CONFORMATIONAL CHARACTERIZATION OF THE INTRINSICALLY

DISORDERED PROTEIN α-SYNUCLEIN: A REPLICA EXCHANGE MOLECULAR

DYNAMICS STUDY

by

CHITRA NARAYANAN

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

And the Graduate School of Biomedical Sciences

University of Medicine and Dentistry of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Biochemistry

Written under the direction of

Prof. Ronald M. Levy

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October 2012

ABSTRACT OF THE DISSERTATION

Conformational Characterization of the Intrinsically Disordered Protein α-Synuclein: A

Replica Exchange Molecular Dynamics Study

By CHITRA NARAYANAN

Dissertation Director:
Prof. Ronald M. Levy

Proteins in their functional forms play a vital role in all major processes in the cell. Protein misfolding has been associated with a large number of diseased states. Intrinsically Disordered Proteins (IDPs) have gained much attention because of their involvement in key cellular processes and predominance in diseased states. A number of neurodegenerative diseases including Alzheimer's, Parkinson's and Huntington's diseases have been correlated with the aggregation of IDPs.

α-synuclein is a 140 amino acid archetypal IDP implicated in the pathology of Parkinson's disease. Aggregation of α-synuclein is sensitive to changes in amino acid substitutions along the sequence and changes in chemical environments. Characterizing the monomeric form is essential to understanding the conformational changes leading to the aggregated state. In this work, ensembles of structures generated from Replica Exchange Molecular Dynamics simulations were used to characterize aggregation-prone states of monomeric form of α-synuclein.

The conformational characteristics of α-synuclein were evaluated in terms of the statistical properties of the chain over a range of solvent conditions and comparing with predictions from polymer theory, using temperature as a proxy for solvent quality. Results of this work indicate that α-synuclein behaves as expected for a homopolymer chain at extremes of solvent quality while at intermediate values, the identity of the monomeric units along the sequence significantly influence the polymeric and conformational characteristics of the chain. Comparison of the back-calculated experimental parameters for the simulation ensemble with that of the NMR observation shows that ensembles that fit to experimental parameters describing both local and long-range characteristics, represented by the experimental Residual Dipolar Couplings (RDC) and Paramagnetic Relaxation Enhancements respectively, provides a better representation of the experimental ensemble.

The conformations of the neutral and low pH ensembles of α-synuclein were characterized by integrating molecular simulations with experimental NMR observations, to elucidate the effect of the altered charge distribution with change in pH on the structural properties of these ensembles. The results from this study indicate a significant structural reorganization with change in pH in terms of the long-range interactions, compaction of the C-terminal region at low pH leading to faster aggregation at low pH.

Acknowledgement

I would first like to express my sincere gratitude to my adviser Dr. Ronald M. Levy for his guidance, support and encouragement during this time. I would also like to extend my sincere thanks to Dr. Jean Baum for her support and encouragement for all the research projects. I would also like to express my gratitude to Dr. Vikas Nanda for his time and encouragement, and Dr. Emilio Gallicchio for serving on the thesis defense committee, and Dr. David Talaga for serving in the thesis proposal committee.

A number of people have been involved in the successful completion of my research work. Many thanks to Dr. Michael Andrec for all his guidance - his mentoring and help during my lab rotation played a significant role in my choice of the lab. Thanks also to Dr. Anthony Felts for all his help in the work on the minipeptides. Many thanks to Dr. Daniel Weinstock for all his help in the work on alpha-syunclein.

I am indebted to Dr. Kuen-Phon Wu for all his support and constructive criticism. Thanks also to Dr. Yu-Jen Chen, and other present and past members of the Baum and the Levy groups. I would like to extend my gratitude to Dr. N. Srinivasan for his mentoring and confidence in my potential to be a researcher while I worked with him in the Indian Institute of Science.

A lot of friends were involved in making my time at Rutgers very memorable. A special thanks to all my friends in the Mathematics and Physics departments with whom I had many interesting and educational conversations! Thanks also to my two wonderful friends - Dr. Amrita Rai and Hemlata Chawla, who have made my journey through Masters' to Ph.D a very memorable one.

Dedication

I dedicate this thesis to my late parents Mr. R. Narayanan and Mrs R. Kalavati, who were instrumental in shaping my science career and also the successful pursuit at every stage of my education.

Table of Contents

List of Tables

List of Figures

Chapter 1

Background and Motivation

1.1 Introduction

Proteins constitute one of the most complex classes of biomolecules that carry out a diverse range of vital function in all cells. Proteins are involved in almost all of the cellular functions including signaling, catalysis, molecular transport, structural support and movement etc. To perform their function, proteins self-assemble into unique three-dimensional folded conformations, facilitated by non-covalent interactions including electrostatic, Van der Waals, hydrophobic and hydrogen bonding interactions.

A majority of proteins typically fold into stable structural motifs corresponding to the native conformation. Occasionally, proteins tend to fold incorrectly, leading to their misfolding. Protein misfolding leads to the exposure of aggregation-prone hydrophobic residues, which are normally buried in the native conformation, consequently promoting protein aggregation (Figure1.1 Schematic representation of protein misfolding and aggregation, leading to disease. Electron microscopy images were taken from Dobson, Protein folding and misfolding (Dobson 2003).). A variety of human diseases are associated with protein misfolding where proteins convert from their normally soluble form to aggregates that accumulate in affected organs. The final forms of the aggregates in all these diseased states include fibrillar plaques known as amyloid.

Intrinsically Disordered Proteins are implicated in a number of neurodegenerative diseases including Alzheimer's, Parkinson's and Huntington's disease, Prion diseases,

and type II diabetes (Dobson 1999; Lansbury 1999; Perutz 1999; Chiti and Dobson 2006). IDPs are best described as fluctuating ensembles of conformations in solution lacking stable structure under physiological conditions (Dyson and Wright 2005) and are typically characterized by their low sequence complexity, high net charge and low hydrophobicity (Uversky, Gillespie et al. 2000).
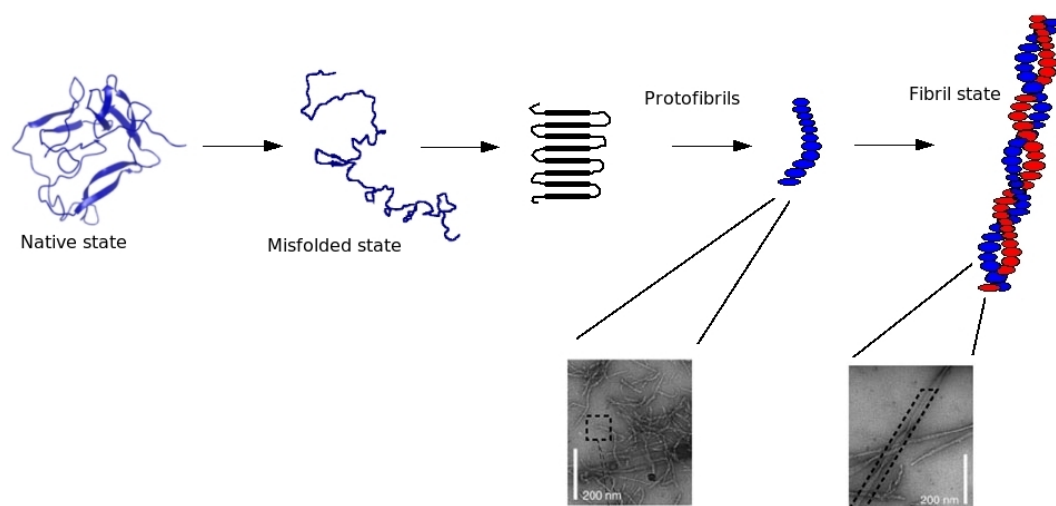
Figure1.1 Schematic representation of protein misfolding and aggregation, leading to disease. Electron microscopy images were taken from Dobson, Protein folding and misfolding (Dobson 2003).

Parkinson's disease (PD) is the second most prevalent of the late onset neurodegenerative diseases. α-synuclein is the primary protein component of the Lewy body deposits that are the diagnostic hallmark of PD (Spillantini, Schmidt et al. 1997). α-synuclein is a highly conserved, intrinsically disordered protein (IDP) that is expressed in

high levels in the brain (Rochet and Lansbury 2000). The monomeric conformation of α-synuclein is described as a 140-residue archetypal IDP. The sequence of α-synuclein has an uneven distribution of charged residues along the chain and is divided into 3 regions – the N-terminal domain (residues 1-60) with a balanced distribution of positively and negatively charged residues corresponding to a polyampholyte chain, the Non Amyloid-beta Component of Alzheimer's disease (NAC, residues 61-95), which forms the hydrophobic core of the protein having minimal charged residues and the highly acidic C-terminal domain (residues 96-140) with a predominance of negative charges characteristic of a polyelectrolyte chain.

α-synuclein can adopt different conformations under various conditions. The N-terminal region of α-synuclein has been shown to form amphipathic helices upon binding to lipid surfaces (Eliezer, Kutluay et al. 2001). Recent reports also suggest that the physiological form of α-synuclein might actually be a stable helical tetramer (Bartels, Choi et al. 2011; Wang, Perovic et al. 2011). A conformational change to the monomeric form of α-synuclein leading to its aggregation into fibrils has been shown to play a key role in the pathology of Parkinson's disease (Spillantini, Schmidt et al. 1997; Goedert 2001; Moore, West et al. 2005). Characterizing the monomeric form is important for understanding the conformational changes leading to the aggregated state.

Conformational characterization of IDPs like α-synuclein is challenging because of their highly flexible nature and the resulting lack of a stable 3D structure. NMR observations and computational simulations have been routinely used to characterize the conformational characteristics of intrinsically disordered proteins (Dyson and Wright

2004; Eliezer 2009). Characterizing the conformations of the monomeric state of IDPs including α-synuclein is important for understanding the conformational changes leading to aggregation.

Peptides in solution, like intrinsically disordered proteins, are best characterized as conformational ensembles. Consequently, peptides can serve as useful test systems for examining methods for combining simulation and experiments to structurally characterize polypeptide ensembles. Chapters 3 and 4 of the thesis explore the approaches to integrating NMR and computation using minipeptide systems to study peptide conformational ensembles. The first peptide system, GB1, is comprised of the C-terminal β-hairpin of the B1 domain of G-protein, which has been shown by NMR and fluorescence measurements to form a β-hairpin in solution (Blanco, Rivas et al. 1994; Munoz, Thompson et al. 1997). Conformational ensembles of this peptide were generated using the AGB-NP implicit solvent model and temperature Replica Exchange Molecular Dynamics (270-690 K). Results show that peptide ensembles in the middle range of temperatures near 400 K provide the best fit to the low temperature (278 K) experimental NMR data, thereby providing a set of models for visualizing the heterogeneity present in the experimental ensemble and also an approach that can be used to help calibrate the effective potential. The second minipeptide study explores the correlation between Cα chemical shifts and helical content of the S-peptide ensembles. Results show that while in general there is a correspondence between helicity and the chemical shift deviation (CSD), on the edges of the helical segments, CSD can under- and over-estimate the percent helicity of the polypeptide ensemble.

Chapters 8 and 9 of the thesis investigate the conformational properties of α-

synuclein under a variety of conditions including temperature, electrostatics and the sequence identity. In chapter 8, a combination of computational simulations and polymer theory are used to characterize the statistical properties of α-synuclein under different solvent conditions, using temperature as a modulator for solvent quality. Results indicate that at extremes of solvent quality, α-synuclein scales as expected for homopolymer chains while the effect of sequence identity impacts the scaling at intermediate solvent conditions.

Rate of fibrillation of α-synuclein is highly dependent on the sequence identity and solution conditions (Conway, Harper et al. 2000; Conway, Lee et al. 2000). The primary sequence shows an uneven distribution of charged residues with a net charge of -9 at neutral pH. Change in pH from neutral to low pH has been shown to accelerate fibrillation of α-synuclein. The second study described in chapter 9 investigates the effect of pH on the conformational properties of α-synuclein by comparing the conformational characteristics at neutral and low pH conditions. The results suggest that there is a significant structural reorganization in terms of the long-range intra-chain contacts with change in pH from neutral to low pH conditions. The structural changes observed at low pH highlight the effect of very different charge distributions in the N, NAC and C-terminal domains on the structural reorganization observed with change in pH.

