index (residue F4-V15)

N	0.99/ 0.87	0.97/ 0.82	1.17	2.52	0.97/ 0.53	0.95/ 0.66	1.95	2.88	0.99/ 0.92	0.98/ 0.89	1.81	3.16
H _N	0.84/ 0.79	0.86/ 0.82	0.12	0.14	0.27/ 0.26	0.65/ 0.56	0.26	0.17	0.79/ 0.72	0.87/ 0.83	0.19	0.21

³J_{HNHA}

ILDN PCC	ILDN RMSD (Hz)	OPLS PCC	OPLS RMSD (Hz)	ILDN, 2500 bar PCC	ILDN, 2500 bar RMSD (Hz)
0.11	1.04	-0.23	1.18	-0.11	1.14

C

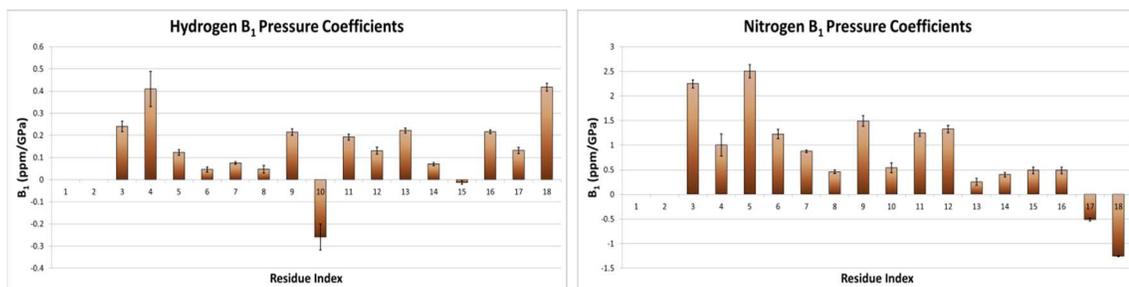


Figure 5.10. Correlation and RMSD of calculated chemical shifts and $^3J_{\text{HNHA}}$ couplings for $N_{\alpha\text{-syn}}$ simulations compared to experimentally determined values. (A) Plot of experimentally and computationally derived $^3J_{\text{HNHA}}$ couplings as a function of residue index. (B) Direct comparison of experimental NMR observables with computationally back-calculated values. “Primary” refers to the comparison of the chemical shifts as they are, while “secondary” correlation is measured once both the predicted and experimental chemical shifts are corrected for their random coil values. Even in the full peptide case, resonances for M1 and D2 are not observed, while the final residue, D18, is unreliably predicted by both calculators and is ignored in the analysis of chemical shifts. (C) First order pressure coefficients for N and H_N atoms based on the chemical shifts as a function of pressure, fitting the data according to the method described in Chapter 4. $^3J_{\text{HNHA}}$ couplings were back-calculated with Karplus parameters published by Vuister and Bax(362). Experimental data provided by Dr. Julien Roche and Prof. Ad Bax at the Laboratory of Chemical Physics of NIDDK and NIH.

$^3J_{\text{HNHA}}$ compared to experiment (except in regions where the experimental $^3J_{\text{HNHA}}$ was highest), but the trend for these values nevertheless matched well with experiment, yielding good correlation. For all $N_{\alpha\text{-syn}}$ values, however, there is both an overall overestimation of $^3J_{\text{HNHA}}$, reduced range of values sampled in simulation, and a poor reproduction of the trends of the data. Also unlike $A\beta$, OPLS and ILDN have rather different profiles, with OPLS assuming a flatter profile with overall higher $^3J_{\text{HNHA}}$ and sampling some different couplings, particularly at residue V3. Both OPLS and ILDN 2500 bar yield an overall negative correlation to experiment, while ILDN 1 bar yields a weak positive correlation. Particular outliers include regions where the experimental $^3J_{\text{HNHA}}$ is lowest, particularly at the C-terminus (E13, V16), as well as an anticorrelation to experiment in the center of the peptide at A11. The effect of pressure manifests, in general, as very subtle decreases in $^3J_{\text{HNHA}}$ for values from both ILDN simulations and experiment, however, the residues that are substantially affected by this change are different for both sets. That PCC is reduced and RMSD is increased at 2500 bar (most likely driven by the changes in the anticorrelation caused by pressure at A11), again, suggests that we may not be accurately simulating the effect of pressure on the ensemble.

In Chapter 4, chemical shifts across HSQCs collected at different pressures were fit to a second order Taylor polynomial (Equation 4.4) for $A\beta$. The first order pressure coefficients (B_I) that comes from this fit were calculated for each residue and compared across the sequence. Using the same methodology, pressure coefficients were also calculated for $N_{\alpha\text{-syn}}$ here as a function of residue, for nitrogen and hydrogen resonances of each amide group. Examination of these values and other NMR data lends some insight as to why the simulation may be in disagreement with experiment at the C-terminus. This region is characterized by unusually low B_I values for nitrogen and high B_I values for hydrogen; an unusual combination that was not observed for the pressure coefficients of the $A\beta$ peptides in Chapter 4 (see Figure 4.8A). Coupled with this is the observation that even at ambient pressure, residue V16-D18's nitrogen chemical shifts are each 5-7 ppm downstream of random coil values the respective residues, while substantially lower R_I relaxation rates and heteronuclear steady state ^1H - ^{15}N NOE ratios were also observed in this region compared to the rest of the peptide (*data not published*;

reported here with permission). The C-terminal region may also be biased to helical, coil, or turn based structure, since the $^3J_{\text{HNHA}}$ is also quite low in this region. Either way, these experimental data suggest that the C-terminus may be deshielded, highly flexible, and resistant to the denaturing changes imposed by pressure. An inability of simulations to model these phenomena could underlie the poor agreement between simulation and experiment seen in this region.

5.5 Discussion

Both of the projects presented in this chapter represent our efforts to assess the effectiveness of our simulation methods, presented in Chapter 2 and Chapter 3, in the *de novo* simulation of other small, disease-relevant peptides, several of which are known IDPs. Unfortunately, comparison of the simulation results with experiment for either of these projects show that our simulations do not reproduce all of the collected experimental data with the same fidelity we report for A β (as seen in Chapter 4), which may present a sobering depiction of the flaws of our technique.

In the case of SOD1 peptides, a comparison of radius of gyration and end to end distance suggests that, while simulations may be capturing the right size distributions of collapsed peptides such as the CTP (at 0M urea), they are very inaccurate for known extended peptides like the NTP, strongly biasing to collapsed states. Ultimately, this means the simulation method is very weak to distinguish between these different categories of peptide. Poor reproducibility/high standard deviations in the simulated values are another disquieting concern for the measurements made here. The most popularly used force fields, including those used for the OPLS and ILDN simulations in this thesis, are largely parameterized to collapse and fold globular proteins (see Chapter 1). It has recently been reported that these force fields (and others like them) overemphasize collapsed states for theoretically extended peptides, such as intrinsically disordered or unfolded peptides(46-48, 52, 53). Further, that radii of gyration of unfolded proteins and the equivalent folded forms were barely distinguishable even through long timescale simulations with our current force fields(53). The underestimation of the NTP radius of gyration, and inability to distinguish the collapsed CTP from the extended NTP are failings of this study that are consistent with these

observations. These deficiencies may be addressed by changing the force field to one that has been shown to better reproduce these experimental parameters (CHARMM22* or AMBER03w-ILDN have both been suggested to produce more extended states and better agreement with experiment for IDPs(46, 47, 488)), though even these (or similar) force fields have been suggested to underestimate radii of gyration of unfolded proteins compared to experimental measures(52, 53). Other approaches include strengthening short-range protein-water pair interactions(48, 489), modifying the weight of other nonbonded interactions in pre-existing force fields to deemphasize the nonspecific protein-protein interactions that promote collapsed states, or employing experimental-based constraints on either simulation or the analysed ensemble. Parameter changes can be optimized to improve agreement with experimental FRET or SAXS data, or the osmotic pressures of individual amino acids. We discuss these strategies in more detail in the Future Directions section (Section 6.2.1).

Beyond these global shape parameters, the NMR analysis of $N_{\alpha\text{-syn}}$ suggests that our simulation method may not even be capturing the right local structure for peptides with weak structural biases. Comparisons reveal that the simulations were able to reproduce experimental chemical shift data but not $^3J_{\text{HNHA}}$ using the ILDN force field, but neither comparison is reproduced by OPLS. Further, even with ILDN, the effect of the pressure perturbation on $N_{\alpha\text{-syn}}$ did not seem to be well represented by our comparison with NMR observables under the same condition. The bias toward bend structure and strong ionic interactions in the OPLS simulations of $N_{\alpha\text{-syn}}$ compared to that of ILDN is reminiscent of the differences in the central turn in the simulations of A β seen between Chapter 2 and 3, where OPLS favored a highly-populated bend region stabilized by highly populated side chain electrostatic interactions, where ILDN supported a mixture of turn and bend like structures that depleted these interactions in favor of helical-like hydrogen bonds. For the smaller $N_{\alpha\text{-syn}}$ peptide, the effect is more pronounced and ultimately results in a very different ensemble compared to ILDN, with a drastically reduced radius of gyration. As little is in common between the two simulation studies and experiment for $N_{\alpha\text{-syn}}$, we are unable to make generalizable conclusions about the system, unlike A β .

One of the goals of the SOD1 project was to clarify the molecular basis underlying the CTP's unique changes in end to end distance with urea, and its possible ties to full,

disulfide-reduced SOD1's observed FLF. Unfortunately, the monomeric simulations of peptide reveal a largely disordered ensemble, and seemingly offer little insight into these phenomena. These experimental observations could, however, be explained if the CTP peptide is prone to aggregation, while the other peptides are not. Self-association could produce a high intermolecular FRET signal, and disaggregation by urea could reduce that signal. No visible aggregates were seen during experiments, and an indirect control for aggregation was run by observing the FRET as a function of varying peptide concentrations (see Methods). These experiments revealed that FRET efficiencies did not substantially change with [peptide], suggesting transfer does not come from intermolecular sources. Nevertheless, we suggest more direct controls by analyzing the solution for soluble oligomers using analytical ultracentrifugation (AUC). Further, we also propose a set of control FRET runs using equivalent experimental setup and peptides, but each having one chromophore rather than two. Half of the peptides should have tryptophan (donor-only), while others uniquely have E-EDANS without tryptophan (acceptor-only). FRET efficiency using those constructs must come from intermolecular association, so that component of the efficiency seen in the donor-acceptor experiments presented in this section can be isolated. Substantial intermolecular FRET would lend support to this aggregation-centric hypothesis proposed here. Lastly, we saw through REMD simulation that the CTP represented a very disordered ensemble, producing little secondary or tertiary structure bias, and resulting in many more clusters of low ensemble representation when clustered with the same criteria used to characterize other peptides (NTP and A β). Although possible structured seeds of aggregation were not identified through the REMD simulation like they were for A β , the sheer diversity of the CTP's landscape may be critical to its unique aggregation prone nature, if that hypothesis indeed turned out to be true.

6. Concluding Remarks and Future Directions

6.1 Concluding Remarks

The work in this thesis aimed to accurately define the conformational states of certain disease-relevant intrinsically disordered peptides through the union of experimental data and simulations. The rough, diverse free energy profiles of these peptides both represent a fringe case that tests the limit of the validity of the current state of our computational modeling techniques and preclude the utility of many experimental techniques we currently use to efficiently characterize globular proteins.

To attack the problem, we have represented the systems in the simulations in this thesis in a highly accurate manner by using all atom models, explicit water, and state of the art force fields. Further, to model as much of the conformational landscapes of these peptides as possible, we have employed enhanced sampling through the replica exchange molecular dynamics technique, and simulated these systems on longer timescales than had been done in the past using this methodology. We saw the need for this extensive sampling in order to approach convergence for the emerging β structure of the A β monomers simulated in Chapter 2 and 3.

To assess the generalizability of our conclusions, we have simulated these systems with multiple force fields and models, such as the OPLS model in Chapter 2 and the ILDN model in Chapter 3. Further, we have collected atomic resolution structural data on these peptides through NMR experiment, which was the focus of Chapter 4. These values typically only represent an ensemble averaged property of the system, but nevertheless, are used as a means for assessing the validity of the simulations described in this thesis. We found it instructive to point out where discrepancies lie between experimental data and simulations of multiple models, but ultimately lend greater weight to conclusions derived from where these data actually reach agreement. The insight from the union of these models has provided an understanding of the ensemble biases of A β peptide that is both supported by several different methods and that correlates to experimentally-determined intrapeptide models of higher order aggregates. Finally, we have attempted to test the rigor of our simulation techniques by using them to study other IDP systems that we also had experimental data for, described in Chapter 5.

6.2 Future Directions

On that last point, we met some difficulty in recapturing the experimental properties of other IDPs, and even for A β , where a consistent backbone secondary and tertiary character seemed to emerge, the comparison between methods was far from perfect. We saw, for example, that OPLS and ILDN produced very different representations for the central region of A β and the effects of familial Alzheimer's disease mutants, as well as very dissimilar ensembles for the whole N $_{\alpha}$ -syn system. Ultimately, the experimental data appeared to disfavor the OPLS model of electrostatics more of the two, but the ILDN model was not necessarily capturing the local interactions properly either (in particular, see the N $_{\alpha}$ -syn project in Chapter 5).

6.2.1 Simulations

Deficiencies in our current simulation technique may arise from a number of sources. First, fundamentally, we may be using inappropriate force fields for representing the more disordered or extended states of the ensemble. We saw evidence of this through the SOD1 peptide project in Chapter 5, where shape parameters for the experimentally extended NTP system were not captured through simulation, but more collapsed peptides like the CTP seemed to be represented better. OPLS-AA(19) turned out to universally produce more collapsed ensembles in this thesis than AMBER99sb-ILDN(21), while AMBER99sb-ILDN itself has been reported in the literature to be more biased to lower radii of gyration than other force fields in the simulation of IDPs, including CHARMM22*(28) and AMBER03w(490) (with and without ILDN modifications); these latter choices are better in reproducing experimental distributions(46, 47, 488). Additionally, a recent force field based on AMBER99sb-ILDN, termed AMBER99-IDPs, has also been designed specifically to model the ensembles of intrinsically disordered proteins, improving the agreement of simulation data with secondary chemical shifts and residual dipolar couplings for two IDP systems(491). However, it is possible these force fields may not reproduce the NMR data with reasonable accuracy, like OPLS and ILDN have for A β (see Chapter 4). Further, this approach may not even fix the issue at hand, since even these force fields

(or similar) have been suggested to largely underestimate radii of gyration compared to experiment for unfolded proteins(52, 53).

To begin to test whether this is the case, we have started to simulate A β 40 using REMD using the CHARMM22* force field(28) and the modified CHARMM TIP3P (TIP3SP) water model(20) (hereafter called the “CHARMM” condition). As of the writing of this thesis, two separate REMD runs have been simulated to 500 ns/replica; both simulations are illustrated with a summary figure (Figure 7.8 in the Appendix). The first run uses an identical methodology as the AMBER99sb-ILDN REMD simulations described in Chapter 3 (we call this “same box” or SB), while the second (denoted as “big box” or BB) uses a larger box size of 5.9 nm and 6846 water molecules as opposed to a box size of 5.4 nm and 4947 waters used in all other A β REMD runs (after NPT equilibration). The BB version was run with 66 replicas, while the SB version uses 52 replicas, both encompassing the temperature range ~270-600 K. Both were analyzed over the period of 150-500 ns/replica for 7 different temperatures encompassing the range ~290-310 K: BB is analyzed over temperatures 7-13, while SB is analyzed over temperatures 6-12. The BB run was done to accommodate the more extended structures expected to be produced by CHARMM22* and reduce the likelihood of periodic boundary contacts, while the SB run was done to observe how that change in box size (and number of replicas) might affect the ensemble. Indeed, it was observed that the BB run had 0.1% of frames in the production period as being less than or equal to 0.5 nm away from its periodic image, while SB had 1.5% of frames fall into this category.

Interestingly, both CHARMM simulations produce an ensemble-averaged contact map similar in profile to A β 40 simulated with other force fields (compare Figure 7.8A with Figure 2.3 and Figure 3.3), all possessing contacts between the central hydrophobic cluster (CHC, L17-A21) and C-terminus and a central region that is isolated from rest of peptide. This lends further support to the ensemble characterization of A β made in this thesis. However, unlike OPLS and ILDN, these contacts only weakly follow an antiparallel, with a stronger bias for the CHC to interact with more C-terminal residues (L34-V36). Next, we observe that CHARMM SB is characterized with stronger contacts between the N-terminal and C-terminal regions than CHARMM BB (possibly indicating a crowding effect due to the smaller box). Nonetheless, either CHARMM simulation is

less prone to form intrapeptide contacts overall relative to those simulated with other force fields; in terms of probability of intrapeptide contacts, the trend is OPLS > ILDN > CHARMM. These characterizations are in line with the observation that either CHARMM simulation samples a much more extended ensemble, as seen when we compare the free energy surface using radius of gyration and end to end distance as order parameters (Figure 7.8B) with those typical of OPLS (Figure 2.9) and ILDN (Figure 3.7), and the histograms for these values for A β 40 across force fields (Figure 7.8C and 7.8D). We observe that box size did have an effect on these parameters, with CHARMM SB showing a substantial bias to sample lower R_{ee} , matching with the increase in N- to C-terminal contact seen in the contact maps. The effect of box size on R_g was more subtle, with CHARMM SB sampling a slightly broader distribution, particularly at low R_g . Despite this, for both CHARMM simulations, the ordering for the ensemble average of both R_g and R_{ee} follow the trends CHARMM > ILDN > OPLS (ignoring the small population with $R_g \sim 1.8$ nm for OPLS that are known to be biased by periodic boundary contacts, as described in Chapter 2). Notably, both CHARMM simulations have a positively skewed R_g distribution with a long right tail, indicating that CHARMM samples configurations with R_g between 1.35-1.7 nm that both OPLS and ILDN sample sparingly. This was the expected result in choosing this field, and it does suggest that it (along with choosing a large box size) may help avoid a bias toward overestimating ensemble compaction, if one should indeed exist for A β monomer simulations. However, since there are no experimental values for R_g/R_{ee} of A β monomers (to our knowledge) in the literature (presumably due to its tendency to aggregate), we plan to simulate other peptides with known measurements of R_g/R_{ee} through FRET/SAXS (for example, the SOD1 peptides described in Chapter 5) in order to assess whether CHARMM simulation parameters can indeed reproduce these experimental data.

Despite CHARMM simulations sharing strong commonalities to ILDN and OPLS simulations of A β 40 in terms of biases in tertiary structure, DSSP(347, 348) annotation reveals a somewhat different characterization of secondary structure (compare Figure 7.8E and Figure 7.8F to Figure 2.2 and Figure 3.2). Both CHARMM simulations produce a profile for the first half of the peptide that is similar to the other force fields, particularly: a CHC that samples substantial β character (up to $\sim 45\%$ at the apex at F19

for CHARMM BB, which is less than the ~60% for both OPLS and ILDN, but still comparable), some β bias in E3-H6 and E11-H13, and turn character particularly in G9-Y10 and H14-Q15. CHARMM SB formed substantially less β character in these CHC and E11-H13 peaks, but otherwise possesses very similar secondary structure to the BB variant throughout the peptide. For both CHARMM simulations, the central motif is also prone to forming irregular, turn, and coil structure, with G25-S26 substantially biased toward turn like the ILDN simulation (though V24 is also characterized by turn for ILDN, but not CHARMM). Where CHARMM simulations primarily differ from OPLS and ILDN is the characterization of more C-terminal regions. In addition to the G25-S26 turn, there are new turns at G29-A30, G33-L34, and G37-G38 not present in ILDN and OPLS, which are more prone to characterize these regions with bend (particularly for G29-A30 and G37-G38) or β character (G33-L34). The new turn regions in CHARMM are predominantly interspersed with irregular/coil character. The presence of strongly populated turn regions in the C-terminus imply the presence of helical hydrogen bonds, and the spacing of the apex of these turn regions occur in an i to $i+4$ pattern, suggesting a bias to helical-like sampling. However, the presence of canonical α -helices or 3-helices is sparing, with no residue sampling more than 10% of either character in the BB simulation. Lastly, we note that the pattern for these four C-terminal turns align to three GxxxG motifs in the A β sequence, with turns starting at G25, G29, G33, and G37. The exception to all of these trends in the C-terminus is a substantial peak of β character, which is strongest in M35-G37 at ~30%. This C-terminal peak is very different from that of ILDN and OPLS, however, which both strongly characterize A30-L34 with β character, reaching ~50-60% at I31-I32 (these residues are <20% β character in both CHARMM simulations, in contrast). The similarity of contact maps but differences in secondary structure biases suggest that the nature of backbone hydrogen bonding is substantially different in the CHARMM simulations.

As previously mentioned, we are also interested if CHARMM simulations are able to reproduce the NMR data to the same degree as ILDN and OPLS simulations, using the same methods described in Chapter 4. Figure 7.8G correlates experimental chemical shift to those calculated from simulation frames with the SPARTA+(433) and SHIFTX+(434) chemical shift predictors. Comparing these to predictions made from

simulations of other force fields in Figure 4.5A reveals that both CHARMM simulations predict all CSs with nearly uniformly lower RMSD than OPLS or ILDN for all nuclei and with both predictors. Indeed, N shifts are predicted with nearly 1 ppm smaller RMSD for CHARMM BB/SB versus either force field. Further, in general, both primary and secondary PCCs to the experimental data are substantially improved by CHARMM, particularly for amide nitrogen and amide hydrogen shifts. The only exceptions to this trend are secondary PCCs for H α and C β shifts: ILDN and particularly OPLS do better for H α shifts, while ILDN is by far the best predictor for C β shifts when it comes to secondary PCC. Nevertheless, in general, CHARMM appears to substantially improve the prediction of chemical shifts relative to other force fields. Next, when we compare $^3J_{\text{HNHA}}$ couplings backcalculated from the Karplus equation(435) with Vuister and Bax(362) parameters to experimental values in Figure 7.8H and Figure 7.8I and compare to Figure 4.6A and Figure 4.6B, we observe that CHARMM reproduces $^3J_{\text{HNHA}}$ couplings with roughly the same fidelity of OPLS and ILDN, with the BB variant producing a slightly smaller PCC of 0.62 versus the ILDN value of 0.71 and the OPLS value of 0.72, but also a smaller RMSD of 0.61 Hz versus ILDN's 0.68 Hz and OPLS's 0.77 Hz. CHARMM seems to improve agreement in the N-terminus and central region relative to ILDN/OPLS, sharing those fields' problem of overestimating certain J-couplings in the N-terminus and central region (such as E11, K16, A21-E22). That said, agreement at the C-terminus is noticeably worse, and regions that anticorrelates to the experimental trends, such as D23 and I32-V36, are observed. However, all 3 force fields, importantly, agree with experiment in characterizing the CHC (V18-F20) and V40 with higher than average J-coupling (> 8 Hz). In general, box size seems to have a very subtle effect on the overall $^3J_{\text{HNHA}}$ couplings for CHARMM, but the apparent changes between SB and BB exacerbate some of the aforementioned anticorrelations, worsening the overall agreement (lowered PCC of 0.47, raised RMSD of 0.69 Hz). In general, we conclude the ability of CHARMM to reproduce experimental $^3J_{\text{HNHA}}$ is comparable overall to that of the other force fields, while its ability to predict experimental chemical shifts is decidedly superior.

Despite the apparent successes of the CHARMM22* A β 40 simulation described above, should simulations with this (and the other force fields described above) fail to

adequately reproduce experimental data, we may also opt to directly modify a pre-existing force field to try to remove biases toward collapsed states. One proposed solution to this problem is to increase the strength of short-range protein-water pair interactions(48, 489, 492). In one example, the AMBER03w force field(490) and TIP4P/2005 water model(493) were modified so that the Lennard-Jones interactions between the water oxygen and all protein atoms were scaled 1.1 fold, tuned to best reproduce the distribution of radii of gyration for the REMD ensemble of a 34-residue fragment of cold-shock protein (Csp M34) to experimental Förster resonance energy transfer (FRET) data(489). This change was observed to be sufficient to reproduce dimensional data from small-angle X-ray scattering (SAXS) and FRET for an IDP (activator for thyroid hormone and retinoid receptors or ACTR), without needing to change water-water and protein-protein parameters and without adversely affecting the stability of the folded state of other ordered proteins (ubiquitin, CspTm, human lysozyme, and spectrin R15)(489). Another approach for addressing these issues come from our lab's previous success at improving the behavior of the osmolyte trimethylamine N-oxide (TMAO) by tuning the charge and cosolvent-cosolvent Lennard-Jones parameters to better match experimental values of osmotic pressure(458, 494). Our lab is currently applying similar techniques to improve agreement with osmotic data for each of the individual amino acids. When this project is complete, we expect this will result in more realistic interactions of protein with water that will correct the inherent susceptibility of simulations to sample unfolded or disordered ensembles that are too collapsed.

Further, we can consider exploring constrained molecular dynamics simulations rather than the current unbiased approach. Constraints may be derived from distance limitations from NOESY/ROESY data and/or angular restrictions given by RDCs or J-couplings. We could also introduce this bias by modify pre-existing force fields with empirically-derived parameters or added restraints to improve simulation behavior. Biasing potentials for simulations based on ensemble-averaged experimental NMR data such as nuclear Overhauser effect (NOE) intensities, hydrogen exchange rates, J-couplings, and S^2 order parameters have previously been introduced to complement force fields in atomistic simulations, penalizing NMR observables or distances averaged

over simulation time(495-498) and/or multiple concurrent simulations(499, 500). Another possibility, as Ball *et al.*(17) have already explored to some success for A β , is to continue to use unbiased force fields to explore the conformational space of IDPs but use the NMR data to select a better weighted ensemble from a pool of simulated conformations, for example, by using the ENSEMBLE package(228-230).