In the studies presented, NMR and computational simulation approaches have been integrated to characterize and explicitly visualize conformational properties of partially stable minipeptides and the intrinsically disordered ensemble of α-synuclein under different environmental conditions.

1.2 References

Bartels, T., J. G. Choi, et al. (2011). "alpha-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation." Nature **477**(7362): 107-110.

Blanco, F. J., G. Rivas, et al. (1994). "A short linear peptide that folds intoa native stable β-hairpin in aueous solution." Structural Biology **1**(9): 584-590.

Chiti, F. and C. M. Dobson (2006). "Protein misfolding, functional amyloid, and human disease." Annu Rev Biochem **75**: 333-366.

Conway, K. A., J. D. Harper, et al. (2000). "Fibrils formed in vitro from alpha-synuclein and two mutant forms linked to Parkinson's disease are typical amyloid." Biochemistry **39**(10): 2552-2563.

Conway, K. A., S. J. Lee, et al. (2000). "Accelerated oligomerization by Parkinson's disease linked alpha-synuclein mutants." Ann N Y Acad Sci **920**: 42-45.

Dobson, C. M. (1999). "Protein misfolding, evolution and disease." Trends Biochem Sci **24**(9): 329-332.

Dobson, C. M. (2003). "Protein folding and misfolding." Nature **426**: 884-890.

Dyson, H. and P. Wright (2005). "Intrinsically Unstructured Proteins and their functions." Mol. Cell Biol. **6**: 197-208.

Dyson, H. J. and P. E. Wright (2004). "Unfolded proteins and protein folding studied by NMR." Chem Rev **104**(8): 3607-3622.

Eliezer, D. (2009). "Biophysical characterization of intrinsically disordered proteins." Curr Opin Struct Biol **19**(1): 23-30.

Eliezer, D., E. Kutluay, et al. (2001). "Conformational properties of alpha-synuclein in its free and lipid-associated states." J Mol Biol **307**(4): 1061-1073.

Goedert, M. (2001). "Alpha-Synuclein and Neurodegernative Diseases." <u>Nature</u> **2**: 492-501.

Lansbury, P. T., Jr. (1999). "Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease." <u>Proc Natl Acad Sci U S A</u> **96**(7): 3342-3344.

Moore, D. J., A. B. West, et al. (2005). "Molecular pathophysiology of Parkinson's disease." <u>Annu Rev Neurosci</u> **28**: 57-87.

Munoz, V., P. A. Thompson, et al. (1997). "Folding Dynamics and Mechanism of β-haipin Formation." <u>Nature</u> **390**: 196-199.

Perutz, M. F. (1999). "Glutamine repeats and neurodegenerative diseases: molecular aspects." <u>Trends Biochem Sci</u> **24**(2): 58-63.

Rochet, J. C. and P. T. Lansbury, Jr. (2000). "Amyloid fibrillogenesis: themes and variations." <u>Curr Opin Struct Biol</u> **10**(1): 60-68.

Spillantini, M. G., M. L. Schmidt, et al. (1997). "Alpha-synuclein in Lewy bodies." <u>Nature</u> **388**(6645): 839-840.

Uversky, V. N., J. R. Gillespie, et al. (2000). "Why are "natively unfolded" proteins unstructured under physiologic conditions?" <u>Proteins</u> **41**(3): 415-427.

Wang, W., I. Perovic, et al. (2011). "A soluble alpha-synuclein construct forms a dynamic tetramer." <u>Proceedings of the National Academy of Sciences of the United States of America</u> **108**(43): 17797-17802.

Chapter 2

Conformational Sampling using Replica Exchange Molecular Dynamics Simulations

2.1 Introduction

The energy landscape of proteins is characterized by rugged surfaces with conformational multiple minima. The barriers surrounding the low energy minima make it difficult to achieve adequate conformational sampling using the traditional molecular dynamics approaches. These simulations often get trapped in one of a large number of local energy minima posing a major challenge for simulation approaches. To overcome this problem, Sugita and Okamoto introduced the replica exchange molecular dynamics method (REMD) based on a new generalized ensemble algorithm (Sugita and Okamoto 1999).

REMD is an advanced sampling technique used for enhanced sampling of the rough energy landscapes of biomolecules such as proteins where random walks are performed in energy space. This allows the simulations to sample more conformational space and pass energy barriers compared with traditional MD approaches. This technique is a result of application of the parallel tempering method to molecular dynamics simulations.

Comparison of REMD with regular MD simulations has shown that REMD samples about five times more conformational space within the same time scales than MD (Sanbonmatsu and Garcia 2002). Another important advantage is that REMD provides estimates of free energies since it samples from the canonical ensemble over each of the different simulation temperatures.

Replica exchange sampling (Figure 2.1) involves running independent simulations of systems with similar potential energies simultaneously over a specified temperature range. Each simulation is called a replica and each conformation that is simulated at the different temperatures is termed a walker. For a system of N atoms, the Hamiltonian of the system H(X) (X represents the state of the system or the point in phase space) is the sum of the potential and kinetic energies. In the NVT ensemble (or canonical) ensemble, each state (x) with the Hamiltonian H(q,p) is weighted by a Boltzmann factor as

$$W_B(x;T) = e^{-\beta H(q,p)}$$

Here, q and p represent the coordinate and momentum vectors respectively, β corresponds to the inverse temperature.

Since the replicas (temperatures) are non-interacting, the weight factor for state X is given by the product of the Boltzmann factors. For the convergence of the exchange between replicas, satisfying the detailed balance condition for the transition probability w(X- X') is sufficient. Here, P_{eq} represents the equilibrium probability of state X.

$$P_{eq}(X,T)w(X \to X') = P_{eq}(X',T)w(X' \to X)$$

Rescaling of the momenta ($p'_i = \sqrt{(T_j/T_i)p_i}$) following exchange of temperatures results in the kinetic energies to cancel out. Periodic exchange of walkers between replicas at adjacent temperatures is possible with this exchange being accepted based on the Metropolis criterion.

$$P(\{T_i,T_j\} \to \{T_j,T_i\}) = \min(1, \exp[-(\beta_j - \beta_i)(U_i - U_j)])$$

where the inverse temperature $\beta_{i(j)} = 1/KT_{i(j)}$, and $E_{i(j)}$ is the potential energy of the i_{th} (j_{th}) replica. The acceptance probability ratio is dependent on the temperature differences between the adjacent replicas. Exchange between replicas is accepted if $U_i \geq U_j$ (when T_i

< T_j). The exchange probabilities will be zero if the energy distributions of the adjacent replicas do not overlap with each other.

The exchange between replicas at different temperatures provides better sampling of the conformational energy landscapes of molecules for shorter simulation time scales. The choice of the temperature range and the interval between adjacent temperatures is crucial since this affects the acceptance probability for exchange of walkers between replicas. An acceptance ratio of 15-20% for exchanges across all temperatures indicates a valid choice of temperatures.
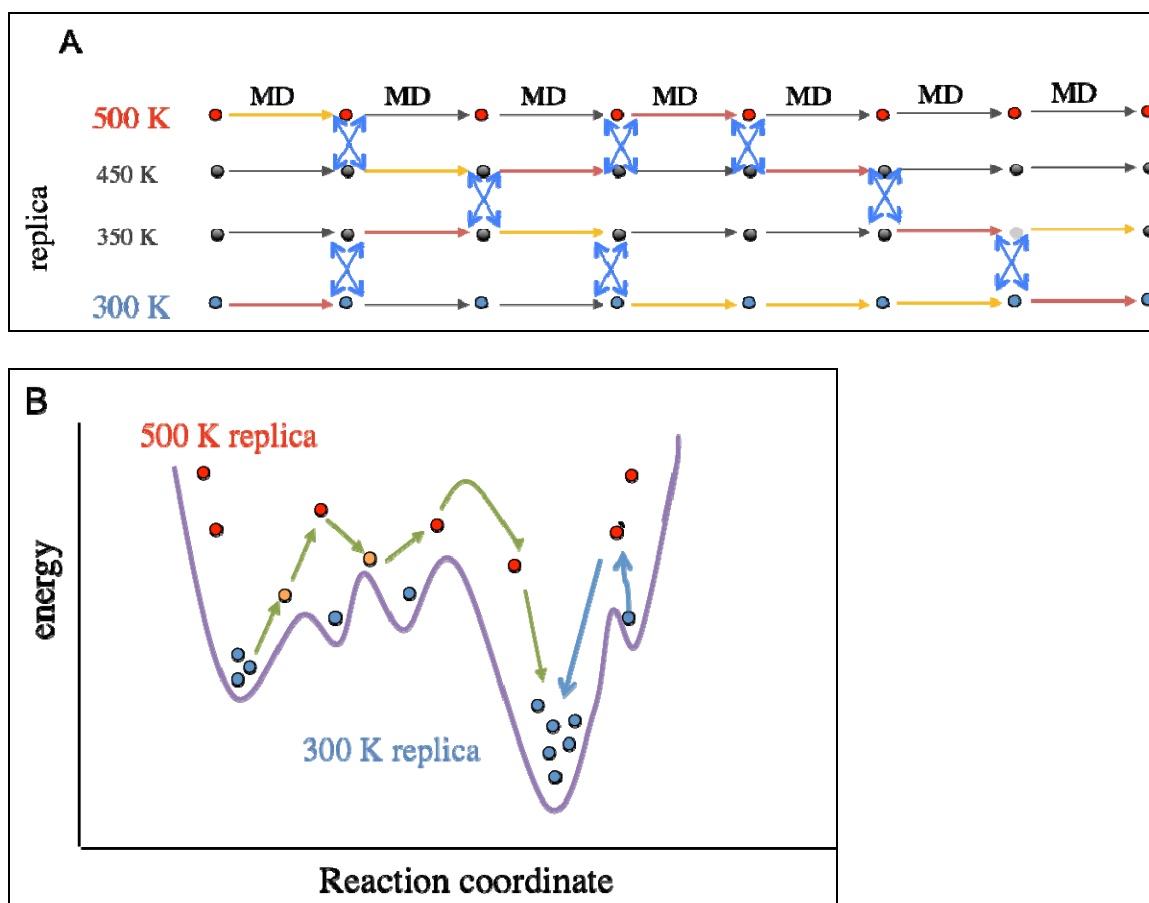
Figure 2.1 (A) Schematic representation of the replica exchange molecular dynamics method. (B) REMD allows the exchange of low temperature replicas that are trapped in

local minima and exchange with other replicas to overcome energy barriers and move towards the global minimum.

2.2 Advantages and potential issues of the REMD method

In the REMD approach, the move upward of one replica is coupled with the move downward of another replica. The probability weight factors are calculated from this coupled exchange as a product of the Boltzmann factors for each temperature or replica. This provides a significant advantage as it does not require any initial simulations for determining the weight factors that are required for the generalized ensemble algorithm (Hansmann and Okamoto 1997). Another major advantage of the REMD approach is the enhanced exploration of the conformational landscapes of biomolecules. REMD samples more conformational space than MD under comparable simulation times.

One of the problems of the replica exchange approach is that the number of temperatures required is dependent on the number of replicas, which in turn is largely influenced by the size of the system (number of atoms). Larger systems therefore require more replicas. The efficiency of the replica exchange approach is another important challenge – a number of studies have addressed this issue (Zheng, Andrec et al. 2007; Nadler and Hansmann 2008; Sindhikara, Meng et al. 2008; Zhang and Ma 2008; Zheng, Andrec et al. 2008). It was shown that the efficiency of complex systems is limited by the timescale of conformational relaxation within the free energy basins (Zheng, Andrec et al. 2007; Zheng, Andrec et al. 2008).

The use of implicit solvent models to simulate protein structures greatly reduces the system size and therefore enhances the efficiency of the simulation. Implicit solvent

models have been used successfully to model minipeptides and small IDPs (Weinstock, Narayanan et al. 2007; Weinstock, Narayanan et al. 2008; Ganguly and Chen 2009; Vitalis and Pappu 2009; Mao, Crick et al. 2010). Replica exchange simulations using implicit solvent models have been used to study intrinsically disordered proteins (Jonsson, Mohanty et al. 2012; Zhang, Ganguly et al. 2012).