Further, it is very possible our systems are still not converged, and pursuing even longer sampling times and devising superior methods for assessing convergence may be needed. Convergence is often measured by comparison to an experimentally-determined structure, however, that option is not viable for intrinsically disordered peptides. The primary method we have been using throughout the thesis is measurement of overall secondary structure character as a function of simulation time, ensemble averaged over multiple replicas and a set bin of simulation time. This metric is a discerning physical feature of the ensemble, but it could be inherently flawed because different structures could have similar overall secondary profiles but very different conformations (particularly true in the case of intrinsically disordered peptides). An unambiguous measure of convergence based on tertiary structure profile would be better for our purposes. With that in mind, we propose the following method, based on ideas suggested by fellow colleague Anthony Bishop: break the trajectory into equal segments of simulation time, analyze first segment with an appropriate clustering algorithm (like single linkage(349) or Daura(351) clustering) and RMSD cutoff to create a series of reference centroid structures, then, in all subsequent segments, group each of the structures of the ensemble to the reference structure that has the lowest RMSD. When this analysis was applied to the OPLS A β 42 simulation described in Chapter 2 (Figure 7.9 in the Appendix), it was observed that the highest populated clusters seen in the first analyzed segment gained an even greater representation of the population in future simulation times up to a point (ie, between 500-633 ns and 633-866 ns), then the distribution of populations did not change substantially (between 633-866 ns and 866-1000 ns). The increased population of higher ranked centroids suggests the deeper sampling of the affiliated energy minima, while the limited changes in distribution seen at later time points may indicate that the ensemble is thermodynamic equilibrium and convergence has been reached.

Lastly, we have observed in this thesis (and elsewhere) interesting behaviors for the A β 40-E22G monomer in the NMR tube, including: very large chemical shift perturbations in ^1H - ^{15}N HSQC surrounding the mutation (Figure 4.3C), a highly shielded G22 resonance in these spectra (Figure 7.4 in the Appendix), and both reduced R_2 rates(54) and diminished sensitivity of chemical shift movement to pressure (Figure 4.8A) in residues near the mutation compared to wild type A β 40. To clarify the molecular basis behind these observations, we have simulated the A β 40-E22G monomer using the AMBER99sb-ILDN(21) force field and the TIP4P-Ew(22) water model, using an identical methodology as the REMD simulations described in Chapter 3. This simulation has only very recently (as of the writing of this thesis) completed its run to 1000 ns/replica, so a figure summarizing the results of this simulation is presented in Figure 7.10 in the Appendix. We observe that the E22G has a prominent effect on the β distribution of the peptide, primarily shifting the peak in the central hydrophobic cluster (CHC) N-terminally, while biasing the peak normally amongst residues A30-V36 toward the C-terminus (Figure 7.10D). In other words, E11-Q15 and L34-V39 are more prone to be involved in β sheets, while F18-D23 and N27-I32 lose β structure. The difference contact map (Figure 7.10A) confirms that E22G favors a register for the main antidiagonal where more N-terminal partners in the CHC interact with more C-terminal partners, relative to wild type. This map also illustrates an increased interaction of the N-terminus with more global parts of the peptide (an effect seen with all of the point mutants of A β , seen in Figure 3.3). In general, the central region is broadened (Figure 7.10C): in fringe regions (F19-D23 and G29-A30) β -sheet character is replaced by coil/irregular character, while the center of the region (V24-S26) supplants bend character for an increase in α -helix annotations. Interestingly, the distribution of turn is mostly unaffected. Figure 7.10B further reveals that A β 40-E22G samples a broader distribution of radii of gyration than A β 40, including a sizable population with R_g of 1.3-1.6 nm that is otherwise sparsely populated in all of the other A β simulations with this force field (Figure 3.7A).

Figure 7.10E plots S^2 order parameters for each backbone amide group, values which relate to the angular flexibility of the nitrogen-hydrogen internuclear vector. In addition to reduced β character and increased coil/irregular annotation in residues near

the site of the mutation as described in the previous site, S^2 order parameters are also greatly diminished near the E22G site, revealing that the backbone amides are more dynamic. This suggests that this region is substantially destabilized by the mutation. Observations of experimentally reduced R_2 relaxation rates(54) and reduced B_1 pressure coefficient (Figure 4.8A) near the E22G substitution further sustain this conclusion. However, as seen in Figure 7.10F, NMR reveals very low nitrogen and hydrogen chemical shifts for the G22 residue compared to other glycines in the system, but chemical shift prediction from the frames of the simulation fails to capture a large degree of the magnitude of this effect. This is quite possibly a failing of the empirical-based chemical shift predictors (which are based largely on globular proteins possessing both NMR chemical shift and X-ray crystal data) to properly back-calculate chemical shifts from shielding confirmations that may not occur in the training sets they were parameterized against, but it is also possible that the simulation frames are not capturing some structural or electronic phenomena responsible for the unique NMR properties of this resonance.

Beyond this A β 40-E22G simulation, we plan to continue to run simulations for other variants of A β , including those investigating the residue H6-D7 hotspot involved in familial AD (see Table 1.1).

6.2.2 Experiment

Poor agreement between simulation and experiment may also stem from the possibility that the experimental observables studied here only capture a limited or biased portion of the structural ensemble. More global NMR observables, such as nuclear/rotating-frame Overhauser effect (NOE/ROE) values, S^2 order parameter and relaxation rate estimations, and paramagnetic relaxation enhancement (PRE) line broadening of cysteine mutants, may be more challenging to acquire and interpret but may be able to circumvent these limitations. We are particularly interested in PRE, a technique that broadens peaks within a certain proximity to a spin label that affiliates with cysteine residues in the protein(501). The relaxation is dependent on the square of the gyromagnetic ratio of the nucleus of detection(502). Using normal ^1H detection, this radius would be 11 Å(502), which would be likely to broaden the majority of resonances

in the A β monomer, whose radius of gyration was observed to be \sim 8-13 Å in simulation (see Chapter 2 and 3). However, ^{13}C detection experiments have also been successfully used to reduce the PRE broadening range; theoretically, this change in detection nucleus decreases the radius \sim 16-fold, but in practice, it reduces the extent to 6 Å(502). Our lab is currently working on studying PRE broadening using ^{13}C detection to probe intramolecular distances in spin-label affiliated cysteine mutants of A β monomers. Constructs of the D1C, S26C, and V40C single mutants of A β 40 have already been created by our lab for this purpose. Additionally, ^{13}C direct detection spectroscopy is also inherently far less prone to spectral overlap and solvent exchange, making it well suited toward the study of intrinsically disordered proteins(503-505). We have begun implementing pulse sequences that make use of this technique and plan to use them to supplement the chemical shift information we have for A β in the future.

On that note, the experimental data we have right now for A β are almost exclusively specific to the structural states of the backbone, which are insufficient to clarify some of the discrepancies discussed in this thesis, such as the role of the pathogenically-linked residues E22 and D23 in the central region. Overexpression and purification of ^{13}C labeled A β peptide has become practical and cost-effective with advent of the protocol described in Chapter 4. Experiments including (H)C(CCO)NH-TOCSY and H(CCCO)NH-TOCSY experiments(429, 430) as well as ^{15}N -TOCSY-HSQC(506) and HCCH-TOCSY(507, 508) will be used in tandem with the backbone assignment experiments discussed in Chapter 4 to get as close to a full assignment for each mutant. The experiments in Chapter 4 have demonstrated that non-uniform sampling (NUS) techniques are highly effective at reducing acquisition time while preserving the integrity of the NMR data; use of NUS will help us overcome limitations in acquisition time due to sample aggregation. With these assignments, we intend to obtain NOE constraints using NOESY experiments. Analysis of ^1H - ^1H NOESY(509) or ^{15}N NOESY-HSQC(506, 510) by our group (*unpublished data*) and by others(13, 202, 511) for A β peptides has revealed very few unambiguous long range NOEs that would serve as useful distance constraints. The homonuclear NOESY is very difficult to assign because of spectral overlap (ie, only \sim 25% of crosspeaks were able to be assigned by Ball *et al.*(13, 511)), while the ^{15}N NOESY-HSQC only reveals NOEs that involve amide

groups, excluding much of the side chain-specific information that may be more prone to adopt interactions on a longer range or timescale. The ^{13}C NOESY-HSQC(506, 512) represents a compromise between these techniques, granting the increased spectral resolution that comes from extra dimensionality while (in theory) allowing determination of any NOE involving an aliphatic group, representing a much broader set of NOEs than those theoretically available through the ^{15}N NOESY-HSQC. Constraints determined from these NOEs may allow us to validate and discern between the simulation models proposed in this thesis.

We plan to continue the high pressure NMR studies described in Chapter 4. In line with the philosophy of the previous paragraph, we propose extending the analysis to encompass ^{13}C HSQCs in order to describe the sensitivity of side chains to pressure. One major challenge with this project, however, is that, unlike that for backbone amides(445), the changes for random coil shifts for ^{13}C correlated resonances over pressure for each residue type have not yet been measured our knowledge; these are necessary to remove residue specificity in our calculation of pressure coefficients. Though we could synthesize ^{13}C labeled model peptides for each amino acid to acquire these data, we plan instead to perform these experiments on unlabeled peptides, relying on the natural abundance of ^{13}C in these samples. Synthesizing these unlabeled peptides is much more cost-effective, but more concentrated samples will be needed to get reliable NMR data (only 1.1% of the carbons in the sample will be NMR active ^{13}C isotopes(394)).

Next, we have observed in Chapter 4 that high pressure can be used to manipulate the aggregation of A β . The resulting data heights in our experiments, however, were affected by both pressure and the acquisition time in between collection of HSQCs, since these spectra were sequentially collected on one sample. In the case of the A β 40-E22G peptide, these conditions led to observed monomer loss due to aggregation with time at lower pressures, but signal recovery at later data points, presumably due to the disaggregating effect of high pressure (Figure 4.8B). To properly study the effect of pressure on the association/disassociation of A β , we need to either remove the confounding variable of time or, ideally, deconvolute this variable from pressure to form a third axis. Kamatari *et al.*(416) used cycles of upward and downward pressure-jump

NMR experiments to control the aggregation of another IDP, disulfide-deficient hen lysozyme. Following their example, we plan to prepare multiple HFIP-NaOH disaggregated samples of A β at the same concentration and buffer conditions as our other pressure experiments (see Chapter 4, Methods), then measure their NMR spectra at the same initial time $t=0$ (or, as close as practically possible), but with each sample at a different pressure. Multiple HSQCs will then be collected maintaining the same pressure, effectively plotting the each residue's aggregation curve as a function of pressure and time in a three dimensional scheme. On the other side, we can also start with peptide aggregated at ambient pressure on a long timescale (to approach thermodynamic equilibrium), then employ a rapid jump to higher pressures and collect successive HSQCs as pressure dissociates the peptide until equilibrium is reached under this new condition. Both the experiments starting from monomeric A β and those starting from mature aggregate described above will inform the rate constants of aggregation (k_{on}) and disaggregation (k_{off}) for A β for a given target pressure. Changing the target pressure would be expected to change the resulting rate constants, so we can run multiple pressure-jump experiments from both starting conditions to fully define the pressure-dependence kinetics of the aggregation/disaggregation process between monomer and the aggregates that form the NMR dark state. This can be done for each A β species that we have previously investigated with NMR. Further, using the resolution provided by ^1H - ^{15}N HSQCs, we can also observe if these constants may change for different amides, which may inform which residues are involved in aggregation/disaggregation. Lastly, we can also how these rate constants change with different temperatures, or the introduction of small compounds or proteins that modulate A β aggregation(21, 233-235, 355).

Another project planned by our lab in an effort to validate our simulations is to explore the conformation of different A β variants through titrations with the Z_{A β 3} Affibody, described by Hoyer *et al.*(355). To summarize, a series of 16 proteins, called Affibodies, were engineered from a combinatorial library of proteins based on the staphylococcal protein A-derived Z domain and selected for their affinity for A β 40 using phage display methods(513). The structure of the strongest of these binders, Z_{A β 3}, in complex with monomeric A β 40, was solved by solution NMR (PDB ID: 2OTK)(355).

The $Z_{A\beta 3}$ Affibody is a disulfide-linked homodimer that binds $A\beta 40$ monomers with nanomolar affinity and stoichiometrically inhibits its fibrillation(355). The $Z_{A\beta 3}$ -bound $A\beta 40$ monomer includes a β hairpin from residues 17-36, between the central hydrophobic cluster (CHC) and C-terminus(355). We have observed through our simulations in Chapter 2 and Chapter 3 (Figure 2.7 and Figure 3.6) that different alloforms and mutants of $A\beta$ have varying propensity to sample conformations with low RMSD to this trapped $A\beta 40$ monomer structure. We therefore posit that these different forms of $A\beta$ may interact with $Z_{A\beta 3}$ with different binding affinities and/or complexed conformations. To test this hypothesis, we plan to overexpress and purify the $Z_{A\beta 3}$ protein and use it in order to assess the affinity, site of binding, and effect on aggregation for different variants of $A\beta$. Experiments that will be conducted in order to obtain this information include NMR, isothermal titration calorimetry (ITC), and fluorimetric analysis. We anticipate that the binding of the Affibody will serve as a conformational probe, providing insight as to the degree to which modifications of $A\beta$ disturb the CHC to C-terminal β hairpin conformations that we report in Chapter 2 and 3 of this thesis. By correlating measured binding affinities with the propensity of each variant to form the C-terminal β hairpin in the REMD ensembles, we predict that this information may be useful for simulation validation. As of the writing of this thesis, our lab has requested and received a construct expressing $Z_{A\beta 3}$, courtesy of Prof. Fredrik Frejd, Prof. Torleif Hård, Prof. John Löfblom, and the Affibody AB company. Further, we have overexpressed and purified unlabeled $Z_{A\beta 3}$ protein and titrated it into each of three different ^{15}N labeled variants of $A\beta$ ($A\beta 40$, $A\beta 42$, and $A\beta 40\text{-E22G}$), while collecting high quality $^1\text{H}\text{-}^{15}\text{N}$ HSQCs at different stoichiometries, allowing us to infer the binding affinity and binding site of the resulting complexes. Unfortunately, the details and results of these experiments cannot be reported in this thesis document due to the stipulations of the Materials Transfer Agreement between Affibody AB and Rensselaer Polytechnic Institute, but a publication including these experiments is planned for the future.

REFERENCES

1. V. N. Uversky, J. R. Gillespie, A. L. Fink, Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins Struct. Funct. Genet.* **41**, 415-427 (2000).
2. T. Härd, C. Lendel, Inhibition of amyloid formation. *J. Mol. Biol.* **421**, 441-465 (2012).
3. T. P. J. Knowles, M. Vendruscolo, C. M. Dobson, The amyloid state and its association with protein misfolding diseases. *Nature Rev. Mol. Cell Biol.* **15**, 384-396 (2014).
4. L. Hou *et al.*, Solution NMR studies of the A β (1-40) and A β (1-42) peptides establish that the Met35 oxidation state affects the mechanism of amyloid formation. *J. Am. Chem. Soc.* **126**, 1992-2005 (2004).
5. N. Batsch, M. Mittelman, World Alzheimer report 2012: overcoming the stigma of dementia. *London Alzheimers Dis. Int.* **75**, (2012).
6. G. Bitan, S. S. Vollers, D. B. Teplow, Elucidation of primary structure elements controlling early amyloid β -protein oligomerization. *J. Biol. Chem.* **278**, 34882-34889 (2003).
7. G. Bitan *et al.*, Amyloid β -protein (A β) assembly: A β 40 and A β 42 oligomerize through distinct pathways. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 330-335 (2002).
8. D. J. Selkoe, M. B. Podlisny, Deciphering the genetic basis of Alzheimer's disease. *Annu. Rev. Genomics Hum. Genet.* **3**, 67-99 (2002).
9. N. G. Sgourakis, Y. Yan, S. A. McCallum, C. Wang, A. E. Garcia, The Alzheimer's peptides A β 40 and 42 adopt distinct conformations in water: a combined MD / NMR study. *J. Mol. Biol.* **368**, 1448-1457 (2007).
10. S. Mitternacht, I. Staneva, T. Härd, A. Irbäck, Comparing the folding free-energy landscapes of A β 42 variants with different aggregation properties. *Proteins Struct. Funct. Genet.* **78**, 2600-2608 (2010).
11. N. G. Sgourakis *et al.*, Atomic-level characterization of the ensemble of the A β (1-42) monomer in water using unbiased molecular dynamics simulations and spectral algorithms. *J. Mol. Biol.* **405**, 570-583 (2011).
12. S. Côté, P. Derreumaux, N. Mousseau, Distinct morphologies for amyloid β protein monomer: A β ₁₋₄₀, A β ₁₋₄₂, and A β ₁₋₄₀(D23N). *J. Chem. Theory Comput.* **7**, 2584-2592 (2011).
13. K. A. Ball *et al.*, Homogeneous and heterogeneous tertiary structure ensembles of amyloid- β peptides. *Biochemistry* **50**, 7612-7628 (2011).
14. Y. S. Lin, G. R. Bowman, K. A. Beauchamp, V. S. Pande, Investigating how peptide length and a pathogenic mutation modify the structural ensemble of amyloid β monomer. *Biophys. J.* **102**, 315-324 (2012).
15. Y. S. Lin, V. S. Pande, Effects of familial mutations on the monomer structure of A β 42. *Biophys. J.* **103**, L47-49 (2012).
16. D. Granata, Ph.D. Thesis, SISSA (2013).
17. K. A. Ball, A. H. Phillips, D. E. Wemmer, T. Head-Gordon, Differences in β -strand populations of monomeric A β 40 and A β 42. *Biophys. J.* **104**, 2714-2724 (2013).

18. M. Zhu *et al.*, Identification of small-molecule binding pockets in the soluble monomeric form of the A β 42 peptide. *J. Chem. Phys.* **139**, 035101 (2013).
19. G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **105**, 6474-6487 (2001).
20. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-926 (1983).
21. K. Lindorff-Larsen *et al.*, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Genet.* **78**, 1950-1958 (2010).
22. H. W. Horn *et al.*, Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **120**, 9665-9678 (2004).
23. A. T. Petkova, W.-M. Yau, R. Tycko, Experimental Constraints on Quaternary Structure in Alzheimer's β -Amyloid Fibrils. *Biochemistry* **45**, 498-512 (2006).
24. A. K. Paravastu, R. D. Leapman, W.-M. Yau, R. Tycko, Molecular structural basis for polymorphism in Alzheimer's β -amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18349-18354 (2008).
25. L. Yu *et al.*, Structural characterization of a soluble amyloid β -peptide oligomer. *Biochemistry* **48**, 1870-1877 (2009).
26. M. Ahmed *et al.*, Structural conversion of neurotoxic amyloid- β_{1-42} oligomers to fibrils. *Nature Struct. Biol.* **17**, 561-567 (2010).
27. I. Bertini, L. Gonnelli, C. Luchinat, J. Mao, A. Nesi, A new structural model of A β 40 fibrils. *J. Am. Chem. Soc.* **133**, 16013-16022 (2011).
28. S. Piana, K. Lindorff-Larsen, D. E. Shaw, How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **100**, L47-L49 (2011).
29. L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradović, A. K. Dunker, Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **323**, 573-584 (2002).
30. V. N. Uversky, C. J. Oldfield, A. K. Dunker, Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* **37**, 215-246 (2008).
31. K. S. Sandhu, Intrinsic disorder explains diverse nuclear roles of chromatin remodeling proteins. *J. Mol. Recogn.* **22**, 1-8 (2009).
32. A. Tantos, K.-H. Han, P. Tompa, Intrinsic disorder in cell signaling and gene transcription. *Mol. Cell. Endocrinol.* **348**, 457-465 (2012).
33. P. Romero *et al.*, Sequence complexity of disordered protein. *Proteins Struct. Funct. Genet.* **42**, 38-48 (2001).
34. G. W. Daughdrill, G. J. Pielak, V. N. Uversky, M. S. Cortese, A. K. Dunker, in *Protein Folding Handbook*, J. Buchner, T. Kiefhaber, Eds. (Wiley-VCH Verlag GmbH, Weinheim, Germany, 2008), pp. 275-357.
35. K. Hukushima, K. Nemoto, Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **65**, 1604-1608 (1996).
36. Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141-151 (1999).

37. A. Laio, M. Parrinello, Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562-12566 (2002).
38. A. Laio, F. L. Gervasio, Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **71**, (2008).
39. V. Hornak *et al.*, Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Genet.* **65**, 712-725 (2006).
40. R. B. Best, G. Hummer, Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* **113**, 9004-9015 (2009).
41. H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269-6271 (1987).
42. H. W. Horn *et al.*, Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **120**, 9665-9665 (2004).
43. D. E. Shaw *et al.*, Atomic-level characterization of the structural dynamics of proteins. *Science* **330**, 341-346 (2010).
44. C. K. Fisher, C. M. Stultz, Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **21**, 426-431 (2011).
45. P. Wolynes, J. Onuchic, D. Thirumalai, Navigating the folding routes. *Science* **267**, 1619-1620 (1995).
46. S. Rauscher *et al.*, Structural ensembles of intrinsically disordered proteins depend strongly on force field. *Biophys. J.* **106**, 271a (2014).
47. S. Rauscher, V. Gapsys, B. de Groot, H. Grubmüller, Structural ensembles of intrinsically disordered proteins using molecular dynamics simulation. *Biophys. J.* **108**, 14a (2015).
48. J. Henriques, C. Cragnell, M. Skepö, Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. *J. Chem. Theory Comput.* **11**, 3420-3431 (2015).
49. K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, How fast-folding proteins fold. *Science* **334**, 517-520 (2011).
50. S. Piana, K. Lindorff-Larsen, D. E. Shaw, Protein folding kinetics and thermodynamics from atomistic simulation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17845-17850 (2012).
51. J. E. Kohn *et al.*, Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12491-12496 (2004).
52. K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana, D. E. Shaw, Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J. Am. Chem. Soc.* **134**, 3787-3791 (2012).
53. S. Piana, J. L. Klepeis, D. E. Shaw, Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **24**, 98-105 (2014).
54. C. R. Connors, Ph.D. Thesis, Rensselaer Polytechnic Institute (2013).
55. A. Association, 2014 Alzheimer's disease facts and figures. *Alzheimers Dement.* **10**, 47-92 (2014).

56. L. E. Hebert, J. Weuve, P. A. Scherr, D. A. Evans, Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology* **80**, 1778-1783 (2013).
57. O. Zanetti, S. B. Solerte, F. Cantoni, Life expectancy in Alzheimer's disease (AD). *Archives Gerontol. Geriatr.* **49**, 237-243 (2009).
58. A. Burns, S. Iliffe, Alzheimer's disease. *Br. Med. J.* **338**, (2009).
59. A. Burns, R. Jacoby, R. Levy, Psychiatric phenomena in Alzheimer's disease. IV: Disorders of behaviour. *Br. J. Psychiatry* **157**, 86-94 (1990).
60. A. B. Rodgers, *Alzheimer's Disease: Unraveling the Mystery*. (National Institute of Aging, Bethesda, MD, 2008), vol. 1.
61. M. M. Breteler *et al.*, Cognitive correlates of ventricular enlargement and cerebral white matter lesions on magnetic resonance imaging. The Rotterdam Study. *Stroke* **25**, 1109-1115 (1994).
62. B. J. Cummings, C. W. Cotman, Image analysis of β -amyloid load in Alzheimer's disease and relation to dementia severity. *Lancet* **346**, 1524-1528 (1995).
63. M. R. Emmerling, Water-soluble A β (N-40, N-42) oligomers in normal and Alzheimer disease brains. *J. Biol. Chem.* **271**, 4077-4081 (1996).
64. C. A. McLean *et al.*, Soluble pool of A β amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease. *Ann. Neurol.* **46**, 860-866 (1999).
65. A. Harada *et al.*, Altered microtubule organization in small-calibre axons of mice lacking tau protein. *Nature* **369**, 488-491 (1994).
66. Z. A. Levine, L. Larini, N. E. LaPointe, S. C. Feinstein, J.-E. Shea, Regulation and aggregation of intrinsically disordered peptides. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2758-2763 (2015).
67. N. Shahani, R. Brandt, Functions and malfunctions of the tau proteins. *Cell. Mol. Life Sci.* **59**, 1668-1680 (2002).
68. L. M. Ittner, J. Götz, Amyloid- β and tau — a toxic pas de deux in Alzheimer's disease. *Nature Rev. Neurosci.* **12**, 67-72 (2011).
69. M. L. Billingsley, R. L. Kincaid, Regulated phosphorylation and dephosphorylation of tau protein: effects on microtubule interaction, intracellular trafficking and neurodegeneration. *Biochem. J.* **323 (Pt 3)**, 577-591 (1997).
70. A. Alonso, T. Zaidi, M. Novak, I. Grundke-Iqbal, K. Iqbal, Hyperphosphorylation induces self-assembly of tau into tangles of paired helical filaments/straight filaments. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6923-6928 (2001).
71. H. G. Lee *et al.*, Tau phosphorylation in Alzheimer's disease: pathogen or protector? *Trends Mol. Med.* **11**, 164-169 (2005).
72. M. Kidd, Paired helical filaments in electron microscopy of Alzheimer's disease. *Nature* **197**, 192-193 (1963).
73. S. Yagishita, Y. Itoh, W. Nan, N. Amano, Reappraisal of the fine structure of Alzheimer's neurofibrillary tangles. *Acta Neuropathol.* **54**, 239-246 (1981).
74. C. Bouras, P. R. Hof, P. Giannakopoulos, J. P. Michel, J. H. Morrison, Regional distribution of neurofibrillary tangles and senile plaques in the cerebral cortex of elderly patients: a quantitative evaluation of a one-year autopsy population from a geriatric hospital. *Cereb. Cortex* **4**, 138-150 (1994).
75. J. L. Price, J. C. Morris, Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. *Ann. Neurol.* **45**, 358-368 (1999).