2.4 REMD parameters used in this study

The REMD algorithm, implemented in the IMPACT simulation package (Banks, Beard et al. 2005) was employed to generate the conformational ensembles of the two minipeptides – GB1-peptide and S-peptide; and the intrinsically disordered α-synuclein. Simulations were performed using the AGBNP implicit solvent model (Gallicchio and Levy 2004; Gallicchio, Paris et al. 2009) and the OPLS-AA force field (Jorgenson, Maxwell et al. 1996; Kaminsky, Friesner et al. 2001). All simulations were initiated with a fully extended conformation of the peptide and protein systems. The simulations typically start with a short minimization, using the conjugate gradient method, followed by a production run for a total of 10-25 ns (for the different systems studied) over a range of 20 temperatures. The molecular simulation time-step was set to 1.5 fs and exchanges were attempted every 1000fs (1ps).

2.5 References

Banks, J. L., H. S. Beard, et al. (2005). "Integrated Modeling Program, Applied Chemical Theory (IMPACT)." Journal of Computational Chemistry **26**: 1752-1780.

Gallicchio, E. and R. M. Levy (2004). "AGBNP: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling." Journal of Computational Chemistry **25**(4): 479-499.

Gallicchio, E., K. Paris, et al. (2009). "The AGBNP2 Implicit Solvation Model." Journal of Chemical Theory and Computation **5**(9): 2544-2564.

Ganguly, D. and J. Chen (2009). "Atomistic details of the disordered states of KID and pKID. Implications in coupled binding and folding." Journal of the American Chemical Society **131**(14): 5214-5223.

Hansmann, U. H. E. and Y. Okamoto (1997). "Generalized-ensemble Monte Carlo method for systems with rough energy landscape." Physical Review E **56**(2): 2228-2233.

Jonsson, S. A., S. Mohanty, et al. (2012). "Distinct phases of free alpha-synuclein-a monte carlo study." Proteins-Structure Function and Genetics.

Jorgenson, W. L., D. S. Maxwell, et al. (1996). "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids." J. Am. Chem. Soc. **118**(45): 11225-11236.

Kaminsky, G., R. Friesner, et al. (2001). "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.

Mao, A. H., S. L. Crick, et al. (2010). "Net charge per residue modulates conformational ensembles of intrinsically disordered proteins." Proc Natl Acad Sci U S A **107**(18): 8183-8188.

Nadler, W. and U. H. E. Hansmann (2008). "Optimized explicit-solvent replica exchange molecular dynamics from scratch." Journal of Physical Chemistry B **112**(34): 10386-10387.

Sanbonmatsu, K. Y. and A. E. Garcia (2002). "Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics." Proteins-Structure Function and Genetics **46**(2): 225-234.

Sindhikara, D., Y. L. Meng, et al. (2008). "Exchange frequency in replica exchange molecular dynamics." Journal of Chemical Physics **128**(2).

Sugita, Y. and Y. Okamoto (1999). "Replica-exchange molecular dynamics method for protein folding." <u>Chemical Physics Letters</u> **314**: 141-151.

Vitalis, A. and R. V. Pappu (2009). "ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions." <u>J Comput Chem</u> **30**(5): 673-699.

Weinstock, D. S., C. Narayanan, et al. (2008). "Correlation Between 13C-alpha Chemical Shifts and Helix Content of Peptide Ensembles." <u>Protein Science</u>.

Weinstock, D. S., C. Narayanan, et al. (2007). "Distinguishing among structural ensembles of the GB1 peptide: REMD simulations and NMR experiments." <u>J Am Chem Soc</u> **129**(16): 4858-4859.

Zhang, C. and J. P. Ma (2008). "Comparison of sampling efficiency between simulated tempering and replica exchange." <u>Journal of Chemical Physics</u> **129**(13).

Zhang, W., D. Ganguly, et al. (2012). "Residual structures, conformational fluctuations, and electrostatic interactions in the synergistic folding of two intrinsically disordered proteins." <u>PLoS Computational Biology</u> **8**(1): e1002353.

Zheng, W. H., M. Andrec, et al. (2007). "Simulating replica exchange simulations of protein folding with a kinetic network model." <u>Proceedings of the National Academy of Sciences of the United States of America</u> **104**(39): 15340-15345.

Zheng, W. H., M. Andrec, et al. (2008). "Simple continuous and discrete models for simulating replica exchange simulations of protein folding." <u>Journal of Physical Chemistry B</u> **112**(19): 6083-6093.

Chapter 3

Distinguishing among structural ensembles of the GB1 peptide: REMD simulations and NMR Experiments

3.1 Introduction

Peptides in solution, like intrinsically disordered proteins, are best characterized as conformational ensembles. Consequently, peptides can serve as useful test systems for examining methods for combing simulation and experiments to structurally characterize polypeptide ensembles. Structural characterization of these dynamic systems is critical to understanding the principles of protein folding and to understand the basis of protein misfolding that results in proteins aggregation and disease. One approach to characterizing these conformational ensembles is to integrate computational approaches with NMR data (Lindorff-Larsen, Kristjansdottir et al. 2004).
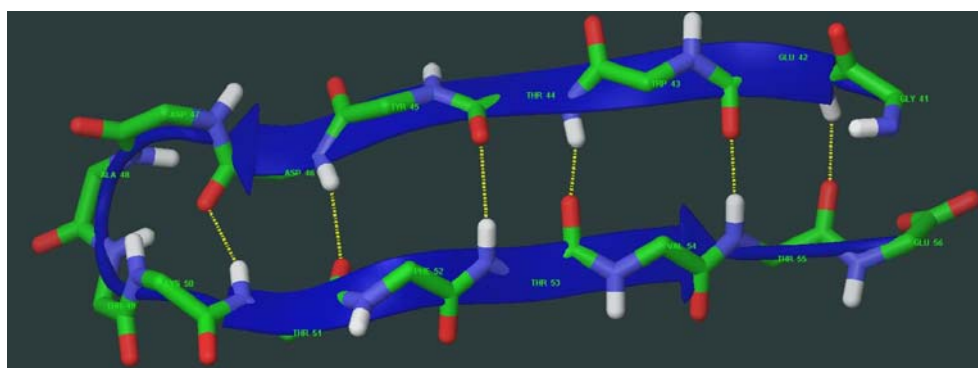
Figure 3.1 Representative conformation of GB1 peptide

GB1 peptide (Figure 3.1) is a small dynamic peptide system, which has been studied extensively using both experimental and computational approaches (Munoz, Thompson

et al. 1997; Ferrara, Apostolakis et al. 2000; Zagrovic, Snow et al. 2002; Gnanakaran, Nymeyer et al. 2003; Felts, Harano et al. 2004a). It is comprised of the C-terminal β-hairpin of the B1 domain of G-protein, which has been shown by NMR and fluorescence measurements to form a β-hairpin in the intact protein (Gronenborn, Filpula et al. 1991) and also in the isolated form in solution (Blanco, Rivas et al. 1994; Munoz, Thompson et al. 1997). Blanco et. al., determined the population of β-hairpin to be ~42% at 273K based on NMR measurements. The GB1 peptide has been used in a number of simulation studies for a variety of reasons including the determination of the free energy landscape of the peptide, and also to test the effectiveness of the implicit solvent model for simulations (Garcia and Sanbonmatsu 2001; Zhou, Berne et al. 2001; Felts, Harano et al. 2004a).

In this study, replica exchange molecular dynamics simulation of the GB1 peptide has been performed in which peptide ensembles were generated over a series of 20 temperatures. The peptide ensembles ranged from being most ordered at 270K to being most disordered at 690K (Felts, Harano et al. 2004b; Andrec, Felts et al. 2005). The structural ensembles obtained over each of the 20 temperatures were compared with experimental NMR measurements including HA and HN chemical shifts, J scalar couplings and NOEs obtained for the peptide at 278K, to determine which of the simulation ensembles best fit the experimental data.

NMR parameters back calculated from the simulation ensembles were compared with those experimental parameters described above all show a clear minimum in the intermediate temperature range of 381 - 442 K. These results indicate that the intermediate temperature ensembles are most similar to the experimental ensemble. The

identification of ensembles best fitting to experiment provides a set of models for visualizing the heterogeneity present in the experimental ensemble, in addition to an approach that can be utilized to calibrate the effective potential.

3.2 Methods and Results

The procedures and results for this study are presented as a reprint of the paper published in the *Journal of American Chemical Society* 2007, 129: 4858-4859.

3.3 Publication Reprint

Distinguishing among Structural Ensembles of the GB1 Peptide: REMD Simulations and NMR Experiments

Daniel S. Weinstock, Chitra Narayanan, Anthony K. Felts, Michael Andrec, Ronald M. Levy,*
Kuen-Phon Wu, and Jean Baum*

*BIOMAPS Institute and Department of Chemistry & Chemical Biology, Rutgers University,
610 Taylor Road, Piscataway, New Jersey 08854-8087*

Received October 30, 2006; E-mail: levy@lutece.rutgers.edu; jean.baum@rutgers.edu

Unlike globular proteins with stable native structures, peptides in solution, partially folded proteins, and natively unfolded proteins are best characterized as conformational ensembles of rapidly inter-converting structures. Structural characterization of these dynamic systems is critical to understand principles of protein folding and to understand the basis of protein misfolding that results in protein aggregation and disease. NMR provides a method for the structural description of peptides and proteins in solution; however, observables such as NOEs and chemical shifts reflect averages of the properties of individual conformations, and the structural characterization of disordered systems cannot be done using conventional static constraints typically used for globular proteins.

One approach to the characterization of these conformational ensembles is to integrate NMR and computational strategies.[1a] A number of computational methods have been suggested based on molecular dynamics with time- or ensemble-averaged NMR restraints introduced as an additional term in the effective potential used for structure refinement.[1a−e] If the effective potentials are accurate enough and the sampling is extensive enough, it should not be necessary to add NMR restraint terms to the potential function to generate structural ensembles which fit the NMR data. However, there have only been a few comparisons published of the results of "unbiased" MD simulations of peptide ensembles and corresponding NMR experiments.[2a,b] We recently reported the results of replica exchange molecular dynamics (REMD) simulations of GB1 peptide,[3a,b] in which 20 peptide ensembles were generated as a function of temperature, spanning a range from the most ordered ensemble at 270 K to the most disordered at 690 K. In this communication, we compare each of these structural ensembles with experimental NMR data we have recorded for the peptide. We show that the peptide ensembles in the middle range of temperatures near 400 K provide the best fit to the low temperature (278 K) experimental NMR data, thereby providing a set of models for visualizing the heterogeneity present in the experimental ensemble and also an approach that can be used to help calibrate the effective potential.

Ensembles of peptide conformations were generated using REMD,[4] the OPLS-AA/AGBNP effective potential[5] within the IMPACT[6] molecular simulation package. The current generation of implicit solvent models, including OPLS-AA/AGBNP, is not parametrized to accurately model temperature effects and tends to predict overly structured molecules at low temperature,[3a] thus leaving open the question of which of the ensembles best matches the experimental chemical shifts, *J*-couplings (Table S1), and NOE NMR parameters (Figure S1) obtained on the GB1 peptide at pH 7.0 and 278 K.

The C-terminal β-hairpin of the B1 domain of G protein is a small dynamic system. This peptide has been shown by NMR and fluorescence measurements to form a β-hairpin in solution.[7a,b] One measure of the extent of hairpin formation within the peptide
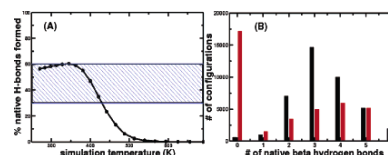
Figure 1. (A) Percentage of native β hydrogen bonds formed relative to the total number possible. (B) Distributions of configurations with 0−6 native β hydrogen bonds in the 270 K (black) and 421 K (red) ensembles.
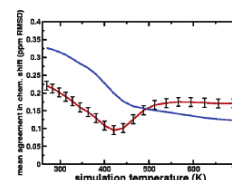
Figure 2. Mean agreement in HA chemical shifts for each simulation ensemble relative to the experimental shifts at 278 K (red curve) and random coil (blue curve). See Supporting Information for details.

ensemble is the percent of native H bonds formed, where native refers to the H bonds formed in the β-hairpin within the intact protein. The REMD ensembles, calculated as a function of temperature, indicate that the system is 35−60% β-hairpin over temperatures ranging from 270 to 421 K (Figure 1a). All of the peptide ensembles below 421 K are consistent with previous estimates of percent β-hairpin derived from NMR chemical shift analysis as well as computer simulations.[7a,8a−c] However, detailed analysis of two of these ensembles (270 and 421 K) shows that the populations of the six native H bonds within the peptide have very different H-bond distributions (Figure 1b). In the low temperature 270 K ensemble, 99% of the structures have at least one native H bond formed. In contrast, the higher temperature ensemble shows a very heterogeneous distribution of H bond populations, with approximately 50% of the structures having no native H bonds.