76. J. J. Kril, S. Patel, A. J. Harding, G. M. Halliday, Neuron loss from the hippocampus of Alzheimer's disease exceeds extracellular neurofibrillary tangle formation. *Acta Neuropathol.* **103**, 370-376 (2002).
77. R. J. Castellani, A. Nunomura, H. G. Lee, G. Perry, M. A. Smith, Phosphorylated tau: toxic, protective, or none of the above. *J. Alzheimers Dis.* **14**, 377-383 (2008).
78. T. Iwatsubo *et al.*, Visualization of A β 42(43) and A β 40 in senile plaques with end-specific A β monoclonals: Evidence that an initially deposited species is A β 42(43). *Neuron* **13**, 45-53 (1994).
79. T. Iwatsubo, T. C. Saido, D. M. Mann, V. M. Lee, J. Q. Trojanowski, Full-length amyloid- β (1-42(43)) and amino-terminally modified and truncated amyloid- β 42(43) deposit in diffuse plaques. *Am. J. Pathol.* **149**, 1823-1830 (1996).
80. C. S. Atwood, R. N. Martins, M. A. Smith, G. Perry, Senile plaque composition and posttranslational modification of amyloid- β peptide and associated proteins. *Peptides* **23**, 1343-1350 (2002).
81. R. N. Kalaria, S. N. Kroon, I. Grahovac, G. Perry, Acetylcholinesterase and its association with heparan sulphate proteoglycans in cortical amyloid deposits of Alzheimer's disease. *Neuroscience* **51**, 177-184 (1992).
82. L. S. Perlmutter, H. C. Chui, D. Saperia, J. Athanikar, Microangiopathy and the colocalization of heparan sulfate proteoglycan with amyloid in senile plaques of Alzheimer's disease. *Brain Res* **508**, 13-19 (1990).
83. A. D. Snow *et al.*, The presence of heparan sulfate proteoglycans in the neuritic plaques and congophilic angiopathy in Alzheimer's disease. *Am. J. Pathol.* **133**, 456-463 (1988).
84. J. M. Candy *et al.*, Aluminosilicates and senile plaque formation in Alzheimer's disease. *Lancet* **1**, 354-357 (1986).
85. M. A. Lovell, J. D. Robertson, W. J. Teesdale, J. L. Campbell, W. R. Markesbery, Copper, iron and zinc in Alzheimer's disease senile plaques. *J. Neurol. Sci.* **158**, 47-52 (1998).
86. R. Strohmeyer, J. Rogers, Molecular and cellular mediators of Alzheimer's disease inflammation. *J. Alzheimers Dis.* **3**, 131-157 (2001).
87. T. Duong, E. C. Pommier, A. B. Scheibel, Immunodetection of the amyloid P component in Alzheimer's disease. *Acta Neuropathol.* **78**, 429-437 (1989).
88. K. Ueda *et al.*, Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11282-11286 (1993).
89. P. H. Weinreb, W. Zhen, A. W. Poon, K. A. Conway, P. T. Lansbury, Jr., NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* **35**, 13709-13715 (1996).
90. A. Iwai, M. Yoshimoto, E. Masliah, T. Saitoh, Non-A β component of Alzheimer's disease amyloid (NAC) is amyloidogenic. *Biochemistry* **34**, 10139-10145 (1995).
91. K. Vamvaca, M. J. Volles, P. T. Lansbury, Jr., The first N-terminal amino acids of α -synuclein are essential for α -helical structure formation in vitro and membrane binding in yeast. *J. Mol. Biol.* **389**, 413-424 (2009).

92. M. A. Smith, in *International Review of Neurobiology*, R. A. H. Ronald J. Bradley, J. Peter, Eds. (Academic Press, San Diego, CA, 1998), vol. 42, pp. 1-54.
93. H. Yamaguchi *et al.*, Alzheimer type dementia: diffuse type of senile plaques demonstrated by β protein immunostaining. *Prog. Clin. Biol. Res.* **317**, 467-474 (1989).
94. A. Lorenzo, B. A. Yankner, β -amyloid neurotoxicity requires fibril formation and is inhibited by congo red. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12243-12247 (1994).
95. L. Mucke *et al.*, High-level neuronal expression of A β 1-42 in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation. *J. Neurosci.* **20**, 4050-4058 (2000).
96. A. Y. Hsia *et al.*, Plaque-independent disruption of neural circuits in Alzheimer's disease mouse models. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3228-3233 (1999).
97. W. L. Klein, G. A. Krafft, C. E. Finch, Targeting small A β oligomers: the solution to an Alzheimer's disease conundrum? *Trends Neurosci.* **24**, 219-224 (2001).
98. M. D. Kirkitadze, G. Bitan, D. B. Teplow, Paradigm shifts in Alzheimer's disease and other neurodegenerative disorders: the emerging role of oligomeric assemblies. *J. Neurosci. Res* **69**, 567-577 (2002).
99. C. Haass, D. J. Selkoe, Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β -peptide. *Nature Rev. Mol. Cell Biol.* **8**, 101-112 (2007).
100. J. Busciglio, A. Lorenzo, J. Yeh, B. A. Yankner, β -Amyloid fibrils induce tau phosphorylation and loss of microtubule binding. *Neuron* **14**, 879-888 (1995).
101. J. Götz, F. Chen, J. van Dorpe, R. M. Nitsch, Formation of neurofibrillary tangles in P301L tau transgenic mice induced by A β 42 fibrils. *Science* **293**, 1491-1495 (2001).
102. F. G. De Felice *et al.*, Alzheimer's disease-type neuronal tau hyperphosphorylation induced by A β oligomers. *Neurobiol. Aging* **29**, 1334-1347 (2008).
103. S. Oddo, L. Billings, J. P. Kesslak, D. H. Cribbs, F. M. LaFerla, A β immunotherapy leads to clearance of early, but not late, hyperphosphorylated tau aggregates via the proteasome. *Neuron* **43**, 321-332 (2004).
104. M. Rapoport, H. N. Dawson, L. I. Binder, M. P. Vitek, A. Ferreira, Tau is essential to β -amyloid-induced neurotoxicity. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6364-6369 (2002).
105. L. M. Ittner *et al.*, Dendritic function of tau mediates amyloid- β toxicity in Alzheimer's disease mouse models. *Cell* **142**, 387-397 (2010).
106. G. P. Morris, I. A. Clark, B. Vissel, Inconsistencies and controversies surrounding the amyloid hypothesis of Alzheimer's disease. *Acta Neuropathol. Commun* **2**, 135 (2014).
107. E. Karran, M. Mercken, B. De Strooper, The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. *Nature Rev. Drug Discov.* **10**, 698-712 (2011).
108. J. Hardy, D. J. Selkoe, The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353-356 (2002).

109. K. Duff *et al.*, Increased amyloid- β 42(43) in brains of mice expressing mutant presenilin 1. *Nature* **383**, 710-713 (1996).
110. M. Citron *et al.*, Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid β -protein in both transfected cells and transgenic mice. *Nature Med.* **3**, 67-72 (1997).
111. S. Kumar-Singh *et al.*, Mean age-of-onset of familial Alzheimer disease caused by presenilin mutations correlates with both increased A β 42 and decreased A β 40. *Hum. Mutat.* **27**, 686-695 (2006).
112. L. Bertram, R. E. Tanzi, The genetic epidemiology of neurodegenerative disease. *J. Clin. Invest.* **115**, 1449-1457 (2005).
113. D. Van Dyke, D. Harper, M. Dyken, Alzheimer's disease and Down syndrome. *Down Syndrome Quart.* **3**, 1-11 (1998).
114. Y. Yan, C. Wang, A β 40 protects non-toxic A β 42 monomer from aggregation. *J. Mol. Biol.* **369**, 909-916 (2007).
115. S. T. Ferreira, W. L. Klein, The A β oligomer hypothesis for synapse failure and memory loss in Alzheimer's disease. *Neurobiol. Learn. Mem.* **96**, 529-543 (2011).
116. C. Kaether, C. Haass, A lipid boundary separates APP and secretases and limits amyloid β -peptide generation. *J. Cell Biol.* **167**, 809-812 (2004).
117. G. Thinakaran, E. H. Koo, Amyloid precursor protein trafficking, processing, and function. *J. Biol. Chem.* **283**, 29615-29619 (2008).
118. W. Tharp, I. Sarkar, Origins of amyloid- β . *BMC Genomics* **14**, 290 (2013).
119. R. Sandbrink, C. L. Masters, K. Beyreuther, APP gene family alternative splicing generates functionally related isoforms. *Ann. N. Y. Acad. Sci.* **777**, 281-287 (1996).
120. R. Sandbrink, C. L. Masters, K. Beyreuther, β -amyloid protein precursor mRNA isoforms without exon 15 are ubiquitously expressed in rat tissues including brain, but not in neurons. *J. Biol. Chem.* **269**, 1510-1517 (1994).
121. S. O. Dahms *et al.*, Structure and biochemical analysis of the heparin-induced E1 dimer of the amyloid precursor protein. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5381-5386 (2010).
122. J. N. Octave, N. Pierrot, S. Ferao Santos, N. N. Nalivaeva, A. J. Turner, From synaptic spines to nuclear signaling: nuclear and synaptic actions of the amyloid precursor protein. *J. Neurochem.* **126**, 183-190 (2013).
123. C. Kaether, C. Haass, A lipid boundary separates APP and secretases and limits amyloid β -peptide generation. *J. Cell Biol.* **167**, 809-812 (2004).
124. V. W. Chow, M. P. Mattson, P. C. Wong, M. Gleichmann, An overview of APP processing enzymes and products. *Neuromol. Med.* **12**, 1-12 (2010).
125. M. Lalowski *et al.*, The nonamyloidogenic p3 fragment (amyloid β 17-42) is a major constituent of Down's syndrome cerebellar preamyloid. *J. Biol. Chem.* **271**, 33623-33631 (1996).
126. L. S. Higgins, G. M. Murphy, L. S. Forno, R. Catalano, B. Cordell, P3 β -amyloid peptide has a unique and potentially pathogenic immunohistochemical profile in Alzheimer's disease brain. *Am. J. Pathol.* **149**, 585-596 (1996).
127. F. Dulin *et al.*, P3 peptide, a truncated form of A β devoid of synaptotoxic effect, does not assemble into soluble oligomers. *FEBS Lett.* **582**, 1865-1870 (2008).

128. Y. Miller, B. Ma, R. Nussinov, Polymorphism of Alzheimer's A β 17-42 (p3) oligomers: the importance of the turn location and its conformation. *Biophys. J.* **97**, 1168-1177 (2009).
129. R. Vassar, BACE1: the β -secretase enzyme in Alzheimer's disease. *J. Mol. Neurosci.* **23**, 105-114 (2004).
130. X. Sun, G. He, W. Song, BACE2, as a novel APP θ -secretase, is not responsible for the pathogenesis of Alzheimer's disease in Down syndrome. *FASEB J.* **20**, 1369-1376 (2006).
131. S. L. Cole, R. Vassar, The Alzheimer's disease β -secretase enzyme, BACE1. *Mol. Neurodegener.* **2**, 22-22 (2007).
132. D. Dominguez *et al.*, Phenotypic and biochemical analyses of BACE1- and BACE2-deficient mice. *J. Biol. Chem.* **280**, 30797-30806 (2005).
133. G. K. Gouras *et al.*, Generation and regulation of β -amyloid peptide variants by neurons. *J. Neurochem.* **71**, 1920-1925 (1998).
134. Y. Deng *et al.*, Amyloid- β protein (A β) Glu11 is the major β -secretase site of β -site amyloid- β precursor protein-cleaving enzyme 1(BACE1), and shifting the cleavage site to A β Asp1 contributes to Alzheimer pathogenesis. *Eur. J. Neurosci.* **37**, 1962-1969 (2013).
135. C. Kaether, C. Haass, H. Steiner, Assembly, trafficking and function of γ -secretase. *Neurodegener. Dis.* **3**, 275-283 (2006).
136. K. Shirotani, D. Edbauer, S. Prokop, C. Haass, H. Steiner, Identification of distinct γ -secretase complexes with different APH-1 variants. *J. Biol. Chem.* **279**, 41340-41345 (2004).
137. S. Zhou, H. Zhou, P. J. Walian, B. K. Jap, CD147 is a regulatory subunit of the γ -secretase complex in Alzheimer's disease amyloid β -peptide production. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7499-7504 (2005).
138. M. S. Wolfe *et al.*, Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and gamma-secretase activity. *Nature* **398**, 513-517 (1999).
139. S. Prokop, K. Shirotani, D. Edbauer, C. Haass, H. Steiner, Requirement of PEN-2 for stabilization of the presenilin N-/C-terminal fragment heterodimer within the γ -secretase complex. *J. Biol. Chem.* **279**, 23255-23261 (2004).
140. S.-F. Lee *et al.*, A conserved GXXXG motif in APH-1 is critical for assembly and activity of the γ -secretase complex. *J. Biol. Chem.* **279**, 4144-4152 (2004).
141. Y.-W. Zhang *et al.*, Nicastrin is critical for stability and trafficking but not association of other presenilin/ γ -secretase components. *J. Biol. Chem.* **280**, 17020-17026 (2005).
142. P. Lu *et al.*, Three-dimensional structure of human γ -secretase. *Nature* **512**, 166-170 (2014).
143. X. Li *et al.*, Structure of a presenilin family intramembrane aspartate protease. *Nature* **493**, 56-61 (2013).
144. S. Sobhanifar *et al.*, Structural investigation of the C-terminal catalytic fragment of presenilin 1. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9644-9649 (2010).
145. M. Takami *et al.*, γ -Secretase: successive tripeptide and tetrapeptide release from the transmembrane domain of β -carboxyl terminal fragment. *J. Neurosci.* **29**, 13042-13052 (2009).

146. M. S. Wolfe *et al.*, A substrate-based difluoro ketone selectively inhibits Alzheimer's γ -secretase activity. *J. Med. Chem.* **41**, 6-9 (1998).
147. M. A. Findeis, The role of amyloid β peptide 42 in Alzheimer's disease. *Pharmacol. Ther.* **116**, 266-286 (2007).
148. J. Kim *et al.*, A β 40 inhibits amyloid deposition In vivo. *J. Neurosci.* **27**, 627-633 (2007).
149. T. Sato *et al.*, Potential link between amyloid β -protein 42 and C-terminal fragment γ 49–99 of β -amyloid precursor protein. *J. Biol. Chem.* **278**, 24294-24301 (2003).
150. Y. Qi-Takahara *et al.*, Longer forms of amyloid β protein: implications for the mechanism of intramembrane cleavage by γ -secretase. *J. Neurosci.* **25**, 436-445 (2005).
151. W. Chen *et al.*, Familial Alzheimer's mutations within APPTM increase A β 42 production by enhancing accessibility of ϵ -cleavage site. *Nature Commun.* **5**, (2014).
152. S. Funamoto *et al.*, Truncated carboxyl-terminal fragments of β -amyloid precursor protein are processed to amyloid β -proteins 40 and 42. *Biochemistry* **43**, 13532-13540 (2004).
153. M. Hiltunen, T. van Groen, J. Jolkkonen, Functional roles of amyloid- β protein precursor and amyloid- β peptides: evidence from experimental studies. *J. Alzheimers Dis.* **18**, 401-412 (2009).
154. U. Muller *et al.*, Behavioral and anatomical deficits in mice homozygous for a modified β -amyloid precursor protein gene. *Cell* **79**, 755-765 (1994).
155. H. Zheng *et al.*, β -amyloid precursor protein-deficient mice show reactive gliosis and decreased locomotor activity. *Cell* **81**, 525-531 (1995).
156. P. Tremml, H.-P. Lipp, U. Müller, L. Ricceri, D. P. Wolfner, Neurobehavioral development, adult openfield exploration and swimming navigation learning in mice with a modified β -amyloid precursor protein gene. *Behav. Brain Res.* **95**, 65-76 (1998).
157. P. R. Turner, K. O'Connor, W. P. Tate, W. C. Abraham, Roles of amyloid precursor protein and its fragments in regulating neural activity, plasticity and memory. *Prog. Neurobiol.* **70**, 1-32 (2003).
158. C. Priller *et al.*, Synapse Formation and Function Is Modulated by the Amyloid Precursor Protein. *J. Neurosci.* **26**, 7212-7221 (2006).
159. S. Gunawardena, L. S. Goldstein, Disruption of axonal transport and neuronal viability by amyloid precursor protein mutations in Drosophila. *Neuron* **32**, 389-401 (2001).
160. J. A. Duce *et al.*, Iron-export ferroxidase activity of β -amyloid precursor protein is inhibited by zinc in Alzheimer's disease. *Cell* **142**, 857-867 (2010).
161. J. E. Morley *et al.*, A physiological role for amyloid- β protein: enhancement of learning and memory. *J. Alzheimers Dis.* **19**, 441-449 (2010).
162. E. Abramov *et al.*, Amyloid- β as a positive endogenous regulator of release probability at hippocampal synapses. *Nature Neurosci.* **12**, 1567-1576 (2009).
163. H. A. Pearson, C. Peers, Physiological roles for amyloid β peptides. *J. Physiol.* **575**, 5-10 (2006).

164. Z. X. Yao, V. Papadopoulos, Function of β -amyloid in cholesterol transport: a lead to neurotoxicity. *FASEB J.* **16**, 1677-1679 (2002).
165. K. Zou, J. S. Gong, K. Yanagisawa, M. Michikawa, A novel function of monomeric amyloid β -protein serving as an antioxidant molecule against metal-induced oxidative damage. *J. Neurosci.* **22**, 4833-4841 (2002).
166. M. Tabaton, X. Zhu, G. Perry, M. A. Smith, L. Giliberto, Signaling effect of amyloid- β (42) on the processing of A β PP. *Exp. Neurol.* **221**, 18-25 (2010).
167. S. J. Soscia *et al.*, The Alzheimer's disease-associated amyloid β -protein is an antimicrobial peptide. *PLoS ONE* **5**, e9505 (2010).
168. M. P. Mattson *et al.*, Evidence for excitoprotective and intraneuronal calcium-regulating roles for secreted forms of the β -amyloid precursor protein. *Neuron* **10**, 243-254 (1993).
169. S. Ring *et al.*, The secreted β -amyloid precursor protein ectodomain APPs α is sufficient to rescue the anatomical, behavioral, and electrophysiological abnormalities of APP-deficient mice. *J. Neurosci.* **27**, 7817-7826 (2007).
170. W. T. Kimberly, J. B. Zheng, S. Y. Guénette, D. J. Selkoe, The intracellular domain of the β -amyloid precursor protein is stabilized by Fe65 and translocates to the nucleus in a Notch-like manner. *J. Biol. Chem.* **276**, 40288-40292 (2001).
171. T. Nakaya, T. Suzuki, Role of APP phosphorylation in Fe65-dependent gene transactivation mediated by AICD. *Genes Cells* **11**, 633-645 (2006).
172. J. Stöhr *et al.*, Purified and synthetic Alzheimer's amyloid beta (A β) prions. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11025-11030 (2012).
173. R. Khurana *et al.*, Mechanism of thioflavin T binding to amyloid fibrils. *J. Struct. Biol.* **151**, 229-238 (2005).
174. S. Kumar, J. Walter, Phosphorylation of amyloid β (A β) peptides - a trigger for formation of toxic aggregates in Alzheimer's disease. *Aging* **3**, 803-812 (2011).
175. M. E. van Raaij, J. van Gestel, I. M. Segers-Nolten, S. W. de Leeuw, V. Subramaniam, Concentration dependence of α -synuclein fibril length assessed by quantitative atomic force microscopy and statistical-mechanical theory. *Biophys. J.* **95**, 4871-4878 (2008).
176. T. Luhrs *et al.*, 3D structure of Alzheimer's amyloid- β (1-42) fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17342-17347 (2005).
177. R. Tycko, Solid-state NMR studies of amyloid fibril structure. *Annu. Rev. Phys. Chem.* **62**, 279-299 (2011).
178. R. Tycko, in *Proteopathic Seeds and Neurodegenerative Disease*. (Springer Science & Business Media, New York, NY, 2013), pp. 19-31.
179. R. Tycko, Physical and structural basis for polymorphism in amyloid fibrils. *Protein Sci.* **23**, 1528-1539 (2014).
180. K. L. Sciarretta, D. J. Gordon, A. T. Petkova, R. Tycko, S. C. Meredith, A β 40-lactam(D23/K28) models a conformation highly favorable for nucleation of amyloid. *Biochemistry* **44**, 6003-6014 (2005).
181. A. K. Paravastu, I. Qahwash, R. D. Leapman, S. C. Meredith, R. Tycko, Seeded growth of β -amyloid fibrils from Alzheimer's brain-derived fibrils produces a distinct fibril structure. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7443-7448 (2009).

182. G. Bitan, D. B. Teplow, Rapid photochemical cross-linking--a new tool for studies of metastable, amyloidogenic protein assemblies. *Acc. Chem. Res.* **37**, 357-364 (2004).
183. A. R. A. Ladiwala, J. S. Dordick, P. M. Tessier, Aromatic small molecules remodel toxic soluble oligomers of amyloid β through three independent pathways. *J. Biol. Chem.* **286**, 3209-3218 (2011).
184. A. R. A. Ladiwala *et al.*, Polyphenolic glycosides and aglycones utilize opposing pathways to selectively remodel and inactivate toxic oligomers of amyloid β . *ChemBioChem* **12**, 1749-1758 (2011).
185. B. Caughey, P. T. Lansbury, Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26**, 267-298 (2003).
186. D. M. Walsh *et al.*, Amyloid β -protein fibrillogenesis: structure and biological activity of protofibrillar intermediates. *J. Biol. Chem.* **274**, 25945-25952 (1999).
187. M. Schmidt *et al.*, Comparison of Alzheimer A β (1-40) and A β (1-42) amyloid fibrils reveals similar protofilament structures. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19813-19818 (2009).
188. D. M. Hartley *et al.*, Protofibrillar intermediates of amyloid β -protein induce acute electrophysiological changes and progressive neurotoxicity in cortical neurons. *J. Neurosci.* **19**, 8876-8884 (1999).
189. R. Roychaudhuri, M. Yang, M. M. Hoshi, D. B. Teplow, Amyloid β -protein assembly and Alzheimer disease. *J. Biol. Chem.* **284**, 4749-4753 (2008).
190. G. M. Shankar *et al.*, Amyloid- β protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nature Med.* **14**, 837-842 (2008).
191. M. Townsend, Effects of secreted oligomers of amyloid β -protein on hippocampal synaptic plasticity: a potent role for trimers. *J. Physiol.* **572**, 477-492 (2006).
192. M. P. Lambert *et al.*, Diffusible, nonfibrillar ligands derived from A β 1-42 are potent central nervous system neurotoxins. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6448-6453 (1998).
193. S. Lesné *et al.*, A specific amyloid- β protein assembly in the brain impairs memory. *Nature* **440**, 352-357 (2006).
194. S. Barghorn *et al.*, Globular amyloid β -peptide1-42 oligomer - a homogenous and stable neuropathological protein in Alzheimer's disease. *J. Neurochem.* **95**, 834-847 (2005).
195. G. P. Gellermann *et al.*, A β -globulomers are formed independently of the fibril pathway. *Neurobiol. Dis.* **30**, 212-220 (2008).
196. A. Deshpande, E. Mina, C. Glabe, J. Busciglio, Different conformations of amyloid β induce neurotoxicity by distinct mechanisms in human cortical neurons. *J. Neurosci.* **26**, 6011-6018 (2006).
197. M. Hoshi, Spherical aggregates of β -amyloid (amylospheroid) show high neurotoxicity and activate tau protein kinase I/glycogen synthase kinase-3 β . *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6370-6375 (2003).
198. W. M. Tay, D. Huang, T. L. Rosenberry, A. K. Paravastu, The Alzheimer's amyloid- β (1-42) peptide forms off-pathway oligomers and fibrils that are

- distinguished structurally by intermolecular organization. *J. Mol. Biol.* **425**, 2494-2508 (2013).
199. K. N. Dahlgren *et al.*, Oligomeric and fibrillar species of amyloid- β peptides differentially affect neuronal viability. *J. Biol. Chem.* **277**, 32046-32053 (2002).
 200. Y. H. Chong *et al.*, ERK1/2 activation mediates A β oligomer-induced neurotoxicity via caspase-3 activation and tau cleavage in rat organotypic hippocampal slice cultures. *J. Biol. Chem.* **281**, 20315-20325 (2006).
 201. C. L. Masters, R. Cappai, K. J. Barnham, V. L. Villemagne, Molecular mechanisms for Alzheimer's disease: implications for neuroimaging and therapeutics. *J. Neurochem.* **97**, 1700-1725 (2006).
 202. L. Hou, I. Kang, R. E. Marchant, M. G. Zagorski, Methionine 35 oxidation reduces fibril assembly of the amyloid A β -(1-42) peptide of Alzheimer's disease. *J. Biol. Chem.* **277**, 40173-40176 (2002).
 203. G. Bitan *et al.*, A molecular switch in amyloid assembly: Met(35) and amyloid β -protein oligomerization. *J. Am. Chem. Soc.* **125**, 15359-15365 (2003).
 204. G. Bitan, D. B. Teplow, Rapid Photochemical Cross-Linking - A New Tool for Studies of Metastable, Amyloidogenic Protein Assemblies. *Acc. Chem. Res.* **37**, 357-364 (2004).
 205. R. Liu, C. McAllister, Y. Lyubchenko, M. R. Sierks, Residues 17-20 and 30-35 of β -amyloid play critical roles in aggregation. *J. Neurosci. Research* **75**, 162-171 (2004).
 206. A. D. Williams *et al.*, Mapping A β amyloid fibril secondary structure using scanning proline mutagenesis. *J. Mol. Biol.* **335**, 833-842 (2004).
 207. S. L. Bernstein *et al.*, Amyloid β -protein: monomer structure and early aggregation states of A β 42 and its Pro 19 alloform. *J. Am. Chem. Soc.* **127**, 2075-2084 (2005).
 208. Y. Masuda *et al.*, Identification of Physiological and Toxic Conformations in A β 42 Aggregates. *ChemBioChem* **10**, 287-295 (2009).
 209. S. Zhang *et al.*, The Alzheimer's peptide A β adopts a collapsed coil structure in water. *J. Struct. Biol.* **130**, 130-141 (2000).
 210. K. H. Lim, H. H. Collver, Y. T. H. Le, P. Nagchowdhuri, J. M. Kenney, Characterizations of distinct amyloidogenic conformations of the A β (1-40) and (1-42) peptides. *Biochem. Biophys. Res. Commun.* **353**, 443-449 (2007).
 211. Y. Yan, S. A. McCallum, C. Wang, M35 Oxidation Induces A β 40-like Structural and Dynamical Changes in A β 42. *J. Am. Chem. Soc.* **130**, 5394-5395 (2008).
 212. D. J. Rosenman, C. R. Connors, W. Chen, C. Wang, A. E. Garcia, A β monomers transiently sample oligomer and fibril-like configurations: ensemble characterization using a combined MD/NMR approach. *J. Mol. Biol.* **425**, 3338-3359 (2013).
 213. L. Gu, S. Ngo, Z. Guo, Solid-support electron paramagnetic resonance (EPR) studies of A β 40 monomers reveal a structured state with three ordered segments. *J. Biol. Chem.* **287**, 9081-9089 (2012).
 214. K. Ono, M. M. Condrón, D. B. Teplow, Structure-neurotoxicity relationships of amyloid β -protein oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14745-14750 (2009).