To determine which ensembles within the range of temperatures from 270 to 690 K best fit the experimental data, comparisons are made between predicted and measured values of HA (Figure 2) and HN (Figure S2) chemical shifts and J_{HAHN} scalar couplings (Figure S3). At each simulation temperature, chemical shifts and scalar couplings are calculated for each of the residues in each of the 40 000 structures constituting an ensemble using SHIFTX[9] for the chemical shifts and a Karplus relationship for the *J*-couplings.[10] The HA and HN chemical shift and J_{HAHN} scalar coupling plots all have minima within the range of temperatures from 381 to 442 K, indicating that the simulation ensembles in this middle temperature range are the most similar to the ensemble represented by the NMR
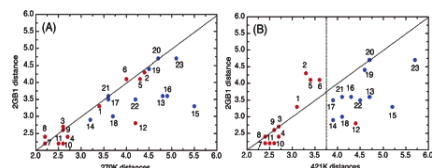
Figure 3. Backbone–backbone NOEs (obtained from the BMRB) observed in the G-protein which are either observed (red) or unobserved (blue) in the GB1 peptide, plotted with the distance in the protein (pdb structure 2gb1) versus the average distance in the low temperature (A, 270 K) and a representative middle range temperature (B, 421 K) simulation ensembles. NOE list is provided in Supporting Information.

data. Additionally, these ensembles are substantially different from those near 270 K and those near 690 K. The correspondence between the simulation ensembles in this middle range and the low temperature experimental results is further supported by comparing the temperature dependence of predicted and experimental HN chemical shifts.[11a,b] The experimental trends are most closely reproduced by matching the predicted temperature-dependent HN chemical shifts for the ensembles in the middle of the simulation temperature range to the experimental low temperature values (Figure S4), supporting the correlation between the experimental system and the heterogeneous hydrogen bonding pattern described by the simulations.

Comparison of distances derived from NOEs provides another approach for selecting the ensembles which best fit the NMR data at low temperature. The NOEs that are included (Figure 3a,b) are those observed for residues 41–56 in the intact G-protein,[12] with the subset seen in the peptide NMR experiments labeled in red. Distances derived from these NOEs, taken from the NMR structure of the protein, are plotted as a control against predicted interproton distances from two ensembles: one at low temperature (270 K) and one from the middle simulation temperature range (421 K). Predicted NOEs are obtained from the simulations by assuming the members of the ensembles interconvert very rapidly and by averaging $\langle r^{-3} \rangle$ over the ensembles.[13a,b] At low temperature, most of the points lie along the diagonal, indicating that the conformational ensemble at 270 K is similar to the NMR structure obtained for the β-hairpin region in the G-protein. However, the peptide experimental NOEs observed at 278 K are not consistent with either the NMR structure of the G-protein or the simulated ensemble at low temperature; certain short- and long-range NOEs are observed, but many expected short- and long-range NOEs are missing in an unsystematic way.

In contrast, at 421 K, there is a clear systematic separation of NOEs, with all those observed in the peptide at short ensemble-averaged distances. The NOEs that have moved most significantly relative to the 270 K ensemble are sequential HN–HN NOEs corresponding to points 6 (55HN–56HN), 5 (51HN–52HN), and 2 (45HN–46HN). Neighboring amide protons in the strand of a β-hairpin are constrained to be maximally separated, and any change in conformation will bring these protons closer together. The opposite effect takes place for HA$_i$–HN$_{i+1}$ protons in β-strands, which are constrained to be close. Therefore, the introduction of new non-hairpin populations will significantly affect the predicted distances of the sequential amide protons while not greatly changing the averaged distances between sequential HA–HN pairs due to the $\langle r^{-3} \rangle$ averaging, which is biased toward short distances.

That only short interproton distances are observed in the peptide is due to the fact that the small size of the peptide gives rise to a smaller rotational correlation time relative to the same residues in protein G leading to a reduction in the NOE intensity. This smaller NOE intensity in turn means that, for a given noise level, signals corresponding to long distances are lost in the noise. The one NOE that does not follow this pattern is a cross-strand NOE (54HA–44HN) which is observed even though the protons are separated on average by more than 4 Å in the high temperature ensemble. This reflects the fact that, apparently, the simulation does not align the strands of the hairpin as closely as is observed in the peptide.

In conclusion, detailed comparison of REMD simulations to numerous experimental NMR parameters, including HA, HN chemical shifts, J_{HAHN} scalar coupling, and NOEs has provided an approach for identifying the set of simulated ensembles that best match the low temperature experimental data. These ensembles allow the visualization of distinct populations within the conformational ensemble of the GB1 peptide, thereby providing access not only to averages of conformational properties, such as the number of hydrogen bonds, but also to the distributions which make up those averages. What is striking is that the ensembles that best match the experimental NMR data are very inhomogeneous and the fact that they include a large population of conformations with no native hydrogen bonds.

Supporting Information Available: NMR chemical shifts, J-couplings, and NOE parameters; computational details; full ref 6. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Lindorff-Larsen, K.; Kristjansdottir, S.; Teilum, K.; Fieber, W.; Dobson, C. M.; Poulsen, F. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 3291. (b) Choy, W.-Y.; Forman-Kay, J. D. *J. Mol. Biol.* **2001**, *308*, 1011. (c) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1989**, *157*, 289. (d) Best, R. B.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 8090. (e) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 476.
(2) (a) Satoh, D.; Shimizu, K.; Nakamura, S.; Terada, T. *FEBS Lett.* **2006**, *590*, 3422. (b) Zagrovic, B.; Sorin, E. J.; Pande, V. *J. Mol. Biol.* **2001**, *313*, 151.
(3) (a) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins* **2004**, *56*, 310. (b) Andrec, M.; Felts, A. K.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6801.
(4) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
(5) Gallicchio, E.; Levy, R.; Levy, M. *J. Comput. Chem.* **2004**, *25*, 479.
(6) Banks, J. L.; et al. *J. Comput. Chem.* **2005**, *26*, 1752.
(7) (a) Blanco, F. J.; Rivas, G.; Serrano, L. *Struct. Biol.* **1994**, *1*, 584. (b) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196.
(8) (a) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068. (b) Garcia, A. E.; Sonbonmatsu, K. Y. *Proteins* **2001**, *42*, 345. (c) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. *J. Am. Chem. Soc.* **2004**, *126*, 7238.
(9) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J. Biomol. NMR* **2003**, *26*, 215.
(10) Lindeorff-Larsen, K.; Best, R. B.; Vendruscolo, M. *J. Biomol. NMR* **2005**, *32*, 273.
(11) (a) Case, D. A. *Curr. Opin. Struct. Biol.* **2000**, *10*, 197. (b) Baxter, N. J.; Williamson, M. P. *J. Biomol. NMR* **1997**, *9*, 359.
(12) Gronenborn, A. M.; Fipula, D. R.; Essig, N. Z.; Archari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657.
(13) (a) Olejniczak, E. T.; Dobson, C. M.; Karplus, M.; Levy, R. M. *J. Am. Chem. Soc.* **1984**, *106*, 1923. (b) Kessler, H.; Greisinger, C.; Lautz, J.; Muller, A.; van Gunsteren, W. F.; Berendsen, H. J. *J. Am. Chem. Soc.* **1988**, *110*, 3393.

JA0677517

Supporting information

Distinguishing among structural ensembles of the GB1 peptide:
REMD simulations and NMR experiments

Methods and Materials

1) ^{1}H-NMR spectroscopy

   GB1 peptide (Ac-GEWTYDDATKTFTVTE-NH$_2$) NMR samples were prepared by
dissolving lyophilized GB1 peptide directly in a solution of 10 mM phosphate buffer with
10% (v/v) D$_2$O. NMR sample concentrations were 0.3 to 0.5 mM at pH 7.0. All NMR
experiments were performed on a Varian Inova spectrometer operating at a proton
frequency of 600 MHz. TOCSY spectra were acquired with an 80 ms DIPSI2 spin lock
mixing time and WATERGATE solvent suppression from 278K to 318K with 5 degree
increments. 2D NOESY spectra were recorded at 150 ms and 200 ms mixing times with
WATERGATE solvent suppression at 278K. Proton chemical shifts are presented in
TableS1. The summary of intra-, sequential and long-range NOE connectivities is shown
in figure S1.

   [15]N labeled GB1 peptide was expressed in a GST fusion protein and cleaved by
AcTEV protease (Invitrogen Inc.). The sequence (ENLYFQGEWTYDDATKTFTVTE)
containing an AcTEV protease cutting site (cleaving between Q and G) and GB1 peptide
sequence were introduced into pGEX 4T-1 GST expression system. GST-GB1 fusion
protein was over-expressed by 1 mM IPTG induction for 4 hours at 37 °C and was
purified by FPLC (GE Healthcare, Piscataway, NJ) using a GST affinity column. Pure
GST-GB1 fusion protein was digested by AcTEV protease for 3 hours at room
temperature; then GST protein, AcTEV protease and GB1 peptide were quickly separated
by Superdex-100 size exclusion gel filtration column. 0.5 mM ^{15}N labeled GB1 peptide
was prepared for HNHA experiments[1,2]. [3]J$_{HNHA}$ coupling constants were calculated using
the following equation: $I_{cross} / I_{dia} = - \tan^2[\pi J (\Delta1+\Delta2)]$ where I_{cross} and I_{dia} are
intensities of cross and diagonal peaks, respectively; $\Delta1$ and $\Delta2$ are 12.57 and 12.49 ms,
respectively. The measured [3]J$_{HNHA}$ coupling constants were listed in Table S1.

2) Computational Details
Replica Exchange Molecular Dynamics Simulations
   REMD simulations were performed with 20 replicas run in parallel at the following

20 temperatures: 270, 283, 298, 313, 328, 345, 363, 381, 400, 421, 442, 465, 488, 513, 539, 566, 594, 625, 656, and 690 K - for a total of 10 ns for each replica, with a time step of 1 fs. Transitions between adjacent temperatures were attempted every 250 MD steps using a Metropolis transition probability as detailed in our previous work[3,4]. Configurations were saved prior to every attempted transition, leading to an ensemble at each temperature containing 40,000 structures.

Replica exchange sampling was chosen to generate the peptide ensembles due to the method's ability to sample equilibrium distributions more completely than standard MD sampling. Since we are not focused on dynamics of the solvent, we employed an implicit solvent model, which is a more natural fit to replica exchange simulations than an explicit water model. The implicit solvent model we used, AGBNP[5], is based on a novel pairwise descreening implementation of the generalized Born model and a recently proposed non-polar hydration free energy estimator.

Prediction of HA and HN chemical shifts

SHIFTX was used to generate predicted shifts for all the residues in all 40,000 structures in the ensembles at each simulation temperature. SHIFTX[6] is a hybrid predictive technique that makes use of classical or semi-classical calculations for the effects of ring currents, electric field, hydrogen bond and solvent effects combined with empirical hypersurfaces that capture dihedral angle, sidechain, and nearest neighbor effects that cannot be predicted by classical means. The developers of SHIFTX have estimated correlation coefficients and RMS errors of 0.911 and 0.23 ppm for HA and 0.741 and 0.49 ppm for HN shifts. The lower accuracies for HN shifts may be due to an incomplete quantitative understanding of all of the factors which influence the variability of that shift.

Prediction of J-coupling constants

J_{HA-HN} coupling constants were calculated from dihedral angles from each residue in each of the peptide structures through a Karplus relationship
$J(theta) = Acos^2(theta) + Bcos(theta) + C$. The values for the parameters A,B, and C were taken from a recent study of coupling constants using ensembles of proteins conducted by Vendruscolo and coworkers[7].

Comparison of predicted HA and HN chemical shifts and J-coupling constants with

experimental values

Mean agreement of the various physical parameters between simulation ensembles and the experimental measurements , as shown in Figure 2 of the communication for HA chemical shifts and in Figures S2 and S3 for HN chemical shifts and J_{HA-HN} coupling constants, were calculated by 1) using SHIFTX(or the Karplus equation) to generate predicted shifts(coupling constants) for all the residues in all 40,000 structures in the ensembles at each simulation temperature, 2) determining the deviation of those shifts(couping constants) from experimental values for those residues, and then 3)taking the RMSD of the means of those distributions. Error bars in these figures were generated by using standard error propagation methods. They are dominated by experimental error, so the error associated with the random coil curve are not perceptible on the scale presented.

Figure S4 shows a comparison of the temperature dependence of the predicted and experimental HN chemical shifts of individual residues. The experimental temperature range, 278-318K is compared to average values from the simulation ensembles for simulation temperatures 381-561K. 381K is matched with experimental low temperature because that is where the HN rmsd plot has its minimum. The temperature dependence of HN chemical shifts over this range of simulation temperatures matches the trends in the experimental temperature dependence of the chemical shifts much more closely than the range starting from the simulation ensemble at 270K (data not shown).

Quantitation of ^1H-^1H NOEs

^1H-^1H NOE intensity, because it arises from a dipole-dipole interaction is related to r^{-6} ,where r is the interproton distance. In the fast limit, assuming the members of the ensemble interconvert very rapidly (e.g. a time scale of ~100 ps or faster), NOE intensities scale with distance as $<r^{-3}>^2$.[8,9] Therefore, $<r^{-3}>$ averaged interproton distances were calculated over the simulation ensembles when comparing NOE derived distances in the simulation ensembles and in the NMR structure of the G-protein.

There is an angular average involved in the exact calculation of the spectral density and the fast motional averaging order parameter.[8,10-13] However, it is common practice in the field, as documented by a long line of papers in the literature[9,14-19], to ignore the angular averaging effects on the calculation of NOEs from molecular dynamics trajectories.   The rationale is that the effect of fast angular averaging on NOE intensities is likely to be small compared with that of distance averaging. Even if angular averaging

is not ignored, the argument can be made that the effect is likely to be small because it enters as a root sixth power. The order parameter, S^2, introduced by Lipari and Szabo incorporates the angular term. The effect on NOEs enters the calculation as $(S^2)^{-1/6}$. This means that for example, even for values of the order parameter of $S^2 \sim 0.6$, which represent a significant amount of internal angular motion, the order parameter term $(S^2)^{-1/6}$ does not deviate greatly from 1, minimizing the influence of this term on the effective distances extracted from NOE measurements on structural ensembles.

Selection of NOE intensities

All interresidue backbone-backbone NOEs observed in the peptide were included in Figure 3. Backbone-sidechain and sidechain-sidechain NOEs were omitted because of the difficulties in averaging sidechain protons.

residues	H^N	H^α	H^β	H^γ	H^δ	Others	$^3J_{HNHA}$ (Hz)
Gly 41		3.72, 3.82					6.89
Glu 42	8.67	4.40	1.83, 1.97	2.14, 2.20			7.79
Trp 43	8.65	4.91	3.20			$H^{\delta 1}$: 7.15 $H^{\varepsilon 1}$: 10.10 $H^{\varepsilon 3}$: 7.43 $H^{\zeta 2/3}$: 7.20, 7.48 H^{η}: 7.10	6.73
Thr 44	8.52	4.38	1.20				N/A
Tyr 45	8.56	4.33	2.77, 2.86			$H^{\delta 1/2}$: 6.66 $H^{\varepsilon 1/2}$: 6.66	8.56
Asp 46	8.17	4.58	2.47, 2.69				8.00
Asp 47	8.44	4.32	2.65				5.84
Ala 48	8.40	4.25	1.47				5.38
Thr 49	7.80	4.25	4.25	1.17			N/A
Lys 50	8.11	4.08	1.86, 1.93	1.35	1.65	H^{ε}: 2.97	7.16
Thr 51	7.69	4.47	4.12	1.15			8.92
Phe 52	8.61	4.97	2.99			$H^{\delta 1/2}$: 7.21 $H^{\varepsilon 1/2}$: 7.32 H^{ζ}: 7.21	6.80
Thr 53	8.63	4.47	4.06	1.15			8.89
Val 54	8.47	4.21	1.78	0.73, 0.81			7.01
Thr 55	8.47	4.38	4.21	1.2			8.12
Glu 56	8.27	4.13	1.90, 2.05	2.21			7.23

Table S1. Proton chemical shift assignments and $^3J_{HNHA}$ measurements at 278K.

Number	NOE		Number	NOE		Number	NOE
1	47HA-48HN		10	53HA-54HN		19	48HA-50HN
2	45HN-46HN		11	55HA-56HN		20	47HN-49HN
3	47HN-48HN		12	54HA-44HN		21	48HN-50HN
4	48HN-49HN		13	43HA-55HN		22	49HN-51HN
5	51HN-52HN		14	52HA-46HN		23	46HA-51HN
6	55HN-56HN		15	56HA-42HN			
7	45HA-46HN		16	42HN-55HN			
8	46HA-47HN		17	44HN-53HN			
9	50HA-51HN		18	46HA-48HA			

Table S2 A list of the NOEs included in Figure 3 of the paper.

Figure S1. Summary of intra-, sequential and long-range NOE connectivities

Two β-strands are illustrated as arrows according to the protein G B1 structure (PDB: 2GB1). NOE intensities of GB1 peptide were obtained from ^{1}H-^{1}H NOESY experiment. NOEs were classified into three groups, strong, medium and weak, based on the signal strengths and are indicated by the thickness of the lines in this figure. Some weak long-range NOEs (i.e. H^α to side-chain proton or side-chain proton to side-chain proton) were observed and presented. Asterisks are NOEs that are overlapped with other peaks and cause uncertainty in categorizing NOE strengths.
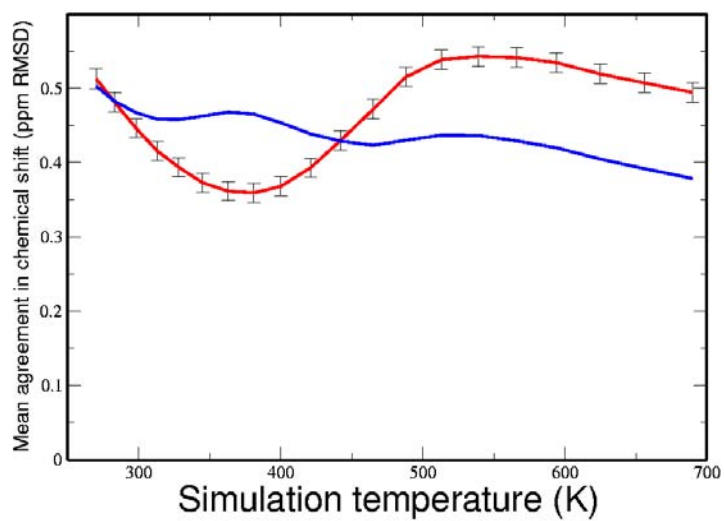
Figure S2 Mean agreement in HN chemical shifts for each simulation ensemble relative to the experimental shifts at 278K(red curve) and random coil(blue curve).
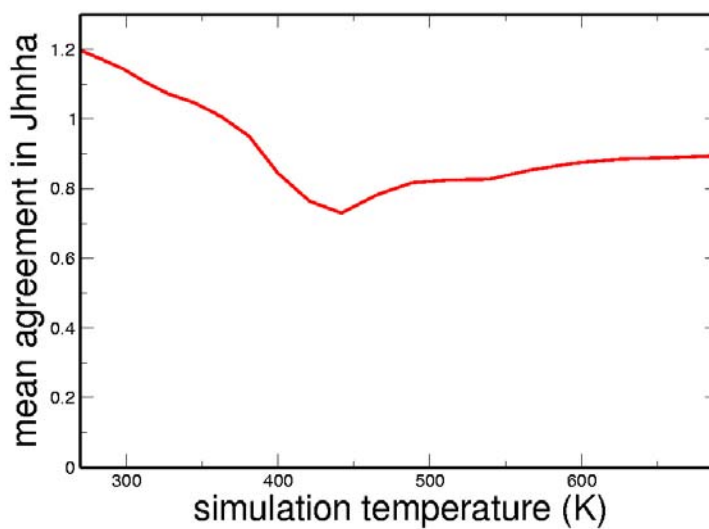
Figure S3 Mean agreement in J_{HAHN} scalar coupling for each simulation ensemble relative to the experimental couplings at 278K.
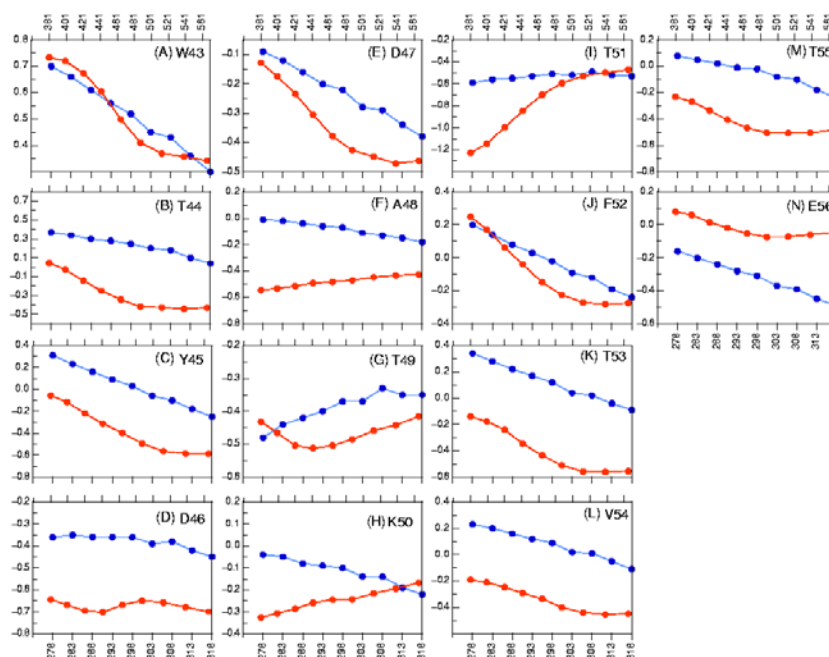
Figure S4 Boxes A to N show the temperature dependent H^N chemical shift deviations of 14 residues (W43 to E56) of GB1 peptide. Experimental NMR data and REMD data calculated from SHIFTX are showing in blue and orange dots in each box, respectively. In each box, Y axis shows the scale of H^N chemical shift deviation (ppm) of the corresponded residue and two X axes are used to present the temperatures of experimental NMR data (278K to 318K, bottom axis) and REMD data (381K to 566K, top axis).