215. S. Vivekanandan, J. R. Brender, S. Y. Lee, A. Ramamoorthy, A partially folded structure of amyloid- β (1-40) in an aqueous environment. *Biochem. Biophys. Res. Commun.* **411**, 312-316 (2011).
216. N. D. Lazo, M. A. Grant, M. C. Condrón, A. C. Rigby, D. B. Teplow, On the nucleation of amyloid β -protein monomer folding. *Protein Sci.* **14**, 1581-1596 (2005).
217. Y. Yan, C. Wang, A β 42 is More Rigid than A β 40 at the C Terminus: Implications for A β Aggregation and Toxicity. *J. Mol. Biol.* **364**, 853-862 (2006).
218. Y. Yan, J. Liu, S. A. McCallum, D. Yang, C. Wang, Methyl dynamics of the amyloid- β peptides A β 40 and A β 42. *Biochem. Biophys. Res. Commun.* **362**, 410-414 (2007).
219. R. Roychaudhuri *et al.*, C-terminal turn stability determines assembly differences between A β 40 and A β 42. *J. Mol. Biol.* **425**, 292-308 (2013).
220. T. M. Doran, E. A. Anderson, S. E. Latchney, L. A. Opanashuk, B. L. Nilsson, Turn nucleation perturbs amyloid β self-assembly and cytotoxicity. *J. Mol. Biol.* **421**, 315-328 (2012).
221. R. Roychaudhuri *et al.*, Gly25-Ser26 amyloid β -protein structural isomorphs produce distinct A β 42 conformational dynamics and assembly characteristics. *J. Mol. Biol.* **426**, 2422-2441 (2014).
222. A. Zhang, W. Qi, T. A. Good, E. J. Fernandez, Structural differences between A β (1-40) intermediate oligomers and fibrils elucidated by proteolytic fragmentation and hydrogen/deuterium exchange. *Biophys. J.* **96**, 1091-1104 (2009).
223. T. L. Hwang, S. Mori, A. J. Shaka, P. C. M. vanZijl, Application of phase-modulated CLEAN chemical EXchange spectroscopy (CLEANEX-PM) to detect water-protein proton exchange and intermolecular NOEs. *J. Am. Chem. Soc.* **119**, 6203-6204 (1997).
224. T. L. Hwang, P. C. van Zijl, S. Mori, Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR* **11**, 221-226 (1998).
225. N. Rezaei-Ghaleh, E. Andreetto, L.-M. Yan, A. Kapurniotu, M. Zweckstetter, Interaction between Amyloid β Peptide and an Aggregation Blocker Peptide Mimicking Islet Amyloid Polypeptide. *PLoS ONE* **6**, e20289-e20289 (2011).
226. A. Abelein, B. Bolognesi, C. M. Dobson, A. Gräslund, C. Lendel, Hydrophobicity and conformational change as mechanistic determinants for nonspecific modulators of amyloid β self-assembly. *Biochemistry* **51**, 126-137 (2012).
227. M. A. Grant *et al.*, Familial Alzheimer's disease mutations alter the stability of the amyloid β -protein monomer folding nucleus. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16522-16527 (2007).
228. W. Y. Choy, J. D. Forman-Kay, Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J. Mol. Biol.* **308**, 1011-1032 (2001).

229. J. A. Marsh, J. D. Forman-Kay, Structure and disorder in an unfolded state under non-denaturing conditions from ensemble models consistent with a large number of experimental restraints. *J. Mol. Biol.* **391**, 359-374 (2009).
230. M. Krzeminski, J. A. Marsh, C. Neale, W. Y. Choy, J. D. Forman-Kay, Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* **29**, 398-399 (2013).
231. R. Brenke *et al.*, Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **25**, 621-627 (2009).
232. M. McGann, FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **51**, 578-596 (2011).
233. J. M. Lopez del Amo *et al.*, Structural properties of EGCG-induced, nontoxic Alzheimer's disease A β oligomers. *J. Mol. Biol.* **421**, 517-524 (2012).
234. H. Kroth *et al.*, Discovery and structure activity relationship of small molecule inhibitors of toxic β -amyloid-42 fibril formation. *J. Biol. Chem.* **287**, 34786-34800 (2012).
235. C. R. Connors *et al.*, Tranilast binds to A β monomers and promotes A β fibrillation. *Biochemistry*. **52**, 3995-4002 (2013).
236. D. Granata, C. Camilloni, M. Vendruscolo, A. Laio, Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6817-6822 (2013).
237. D. R. Thal, W. S. Griffin, R. A. de Vos, E. Ghebremedhin, Cerebral amyloid angiopathy and its relationship to Alzheimer's disease. *Acta Neuropathol.* **115**, 599-609 (2008).
238. Y. Tomidokoro *et al.*, Iowa variant of familial Alzheimer's disease: accumulation of posttranslationally modified A β D23N in parenchymal and cerebrovascular amyloid deposits. *Am. J. Pathol.* **176**, 1841-1854 (2010).
239. Y. Yoshiike, D.-H. Chui, T. Akagi, N. Tanaka, A. Takashima, Specific compositions of amyloid- β peptides as the determinant of toxic β -aggregation. *J. Biol. Chem.* **278**, 23648-23655 (2003).
240. A. Jan, O. Gokce, R. Luthi-Carter, H. A. Lashuel, The ratio of monomeric to aggregated forms of A β 40 and A β 42 is an important determinant of amyloid- β aggregation, fibrillogenesis, and toxicity. *J. Biol. Chem.* **283**, 28176-28189 (2008).
241. I. Kuperstein *et al.*, Neurotoxicity of Alzheimer's disease A β peptides is induced by small changes in the A β 42 to A β 40 ratio. **29**, 3408-3420 (2010).
242. C. Van Broeckhoven *et al.*, Amyloid β protein precursor gene and hereditary cerebral hemorrhage with amyloidosis (Dutch). *Science* **248**, 1120-1122 (1990).
243. E. Levy *et al.*, Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type. *Science* **248**, 1124-1126 (1990).
244. L. Miravalle *et al.*, Substitutions at codon 22 of Alzheimer's A β peptide induce diverse conformational changes and apoptotic effects in human cerebral endothelial cells. *J. Biol. Chem.* **275**, 27110-27116 (2000).
245. T. J. Grabowski, H. S. Cho, J. P. Vonsattel, G. W. Rebeck, S. M. Greenberg, Novel amyloid precursor protein mutation in an Iowa family with dementia and severe cerebral amyloid angiopathy. *Ann. Neurol.* **49**, 697-705 (2001).

246. K. Ozawa, T. Tomiyama, M. L. Maat-Schieman, R. A. Roos, H. Mori, Enhanced A β 40 deposition was associated with increased A β 42-43 in cerebral vasculature with Dutch-type hereditary cerebral hemorrhage with amyloidosis (HCHWA-D). *Ann. N. Y. Acad. Sci.* **977**, 149-154 (2002).
247. R. Tycko, K. L. Sciarretta, J. P. R. O. Orgel, S. C. Meredith, Evidence for novel β -sheet structures in Iowa mutant β -amyloid fibrils. *Biochemistry* **48**, 6072-6084 (2009).
248. A. L. Cloe, J. P. Orgel, J. R. Sachleben, R. Tycko, S. C. Meredith, The Japanese mutant A β (Δ E22-A β (1-39)) forms fibrils instantaneously, with low-thioflavin T fluorescence: seeding of wild-type A β (1-40) into atypical fibrils by Δ E22-A β (1-39). *Biochemistry* **50**, 2026-2039 (2011).
249. O. Y. Ovchinnikova, V. H. FINDER, I. Vodopivec, R. M. Nitsch, R. Glockshuber, The Osaka FAD mutation E22 Δ leads to the formation of a previously unknown type of amyloid β fibrils and modulates A β neurotoxicity. *J. Mol. Biol.* **408**, 780-791 (2011).
250. M. Inayathullah, D. B. Teplow, Structural dynamics of the Δ E22 (Osaka) familial Alzheimer's disease-linked amyloid β -protein. *Amyloid* **18**, 98-107 (2011).
251. W. Qiang, W. M. Yau, Y. Luo, M. P. Mattson, R. Tycko, Antiparallel β -sheet architecture in Iowa-mutant β -amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4443-4448 (2012).
252. G. Di Fede *et al.*, A recessive mutation in the APP gene with dominant-negative effect on amyloidogenesis. *Science* **323**, 1473-1477 (2009).
253. M. Messa *et al.*, The peculiar role of the A2V Mutation in amyloid- β (A β) 1-42 molecular assembly. *J. Biol. Chem.* **289**, 24143-24152 (2014).
254. L. Zhou *et al.*, Amyloid precursor protein mutation E682K at the alternative β -secretase cleavage β' -site increases A β generation. *EMBO Mol. Med.* **3**, 291-302 (2011).
255. T. Tomiyama *et al.*, A new amyloid β variant favoring oligomerization in Alzheimer's-type dementia. *Ann. Neurol.* **63**, 377-387 (2008).
256. M. L. Peacock, J. T. Warren, A. D. Roses, J. K. Fink, Novel polymorphism in the A4 region of the amyloid precursor protein gene in a patient without Alzheimer's disease. *Neurology* **43**, 1254-1254 (1993).
257. T. Jonsson *et al.*, A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96-99 (2012).
258. M. Kero *et al.*, Amyloid precursor protein (APP) A673T mutation in the elderly Finnish population. *Neurobiol. Aging* **34**, 1518.e1511-1518.e1513 (2013).
259. J. A. Maloney *et al.*, Molecular mechanisms of Alzheimer disease protection by the A673T allele of amyloid precursor protein. *J. Biol. Chem.* **289**, 30990-31000 (2014).
260. I. Benilova *et al.*, The Alzheimer disease protective mutation A2T modulates kinetic and thermodynamic properties of amyloid- β (A β) aggregation. *J. Biol. Chem.* **289**, 30977-30989 (2014).
261. J. C. Janssen *et al.*, Early onset familial Alzheimer's disease: mutation frequency in 31 families. *Neurology* **60**, 235-239 (2003).

262. Y. Hori *et al.*, The Tottori (D7N) and English (H6R) familial Alzheimer disease mutations accelerate A β fibril formation without increasing protofibril formation. *J. Biol. Chem.* **282**, 4916-4923 (2007).
263. K. Ono, M. M. Condron, D. B. Teplow, Effects of the English (H6R) and Tottori (D7N) familial Alzheimer disease mutations on amyloid β -protein assembly and toxicity. *J. Biol. Chem.* **285**, 23186-23197 (2010).
264. W. T. Chen *et al.*, Amyloid- β (A β) D7H mutation increases oligomeric A β 42 and alters properties of A β -zinc/copper assemblies. *PLoS ONE* **7**, e35807 (2012).
265. Y. Wakutani *et al.*, Novel amyloid precursor protein gene missense mutation (D678N) in probable familial Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* **75**, 1039-1042 (2004).
266. D. Kaden *et al.*, Novel APP/A β mutation K16N produces highly toxic heteromeric A β oligomers. *EMBO Mol. Med.* **4**, 647-659 (2012).
267. L. Hendriks *et al.*, Presenile dementia and cerebral haemorrhage linked to a mutation at codon 692 of the β -amyloid precursor protein gene. *Nature Genet.* **1**, 218-221 (1992).
268. P. Cras *et al.*, Presenile Alzheimer dementia characterized by amyloid angiopathy and large amyloid core type senile plaques in the APP 692Ala-->Gly mutation. *Acta Neuropathol.* **96**, 253-260 (1998).
269. G. Roks *et al.*, Presentation of amyloidosis in carriers of the codon 692 mutation in the amyloid precursor protein gene (APP692). *Brain* **123**, 2130-2140 (2000).
270. C. Nilsberth *et al.*, The 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced A β protofibril formation. *Nature Neurosci.* **4**, 887-893 (2001).
271. S. Kumar-Singh *et al.*, Dense-core senile plaques in the Flemish variant of Alzheimer's disease are vasocentric. *Am. J. Pathol.* **161**, 507-520 (2002).
272. W. Brooks *et al.*, Hemorrhage is uncommon in new Alzheimer family with Flemish amyloid precursor protein mutation. *Neurology* **63**, 1613-1617 (2004).
273. A. R. Wattendorff, G. T. A. M. Bots, L. N. Went, L. J. Endtz, Familial cerebral amyloid angiopathy presenting as recurrent cerebral haemorrhage. *J. Neurol. Sci.* **55**, 121-135 (1982).
274. D. J. Watson, D. J. Selkoe, D. B. Teplow, Effects of the amyloid precursor protein Glu693-->Gln 'Dutch' mutation on the production and stability of amyloid β -protein. *Biochem. J.* **340**, 703-709 (1999).
275. O. Bugiani, G. Giaccone, G. Rossi, et al., Hereditary cerebral hemorrhage with amyloidosis associated with the E693K mutation of APP. *Arch. Neurol.* **67**, 987-995 (2010).
276. K. Kamino *et al.*, Linkage and mutational analysis of familial Alzheimer disease kindreds for the APP gene region. *Am. J. Hum. Genet.* **51**, 998-1014 (1992).
277. N. Norlin *et al.*, Aggregation and fibril morphology of the arctic mutation of Alzheimer's A β peptide by CD, TEM, STEM and in situ AFM. *J. Struct. Biol.* **180**, 174-189 (2012).
278. M. Schöll *et al.*, Low PiB PET retention in presence of pathologic CSF biomarkers in Arctic APP mutation carriers. *Neurology* **79**, 229-236 (2012).

279. H. Takuma, R. Teraoka, H. Mori, T. Tomiyama, Amyloid- β E22 Δ variant induces synaptic alteration in mouse hippocampal slices. *Neuroreport* **19**, 615-619 (2008).
280. Y. Kutoku *et al.*, A Second Pedigree with Amyloid-less Familial Alzheimer's Disease Harboring an Identical Mutation in the Amyloid Precursor Protein Gene (E693 Δ). *Intern. Med.* **54**, 205-208 (2015).
281. S. Greenberg *et al.*, Hemorrhagic stroke associated with the Iowa amyloid precursor protein mutation. *Neurology* **60**, 1020-1022 (2003).
282. W. Qiang, W.-M. Yau, R. Tycko, Structural evolution of Iowa mutant β -amyloid fibrils from polymorphic to homogeneous states under repeated seeded growth. *J. Am. Chem. Soc.* **133**, 4018-4029 (2011).
283. T. Mok *et al.*, Familial cerebral amyloid angiopathy due to the Iowa mutation in an Irish family. *Can. J. Neurol. Sci.* **41**, 512-517 (2014).
284. L. Obici *et al.*, A novel A β PP mutation exclusively associated with cerebral amyloid angiopathy. *Ann. Neurol.* **58**, 639-644 (2005).
285. G. Rossi *et al.*, A family with Alzheimer disease and strokes associated with A713T mutation of the APP gene. *Neurology* **63**, 910-912 (2004).
286. J. Armstrong, M. Boada, M. J. Rey, N. Vidal, I. Ferrer, Familial Alzheimer disease associated with A713T mutation in APP. *Neurosci. Lett.* **370**, 241-243 (2004).
287. L. Bernardi *et al.*, A β PP A713T mutation in late onset Alzheimer's disease with cerebrovascular lesions. *J. Alzheimers Dis.* **17**, 383-389 (2009).
288. R. Jakes, M. G. Spillantini, M. Goedert, Identification of two distinct synucleins from human brain. *FEBS Lett.* **345**, 27-32 (1994).
289. M. Goedert, α -Synuclein and neurodegenerative diseases. *Nature Rev. Neurosci.* **2**, 492-501 (2001).
290. M. G. Spillantini *et al.*, α -synuclein in Lewy bodies. *Nature* **388**, 839-840 (1997).
291. J. J. Zarranz *et al.*, The new mutation, E46K, of α -synuclein causes Parkinson and Lewy body dementia. *Ann. Neurol.* **55**, 164-173 (2004).
292. W. S. Davidson, A. Jonas, D. F. Clayton, J. M. George, Stabilization of α -synuclein secondary structure upon binding to synthetic membranes. *J. Biol. Chem.* **273**, 9443-9449 (1998).
293. T. S. Ulmer, A. Bax, N. B. Cole, R. L. Nussbaum, Structure and dynamics of micelle-bound human α -synuclein. *J. Biol. Chem.* **280**, 9595-9603 (2005).
294. M. Bisaglia *et al.*, A topological model of the interaction between α -synuclein and sodium dodecyl sulfate micelles. *Biochemistry* **44**, 329-339 (2005).
295. E. R. Georgieva, T. F. Ramlall, P. P. Borbat, J. H. Freed, D. Eliezer, Membrane-bound α -synuclein forms an extended helix: long-distance pulsed ESR Measurements using vesicles, bicelles, and rod-like micelles. *J. Am. Chem. Soc.* **130**, 12856-12857 (2008).
296. K. Vamvaca, P. T. Lansbury Jr, L. Stefanis, N-terminal deletion does not affect α -synuclein membrane binding, self-association and toxicity in human neuroblastoma cells, unlike yeast. *J. Neurochem.* **119**, 389-397 (2011).
297. M. J. Volles, P. T. Lansbury Jr, Relationships between the sequence of α -synuclein and its membrane affinity, fibrillization propensity, and yeast toxicity. *J. Mol. Biol.* **366**, 1510-1522 (2007).

298. J. M. McCord, I. Fridovich, Superoxide dismutase: an enzymic function for erythrocyte hemoglobin (hemocyanin). *J. Biol. Chem.* **244**, 6049-6055 (1969).
299. J. D. Crapo, T. Oury, C. Rabouille, J. W. Slot, L. Y. Chang, Copper, zinc superoxide dismutase is primarily a cytosolic protein in human cells. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10405-10409 (1992).
300. C. A. Pardo *et al.*, Superoxide dismutase is an abundant component in cell bodies, dendrites, and axons of motor neurons and in a subset of other neurons. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 954-958 (1995).
301. H. E. Parge, R. A. Hallewell, J. A. Tainer, Atomic structures of wild-type and thermostable mutant recombinant human Cu,Zn superoxide dismutase. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6109-6113 (1992).
302. N. Shibata *et al.*, Intense superoxide dismutase-1 immunoreactivity in intracytoplasmic hyaline inclusions of familial amyotrophic lateral sclerosis with posterior column involvement. *J. Neuropathol. Exp. Neurol.* **55**, 481-490 (1996).
303. S. Kato *et al.*, Familial amyotrophic lateral sclerosis with a two base pair deletion in superoxide dismutase 1 gene: multisystem degeneration with intracytoplasmic hyaline inclusions in astrocytes. *J. Neuropathol. Exp. Neurol.* **55**, 1089-1101 (1996).
304. Y. Kokubo *et al.*, Accumulation of neurofilaments and SOD1-immunoreactive products in a patient with familial amyotrophic lateral sclerosis with I113T SOD1 mutation. *Arch. Neurol.* **56**, 1506-1508 (1999).
305. Y. Takehisa *et al.*, Familial amyotrophic lateral sclerosis with a novel Leu126Ser mutation in the copper/zinc superoxide dismutase gene showing mild clinical features and lewy body-like hyaline inclusions. *Arch. Neurol.* **58**, 736-740 (2001).
306. T. Ohi, K. Nabeshima, S. Kato, S. Yazawa, S. Takechi, Familial amyotrophic lateral sclerosis with His46Arg mutation in Cu/Zn superoxide dismutase presenting characteristic clinical features and Lewy body-like hyaline inclusions. *J. Neurol. Sci.* **225**, 19-25 (2004).
307. Y. Furukawa, R. Fu, H. X. Deng, T. Siddique, T. V. O'Halloran, Disulfide cross-linked protein represents a significant fraction of ALS-associated Cu, Zn-superoxide dismutase aggregates in spinal cords of model mice. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7148-7153 (2006).
308. D. W. Cleveland, J. D. Rothstein, From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS. *Nature Rev. Neurosci.* **2**, 806-819 (2001).
309. M. DiDonato *et al.*, ALS mutants of human superoxide dismutase form fibrous aggregates via framework destabilization. *J. Mol. Biol.* **332**, 601-615 (2003).
310. P. A. Jonsson *et al.*, Minute quantities of misfolded mutant superoxide dismutase-1 cause amyotrophic lateral sclerosis. *Brain* **127**, 73-88 (2004).
311. D. R. Rosen *et al.*, Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362**, 59-62 (1993).
312. P. M. Andersen, Amyotrophic lateral sclerosis associated with mutations in the CuZn superoxide dismutase gene. *Curr. Neurol. Neurosci. Rep.* **6**, 37-46 (2006).
313. P. Pasinelli, R. H. Brown, Molecular biology of amyotrophic lateral sclerosis: insights from genetics. *Nature Rev. Neurosci.* **7**, 710-723 (2006).

314. R. Rakhit *et al.*, Oxidation-induced misfolding and aggregation of superoxide dismutase and its implications for amyotrophic lateral sclerosis. *J. Biol. Chem.* **277**, 47551-47556 (2002).
315. M. E. Gurney *et al.*, Motor neuron degeneration in mice that express a human Cu,Zn superoxide dismutase mutation. *Science* **264**, 1772-1775 (1994).
316. P. C. Wong *et al.*, An adverse property of a familial ALS-linked SOD1 mutation causes motor neuron disease characterized by vacuolar degeneration of mitochondria. *Neuron* **14**, 1105-1116 (1995).
317. L. I. Bruijn *et al.*, ALS-linked SOD1 mutant G85R mediates damage to astrocytes and promotes rapidly progressive disease with SOD1-containing inclusions. *Neuron* **18**, 327-338 (1997).
318. J. S. Valentine, P. J. Hart, Misfolded CuZnSOD and amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3617-3622 (2003).
319. P. M. Andersen *et al.*, Phenotypic heterogeneity in motor neuron disease patients with CuZn-superoxide dismutase mutations in Scandinavia. *Brain* **120**, 1723-1737 (1997).
320. A. Radunovic *et al.*, Copper and zinc levels in familial amyotrophic lateral sclerosis patients with CuZnSOD gene mutations. *Ann. Neurol.* **42**, 130-131 (1997).
321. K. Yamanaka, D. W. Cleveland, Determinants of rapid disease progression in ALS. *Neurology* **65**, 1859-1860 (2005).
322. A. G. Reaume *et al.*, Motor neurons in Cu/Zn superoxide dismutase-deficient mice develop normally but exhibit enhanced cell death after axonal injury. *Nature Genet.* **13**, 43-47 (1996).
323. F. L. Muller, M. S. Lustgarten, Y. Jang, A. Richardson, H. Van Remmen, Trends in oxidative aging theories. *Free. Radic. Biol. Med.* **43**, 477-503 (2007).
324. R. W. Strange *et al.*, Variable metallation of human superoxide dismutase: atomic resolution crystal structures of Cu-Zn, Zn-Zn and as-isolated wild-type enzymes. *J. Mol. Biol.* **356**, 1152-1162 (2006).
325. Y. Furukawa, A. S. Torres, T. V. O'Halloran, Oxygen-induced maturation of SOD1: a key role for disulfide formation by the copper chaperone CCS. *EMBO J.* **23**, 2872-2881 (2004).
326. C. Kayatekin, J. A. Zitzewitz, C. R. Matthews, Disulfide-reduced ALS variants of Cu, Zn superoxide dismutase exhibit increased populations of unfolded species. *J. Mol. Biol.* **398**, 320-331 (2010).
327. K. Toichi, K. Yamanaka, Y. Furukawa, Disulfide scrambling describes the oligomer formation of superoxide dismutase (SOD1) proteins in the familial form of amyotrophic lateral sclerosis. *J. Biol. Chem.* **288**, 4970-4980 (2013).
328. T. J. Lyons *et al.*, Mutations in copper-zinc superoxide dismutase that cause amyotrophic lateral sclerosis alter the zinc binding site and the redox behavior of the protein. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12240-12244 (1996).
329. J. Wang *et al.*, Copper-binding-site-null SOD1 causes ALS in transgenic mice: aggregates of non-native SOD1 delineate a common feature. *Hum. Mol. Genet.* **12**, 2753-2764 (2003).

330. K. S. Molnar *et al.*, A common property of amyotrophic lateral sclerosis-associated variants: destabilization of the copper/zinc superoxide dismutase electrostatic loop. *J. Biol. Chem.* **284**, 30965-30973 (2009).
331. H. J. Dyson, P. E. Wright, Intrinsically unstructured proteins and their functions. *Nature Rev. Mol. Cell Biol.* **6**, 197-208 (2005).
332. B. F. Shaw, J. S. Valentine, How do ALS-associated mutations in superoxide dismutase 1 promote aggregation of the protein? *Trends Biochem. Sci.* **32**, 78-85 (2007).
333. S. M. Lynch, W. Colón, Dominant role of copper in the kinetic stability of Cu/Zn superoxide dismutase. *Biochem. Biophys. Res. Commun.* **340**, 457-461 (2006).
334. F. Chiti, C. M. Dobson, Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **75**, 333-366 (2006).
335. Y. M. Kuo *et al.*, Comparative analysis of amyloid- β chemical structure and amyloid plaque morphology of transgenic mouse and Alzheimer's disease brains. *J. Biol. Chem.* **276**, 12991-12998 (2001).
336. K. Murakami *et al.*, Neurotoxicity and physicochemical properties of A β mutant peptides from cerebral amyloid angiopathy: implication for the pathogenesis of cerebral amyloid angiopathy and Alzheimer's disease. *J. Biol. Chem.* **278**, 46179-46187 (2003).
337. C. Wang, Solution NMR studies of A β monomer dynamics. *Protein Pept. Lett.* **18**, 354-361 (2011).
338. B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435-447 (2008).
339. B. C. Lee, V. N. Gladyshev, The biological significance of methionine sulfoxide stereochemistry. *Free Radic. Biol. Med.* **50**, 221-227 (2011).
340. G. Colombo, M. Meli, G. Morra, R. Gabizon, M. Gasset, Methionine sulfoxides on prion protein Helix-3 switch on the α -fold destabilization required for conversion. *PLoS ONE* **4**, e4296 (2009).
341. Y.-J. Zheng, R. L. Ornstein, A molecular dynamics and quantum mechanics analysis of the effect of DMSO on enzyme structure and dynamics: subtilisin. *J. Am. Chem. Soc.* **118**, 4175-4180 (1996).
342. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684 (1984).
343. B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463-1472 (1997).
344. S. Miyamoto, P. A. Kollman, Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952-962 (1992).
345. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N \cdot log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089-10089 (1993).
346. S. Nosé, Constant temperature molecular dynamics methods. *Prog. Theor. Phys. Supp.* **103**, 1-46 (1991).

347. W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637 (1983).
348. R. P. Joosten *et al.*, A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39**, D411-D419 (2010).
349. A. E. Torda, W. F. van Gunsteren, Algorithms for clustering molecular dynamics configurations. *J. Comput. Chem.* **15**, 1331-1340 (1994).
350. R. A. Jarvis, E. A. Patrick, Clustering using a similarity measure based on shared near neighbors. *IEEE Trans. Comput. C-22*, 1025-1034 (1973).
351. X. Daura *et al.*, Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed.* **38**, 236-240 (1999).
352. V. A. Voelz, M. S. Shell, K. A. Dill, Predicting peptide structures in native proteins from physical simulations of fragments. *PLoS Comput. Biol.* **5**, e1000281 (2009).
353. J. J. Prompers, R. Brüschweiler, 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 (2002).
354. F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, M. Scharf, The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* **16**, 273-284 (1995).
355. W. Hoyer, C. Gronwall, A. Jonsson, S. Stahl, T. Hard, Stabilization of a β -hairpin in monomeric Alzheimer's amyloid- β peptide inhibits amyloid formation. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5099-5104 (2008).
356. J. Nasica-Labouze *et al.*, Amyloid β protein and Alzheimer's disease: when computer simulations complement experimental studies. *Chem. Rev.* **115**, 3518-3563 (2015).
357. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037-10041 (2001).
358. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665-W667 (2004).
359. T. J. Dolinsky *et al.*, PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522-W525 (2007).
360. Schrodinger, LLC, The PyMOL molecular graphics system, version 1.3r1. (2010).
361. D. Frishman, P. Argos, Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Genet.* **23**, 566-579 (1995).
362. G. W. Vuister, A. Bax, Quantitative J correlation: a new approach for measuring homonuclear three-bond $J(\text{H}^{\text{N}}\text{H}^{\alpha})$ coupling constants in ^{15}N -enriched proteins. *J. Am. Chem. Soc.* **115**, 7772-7777 (1993).
363. A. Kolinski, J. Skolnick, Determinants of secondary structure of polypeptide chains: Interplay between short range and burial interactions. *J. Chem. Phys.* **107**, 953-953 (1997).

364. S. Nag *et al.*, Nature of the amyloid- β monomer and the monomer-oligomer equilibrium. *J. Biol. Chem.* **286**, 13827-13833 (2011).
365. G. Rose, A. Geselowitz, G. Lesser, R. Lee, M. Zehfus, Hydrophobicity of amino acid residues in globular proteins. *Science* **229**, 834-838 (1985).
366. J. Tian, A. E. García, Simulations of the confinement of ubiquitin in self-assembled reverse micelles. *J. Chem. Phys.* **134**, 225101-225101 (2011).
367. G. Lipari, A. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules: theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546-4559 (1982).
368. G. Habicht *et al.*, Directed selection of a conformational antibody domain that prevents mature amyloid fibril formation by stabilizing A β protofibrils. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19232-19237 (2007).
369. E. Cerf *et al.*, Antiparallel β -sheet: a signature structure of the oligomeric amyloid β -peptide. *Biochem. J.* **421**, 415-423 (2009).
370. C. Wu *et al.*, The structure of A β 42 C-terminal fragments probed by a combined experimental and theoretical study. *J. Mol. Biol.* **387**, 492-501 (2009).
371. E. Bramucci, A. Paiardini, F. Bossa, S. Pascarella, PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL. *BMC Bioinform.* **13**, S2 (2012).
372. A. Baumketner *et al.*, Structure of the 21-30 fragment of amyloid β -protein. *Protein Sci.* **15**, 1239-1247 (2006).
373. J. M. Borreguero *et al.*, Folding events in the 21-30 region of amyloid β -protein (A β) studied in silico. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6015-6020 (2005).
374. W. Chen, N. Mousseau, P. Derreumaux, The conformations of the amyloid- β (21-30) fragment can be described by three families in solution. *J. Chem. Phys.* **125**, 084911 (2006).
375. N. L. Fawzi *et al.*, Structure and dynamics of the A β 21-30 peptide from the interplay of NMR experiments and molecular simulations. *J. Am. Chem. Soc.* **130**, 6145-6158 (2008).
376. M. Yang, D. Teplow, Amyloid β -protein monomer folding: free-energy surfaces reveal alloform-specific differences. *J. Mol. Biol.* **384**, 450-464 (2008).
377. V. Hornak *et al.*, Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Genet.* **65**, 712-725 (2006).
378. E. Podoly, G. Hanin, H. Soreq, Alanine-to-threonine substitutions and amyloid diseases: Butyrylcholinesterase as a case study. *Chem. Biol. Interact.* **187**, 64-71 (2010).
379. J. Meinhardt *et al.*, Similarities in the thermodynamics and kinetics of aggregation of disease-related A β (1-40) peptides. *Protein Sci.* **16**, 1214-1222 (2007).
380. S. J. Weiner *et al.*, A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765-784 (1984).
381. W. L. Jorgensen, J. Tirado-Rives, The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657-1666 (1988).

382. W. D. Cornell *et al.*, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179-5197 (1995).
383. W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, 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 (1996).
384. Y. Duan *et al.*, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999-2012 (2003).
385. W. L. Jorgensen, J. Tirado-Rives, Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6665-6670 (2005).
386. C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269-10280 (1993).
387. P. Y. Chou, G. D. Fasman, Prediction of protein conformation. *Biochemistry* **13**, 222-245 (1974).
388. P. Prevelige, Jr., G. D. Fasman, in *Prediction of Protein Structure and the Principles of Protein Conformation*, G. D. Fasman, Ed. (Plenum Press, New York, NY, 1989), pp. 391-416.
389. P. Das, B. Murray, G. Belfort, Alzheimer's protective A2T mutation changes the conformational landscape of the A β 1-42 monomer differently than does the A2V mutation. *Biophys. J.* **108**, 738-747 (2015).
390. K. Wüthrich, *NMR of Proteins and Nucleic Acids*. (Wiley New York, New York, NY, 1986), vol. 155.
391. D. S. Wishart, B. D. Sykes, F. M. Richards, The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647-1651 (1992).
392. K. Wüthrich, Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.* **265**, 22059-22062 (1990).
393. J. H. Prestegard, C. M. Bougault, A. I. Kishore, Residual dipolar couplings in structure determination of biomolecules. *Chem. Rev.* **104**, 3519-3540 (2004).
394. G. S. Rule, T. K. Hitchens, *Fundamentals of Protein NMR Spectroscopy*. (Springer, Dordrecht, The Netherlands, 2006), vol. 5.
395. R. Ishima, D. A. Torchia, Protein dynamics from NMR. *Nature Struct. Biol.* **7**, 740-743 (2000).
396. A. G. Palmer, III, C. D. Kroenke, J. P. Loria, in *Methods in Enzymology*, V. D. Thomas L. James, S. Uli, Eds. (Academic Press, San Diego, CA, 2001), vol. 339, pp. 204-238.
397. N. L. Fawzi, J. Ying, R. Ghirlando, D. A. Torchia, G. M. Clore, Atomic-resolution dynamics on the surface of amyloid- β protofibrils probed by solution NMR. *Nature* **480**, 268-272 (2011).
398. L. Hou *et al.*, Solution NMR studies of the A β (1-40) and A β (1-42) peptides establish that the Met35 oxidation state affects the mechanism of amyloid formation. *J. Am. Chem. Soc.* **126**, 1992-2005 (2004).

399. H. J. Dyson, P. E. Wright, Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607-3622 (2004).
400. M. R. Jensen, R. W. H. Ruigrok, M. Blackledge, Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr. Opin. Struct. Biol.* **23**, 426-435 (2013).
401. D. Sahu, M. Bastidas, S. A. Showalter, Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods. *Anal. Biochem.* **449**, 17-25 (2014).
402. R. Konrat, NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson.* **241**, 74-85 (2014).
403. K. Wuthrich, NMR studies of structure and function of biological macromolecules (Nobel Lecture). *J. Biomol. NMR* **27**, 13-39 (2003).
404. H. Iwai, S. Zuger, Protein ligation: applications in NMR studies of proteins. *Biotechnol. Genet. Eng. Rev.* **24**, 129-145 (2007).
405. M. D. Mukrasch *et al.*, Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol.* **7**, e34 (2009).
406. V. H. FINDER, I. Vodopivec, R. M. Nitsch, R. Glockshuber, The recombinant amyloid- β peptide A β 1-42 aggregates faster and is more neurotoxic than synthetic A β 1-42. *J. Mol. Biol.* **396**, 9-18 (2010).
407. Y. Yan, C. Wang, A β 42 is more rigid than A β 40 at the C terminus: implications for A β aggregation and toxicity. *J. Mol. Biol.* **364**, 853-862 (2006).
408. Y. Yan, J. Liu, S. A. McCallum, D. Yang, C. Wang, Methyl dynamics of the amyloid- β peptides A β 40 and A β 42. *Biochem. Biophys. Res. Commun.* **362**, 410-414 (2007).
409. Y. Yan, S. A. McCallum, C. Wang, M35 oxidation induces A β 40-like structural and dynamical changes in A β 42. *J. Am. Chem. Soc.* **130**, 5394-5395 (2008).
410. C. E. Munte, M. B. Erlach, W. Kremer, J. Koehler, H. R. Kalbitzer, Distinct conformational states of the Alzheimer β -amyloid peptide can be detected by high-pressure NMR spectroscopy. *Angew. Chem. Int. Ed.* **52**, 8943-8947 (2013).
411. C. Lendel, P. Damberg, 3D J-resolved NMR spectroscopy for unstructured polypeptides: fast measurement of $^3J_{\text{HNH}\alpha}$ coupling constants with outstanding spectral resolution. *J. Biomol. NMR* **44**, 35-42 (2009).
412. B. J. Bennion, V. Daggett, The molecular basis for the chemical denaturation of proteins by urea. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5142-5147 (2003).
413. D. R. Canchi, D. Paschek, A. E. Garcia, Equilibrium study of protein denaturation by urea. *J. Am. Chem. Soc.* **132**, 2338-2344 (2010).
414. D. R. Canchi, A. E. Garcia, Backbone and side-chain contributions in protein denaturation by urea. *Biophys. J.* **100**, 1526-1533 (2011).
415. C. A. Royer, Revisiting volume changes in pressure-induced protein unfolding. *Biochim. Biophys. Acta* **1595**, 201-209 (2002).
416. Y. O. Kamatari, S. Yokoyama, H. Tachibana, K. Akasaka, Pressure-jump NMR study of dissociation and association of amyloid protofibrils. *J. Mol. Biol.* **349**, 916-921 (2005).
417. Y. S. Kim, T. W. Randolph, M. B. Seefeldt, J. F. Carpenter, in *Methods in Enzymology*, K. Indu, W. Ronald, Eds. (Academic Press, San Diego, CA, 2006), vol. 413, pp. 237-253.

418. F. Meersman, C. M. Dobson, Probing the pressure–temperature stability of amyloid fibrils provides new insights into their molecular properties. *Biochim. Biophys. Acta* **1764**, 452-460 (2006).
419. F. Meersman, C. M. Dobson, K. Heremans, Protein unfolding, amyloid fibril formation and configurational energy landscapes under high pressure conditions. *Chem. Soc. Rev.* **35**, 908-917 (2006).
420. J. Roche, J. Ying, A. S. Maltsev, A. Bax, Impact of hydrostatic pressure on an intrinsically disordered protein: a high-pressure NMR study of α -synuclein. *ChemBioChem* **14**, 1754-1761 (2013).
421. A. A. Watson, D. P. Fairlie, D. J. Craik, Solution structure of methionine-oxidized amyloid β -peptide (1-40). Does oxidation affect conformational switching? *Biochemistry* **37**, 12700-12706 (1998).
422. J. Danielsson, J. Jarvet, P. Damberg, A. Graslund, Translational diffusion measured by PFG-NMR on full length and fragments of the Alzheimer A β (1-40) peptide. Determination of hydrodynamic radii of random coil peptides of varying length. *Magn. Reson. Chem.* **40**, S89-S97 (2002).
423. F. Delaglio *et al.*, NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293 (1995).
424. L. Hou, Ph.D. Thesis, Case Western Reserve University (2003).
425. P. Schanda, H. Van Melckebeke, B. Brutscher, Speeding up three-dimensional protein NMR experiments to a few minutes. *J. Am. Chem. Soc.* **128**, 9042-9043 (2006).
426. E. Lescop, P. Schanda, B. Brutscher, A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.* **187**, 163-169 (2007).
427. T. Luan, V. Jaravine, A. Yee, C. H. Arrowsmith, V. Y. Orekhov, Optimization of resolution and sensitivity of 4D NOESY using multi-dimensional decomposition. *J. Biomol. NMR* **33**, 1-14 (2005).
428. V. Y. Orekhov, V. A. Jaravine, Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. *Prog. Nucl. Magn. Reson. Spectrosc.* **59**, 271-292 (2011).
429. G. T. Montelione, B. A. Lyons, S. D. Emerson, M. Tashiro, An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J. Am. Chem. Soc.* **114**, 10974-10975 (1992).
430. S. Grzesiek, J. Anglister, A. Bax, Correlation of backbone amide and aliphatic side-chain resonances in $^{13}\text{C}/^{15}\text{N}$ -enriched proteins by isotropic mixing of ^{13}C magnetization. *J. Magn. Reson.* **101**, 114-119 (1993).
431. M. Ottiger, F. Delaglio, A. Bax, Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373-378 (1998).
432. S. J. Archer, M. Ikura, D. A. Torchia, A. Bax, An alternative 3D NMR technique for correlating backbone ^{15}N with side chain $\text{H}\beta$ resonances in larger proteins. *J. Magn. Reson.* **95**, 636-641 (1991).
433. Y. Shen, A. Bax, SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* **48**, 13-22 (2010).

434. B. Han, Y. Liu, S. W. Ginzinger, D. S. Wishart, SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50**, 43-57 (2011).
435. M. Karplus, D. H. Anderson, Valence-bond interpretation of electron-coupled nuclear spin interactions; application to methane. *J. Chem. Phys.* **30**, 6-6 (1959).
436. C. Pérez, F. Löhr, H. Rüterjans, J. M. Schmidt, Self-consistent Karplus parametrization of 3J couplings depending on the polypeptide side-chain torsion χ_1 . *J. Am. Chem. Soc.* **123**, 7081-7093 (2001).
437. M. Zweckstetter, A. Bax, 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 (2000).
438. M. Zweckstetter, NMR: prediction of molecular alignment from structure using the PALES software. *Nature Protoc.* **3**, 679-690 (2008).
439. Y. Kitamura, T. Itoh, Reaction volume of protonic ionization for buffering agents. Prediction of pressure dependence of pH and pOH. *J. Solution Chem.* **16**, 715-725 (1987).
440. R. J. Quinlan, G. D. Reinhart, Baroresistant buffer mixtures for biochemical analyses. *Anal. Biochem.* **341**, 69-76 (2005).
441. M. W. Lassalle, K. Akasaka, in *Protein Folding Protocols*, Y. Bai, R. Nussinov, Eds. (Humana Press, Totowa, NJ, 2006), vol. 350, pp. 21-38.
442. D. S. Raiford, C. L. Fisk, E. D. Becker, Calibration of methanol and ethylene glycol nuclear magnetic resonance thermometers. *Anal. Chem.* **51**, 2050-2051 (1979).
443. D. V. Laurents *et al.*, Alzheimer's A β 40 studied by NMR at low pH reveals that sodium 4,4-dimethyl-4-silapentane-1-sulfonate (DSS) binds and promotes β -ball oligomerization. *J. Biol. Chem.* **280**, 3675-3685 (2005).
444. D. S. Wishart *et al.*, H-1, C-13 and N-15 chemical-shift referencing in biomolecular NMR. *J. Biomol. NMR* **6**, 135-140 (1995).
445. J. Koehler *et al.*, Pressure dependence of 15N chemical shifts in model peptides Ac-Gly-Gly-X-Ala-NH₂. *Materials* **5**, 1774-1786 (2012).
446. M. R. Arnold, W. Kremer, H.-D. Lüdemann, H. R. Kalbitzer, 1H-NMR parameters of common amino acid residues measured in aqueous solutions of the linear tetrapeptides Gly-Gly-X-Ala at pressures between 0.1 and 200 MPa. *Biophys. Chem.* **96**, 129-140 (2002).
447. R. C. Neuman, W. Kauzmann, A. Zipp, Pressure dependence of weak acid ionization in aqueous buffers. *J. Phys. Chem.* **77**, 2687-2691 (1973).
448. T. K. Hitchens, R. G. Bryant, Pressure dependence of weak acid ionization in deuterium oxide solutions. *J. Phys. Chem. B* **102**, 1002-1004 (1998).
449. R. E. Iacob, J. R. Engen, Hydrogen exchange mass spectrometry: are we out of the quicksand? *J. Am. Soc. Mass Spectrom.* **23**, 1003-1010 (2012).
450. O. K. Baryshnikova, T. C. Williams, B. D. Sykes, Internal pH indicators for biomolecular NMR. *J. Biomol. NMR* **41**, 5-7 (2008).
451. Y.-Z. Zhang, Ph.D. Thesis, University of Pennsylvania (1995).
452. Y. Bai, J. S. Milne, L. Mayne, S. W. Englander, Primary structure effects on peptide group hydrogen exchange. *Proteins Struct. Funct. Genet.* **17**, 75-86 (1993).

453. A. E. Conicella, N. L. Fawzi, The C-terminal threonine of A β 43 nucleates toxic aggregation via structural and dynamical changes in monomers and protofibrils. *Biochemistry* **53**, 3095-3105 (2014).
454. N. H. Pawley, C. Wang, S. Koide, L. K. Nicholson, An improved method for distinguishing between anisotropic tumbling and chemical exchange in analysis of ¹⁵N relaxation parameters. *J. Biomol. NMR* **20**, 149-165 (2001).
455. S. P. Mielke, V. V. Krishnan, Characterization of protein secondary structure from NMR chemical shifts. *Prog. Nucl. Magn. Reson. Spectrosc.* **54**, 141-165 (2009).
456. S. Neal, A. M. Nip, H. Zhang, D. S. Wishart, Rapid and accurate calculation of protein ¹H, ¹³C and ¹⁵N chemical shifts. *J. Biomol. NMR* **26**, 215-240 (2003).
457. R. P. Barnwal, A. K. Rout, K. R. Chary, H. S. Atreya, Rapid measurement of ³J(HN–H α) and ³J(N–H β) coupling constants in polypeptides. *J. Biomol. NMR* **39**, 259-263 (2007).
458. D. R. Canchi, A. E. Garcia, Cosolvent effects on protein stability. *Annu. Rev. Phys. Chem.* **64**, 273-293 (2013).
459. K. Inoue *et al.*, Pressure-induced local unfolding of the Ras binding domain of RalGDS. *Nature Struct. Biol.* **7**, 547-550 (2000).
460. M. W. Lassalle, H. Yamada, K. Akasaka, The pressure-temperature free energy-landscape of staphylococcal nuclease monitored by ¹H NMR1. *J. Mol. Biol.* **298**, 293-302 (2000).
461. K. Kuwata *et al.*, High pressure NMR reveals a variety of fluctuating conformers in β -lactoglobulin1. *J. Mol. Biol.* **305**, 1073-1083 (2001).
462. R. Kitahara, S. Yokoyama, K. Akasaka, NMR snapshots of a fluctuating protein structure: ubiquitin at 30 bar–3 kbar. *J. Mol. Biol.* **347**, 277-285 (2005).
463. R. Kitahara, K. Hata, H. Li, M. P. Williamson, K. Akasaka, Pressure-induced chemical shifts as probes for conformational fluctuations in proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **71**, 35-58 (2013).
464. R. Kitahara *et al.*, High pressure NMR reveals active-site hinge motion of folate-bound *Escherichia coli* dihydrofolate reductase. *Biochemistry* **39**, 12789-12795 (2000).
465. T. Asakura, K. Taoka, M. Demura, M. P. Williamson, The relationship between amide proton chemical shifts and secondary structure in proteins. *J. Biomol. NMR* **6**, 227-236 (1995).
466. T. K. Hitchens, R. G. Bryant, Pressure dependence of amide hydrogen–deuterium exchange rates for individual sites in T4 lysozyme. *Biochemistry* **37**, 5878-5887 (1998).
467. E. J. Fuentes, A. J. Wand, Local stability and dynamics of apocytochrome b562 examined by the dependence of hydrogen exchange on hydrostatic pressure. *Biochemistry* **37**, 9877-9883 (1998).
468. J. Roche *et al.*, Remodeling of the folding free energy landscape of Staphylococcal nuclease by cavity-creating mutations. *Biochemistry* **51**, 9535-9546 (2012).
469. H. G. Gaston, M. A. Cohen, S. A. Benner, Exhaustive matching of the entire protein sequence database. *Science* **256**, 1443-1445 (1992).