Full Attribution of reference 8 in the manuscript:

(8) J.L. Banks, H.S. Beard, Y. Cao, A.E. Cho, W. Damm, R. Farid, A.K. Felts, T. A. Halgren, D.T. Mainz, J.R. Maple, R. Murphy, D.M. Phillip, M.P. Repasky, L.Y. Zhang, B.J. Berne, R.A. Friesner, E. Gallicchio, and R.M. Levy, *J. Comp. Chem.* 2005, *26*, 1752

References

(1) Vuister, G.W.; Bax, A. *J Am Chem Soc* 1993, 115,7772-7

(2) Grzesiek, S.; Kuboniwa, H.; Hinck, A.P.; Bax, A. *J Am Chem Soc* 1995,117, 5312-5

(3)A.K. Felts, Y. Harano, E. Gallicchio, and R.M. Levy *Proteins* 2004, *56*, 310

(4)M. Andrec, A.K. Felts, E. Gallichio, and R.M. Levy *PNAS* 2005, *102*, 6801

(5) E. Gallicchio and R.M. Levy , *J. Comp. Chem.* 2004, *25*, 479

(6) S. Neal, A.M. Nip, H. Zhang, and D.S. Wishart, *J. Biomol. NMR* 2003, *26*, 215

(7) K. Lindeorff-Larsen, R.B. Best, and M. Vendruscolo *J. Biomol. NMR* 2005, *32*, 273

(8) E.T. Olejniczak, C.M. Dobson, M. Karplus, and R.M. Levy *J. Am. Chem. Soc.* 1984, *106,* 1923

(9) H. Kessler, C. Greisinger, J. Lautz, A. Muller, W.F. van Gunsteren, and H.J. Berendsen *J. Am. Chem. Soc.* 1988, *110,* 3393

(10) J. Tropp *J. Chem. Phys.* 1980, *72*, 6035

(11) G. Lipari and A. Szabo *J. Am. Chem. Soc.* 1982, *104*, 4546

(12) R.M. Levy, M. Karplus and A. McCammon *J. Am. Chem. Soc.* 1981, *103*, 5998

(13) G. Lipari, A. Szabo and R.M. Levy *Nature*, 1982, *300*, 197

(14) B. Richter, J. Gsponer, P. Varnai, X. Salvatella and M. Vendruscolo *J. Biomal NMR* 2007, *37*, 117

(15) W-Y Choy and J.D. Forman-Kay *J. Mol. Biol.* 2001, *308*, 1011

(16) K. L. Constantine, L. Mueller, N.H. Anderson, H. Tong, C.F. Wandler, M.S. Friedrichs, and R.E. Bruccoleri *J. Am. Chem. Soc.* 1995, *117*, 10841

(17) D.A. Pearlman *J. Biomol. NMR* 1994, *4*, 1

(18) A. M. Bonvin and A.T. Brunger *J. Mol. Biol.* 1995, *250*, 80

(19)   P. Cuniasse, I. Raynal, A. Yiotakis and V. Dive *J. Am. Chem. Soc.* 1997, *119*, 5239

3.4 References

Andrec, M., A. K. Felts, et al. (2005). "Protein folding pathways from replica exchange simulations and a kinetic network model." Proceedings of the National Academy of Sciences of the United States of America **102**(19): 6801-6806.

Blanco, F. J., G. Rivas, et al. (1994). "A short linear peptide that folds intoa native stable β-hairpin in aueous solution." Structural Biology **1**(9): 584-590.

Felts, A., Y. Harano, et al. (2004a). "Free-energy surfaces of beta hairpin and alpha-helical peptides generated by Replica Exchange Molecular Dynamics with the AGBNP implicit solvent model." Proteins: Struct., Funct., Bioinf. **56**: 310-321.

Felts, A. K., Y. Harano, et al. (2004b). "Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model." PROTEINS **56**(2): 310-321.

Ferrara, P., J. Apostolakis, et al. (2000). "Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations." Journal of Physical Chemistry B **104**(20): 5000-5010.

Garcia, A. E. and K. Y. Sanbonmatsu (2001). "Exploring the energy landscape of a beta hairpin in explicit solvent." PROTEINS **42**(3): 345-354.

Gnanakaran, S., H. Nymeyer, et al. (2003). "Peptide folding simulations." Current Opinion in Structural Biology **13**(2): 168-174.

Gronenborn, A. M., D. R. Filpula, et al. (1991). "A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G." Science **253**(5020): 657-661.

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

Munoz, V., P. A. Thompson, et al. (1997). "Folding Dynamics and Mechanism of β-haipin Formation." Nature **390**: 196-199.

Zagrovic, B., C. D. Snow, et al. (2002). "Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing." Journal of Molecular Biology **323**(5): 927-937.

Zhou, R., B. J. Berne, et al. (2001). "The free energy landscape for beta hairpin folding in explicit water." Proceedings of the National Academy of Sciences of the United States of America **98**(26): 14931-14936.

Chapter 4

Correlation between 13C chemical shifts and helix content of peptides

4.1 Introduction

NMR parameters provide detailed descriptions of the conformational characteristics of peptide and disordered protein ensembles. Chemical shifts have long been used as primary experimental parameters for the determination of the secondary structure of proteins (Markley, Meadows et al. 1967; Nakamura and Jardetzky 1968; Szilagyi and Jardetzky 1989; Wishart and Sykes 1994) and to nascent secondary structural propensities of disordered polypeptides (Smith, Fiebig et al. 1996; Eliezer, Yao et al. 1998; Dyson and Wright 2001; Marsh, Singh et al. 2006). While the chemical shifts are sensitive to dihedral angles of the backbone, they are insensitive to hydrogen bonding which is an important characteristic for the secondary structure. A number of approaches have been used to characterize the secondary structure using chemical shifts (Eliezer, Yao et al. 1998; Wang and Jardetzky 2002; Marsh, Singh et al. 2006).

Since peptides with propensities to adopt specific secondary structures also exist in solution as conformational ensembles, they serve as good model systems to understand disordered proteins. S-peptide, which constitutes the N-terminal helix of RNase A, was the first peptide shown to adopt transient helical propensity of ~30-40 % under low temperature conditions (Brown and Klee 1969; Bierzynski, Kim et al. 1982; Kim and Baldwin 1984; Nelson and Kallenbach 1989). In this study, an analog of the S-peptide, which adopts a stable helical conformation in solution (Mitchinson and Baldwin 1986), is

used to highlight the relationship between the helicity of the peptide and the chemical shift deviation from random coil. The fact that a valid ensemble will reproduce the ensemble averaged values measured experimentally helps eliminate the inherent ambiguity in the Cα CSDs used for the determination of the secondary structure by explicitly considering the distribution of structures. REMD simulations were used to generate ensembles at low, intermediate and high temperatures. Percent helicities were calculated using STRIDE and compared to that calculated from the Cα CSDs. The results of this study show that while the helicities calculated using the two approaches are largely in quantitative agreement, except at the edges of the helical stretches, where the CSD helicities both under and over estimate the helicities in this region. The underestimation arises due to the averaging of the positive and negative CSDs corresponding to helical and extended conformations respectively, while the overestimation arises because torsion angles in the helical region are not counted by STRIDE because of the lack of the correlated torsion angles of the neighboring residues.

4.2 Methods and Results

The procedures and results for this study are presented as a reprint of the paper published in the *Protein Science*, 2008, 17: 950-954.

4.3 Publication Reprint

Correlation between $^{13}C^{\alpha}$ chemical shifts and helix content of peptide ensembles

DANIEL S. WEINSTOCK, CHITRA NARAYANAN, JEAN BAUM, AND RONALD M. LEVY

BIOMAPS Institute, Rutgers University, Piscataway, New Jersey 08854-8087, USA
Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854-8087, USA

Abstract

Replica exchange molecular dynamics simulations are used to generate three ensembles of an S-peptide analog (AETAAAKFLREHMDS). Percent helicity of the peptide ensembles calculated using STRIDE is compared to percent helicity calculated from $^{13}C^{\alpha}$ chemical shift deviations (CSD) from random coil in order to test the assumption that CSD can be correlated to percent helicity. The two estimates of helicity, one based on structure and the other on CSD, are in close to quantitative agreement, except at the edges of helical stretches where disagreements of as much as 50% can be found. These disagreements can occur by CSDs both as an under- and an overestimate of peptide helicity. We show that underestimation arises due to ensemble averaging of positive CSDs from conformers with torsion angles in the helical region of Ramachandran space with negative CSDs corresponding to conformers of the peptide in the extended region. In contrast, overestimation comes about due to the fact that a large number of conformations with torsion angles in the helical region are not counted as helical by STRIDE due to a lack of correlated helical torsion angles in neighboring residues.

Keywords: chemical shift; peptide ensembles

Supplemental material: see www.proteinscience.org

Chemical shifts have a long history as a tool for determining the secondary structure of proteins (Szilagyi and Jardetzky 1989; Wishart and Sykes 1994a). This methodology was developed to identify secondary structure in native proteins, however, there has increasingly been interest in attempting to find regions of inherent structure within disordered polypeptide systems (Smith et al. 1996; Eliezer et al. 1998; Dyson and Wright 2001; Marsh et al. 2006). In this study we investigate the correlation between ensemble averaged chemical shift deviations (CSD) from random coil and peptide helicity. We demonstrate quantitatively that CSD is a good measure of percent helicity, but that this correspondence can break down at the edges of continuous stretches of helical residues.

A variety of methods have been proposed for characterizing secondary structure using chemical-shift data (Eliezer et al. 1998; Wang and Jardetzky 2002; Marsh et al. 2006). We chose to focus on $^{13}C^{\alpha}$ chemical shifts, which are known to be sensitive mainly to the backbone structure of a polypeptide chain (Spera and Bax 1991). The deviation of the $^{13}C^{\alpha}$ chemical shifts from random coil values can be used as a measure of the population of torsion angles in the α or β regions of Ramachandran space (Eliezer et al. 1998). Small peptides capable of folding into regular secondary structure can serve as useful models for unstructured proteins, since they too exist

in solution as ensembles of inter-converting conformers. Here we use one such peptide, an analog of the S-peptide (Mitchinson and Baldwin 1986) which forms a more stable helix in solution, to illuminate the relationship between CSD and the helicity of the peptide sequence. Using computer simulations, we show that while in general there is a correspondence between helicity and CSD, on the edges of helical segments, CSD can under- and overestimate the percent helicity of the polypeptide ensemble. This is possible both at the ends of the polypeptide sequence and at the edges created when short helical segments are broken by stretches of coil, a motif potentially quite common in natively unstructured proteins.

Materials and Methods

Ensembles of peptide conformations were generated as a function of temperature, ranging from 270 to 690 K, using replica exchange molecular dynamics (REMD) (Sugita and Okamoto 1999) and the OPLS-AA/AGBNP implicit solvent effective potential (Gallicchio and Levy 2004) within the IMPACT molecular simulation package (Banks et al. 2005). The large temperature range used is due to the fact that the helix melts at higher temperatures than observed experimentally because the current generation of implicit solvent models are not parameterized for temperature. This is of little consequence for this study because our goal is to generate ensembles with differing amounts of peptide helicity. Ensemble averaged CSDs were calculated for each ensemble by using SHIFTX (Neal et al. 2003) to calculate chemical shifts for the residues of each peptide conformation, subtracting a reference random coil chemical shift value, and then averaging over all the conformations in the ensemble. For internal consistency in the calculation of CSDs we used simulated random coil values rather than experimental values from the literature. The simulated random coil values were calculated by performing REMD simulations of blocked GGXGG peptides, where X was any of the 20 regular amino acids. These peptides are commonly used experimentally to measure random coil chemical shifts (Wishart et al. 1995; Schwarzinger et al. 2001). For each amino acid, the average chemical shift of a high temperature REMD ensemble was assigned as the simulated random coil chemical shift. Comparisons of the simulation random coil values to two sets of experimental random coil values are included in Supplemental Table S1.

S-peptide, produced by cleaving the first 19 residues off of the N-terminus of ribonuclease A, was the first small peptide shown to form a helix in solution—30% helical at 273 K (Klee 1968). For this study we chose to use a 15-residue analog of S-peptide (AETAAAKFLREHMDS), with increased stability (46% helix at 276 K) (Tirado-Rives and Jorgenson 1991). The analog differs from S-peptide

due to three point mutations—K1A, E9L, Q11E—and the absence of residues 16–19, which do not participate in the helix.

Results and Discussion

Three S-peptide analog ensembles, hereafter referred to as the low (270 K), intermediate (421 K), and high (690 K) temperature ensembles, were selected from the REMD simulations. These three ensembles were chosen because they presented three very different secondary structure profiles. This can be seen by looking at the percent helicity for each ensemble, calculated as the percentage of the ensemble for which each residue was identified as helical by STRIDE (Frishman and Argos 1995). As seen in Figure 1 (red bars), in the low temperature ensemble the peptide is composed of two short helices (residues 2–5 and 10–14) separated by the coil (residues 6–9); the intermediate temperature ensemble has a long helix stretching from residues 3–13; and at high temperature, the helix is fully melted.