470. J. D. Thompson, T. Gibson, D. G. Higgins, Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics.*, 2.3. 1-2.3. 22 (2002).
471. N. Cohen, D. Rosenman, O. Bilsel, A. Garcia, C. R. Matthews, Collapsed structure at the C-terminus of SOD1: implications for ALS. *FASEB J.* **29**, (2015).
472. N.-V. Buchete, G. Hummer, Peptide folding kinetics from replica exchange molecular dynamics. *Phys. Rev. E* **77**, 030902 (2008).
473. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182-7190 (1981).
474. A. E. Garcia, H. Hecce, D. Paschek, Simulations of temperature and pressure unfolding of peptides and proteins with replica exchange molecular dynamics. *Annu. Rep. Comput. Chem.* **2**, 83-95 (2006).
475. C. A. English, A. E. Garcia, Folding and unfolding thermodynamics of the TC10b Trp-cage miniprotein. *Phys. Chem. Chem. Phys.* **16**, 2748-2757 (2014).
476. O. Bilsel, C. Kayatekin, L. A. Wallace, C. R. Matthews, A microchannel solution mixer for studying microsecond protein folding reactions. *Rev. Sci. Instrum.* **76**, 014302 (2005).
477. Y. Wu, E. Kondrashkina, C. Kayatekin, C. R. Matthews, O. Bilsel, Microsecond acquisition of heterogeneous structure in the folding of a TIM barrel protein. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13367-13372 (2008).
478. J. Skilling, R. K. Bryan, Maximum entropy image reconstruction: general algorithm. *Mon. Not. R. Astron. Soc.* **211**, 111-124 (1984).
479. E. Haas, M. Wilchek, E. Katchalski-Katzir, I. Z. Steinberg, Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1807-1811 (1975).
480. J. H. Werner, R. Joggerst, R. B. Dyer, P. M. Goodwin, A two-dimensional view of the folding energy landscape of cytochrome c. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11130-11135 (2006).
481. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33-38, 27-38 (1996).
482. A. R. Fersht, Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* **7**, 3-9 (1997).
483. J. R. Lakowicz, in *Principles of Fluorescence Spectroscopy*. (Springer US, New York, NY, 1999), pp. 367-394.
484. C. Berney, G. Danuser, FRET or no FRET: a quantitative comparison. *Biophys. J.* **84**, 3992-4010 (2003).
485. P. V. Nazarov, R. B. M. Koehorst, W. L. Vos, V. V. Apanasovich, M. A. Hemminga, FRET Study of Membrane Proteins: Simulation-Based Fitting for Analysis of Membrane Protein Embedment and Association. *Biophys. J.* **91**, 454-466 (2006).
486. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*. (Springer Science & Business Media, New York, NY, ed. 3, 2007).
487. N. C. Fitzkee, G. D. Rose, Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12497-12502 (2004).

488. F. Palazzesi, M. K. Prakash, M. Bonomi, A. Barducci, Accuracy of current all-atom force-fields in modeling protein disordered states. *J. Chem. Theory Comput.* **11**, 2-7 (2015).
489. R. B. Best, W. Zheng, J. Mittal, Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* **10**, 5113-5124 (2014).
490. R. B. Best, J. Mittal, Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* **114**, 14916-14923 (2010).
491. W. Wang, W. Ye, C. Jiang, R. Luo, H. F. Chen, New force field on modeling intrinsically disordered proteins. *Chem. Biol. Drug Des.* **84**, 253-269 (2014).
492. P. S. Nerenberg, B. Jo, C. So, A. Tripathy, T. Head-Gordon, Optimizing Solute–Water van der Waals Interactions To Reproduce Solvation Free Energies. *J. Phys. Chem. B* **116**, 4524-4534 (2012).
493. J. L. F. Abascal, C. Vega, A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **123**, 234505 (2005).
494. D. R. Canchi, P. Jayasimha, D. C. Rau, G. I. Makhataдзе, A. E. Garcia, Molecular mechanism for the preferential exclusion of TMAO from protein surfaces. *J. Phys. Chem. B* **116**, 12095-12104 (2012).
495. K. L. Constantine *et al.*, Structural and dynamic properties of a β -hairpin-forming linear peptide. *J. Am. Chem. Soc.* **117**, 10841-10854 (1995).
496. A. Torda, R. Brunne, T. Huber, H. Kessler, W. van Gunsteren, Structure refinement using time-averaged J-coupling constant restraints. *J. Biomol. NMR* **3**, 55-66 (1993).
497. M. Vendruscolo, E. Paci, C. M. Dobson, M. Karplus, Rare fluctuations of native proteins sampled by equilibrium hydrogen exchange. *J. Am. Chem. Soc.* **125**, 15686-15687 (2003).
498. K. Lindorff-Larsen, R. B. Best, M. A. DePristo, C. M. Dobson, M. Vendruscolo, Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128-132 (2005).
499. J. Fennen, A. E. Torda, W. F. van Gunsteren, Structure refinement with molecular dynamics and a Boltzmann-weighted ensemble. *J. Biomol. NMR* **6**, 163-170 (1995).
500. A. Cavalli, C. Camilloni, M. Vendruscolo, Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys.* **138**, 094112 (2013).
501. G. Otting, Protein NMR using paramagnetic ions. *Annu. Rev. Biophys.* **39**, 387-405 (2010).
502. W. Bermel, I. Bertini, I. C. Felli, R. Kümmerle, R. Pierattelli, ¹³C direct detection experiments on the paramagnetic oxidized monomeric copper, zinc superoxide dismutase. *J. Am. Chem. Soc.* **125**, 16423-16429 (2003).
503. S. A. Showalter, Intrinsically disordered proteins: methods for structure and dynamics studies. *eMagRes* **3**, 181-190 (2014).
504. M. Bastidas, E. B. Gibbs, D. Sahu, S. A. Showalter, A primer for carbon-detected NMR applications to intrinsically disordered proteins in solution. *Concepts Magn. Reson. A*, (2015).

505. E. B. Gibbs, S. A. Showalter, Quantitative biophysical characterization of intrinsically disordered proteins. *Biochemistry* **54**, 1314-1326 (2015).
506. D. Marion *et al.*, Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 β . *Biochemistry* **28**, 6150-6156 (1989).
507. A. Bax, G. M. Clore, A. M. Gronenborn, 1H-1H correlation via isotropic mixing of 13C magnetization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins. *J. Magn. Reson.* **88**, 425-431 (1990).
508. E. T. Olejniczak, R. X. Xu, S. W. Fesik, A 4D HCCH-TOCSY experiment for assigning the side chain 1H and 13C resonances of proteins. *J. Biomol. NMR* **2**, 655-659 (1992).
509. J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, N. J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*. (Academic Press, San Diego, CA, 1995).
510. D. Marion, L. E. Kay, S. W. Sparks, D. A. Torchia, A. Bax, Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J. Am. Chem. Soc.* **111**, 1515-1517 (1989).
511. K. A. Ball, Ph.D. Thesis, University of California, Berkeley (2013).
512. E. R. P. Zuiderweg, S. W. Fesik, Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* **28**, 2387-2391 (1989).
513. C. Grönwall *et al.*, Selection and characterization of Affibody ligands binding to Alzheimer amyloid β peptides. *J. Biotechnol.* **128**, 162-183 (2007).

7. Appendix

Table 7.1. Force field parameters for L-methioinine-(S)-sulfoxide used in the OPLS simulation of A β 42-M35ox(S) in Chapter 2. The standard OPLS parameters for methionine were used as a template; this table reviews all of the new or changed definitions made to these methionine parameters. The parameters for Ryckaert-Bellemans dihedrals were set to 0.00000, following the example of Colombo et al.(340) The improper definition was included to strictly enforce S chirality.

Atom Definitions				
<i>Name</i>	<i>Type</i>	<i>Charge</i>	<i>Group</i>	<i>Notes</i>
CG	opls_210	0.040	3	Replaces MET CG: opls_210, 0.048
SD	opls_124	0.139	4	DMSO parameters(341)
OD	opls_125	-0.459	4	DMSO parameters(341)
CE	opls_139	-0.020	5	DMSO parameters(341)
Bond Definitions				
<i>Atom 1</i>	<i>Atom 2</i>			
OD	SD			
Dihedral Definitions				
<i>Atom 1</i>	<i>Atom 2</i>	<i>Atom 3</i>	<i>Atom 4</i>	<i>Value</i>
CA	CB	CG	SD	0.00000 0.00000 0.00000 0.00000 0.00000
CB	CG	SD	OD	0.00000 0.00000 0.00000 0.00000 0.00000
CB	CG	SD	CE	0.00000 0.00000 0.00000 0.00000 0.00000
Improper Definitions				
<i>Atom 1</i>	<i>Atom 2</i>	<i>Atom 3</i>	<i>Atom 4</i>	<i>Value</i>
CG	CE	SD	OD	120.0 43.93200 2

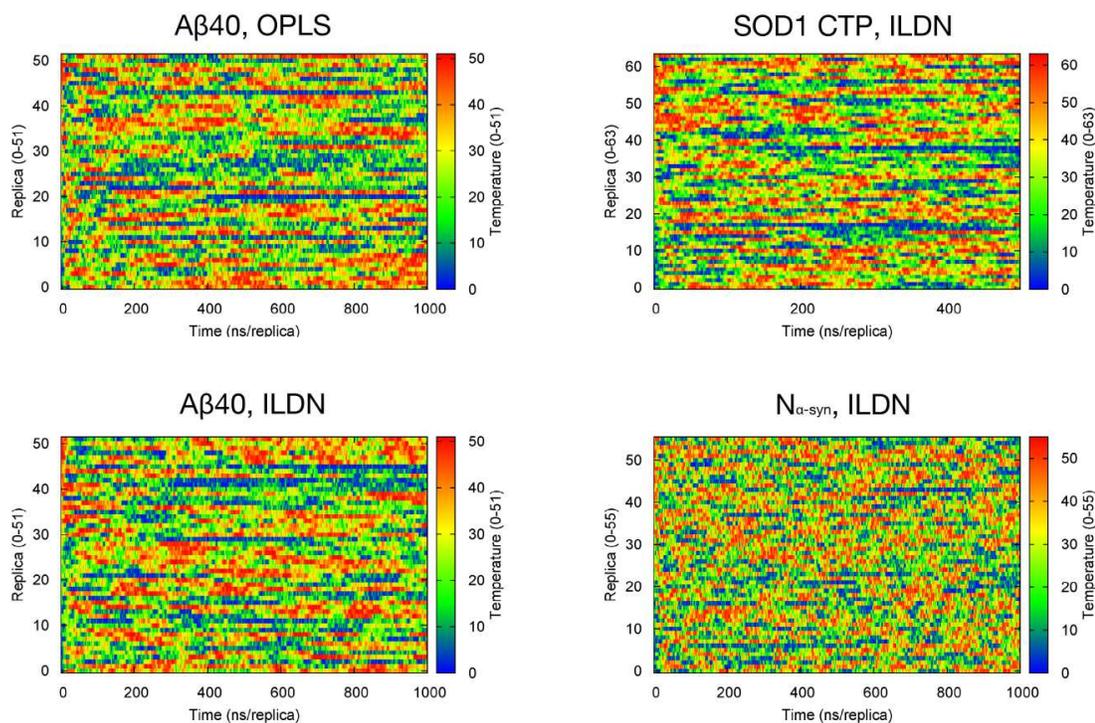


Figure 7.1. Movement of replicas in both temperature and replica space as a function of simulation time for select replica exchange molecular dynamics simulations described in this thesis. “A β 40, OPLS” is described in Chapter 2, “A β 40, ILDN” in Chapter 3, and both “SOD1 CTP, ILDN” and “N α -syn, ILDN” are both described in Chapter 5. While replicas occupy the lowest temperatures for long period of time, particularly for A β 40 simulations and SOD1 CTP, shuffling in and out of the non-boundary production temperatures is frequent. These plots suggest that the random walks are rapid and sufficient enough to properly enhance sampling of our simulations.

Table 7.2. Resonance table for A β 40-D23N and A β 40-A2T based on the assignment experiments performed for each species, as described in Chapter 4.

A β 40-D23N					A β 40-A2T				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift	Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
1	1	D	CA	53.605	1	1	D	CA	53.718
2	1	D	HA	4.189	2	1	D	CB	40.725
3	2	A	C'	177.351	3	1	D	HA	4.269
4	2	A	CA	52.529	4	1	D	HB1	2.703
5	2	A	CB	19.11	5	1	D	HB2	2.715
6	2	A	HA	4.302	6	2	T	C'	174.121
7	3	E	C'	176.053	7	2	T	CA	62.12
8	3	E	CA	56.558	8	2	T	CB	69.689
9	3	E	CB	30.184	9	2	T	CG2	21.631
10	3	E	H	8.503	10	2	T	HA	4.319
11	3	E	HA	4.195	11	2	T	HB	4.182
12	3	E	N	120.585	12	2	T	HG2	1.181
13	4	F	C'	175.262	13	3	E	C'	175.896
14	4	F	CA	57.702	14	3	E	CA	56.454
15	4	F	CB	39.481	15	3	E	CB	30.342
16	4	F	H	8.371	16	3	E	CG	36.072
17	4	F	HA	4.56	17	3	E	H	8.585
18	4	F	N	122.016	18	3	E	HA	4.242
19	5	R	CA	55.507	19	3	E	HB1	1.905
20	5	R	CB	31.059	20	3	E	HB2	1.938
21	5	R	H	8.192	21	3	E	HG1	2.099
22	5	R	HA	4.263	22	3	E	HG2	2.208
23	5	R	HE	7.348	23	3	E	N	124.06
24	5	R	N	123.914	24	4	F	C'	175.256
25	5	R	NE	144.608	25	4	F	CA	57.805
26	6	H	CA	56.358	26	4	F	CB	39.471
27	6	H	HA	4.535	27	4	F	H	8.431
28	7	D	C'	176.312	28	4	F	HA	4.547
29	7	D	CA	54.016	29	4	F	HB	3.002
30	7	D	H	8.409	30	4	F	N	122.377
31	7	D	HA	4.63	31	5	R	C'	174.959
32	7	D	N	121.742	32	5	R	CA	55.468
33	8	S	C'	175.179	33	5	R	CB	31.137
34	8	S	CA	59.115	34	5	R	CD	43.14
35	8	S	CB	63.706	35	5	R	CG	26.905
36	8	S	H	8.479	36	5	R	H	8.211
37	8	S	HA	4.38	37	5	R	HA	4.274
38	8	S	HB1	3.924	38	5	R	HB1	1.633
39	8	S	HB2	3.876	39	5	R	HB2	1.741
40	8	S	N	116.771	40	5	R	HD	3.131
41	9	G	C'	173.951	41	5	R	HG	1.5
42	9	G	CA	45.306	42	5	R	N	123.951
43	9	G	H	8.588	43	6	H	CA	56.375

Aβ40-D23N				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
44	9	G	HA1	3.891
45	9	G	HA2	3.871
46	9	G	N	110.819
47	10	Y	C'	175.822
48	10	Y	CA	58.263
49	10	Y	CB	38.876
50	10	Y	H	7.994
51	10	Y	HA	4.516
52	10	Y	N	120.205
53	11	E	C'	176.253
54	11	E	CA	56.611
55	11	E	CB	30.251
56	11	E	H	8.434
57	11	E	HA	4.195
58	11	E	N	122.694
59	12	V	CA	62.901
60	12	V	CB	32.444
61	12	V	H	8.128
62	12	V	HA	3.936
63	12	V	N	121.261
64	13	H	CA	56.165
65	13	H	HA	4.601
66	14	H	C'	175.168
67	14	H	CA	56.336
68	14	H	HA	4.555
69	15	Q	C'	175.821
70	15	Q	CA	55.974
71	15	Q	H	8.46
72	15	Q	HA	4.26
73	15	Q	HE1	7.616
74	15	Q	HE2	6.935
75	15	Q	N	121.456
76	15	Q	NE	113.014
77	16	K	C'	176.287
78	16	K	CA	56.398
79	16	K	CB	32.909
80	16	K	H	8.437
81	16	K	HA	4.279
82	16	K	N	123.185
83	17	L	C'	176.927
84	17	L	CA	55.175
85	17	L	CB	42.219
86	17	L	H	8.292
87	17	L	HA	4.336
88	17	L	N	123.868

Aβ40-A2T				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
44	6	H	CB	30.602
45	6	H	HA	4.533
46	6	H	HB1	3.055
47	6	H	HB2	3.115
48	7	D	C'	176.281
49	7	D	CA	54.023
50	7	D	CB	41.04
51	7	D	H	8.358
52	7	D	HA	4.632
53	7	D	HB	2.681
54	7	D	N	122.545
55	8	S	C'	175.149
56	8	S	CA	59.1
57	8	S	CB	63.686
58	8	S	H	8.487
59	8	S	HA	4.381
60	8	S	HB	3.909
61	8	S	HB1	3.897
62	8	S	HB2	3.903
63	8	S	N	116.732
64	9	G	C'	173.942
65	9	G	CA	45.279
66	9	G	H	8.594
67	9	G	HA	3.915
68	9	G	HA1	3.89
69	9	G	HA2	3.872
70	9	G	N	110.82
71	10	Y	C'	175.774
72	10	Y	CA	58.143
73	10	Y	CB	38.854
74	10	Y	H	8
75	10	Y	HA	4.518
76	10	Y	HB1	2.954
77	10	Y	HB2	3.038
78	10	Y	N	120.176
79	11	E	C'	176.13
80	11	E	CA	56.523
81	11	E	CB	30.305
82	11	E	CG	36.185
83	11	E	H	8.436
84	11	E	HA	4.202
85	11	E	HB1	1.853
86	11	E	HB2	1.941
87	11	E	HG	2.2
88	11	E	HG1	2.17

Aβ40-D23N				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
89	18	V	C'	175.227
90	18	V	CA	62.103
91	18	V	CB	32.924
92	18	V	H	8.071
93	18	V	HA	4.026
94	18	V	N	121.703
95	19	F	C'	174.922
96	19	F	CA	57.566
97	19	F	CB	40.084
98	19	F	H	8.316
99	19	F	HA	4.579
100	19	F	N	124.451
101	20	F	C'	174.819
102	20	F	CA	57.462
103	20	F	CB	39.796
104	20	F	H	8.244
105	20	F	HA	4.55
106	20	F	N	123.03
107	21	A	C'	177.255
108	21	A	CA	52.329
109	21	A	CB	19.4
110	21	A	H	8.251
111	21	A	HA	4.218
112	21	A	N	126.074
113	22	E	C'	176.216
114	22	E	CA	56.618
115	22	E	CB	30.367
116	22	E	H	8.414
117	22	E	HA	4.196
118	22	E	N	120.054
119	23	N	C'	175.291
120	23	N	CA	53.227
121	23	N	CB	38.676
122	23	N	H	8.549
123	23	N	HA	4.736
124	23	N	HD1	7.658
125	23	N	HD2	6.938
126	23	N	N	119.821
127	23	N	ND	113.255
128	24	V	C'	176.845
129	24	V	CA	62.744
130	24	V	CB	32.418
131	24	V	H	8.196
132	24	V	HA	4.131
133	24	V	N	120.576

Aβ40-A2T				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
89	11	E	HG2	2.21
90	11	E	N	122.625
91	12	V	C'	176.293
92	12	V	CA	62.762
93	12	V	CB	32.521
94	12	V	CG1	20.829
95	12	V	CG2	20.754
96	12	V	H	8.146
97	12	V	HA	3.947
98	12	V	HB	1.953
99	12	V	HG1	0.794
100	12	V	HG2	0.873
101	12	V	N	121.329
102	13	H	CA	56.097
103	13	H	CB	30.602
104	13	H	HA	4.608
105	13	H	HB1	3.024
106	14	H	C'	175.155
107	14	H	CA	56.276
108	14	H	CB	30.58
109	14	H	HA	4.55
110	14	H	HB1	3.011
111	15	Q	C'	175.683
112	15	Q	CA	55.889
113	15	Q	CB	29.366
114	15	Q	CG	33.662
115	15	Q	H	8.476
116	15	Q	HA	4.268
117	15	Q	HB1	1.98
118	15	Q	HB2	2.067
119	15	Q	HE1	7.635
120	15	Q	HE2	6.95
121	15	Q	HG	2.341
122	15	Q	N	121.674
123	15	Q	NE	113.035
124	16	K	C'	176.211
125	16	K	CA	56.377
126	16	K	CB	32.919
127	16	K	CD	29.054
128	16	K	CE	42.023
129	16	K	CG	24.832
130	16	K	H	8.468
131	16	K	HA	4.278
132	16	K	HB	1.786
133	16	K	HB1	1.757

A β 40-D23N				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
134	25	G	C'	174.305
135	25	G	CA	45.301
136	25	G	H	8.596
137	25	G	HA	3.99
138	25	G	N	112.358
139	26	S	C'	174.468
140	26	S	CA	58.413
141	26	S	CB	63.728
142	26	S	H	8.263
143	26	S	HA	4.446
144	26	S	HB1	3.922
145	26	S	HB2	3.875
146	26	S	N	115.658
147	27	N	C'	175.012
148	27	N	CA	53.233
149	27	N	H	8.557
150	27	N	HA	4.745
151	27	N	HD1	7.678
152	27	N	HD2	6.965
153	27	N	N	120.845
154	27	N	ND	113.25
155	28	K	C'	177.14
156	28	K	CA	56.709
157	28	K	CB	32.724
158	28	K	H	8.419
159	28	K	HA	4.278
160	28	K	N	121.996
161	29	G	C'	173.673
162	29	G	CA	45.053
163	29	G	H	8.459
164	29	G	HA	3.931
165	29	G	N	109.723
166	30	A	C'	177.628
167	30	A	CA	52.378
168	30	A	CB	19.355
169	30	A	H	8.092
170	30	A	HA	4.314
171	30	A	N	123.667
172	31	I	C'	176.411
173	31	I	CA	61.067
174	31	I	CB	38.255
175	31	I	H	8.256
176	31	I	HA	4.161
177	31	I	N	120.971
178	32	I	C'	176.619

A β 40-A2T				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
134	16	K	HB2	1.814
135	16	K	HD	1.681
136	16	K	HE	2.978
137	16	K	HG1	1.396
138	16	K	HG2	1.46
139	16	K	N	123.257
140	17	L	C'	176.815
141	17	L	CA	55.062
142	17	L	CB	42.341
143	17	L	CD1	23.445
144	17	L	CD2	24.871
145	17	L	CG	26.91
146	17	L	H	8.333
147	17	L	HA	4.337
148	17	L	HB	1.61
149	17	L	HD1	0.851
150	17	L	HD2	0.925
151	17	L	HG	1.516
152	17	L	N	124.011
153	18	V	C'	175.101
154	18	V	CA	61.987
155	18	V	CB	33.068
156	18	V	CG1	21.196
157	18	V	CG2	20.589
158	18	V	H	8.073
159	18	V	HA	4.039
160	18	V	HB	1.914
161	18	V	HG1	0.759
162	18	V	HG2	0.845
163	18	V	N	121.684
164	19	F	C'	174.716
165	19	F	CA	57.465
166	19	F	CB	40.166
167	19	F	H	8.338
168	19	F	HA	4.589
169	19	F	HB1	2.917
170	19	F	HB2	3
171	19	F	N	124.549
172	20	F	C'	174.775
173	20	F	CA	57.266
174	20	F	CB	40.029
175	20	F	H	8.296
176	20	F	HA	4.575
177	20	F	HB1	2.937
178	20	F	HB2	3.09

Aβ40-D23N				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
179	32	I	CA	61.129
180	32	I	CB	38.384
181	32	I	H	8.355
182	32	I	HA	4.163
183	32	I	N	126.527
184	33	G	C'	173.693
185	33	G	CA	45.084
186	33	G	H	8.535
187	33	G	HA	3.933
188	33	G	N	113.316
189	34	L	C'	177.351
190	34	L	CA	55.123
191	34	L	CB	42.492
192	34	L	H	8.118
193	34	L	HA	4.349
194	34	L	N	121.758
195	35	M	C'	176.154
196	35	M	CA	55.23
197	35	M	CB	32.576
198	35	M	H	8.513
199	35	M	HA	4.54
200	35	M	N	122.166
201	36	V	C'	176.678
202	36	V	CA	62.542
203	36	V	CB	32.714
204	36	V	H	8.312
205	36	V	HA	4.131
206	36	V	N	122.64
207	37	G	C'	174.454
208	37	G	CA	45.141
209	37	G	H	8.666
210	37	G	HA1	4.001
211	37	G	HA2	3.979
212	37	G	N	113.342
213	38	G	C'	173.694
214	38	G	CA	45.012
215	38	G	H	8.313
216	38	G	HA1	4.038
217	38	G	HA2	3.952
218	38	G	N	108.82
219	39	V	C'	175.563
220	39	V	CA	62.506
221	39	V	CB	32.889
222	39	V	H	8.134
223	39	V	HA	4.186

Aβ40-A2T				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
179	20	F	N	123.193
180	21	A	C'	177.245
181	21	A	CA	52.321
182	21	A	CB	19.433
183	21	A	H	8.312
184	21	A	HA	4.224
185	21	A	HB	1.377
186	21	A	N	126.352
187	22	E	C'	176.111
188	22	E	CA	56.567
189	22	E	CB	30.295
190	22	E	CG	36.275
191	22	E	H	8.429
192	22	E	HA	4.206
193	22	E	HB1	1.937
194	22	E	HB2	2.045
195	22	E	HG1	2.276
196	22	E	HG2	2.291
197	22	E	N	120.145
198	23	D	C'	176.579
199	23	D	CA	54.069
200	23	D	CB	40.996
201	23	D	CN	54.057
202	23	D	H	8.495
203	23	D	HA	4.657
204	23	D	HB1	2.652
205	23	D	HB2	2.751
206	23	D	N	121.98
207	24	V	C'	177.084
208	24	V	CA	62.772
209	24	V	CB	32.198
210	24	V	CG2	20.186
211	24	V	H	8.233
212	24	V	HA	4.142
213	24	V	HB	2.199
214	24	V	HG2	0.977
215	24	V	N	120.867
216	25	G	C'	174.479
217	25	G	CA	45.384
218	25	G	H	8.619
219	25	G	HA	3.99
220	25	G	N	111.895
221	26	S	C'	174.476
222	26	S	CA	58.57
223	26	S	CB	63.684