$^{13}C^{\alpha}$ chemical shifts can also be used to determine secondary structure. One method for doing this is the chemical-shift index (CSI) (Wishart and Sykes 1994b), which identifies as a helix any group of three or more consecutive CSDs greater than 0.7 ppm. The secondary structure profile generated in this manner agrees almost entirely with that obtained using STRIDE. Here again we find two short helices separated by residues 6–9 at low temperature, a long helix from residues 3–13 at intermediate temperature, and no helix at high temperature (Fig. 2). There are two discrepancies between STRIDE and the CSI. One is at residue E2 in the low temperature ensemble, on the edge of the first short helix, which is
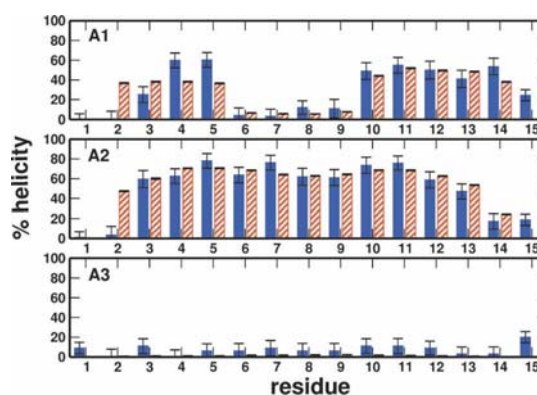
Figure 1. Comparison of percent helicity calculated as the percentage of conformations with a residue assigned as helix by STRIDE (red crosshatches) to percent helicity calculated as % helix (residue i) = Σ_j(CSD$_j$/CSD$_\alpha$), where CSD$_\alpha$ = 3.1 ppm (blue) for the low (*A1*), intermediate (*A2*), and high (*A3*) temperature ensembles.
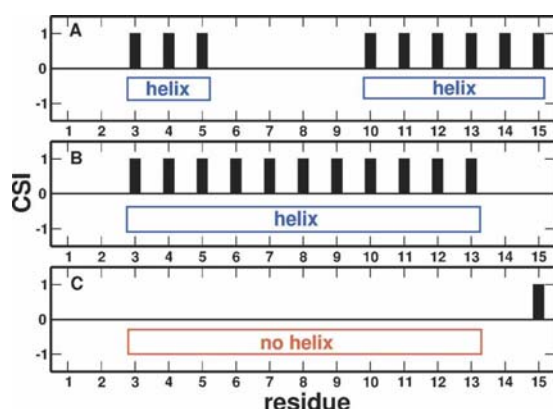
Figure 2. Chemical-shift index (CSI) plotted for each residue of the S-peptide analog for the low (*A*), intermediate (*B*), and high (*C*) temperature ensembles. Segments of the peptide sequence predicted to be helical by CSI are noted.

37% helical according to STRIDE, but not part of the helix according to CSI. The other is at residue S15 in the low temperature ensemble. For this residue, we find the opposite situation—0% helicity according to STRIDE, but part of the helix according to CSI.

An alternative way to generate information about secondary structure from $^{13}C^\alpha$ chemical shifts, one which allows for direct comparison with the percent helicity determined by STRIDE, is to assume a linear relationship between CSD and helicity (Eliezer et al. 1998; Dyson and Wright 2001). The percent helicity of each residue can then be calculated using the expression % helix (residue i) $\Sigma_j(CSD_j/CSD_\alpha)$, where the sum is over all the members of the peptide ensemble. This places each ensemble averaged CSD on a continuum between a maximal CSD value which is equated with 100% helix, and the random coil reference value, which is assumed to be 0% helix. The maximum CSD ($CSD_\alpha = 3.1$) is the average chemical shift deviation of a fully helical residue (Spera and Bax 1991).

Overall, there is close to quantitative agreement between the percent helicities calculated from STRIDE and those derived from CSDs (Fig. 1). Again, both methods predict the same average secondary structure profiles for each of the three ensembles: two short helices separated by coil at low temperature, one long helix at intermediate temperature, and no helix at high temperature, respectively. Differences in the helicities occur at the edges of helical stretches, with the percentages calculated from CSDs able to both under- and overestimate the helicity of a particular residue. Underestimation of the helicity by CSD is seen at residue E2 in both the low temperature ensemble (% helix = 0% by CSD compared to 37% by

STRIDE) and in the intermediate temperature ensemble (% helix = 3% by CSD compared to 47% by STRIDE) (Fig. 1A,B). Overestimation is found at residues A4, A5, and S15 in the low temperature ensemble (% helix = 68%, 68%, and 29% by CSD compared to 38%, 37%, and 0% by STRIDE) (Fig. 1A).

$^{13}C^\alpha$ chemical shifts depend primarily on backbone torsion angles (de Dios et al. 1993), so Ramachandran plots of the residues for which there is a disagreement between the two helicity calculations can help illustrate why the estimate of helicity based on CSD differs from the helicity calculated using STRIDE. Figure 3 presents Ramachandran plots of two residues. In one, E2 in the low temperature ensemble (Fig. 3A,B), CSD underestimates the helicity: 0% versus 37% from STRIDE. In the other, A5 in the low temperature ensemble (Fig. 3C,D), CSD overestimates the helicity predicting the residue to be 68% helical when the STRIDE helicity value is again 37%. Figure 3, A and C show that 3685 out of 10,000 conformers for E2 and 3653 out of 10,000 conformers for A5, or 37% in each case, are identified by STRIDE as helical and that all of these conformers have backbone ϕ/ψ angles for these residues which are within the α region of Ramachandran space.

Though the Ramachandran plots for the conformers identified by STRIDE as being helical at E2 and A5 are almost identical, the torsion angles of these residues for the other members of the low temperature ensemble have two different distributions in ϕ/ψ space (Fig. 3B,D). For E2, where CSD underestimates the helicity, Figure 3B shows that close to 50% of the conformers in the ensemble have E2 torsion angles outside of the helical
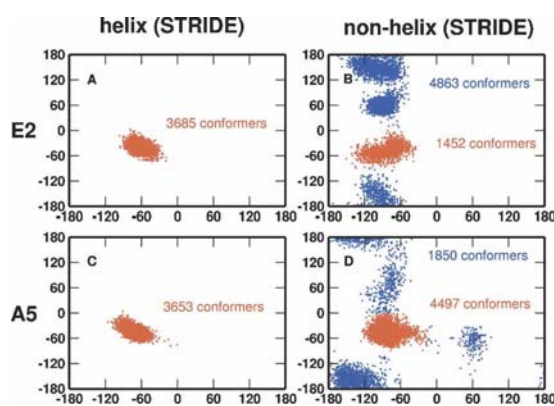
Figure 3. Ramachandran plots of residues E2 and A5 for conformers in the low temperature ensemble which STRIDE identifies as helical (*A,C*) and non-helical (*B,D*). Conformers with torsion angles in the α region of Ramachandran space are shown in red and conformers not in the α region are shown in blue.

section of ϕ/ψ space. Ensemble averaging of the negative CSDs, contributed by the half of the ensemble with E2 torsion angles outside the α region of Ramachandran space, together with the positive chemical shift deviations from the conformations which populate the helical region, yields an average CSD close to 0. Essentially, due to the averaging, the conformations with E2 torsion angles outside of the helical region are masking the presence of a significant fraction (51%) with torsion angles inside the helical region.

While there is still ensemble averaging taking place for residue A5, Figure 3D shows that less than 20% of the conformers in the low temperature ensemble have A5 torsion angles outside of the helical region. This means that the average chemical shift deviation is dominated by contributions from the helical region. In this case, the discrepancy between the percent helicity calculated from STRIDE and that calculated from the CSDs is related to the fact that CSD is a more localized property than helicity. Helicity depends not only on the ϕ/ψ angles of individual residues but also the correlation between consecutive ϕ/ψ pairs and $i,i + 4$ hydrogen bonds, while the chemical shift primarily depends on the ϕ/ψ angles of a particular residue (Eliezer et al. 1998). For residue A5 in the low temperature ensemble, this means that there are a large number of conformers, 4497 out of the 10,000 peptide conformations in the ensemble, with torsion angles in the α region of the Ramachandran plot that are not counted as helical by STRIDE. Even though these conformations are not included in the structural measurement of percent helicity, they all contribute positive CSDs to the ensemble average, thereby inflating the CSD measure of helicity relative to that based on STRIDE.

Chemical shift deviations for each of the amino acid residues were calculated using random coil values obtained from REMD simulations of blocked GGXGG peptides. This is the most consistent way to calculate CSDs for this study; any errors introduced through the simulation protocols and the use of SHIFTX to calculate chemical shifts are equally reflected in both the chemical shifts of the S-peptide and the GGXGG peptides used as references. Using the literature values of the random coil chemical shifts to calculate CSDs does not change the major conclusion of this work—chemical shift deviations are in general a good predictor of peptide helicity for heterogeneous ensembles like the S-peptide. Also, when using random coil chemical shifts from the literature (Wishart et al. 1995; Schwarzinger et al. 2001), we still observe situations, such as residue E2 in the low temperature ensemble, where the CSD underestimates the percent helicity. Overestimation of the percent helicity by CSD at the ends of some helical stretches is, however, diminished when using the literature random coil values, because of a small, systematic increase in the literature random coil values relative to the simulated random coil shifts (see the Supplemental material).

Conclusion

In conclusion, computer simulations of an S-peptide analog have confirmed the correlation between chemical shift deviations from random coil and the percent helicity of a peptide sequence. We have shown that a given $^{13}C^\alpha$ CSD, taken as a fraction of the average chemical shift deviation of a residue in a fully formed helix, can be used as a quantitative measure of the percent helicity at that residue. There are a number of different chemical shift measures, such as CSI (Wishart and Sykes 1994b), SSP (Marsh et al. 2006), and the probability based methods of Jardetzky (Wang and Jardetzky 2002), that make use of this correlation to predict secondary structure propensities. While this correlation holds in most cases, it is possible for the chemical shift deviations to significantly overestimate or underestimate the helicity at the edges of helical segments. These two types of disagreement between CSD and helicity can occur even at edges caused by short turn or coil segments separating two short helical stretches. This is a pattern that is quite likely to be found in parts of large unstructured proteins, where care must be taken in defining the boundaries of segments with helical propensity.

References

Banks, J.L., Beard, H.S., Cao, Y., Cho, A.E., Damm, W., Farid, R., Felts, A.K., Halgren, T.A., Mainz, D.T., Maple, J.R., et al. 2005. Integrated modeling program, applied chemical theory (IMPACT). *J. Comput. Chem.* **26:** 1752–1780.

de Dios, A.C., Pearson, J.G., and Oldfield, E. 1993. Secondary and tertiary structure effects on protein NMR chemical shifts: An ab initio approach. *Science* **260:** 1491–1496.

Dyson, H.J. and Wright, P.E. 2001. Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods Enzymol.* **339:** 259–270.

Eliezer, D., Yao, J., Dyson, H.J., and Wright, P.E. 1998. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* **5:** 148–155.

Frishman, D. and Argos, P. 1995. Knowledge based protein secondary structure assignment. *Proteins* **23:** 566–579.

Gallicchio, E. and Levy, R.M. 2004. AGBNP: An analytic implicit solvent model suitable for molecular dynamics and high-resolution modeling. *J. Comput. Chem.* **25:** 479–499.

Klee, W.A. 1968. Studies of the conformation of ribonuclease S-peptide. *Biochemistry* **7:** 2731–2736.

Marsh, D., Singh, V.K., Zongchao, J., and Forman-Kay, J.D. 2006. Sensitivity of secondary structure propensities to sequence differences between α- and γ-synulcein: Implications for fibrillation. *Protein Sci.* **15:** 2795–2804.

Mitchinson, C. and Baldwin, R.L. 1986. The design and production of semi-synthetic ribonucleases with increased thermostability by incorporation

of S-peptide analogues with enhanced helical stability. *Proteins* **1:** 23–33.

Neal, S., Nip, A.M., Zhang, H., and Wishart, D.S. 2003. Rapid and accurate calculation of protein ^{1}H, ^{13}C, and ^{15}N chemical shifts. *J. Biomol. NMR* **26:** 215–240.

Schwarzinger, S., Kroom, G.J., Foss, T.R., Chung, J., Wright, P.E., and Dyson, H.J. 2001. Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* **123:** 2970–2978.