Aβ40-D23N				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
224	39	V	N	120.038
225	40	V	C'	181.052
226	40	V	CA	63.766
227	40	V	CB	33.276
228	40	V	H	7.878
229	40	V	HA	4.059
230	40	V	N	128.532

Aβ40-A2T				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
224	26	S	H	8.212
225	26	S	HA	4.441
226	26	S	HB	3.903
227	26	S	HB1	3.897
228	26	S	HB2	3.908
229	26	S	N	115.622
230	27	N	C'	175.062
231	27	N	CA	53.273
232	27	N	CB	38.511
233	27	N	H	8.547
234	27	N	HA	4.752
235	27	N	HB1	2.827
236	27	N	HB2	2.873
237	27	N	HD1	7.705
238	27	N	HD2	6.985
239	27	N	N	120.706
240	27	N	ND	113.2
241	28	K	C'	177.213
242	28	K	CA	56.677
243	28	K	CB	32.548
244	28	K	CD	28.948
245	28	K	CE	42.057
246	28	K	CG	24.692
247	28	K	H	8.424
248	28	K	HA	4.275
249	28	K	HB1	1.78
250	28	K	HB2	1.897
251	28	K	HD	1.681
252	28	K	HE	3.001
253	28	K	HG1	1.419
254	28	K	HG2	1.578
255	28	K	N	121.9
256	29	G	C'	173.705
257	29	G	CA	45.103
258	29	G	H	8.487
259	29	G	HA	3.931
260	29	G	N	109.659
261	30	A	C'	177.643
262	30	A	CA	52.369
263	30	A	CB	19.292
264	30	A	H	8.094
265	30	A	HA	4.312
266	30	A	HB	1.379
267	30	A	N	123.651
268	31	I	C'	176.43

Aβ40-A2T				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
269	31	I	CA	61.089
270	31	I	CB	38.349
271	31	I	CD	12.641
272	31	I	CG1	27.324
273	31	I	CG2	17.607
274	31	I	H	8.259
275	31	I	HA	4.157
276	31	I	HB	1.868
277	31	I	HD	0.871
278	31	I	HG11	1.214
279	31	I	HG12	1.518
280	31	I	HG21	0.894
281	31	I	N	120.931
282	32	I	C'	176.643
283	32	I	CA	61.144
284	32	I	CB	38.395
285	32	I	CD	12.641
286	32	I	CG1	27.326
287	32	I	CG2	17.391
288	32	I	H	8.366
289	32	I	HA	4.16
290	32	I	HB	1.872
291	32	I	HD	0.875
292	32	I	HG11	1.22
293	32	I	HG12	1.521
294	32	I	HG21	0.936
295	32	I	N	126.488
296	33	G	C'	173.68
297	33	G	CA	45.097
298	33	G	H	8.547
299	33	G	HA	3.935
300	33	G	N	113.244
301	34	L	C'	177.336
302	34	L	CA	55.168
303	34	L	CB	42.472
304	34	L	CD1	23.613
305	34	L	CD2	24.756
306	34	L	CG	26.868
307	34	L	H	8.126
308	34	L	HA	4.351
309	34	L	HB	1.61
310	34	L	HD	0.903
311	34	L	HD1	0.884
312	34	L	HD2	0.933
313	34	L	HG	1.579

Aβ40-A2T				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
314	34	L	N	121.76
315	35	M	C'	176.136
316	35	M	CA	55.229
317	35	M	CB	32.627
318	35	M	CG	31.926
319	35	M	H	8.523
320	35	M	HA	4.542
321	35	M	HB	2.049
322	35	M	HB1	2.032
323	35	M	HB2	2.073
324	35	M	HG1	2.531
325	35	M	HG2	2.607
326	35	M	N	122.153
327	36	V	C'	176.713
328	36	V	CA	62.544
329	36	V	CB	32.739
330	36	V	CG	21.084
331	36	V	H	8.325
332	36	V	HA	4.127
333	36	V	HB	2.092
334	36	V	HG	0.968
335	36	V	N	122.635
336	37	G	C'	174.433
337	37	G	CA	45.133
338	37	G	H	8.68
339	37	G	HA	3.992
340	37	G	HA1	4.005
341	37	G	HA2	3.98
342	37	G	N	113.314
343	38	G	C'	173.685
344	38	G	CA	45.032
345	38	G	H	8.326
346	38	G	HA	3.97
347	38	G	HA1	4.019
348	38	G	HA2	3.957
349	38	G	N	108.821
350	39	V	C'	175.546
351	39	V	CA	62.441
352	39	V	CB	32.823
353	39	V	CG	21.173
354	39	V	H	8.15
355	39	V	HA	4.182
356	39	V	HB	2.095
357	39	V	HG	0.954
358	39	V	N	120.038

Aβ40-A2T				
Resonance ID	Residue #	Residue Type	Atom ID	Chemical Shift
359	40	V	C'	181.105
360	40	V	CA	63.722
361	40	V	CB	33.313
362	40	V	CG1	20.218
363	40	V	CG2	21.546
364	40	V	H	7.895
365	40	V	HA	4.059
366	40	V	HB	2.067
367	40	V	HG1	0.907
368	40	V	HG2	0.927
369	40	V	N	128.556

Table 7.3. $^3J_{\text{HNHA}}$ values for several A β species, as determined by the J-resolved SOFAST HMQC method(411). Italicized $^3J_{\text{HNHA}}$ listings are ambiguous values derived from glycine. Data generated with the assistance of Dr. Christopher Connors and Nicolina Clemente.

RN	A β 40 1		A β 40 2		A β 40 3		A β 42		A β 42-M35ox		A β 40-E22A		A β 40-D23N 1		A β 40-D23N 2		A β 40-A2T		
	$^3J_{\text{HNHA}}$	Δ^3J_{HNHA}																	
D1																			
A/T2																			
E3	6.8	0.0	6.8	0.1	6.7	0.1	7.0	0.1	6.3	0.1	6.4	0.1	6.8	0.0	6.8	0.0	7.1	0.0	
F4	7.3	0.0	7.3	0.1	7.3	0.1	7.4	0.1	6.9	0.2	6.8	0.1	7.5	0.1	7.4	0.1	7.4	0.0	
R5	8.0	0.0	8.2	0.2	8.1	0.2	8.3	0.1	7.1	0.2	7.6	0.2	8.2	0.1	7.8	0.1	8.1	0.0	
H6																			
D7																			
S8	6.8	0.2	7.0	0.4	6.8	0.4	7.5	0.5	6.3	0.6	6.5	0.9	7.0	0.3	7.0	0.2	7.2	0.1	
G9											5.6	N/A			4.1	N/A			
Y10	7.0	0.0	6.7	0.3	6.6	0.2	7.7	0.1	6.5	0.3	5.3	0.3	6.9	0.2	6.5	0.2	7.2	0.0	
E11	6.9	0.1	7.2	0.4	7.1	0.3	7.2	0.1	6.4	0.4	6.1	0.4	7.0	0.2	7.1	0.1	7.2	0.0	
V12	7.3	0.1					7.3	0.1	6.7	0.4	6.7	0.6			7.3	0.2	6.9	0.0	
H13							7.7	0.1											
H14																			
Q15																	7.3	0.1	
K16	7.2	0.1	7.4	0.6	6.9	0.5	7.5	0.3	6.6	0.8	6.2	0.7	7.6	0.2	7.4	0.2	7.6	0.1	
L17	7.4	0.1	7.8	0.3	7.6	0.2	7.8	0.2	6.7	0.5	6.7	0.4	7.7	0.2	7.5	0.2	7.8	0.1	
V18	9.0	0.1	8.7	0.2	8.6	0.2	9.0	0.1	8.6	0.5	8.1	0.3	8.8	0.2	8.7	0.2	8.6	0.0	
F19	8.3	0.1	8.5	0.2	8.4	0.2	8.6	0.2	7.6	0.4	7.7	0.4	8.3	0.1	8.3	0.1	8.3	0.1	
F20	8.5	0.1	8.4	0.2	8.5	0.2	8.8	0.2	7.6	0.4	7.6	0.4	8.2	0.1	8.3	0.2	8.4	0.1	
A21	6.0	0.1	6.4	0.2	6.3	0.1	6.2	0.2	5.6	0.3	5.5	0.3	6.6	0.1	6.4	0.1	6.5	0.0	
E22	6.4	0.0	6.6	0.1	6.4	0.1	6.6	0.1	6.0	0.3			6.5	0.1	6.6	0.1	6.5	0.0	
D/N23	7.2	0.0	7.2	0.1	7.3	0.1			6.7	0.2	6.5	0.2	8.5	0.2	8.1	0.2	7.6	0.0	
V24	7.3	0.0	7.4	0.1	7.4	0.1	7.5	0.1	6.8	0.2	7.1	0.2	7.6	0.1	7.6	0.1	7.6	0.0	
G25									5.4	N/A									
S26	7.3	0.1	7.4	0.3	7.2	0.2	7.8	0.4	6.6	0.3	6.8	0.5	7.6	0.2	7.4	0.2	6.6	0.1	
N27	7.4	0.2	8.1	1.0	8.5	0.8			7.0	0.5			8.3	0.5	8.6	0.5			
K28																	7.9	0.0	
G29									5.3	N/A	5.3	N/A	5.1	N/A	5.6	N/A			
A30																	6.0	0.0	
I31	8.0	0.0	8.1	0.1	8.0	0.1	7.8	0.1	7.4	0.2	7.6	0.2	8.2	0.1	8.2	0.1	8.2	0.0	
I32	7.8	0.0	8.0	0.1	7.9	0.1	7.9	0.1	7.2	0.3	7.5	0.2	7.9	0.1	7.9	0.1	7.9	0.0	
G33	5.4	N/A	5.3	N/A			5.3	N/A	5.4	N/A			5.5	N/A	4.6	N/A			
L34	7.2	0.0	7.2	0.1	7.1	0.1	7.3	0.1	6.6	0.2	6.6	0.1	7.2	0.0	7.2	0.1	7.1	0.0	
M/X35	7.7	0.0	7.6	0.1	7.3	0.1					7.1	0.1	7.6	0.1	7.5	0.1	7.7	0.0	
V36	8.0	0.0	7.9	0.1	7.8	0.1	8.5	0.5	7.3	0.8			7.9	0.0	7.8	0.0	8.0	0.0	
G37	5.1																		
G38	5.9	N/A	5.9	N/A	5.9	N/A	5.5	N/A	5.8	N/A	5.8	N/A	6.0	N/A	6.0	N/A			
V39	8.2	0.0	8.3	0.0	8.2	0.0			7.5	0.1	7.9	0.1	8.3	0.0	8.3	0.0	8.3	0.0	
V40	8.9	0.0	9.0	0.0	8.9	0.0	8.2	0.0	7.8	0.1	8.7	0.0	9.1	0.0	9.0	0.0	9.1	0.0	
I41							8.8	0.1	8.3	0.1									
A42							7.1	0.0	6.7	0.1									

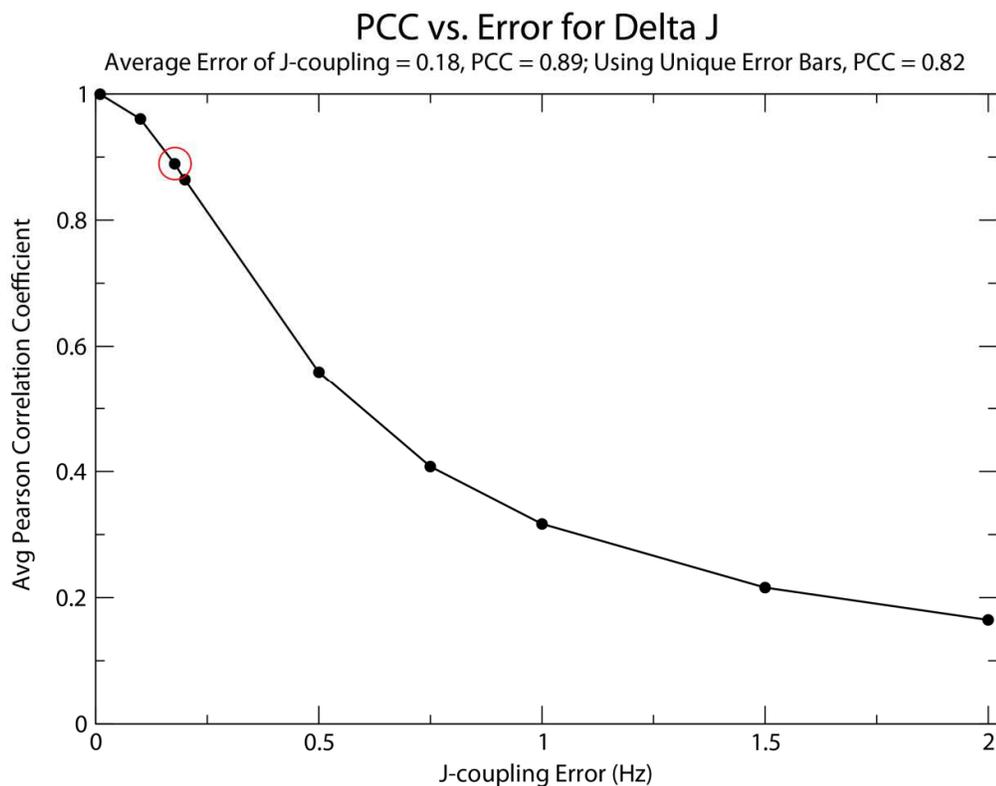


Figure 7.2. Assessment of the significance of the PCC using the Δ^3J_{HNHA} between A β 40-D23N and A β 40 through 100000 Monte Carlo simulations based on the experimental data. Using the experimental values of $^3J_{\text{HNHA}}$ for A β 40 and A β 40-D23N and a given averaged error value to assume for all residues, randomized $^3J_{\text{HNHA}}$ values were generated within the given error for both sets and Δ^3J_{HNHA} was calculated. The PCC was then calculated to the actual Δ^3J_{HNHA} value, and the average of the 100000 PCCs was plotted. Circled is the value when the average of the actual error bars (0.18 Hz) for both data sets was used as input, producing a PCC of 0.89. When the actual experimental error bars for each residue were instead used for this calculation (instead of employing global average, as was used to generate the plotted curve), the resulting averaged PCC was 0.82.

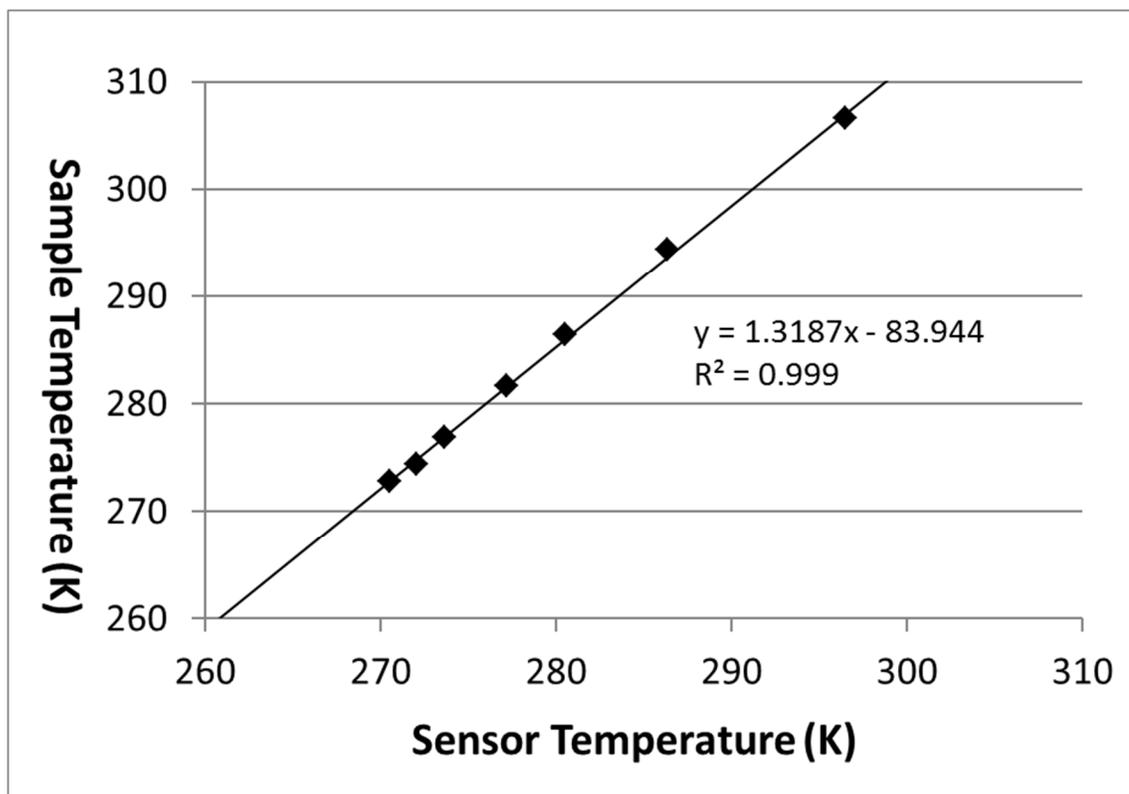


Figure 7.3. Temperature calibration for a sample in the 600 MHz Bruker spectrometer inside the high pressure cell used for the high pressure NMR experiments in Chapter 4. Sample temperature was determined by converting the ^1H chemical shift difference between the methyl and hydroxyl groups of 100% methanol(442).

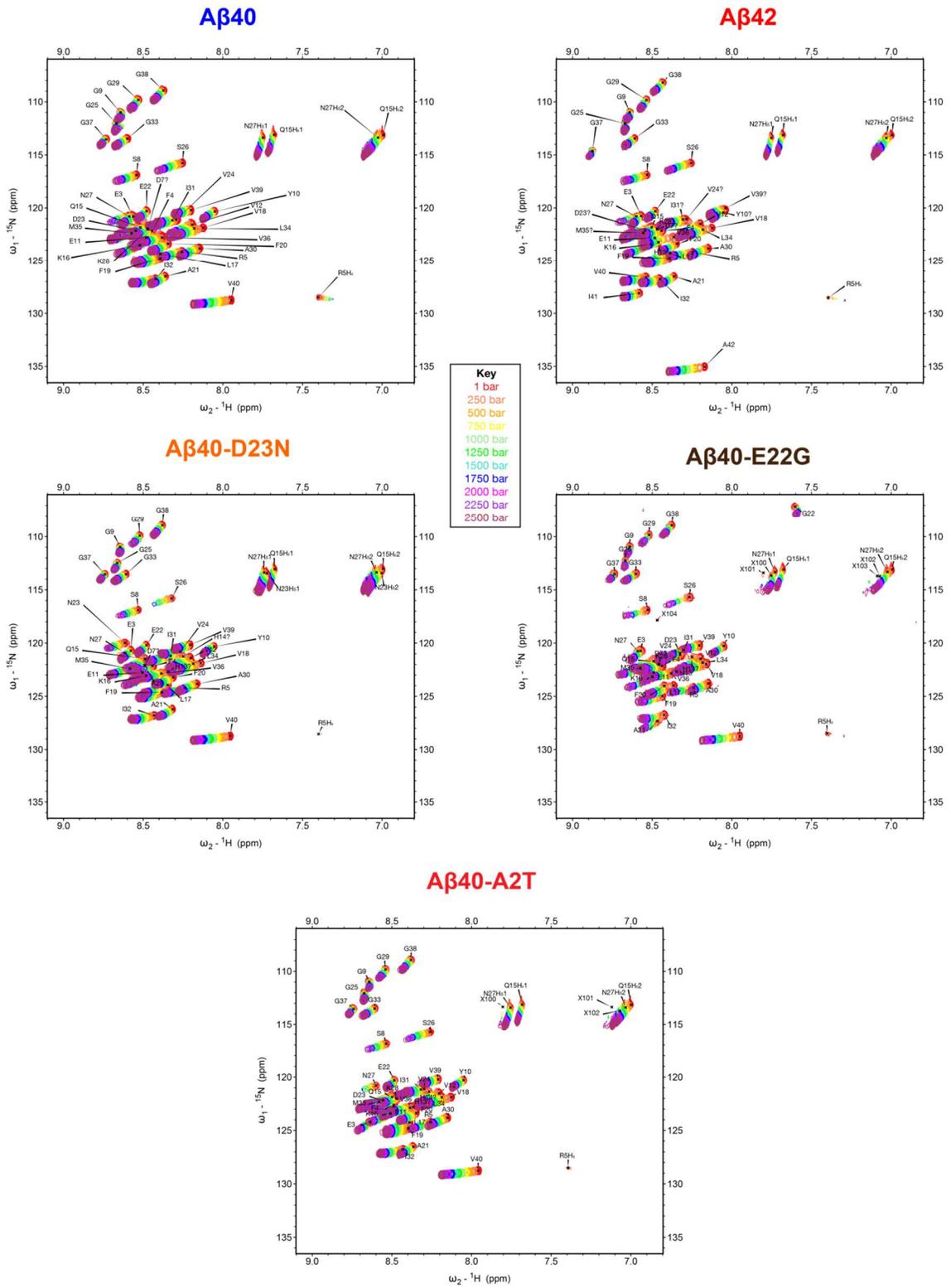


Figure 7.4. ^1H - ^{15}N HSQC spectra for different monomeric $\text{A}\beta$ species collected with high pressure NMR, ranging from ambient pressure to 2500 bar at 250 bar increments. Assignments are shown on the spectra marked at their locations at 1 bar. Unknown peaks are labelled X100-X104. All spectra are shown at the same contour levels and with the same axes to facilitate comparison, both between pressures and between species. Data collected on a 600 MHz spectrometer at 277 K, using the high pressure pump and buffer conditions as described in Chapter 4.

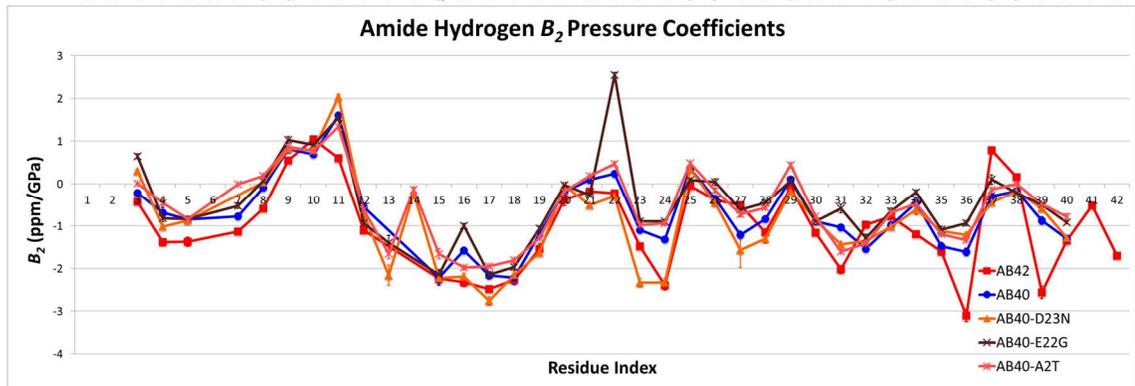
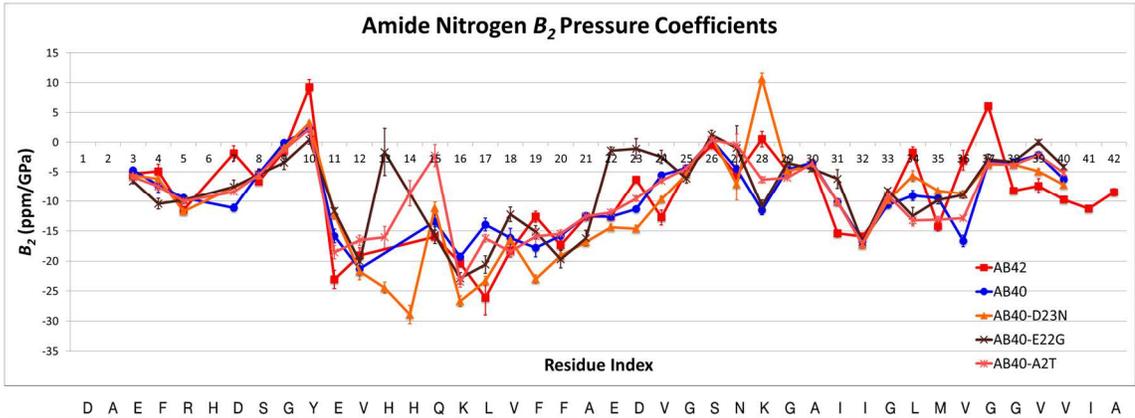
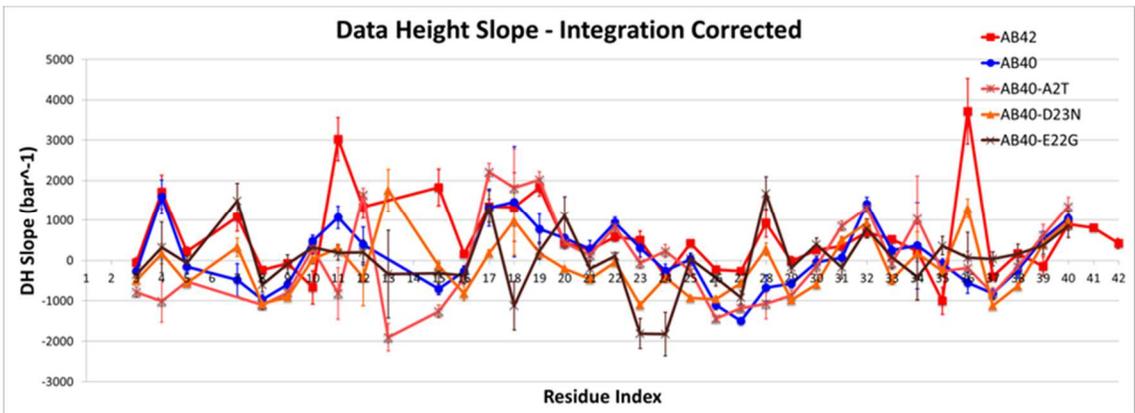
A**B**

Figure 7.5. Additional effects of pressure on the chemical shifts and peak intensities of monomeric A β . (A) Plot of the second order (B_2) pressure coefficients for the nitrogen and hydrogen nuclei for different species of monomeric A β . B_1 coefficients of the second order Taylor polynomial (Equation 4.4) are presented in the main text of Chapter 4 (Figure 4.8A). The sequence of wild type A β 42 is presented in between the two graphs. (B) Slope of the linear fit of data heights for all residues. Data heights were corrected by the overall integration of all amide signals (see Figure 4.8B). All data is presented for the sake of completeness, but the linear fits are not always appropriate, mostly due to spectral overlap. In general, only values with low errors should be trusted. Coefficients and intensities are based on the position and magnitude of peaks, respectively, on ^1H - ^{15}N HSQC spectra collected with high pressure NMR, ranging from ambient pressure to 2500 bar at 250 bar increments. Data collected on a 600 MHz spectrometer at 277 K, using the high pressure pump and buffer conditions as described in Chapter 4. Peak positions were corrected with random coil shifts as a function of pressure, as described in the text.

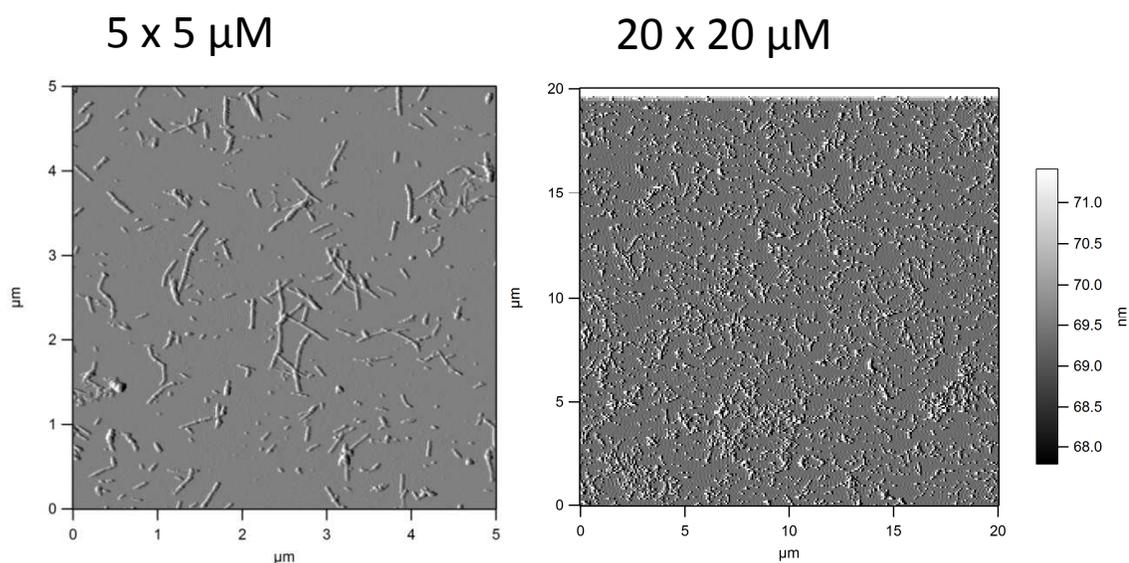


Figure 7.6. Topology of aggregates of the A β 40-E22G sample used for high pressure NMR in Chapter 4, as determined by atomic force microscopy. This data was collected on the sample after high pressure HSQC experiments followed by a subsequent incubation of six hours at ambient pressure and 277 K (experimental buffer matches conditions for high pressure NMR described in Chapter 4). Data collected by Nicolina Clemente and Sherry Xu.

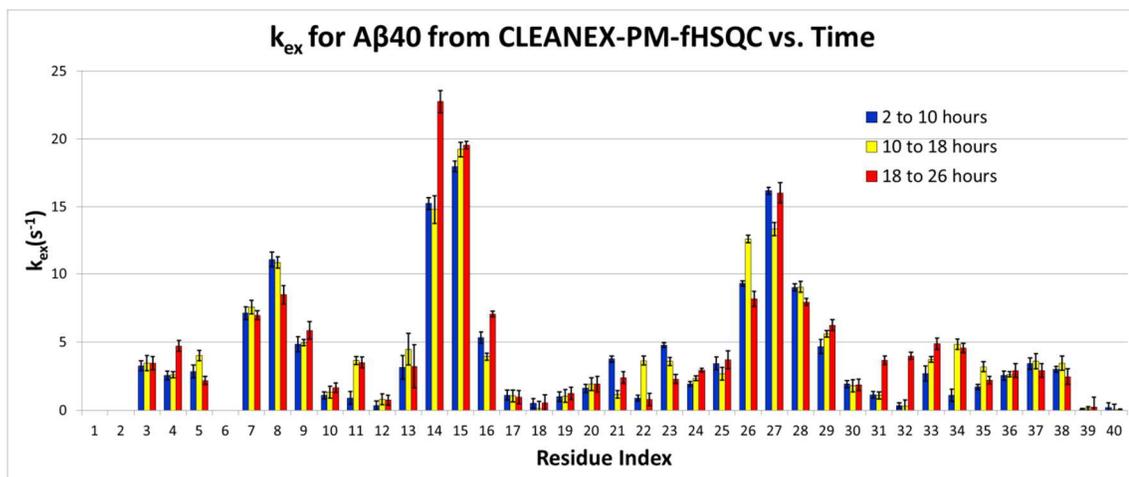
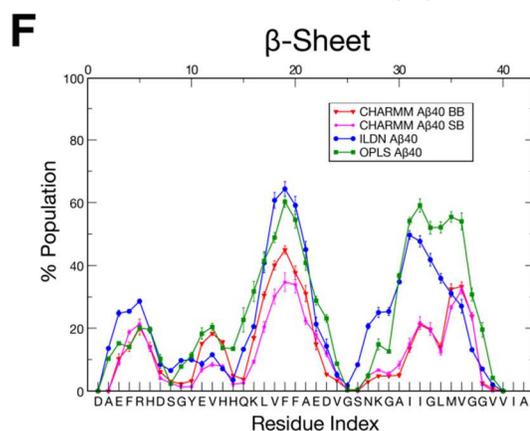
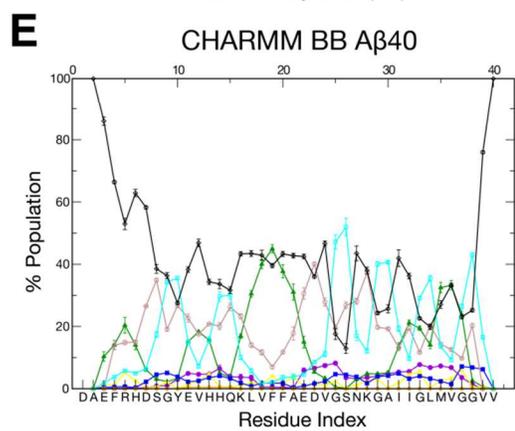
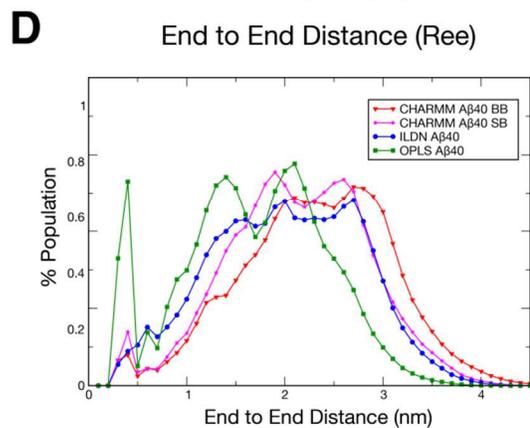
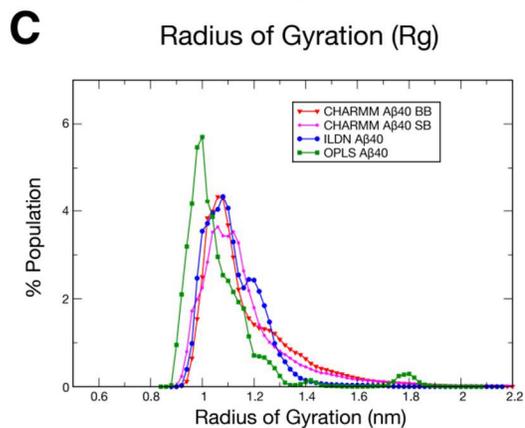
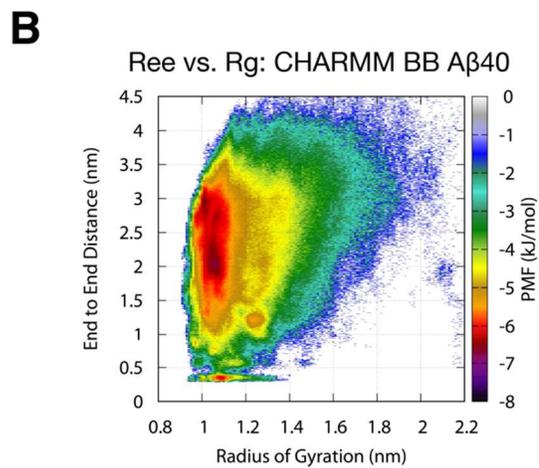
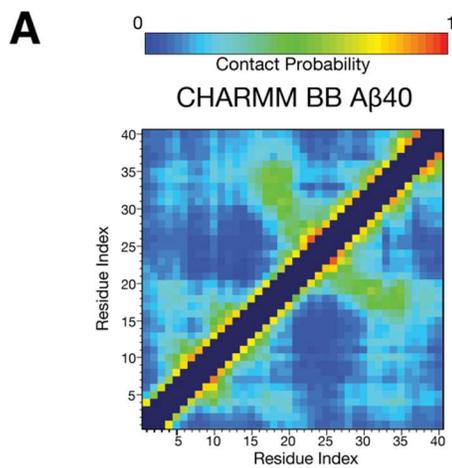


Figure 7.7. Solvent exchange rates (k_{ex}) as a function of time/aggregation for the same sample run at 1 bar. This is a control experiment for high pressure solvent determination (Figure 4.9), which was also done using the same sample over the equivalent amount of experiment time. Substantial aggregation was observed over the duration of the control, gauged by integrating the amide region (8.9-6.0 ppm) of the projection of the HSQC experiments used in the solvent exchange experiments. The HSQC collected at 10-11 hours was 80.74% the intensity of the one at 2-3 hours, the HSQC at 18-19 hours was 64.97% as intense, and the spectrum collected at 26-27 hours was 53.77% as intense. Despite this, there is no observed directional change in observed k_{ex} collected over time because we correct all of the CLEANEX-PM data by the original HSQC data height. Thus, we conclude this experiment is robust against changes in aggregation of the peptide.



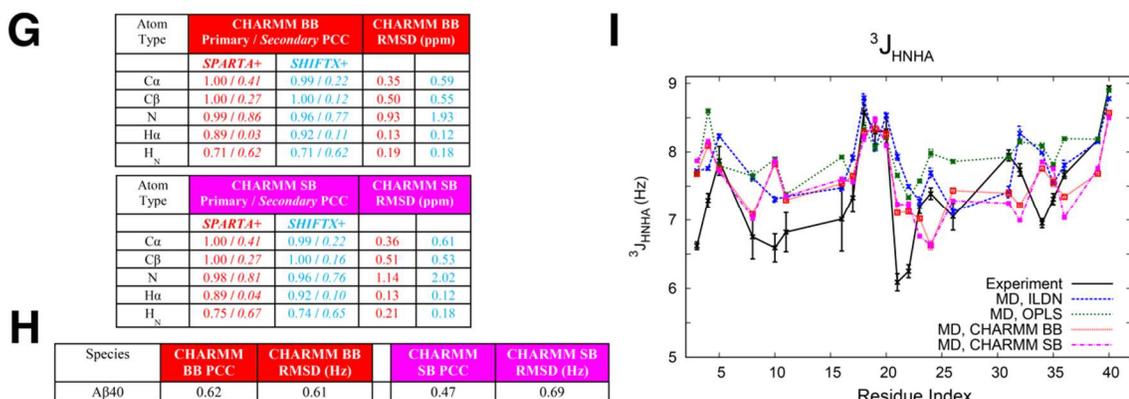


Figure 7.8. Analysis of A β 40, recently simulated using REMD with the CHARMM22* force field(28) and the modified CHARMM TIP3P (TIP3SP) water model(20) (the “CHARMM” condition). Two different REMD simulations were run to 500 ns/replica with this condition. The “BB” version uses a larger box size of 5.9 nm (after NPT equilibration), 6846 water molecules, and 66 replicas. The “SB” version matches the condition used for other REMD simulations in this thesis, with a box size of 5.4 nm (after NPT equilibration), 4947 water molecules, and 52 replicas. REMD was run over the temperature range \sim 270-600 K in either simulation. Analysis was done on the 150-500 ns/replica period, on temperatures 7-13 for BB and temperatures 6-12 for SB, both encompassing the range \sim 290-310 K. Otherwise, the exact same simulation parameters and analysis methods as described for simulations of A β with other force fields in Chapter 2-4 are employed here. (A) Map illustrating the probability of contact over the production ensemble for CHARMM A β 40 BB in the upper corner and CHARMM A β 40 SB in the lower corner. Here, two residues are defined to be in contact if their C α -C α distance is \leq 8 Å(352). Compare with Figure 2.3 and Figure 3.3. (B) Free energy surface along the coordinates of radius of gyration (R_g) and end to end distance (R_{ee}) in units of kJ/mol for the production ensemble of CHARMM A β 40 BB, using 0.02 nm binning in both dimensions. Compare with Figure 2.9 and Figure 3.7. (C) Histogram for R_g for A β 40 simulated with multiple force fields, using 0.02 nm-wide bins. (D) Histogram for R_{ee} , using 0.1 nm-wide bins. (E) Percent composition of secondary structure as determined by DSSP(347, 348) for CHARMM A β 40 BB, averaged over the production ensemble. Compare to Figure 2.2B and Figure 3.2B. (F) Percent composition of β -sheet alone for A β 40 simulated over multiple force fields. The sequence of wild type A β 42 is presented in the lower x -axis. (G) Correlation and RMSD of predicted chemical shifts based on frames of CHARMM simulations with experimentally measured values. In red are the comparisons using shifts predicted by SPARTA+(433), while light blue use values predicted by SHIFTX+(434). “Primary” refers to the comparison of the chemical shifts as they are, while “secondary” correlation is measured once both the predicted and experimental chemical shifts are corrected for their random coil values. The final residue, V40, is unreliably predicted by both calculators and is ignored in all correlation calculations. Stereospecific H α shifts are provided by SPARTA+ for glycine; these were averaged for H α comparisons. Compare with Figure 4.5. (H) Correlation and RMSD of calculated $^3J_{\text{HNHA}}$ couplings from CHARMM simulations compared to

experimentally determined values. Computational values were backcalculated using Karplus parameters published by Vuister and Bax(362), and glycines were not compared due to experimental ambiguity. Compare with Figure 4.6. (I) Plot of experimental $^3J_{\text{HNHA}}$ couplings and those derived from simulations of different force fields for A β 40, as a function of residue index.

AB42 5.0A

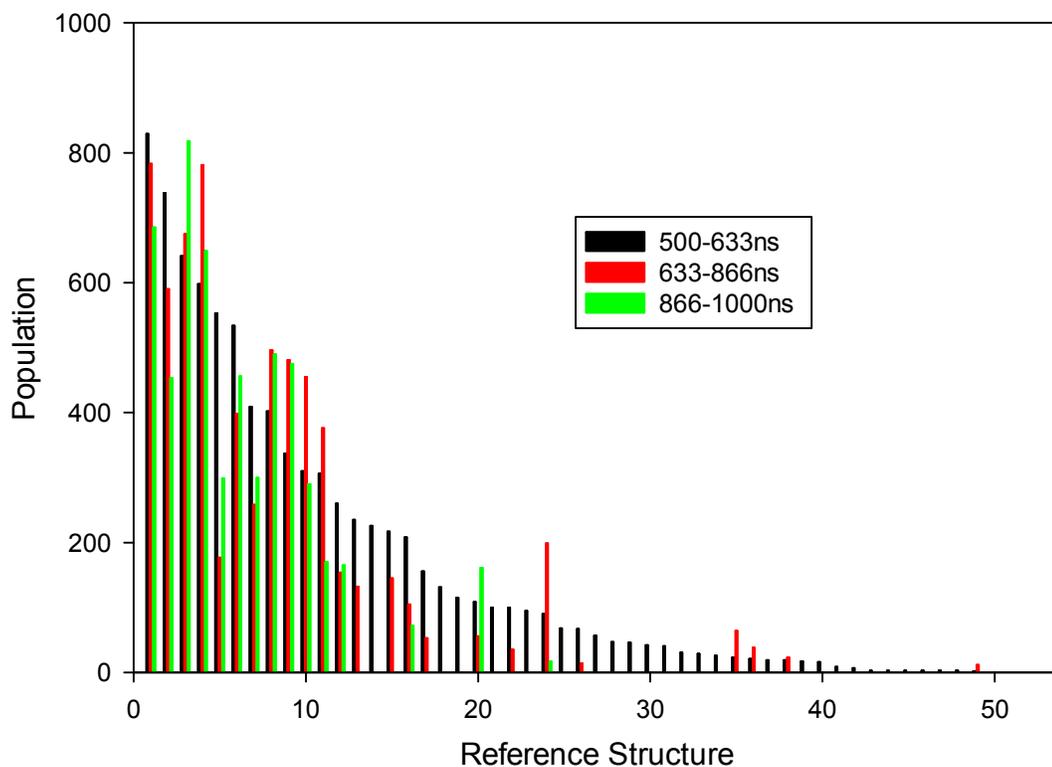
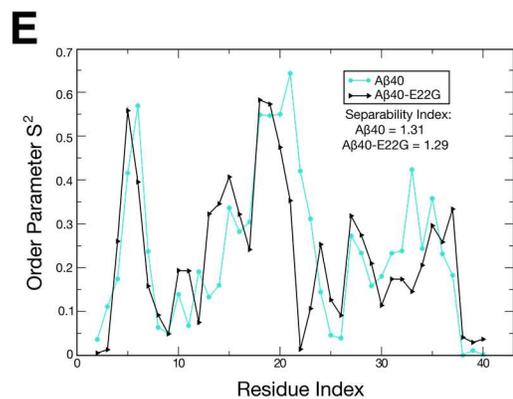
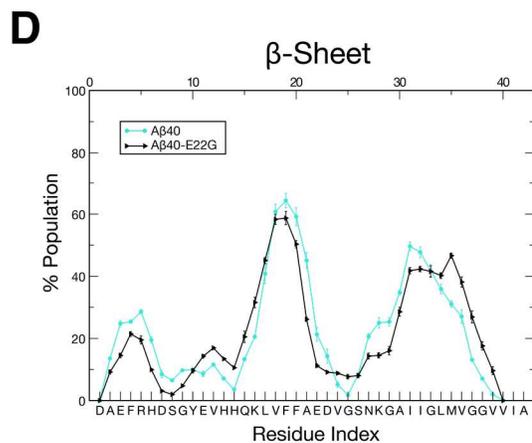
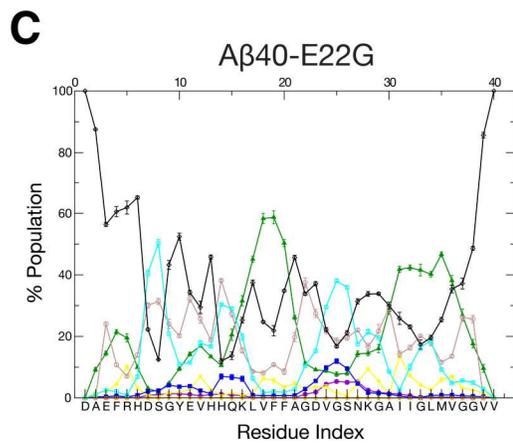
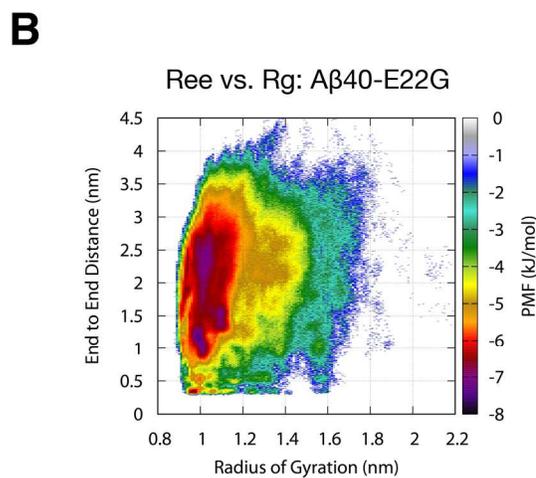
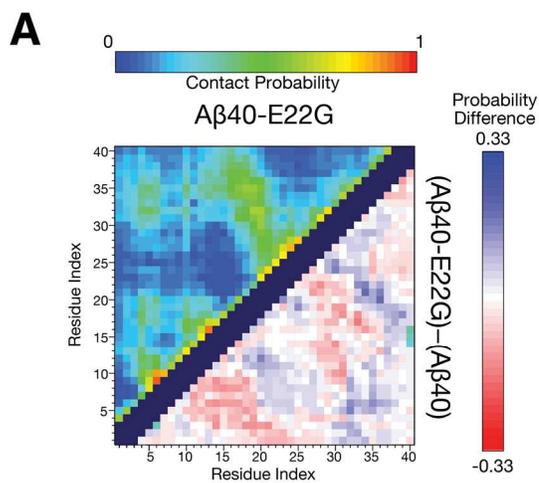


Figure 7.9. Assessment of convergence of the OPLS A β 42 simulation (described in Chapter 2) by analysis of the distribution of structures with respect to those determined from RMSD-based clustering of earlier parts of the trajectory. Frames of temperature 9 (299.6 K) from the time interval 500-633 ns were clustered using Daura clustering(351) with a 5.0 Å RMSD cutoff, then the centroids from this clustering were used as reference structures, while frames in the 633-866 ns and 866-1000 ns segments were grouped to the reference structure with the lowest RMSD. Centroids are numbered in terms of decreasing population of frames represented in the 500-633 ns window. Data provided by Anthony Bishop and presented here with his permission.



F

	Experimental		SPARTA+		SHIFTX+	
	N	H _N	N	H _N	N	H _N
G9	108.87	8.23	109.22	7.93	108.50	8.05
G22	105.04	7.19	108.36	8.03	108.31	7.90
G25	110.18	8.26	109.98	8.27	109.32	8.32
G29	107.76	8.11	108.31	8.21	108.40	8.32
G33	111.42	8.19	111.58	8.28	110.97	8.17
G37	111.44	8.33	110.17	8.22	111.48	8.27
G38	106.86	7.97	107.71	8.10	107.42	8.05

Figure 7.10. Analysis of A β 40-E22G, recently simulated using REMD with the AMBER99sb-ILDN(21) force field and the TIP4P-Ew(22) water model to 1000 ns/replica, using the exact same simulation parameters and analysis methods employed for other A β peptides in Chapter 3 and 4. (A, upper corner) Map illustrating the probability of contact over the production ensemble for A β 40-E22G. Here, two residues are defined to be in contact if their C α -C α distance is ≤ 8 Å(352). (A, lower corner) Difference between this contact probability map for A β 40-E22G compared with A β 40 (see Figure 3.3). (B) Free energy surface along the coordinates of radius of gyration (R_g) and end to end distance (R_{ee}) in units of kJ/mol for the production ensemble of A β 40-E22G, using 0.02 nm binning in both dimensions, compare with A β 40 in Figure 3.7C. (C) Percent composition of secondary structure as determined by DSSP(347, 348) for A β 40-E22G, averaged over the production ensemble. Compare to Figure 3.2B. (D) Percent composition of β -sheet alone for A β 40-E22G versus wild type A β 40. The sequence of wild type A β 42 is presented in the lower x -axis. (E) S^2 order parameters for all backbone N-H internuclear vectors (excluding the N-terminus), calculated with the iRED technique(353) for A β 40-E22G versus wild type A β 40. Compare to Figure 3.8. (F) Chemical shifts for glycines of A β 40-E22G calculated from the REMD ensemble using the SPARTA+(433) and SHIFTX+(434) predictors, versus the actual experimental chemical shifts (measured using ^1H - ^{15}N HSQC with the A β sample preparation described in Chapter 4. Experimental shifts here are referenced so the average shift of all glycines, excluding the G22 outlier, match the predicted values.