Smith, L.J., Fiebig, K.M., Schwalbe, H., and Dobson, C.M. 1996. The concept of a random coil: Residual structure in peptides and denatured proteins. *Fold. Des.* **1:** R95–R106. doi: 10.1016/S1359-0278(96)00046-6.

Spera, S. and Bax, A. 1991. Empirical correlation between protein backbone conformation and Cα and Cβ ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113:** 5490–5492.

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

Szilagyi, L. and Jardetzky, O. 1989. α-Proton chemical shifts and protein secondary structure. *J. Magn. Reson.* **83:** 441–449.

Tirado-Rives, J. and Jorgenson, W.L. 1991. Molecular dynamics simulations of the unfolding of an α-helical analogue of ribonuclease A S-peptide in water. *Biochemistry* **30:** 3864–3871.

Wang, Y. and Jardetzky, O. 2002. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.* **11:** 852–861.

Wishart, D.S. and Sykes, B.D. 1994a. Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239:** 363–392.

Wishart, D.S. and Sykes, B.D. 1994b. The ^{13}C chemical-shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR* **4:** 171–180.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. 1995. ^{1}H, ^{13}C, and ^{15}N random coil NMR chemical shifts of the common amino acids I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* **5:** 67–81.

Supporting Information

Residue	Simulation	Wishart et. al. 1995	Schwarzinger et. al. 2001
ALA	51.83	52.84	52.8
CYS	58.24	57.53	58.6
ASP	53.86	54.18	53
GLU	56.22	56.87	56.1
PHE	57.16	57.98	58.1
GLY	45	45.51	45.4
HIS	55.34	55.86	55.4
ILE	60.63	61.03	61.6
LYS	56.05	56.59	56.7
LEU	54.74	54.92	55.5
MET	54.97	55.67	55.8
ASN	52.9	53.23	55.3
PRO	63.28	63.46	63.7
GLN	55.67	56.12	56.2
ARG	55.59	56.42	56.5

SER	57.92	58.38	58.7
THR	61.38	61.64	62
VAL	61.34	62.06	62.6
TRP	56.46	57.78	57.6
TYR	57.15	57.97	58.3

Table 3.1 Comparison of random coil chemical shift values

4.4 References

Bierzynski, A., P. S. Kim, et al. (1982). "A salt bridge stabilizes the helix formed by isolated C-peptide of RNase A." Proceedings of the National Academy of Sciences of the United States of America **79**(8): 2470-2474.

Brown, J. E. and W. A. Klee (1969). "Conformational studies of a series of overlapping peptides from ribonuclease and their relationship to the protein structure." Biochemistry **8**(7): 2876-2879.

Dyson, H. J. and P. E. Wright (2001). "Nuclear Magnetic Resonance methods for elucidation of structure and dynamics in disordered states." Methods in Enzymology **339**: 259-270.

Eliezer, D., J. Yao, et al. (1998). "Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding." Nat Struct Biol **5**(2): 148-155.

Kim, P. S. and R. L. Baldwin (1984). "A helix stop signal in the isolated S-peptide of ribonuclease A." Nature **307**(5949): 329-334.

Markley, J. L., D. H. Meadows, et al. (1967). "Nuclear magnetic resonance studies of helix-coil transitions in polyamino acids." Journal of Molecular Biology **27**(1): 25-40.

Marsh, J. A., V. K. Singh, et al. (2006). "Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation." Prot. Sci. **15**(12): 2795-2804.

Mitchinson, C. and R. L. Baldwin (1986). "The design and production of semisynthetic ribonucleases with increased thermostability by incoropratino of S-peptide annalogues with enhanced helical stability." Proteins **1**: 23-33.

Nakamura, A. and O. Jardetzky (1968). "Systematic analysis of chemical shifts in the nuclear magnetic resonance spectra of peptide chains. II. Oligoglycines." Biochemistry **7**(3): 1226-1230.

Nelson, J. W. and N. R. Kallenbach (1989). "Persistence of the alpha-helix stop signal in the S-peptide in trifluoroethanol solutions." Biochemistry **28**(12): 5256-5261.

Smith, L. J., K. M. Fiebig, et al. (1996). "The concept of a random coil: residual structure in peptides and denatured proteins." Folding Design **1**: R95-R106.

Szilagyi, L. and O. Jardetzky (1989). "α-Proton chemical-shifts and secondary structure in proteins." J. Magn. Reson. **83**: 441-449.

Wang, Y. and O. Jardetzky (2002). "Probability-based protein secondary structure identification using combined NMR chemical-shift data." <u>Protein science : a publication of the Protein Society</u> **11**(4): 852-861.

Wishart, D. S. and B. D. Sykes (1994). "Chemical Shifts as a tool for structure determination." <u>Methods in Enzymology</u> **239**: 363-392.

Chapter 5

Intrinsically Disordered Proteins

5.1 Introduction

Proteins are essential components of the cell, participating in all major biological pathways, crucial for maintenance of life. Most proteins fold to adopt specific conformations, and these unique structures are believed to determine their specific functions in cells. Protein misfolding has been associated with a large number of disease states. More recently, a large number of proteins, accounting for over 30% of the genome, have been found to entirely lack stable secondary and tertiary structures or have unstructured regions of over 50 residues long (Dunker, Lawson et al. 2001; Uversky 2002). These inherently unstructured (regions of) proteins are described by a variety of names including 'natively unfolded', 'natively unstructured', 'intrinsically disordered', 'intrinsically unstructured' etc. These Intrinsically Disordered Proteins (IDPs) can be best described as randomly fluctuating ensemble of structures that lack well-defined tertiary structures in their physiological form. The sequences of these proteins are characterized by low sequence complexity, low hydrophobicity and high net charge (Arawaka, Saito et al. 1998; Spillantini, Crowther et al. 1998; Uversky 2003).

Figure 5.1 (Uversky, Oldfield et al. 2009) shows the rapid growth in the field focusing on IDPs. These proteins have gained much attention because of the numerous vital roles they play in cells. The functional repertoire of these IDPs includes signal transduction,

regulation of transcription and translation, protein complex assembly etc (Wright and Dyson 1999; Uversky, Gillespie et al. 2000; Iakoucheva, Brown et al. 2002; Tompa 2002; Dyson and Wright 2005a).
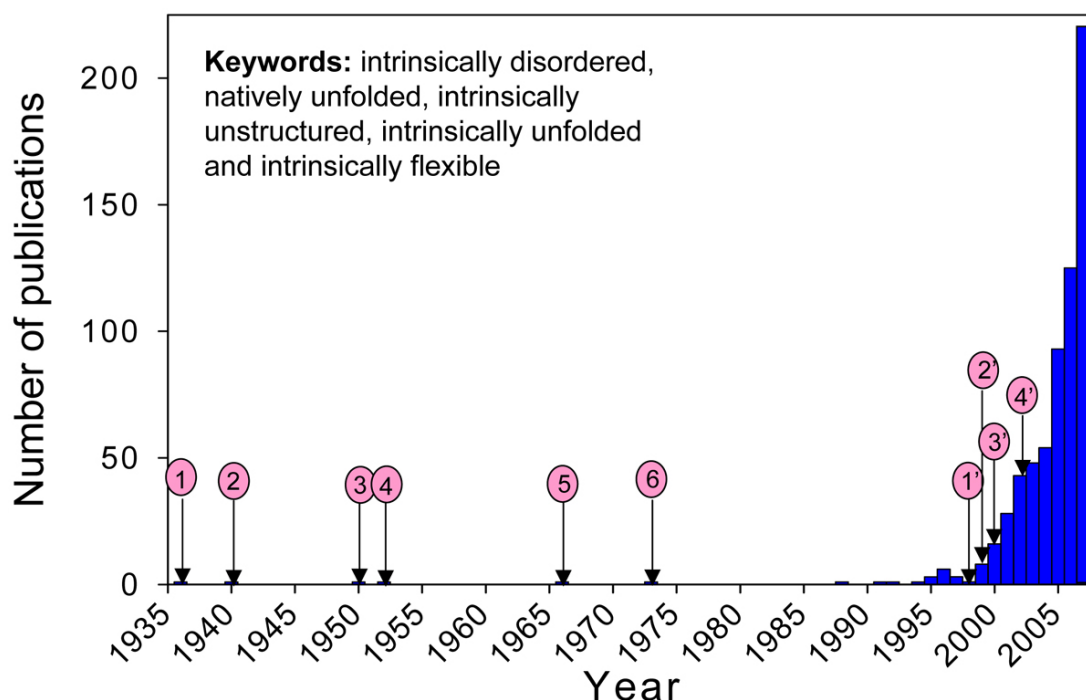
Figure 5.1 Number of publications related to Intrinsically Disordered Proteins, taken from Uversky et al, 2009.

A number of diseased states are associated with protein misfolding, involving the conversion of these proteins from their normally soluble form to insoluble fibres or plaques that accumulate in a variety of organs. The final forms of these aggregated proteins have a well-defined fibrillar structure known as amyloid. Some of the most common diseases involving the formation of amyloid include Alzheimer's, Parkinson's, Huntington's, Prion diseases and type II diabetes (Perutz 1996; Dobson 1999; Lansbury 1999; Chiti and Dobson 2006).

5.2 Methods for characterizing IDPS

Characterizing the conformational characteristics of IDPs is crucial to understanding the relationship between the conformational propensities and biological function and disease. Conformational characterization of IDPs is very challenging due to the lack of persistent secondary and/ or tertiary structures. A number of experimental and computational approaches have been applied to understand the structural characteristics of IDPs. Biophysical techniques including Small angle X-ray scattering (SAXS), Single molecule Flourescence Resonence Energy Transfer (SmFRET), Nuclear Magnetic Resonance (NMR) spectroscopy, near and far ultra-violet Circular Dichroism (UV CD), Fourier Transform Infrared Spectroscopy (FTIR), Raman spectroscopy, Dynamic Light Scattering (Merchant, Best et al. 2007; Eliezer 2009; Tompa 2009; Bernado and Svergun 2012) provide valuable information of the conformational characteristics of the disordered states of IDPs. Of all the experimental techniques, solution NMR provides the highest resolution data for the conformational characterization of IDPs (Dyson and Wright 2004; Eliezer 2009).

5.2.1 NMR Characterization of IDPs

A number of NMR parameters have been used to describe the conformational characteristics of IDPs including the analysis of 13C chemical shifts, Residual Dipolar Couplings (RDCs) and Paramagnetic Relaxation Enhancement (PRE) (Dyson and Wright 2001; Dyson and Wright 2002; Bracken, Iakoucheva et al. 2004; Dyson and Wright 2004; Dyson and Wright 2005b). Deviations of the observed sequence specific chemical shifts

from that of the random coil reference state, provides a semi-quantitative measure for the secondary structural propensies of disordered ensembles. RDCs arise as a consequence of the alignment of a pair of nuclei against the orienting medium. This results in the recoupling of spin-spin dipolar interactions. The orientation dependence of RDCs provides insights into the motional dynamics of chains (Bax 2003). RDCs have recently also been shown to be sensitive to the local secondary structure in disordered proteins (Mohana-Borges R 2004). PRE measurements are used to detect long-range contacts up to 20 Å for disordered ensembles (Bertoncini, Jung et al. 2005; Dedmon, Christodoulou et al. 2005; Vise, Baral et al. 2007; Dancheck, Nairn et al. 2008; Eliezer 2009). Spin label probes are attached to different sites on the polypeptide chain. PREs are measured by calculating the ratio of the intensities of peaks in the presence (paramagnetic) and absence (diamagnetic) of the spin label probe. The intensity ratios range from 0 to 1, corresponding to distances from 0 to 25 Å for the different residues from the spin label probe.

5.2.2 Computational Characterization of IDPs

A number of bioinformatics approaches have been used to predict intrinsically disordered regions and whole proteins. The distinct sequence characteristics of disordered proteins, with higher charged and polar residue content and fewer hydrophobic residues (Dunker, Brown et al. 2002), play a key role in the success of these sequence prediction tools. A number of simple to more sophisticated algorithms, including PONDR (Predictor of Natural Disordered Regions) (Romero, Obradovic et al. 1997; Romero, Obradovic et al. 2001), DisEMBL (Linding, Jensen et al. 2003), GlobPlot (Linding, Russell et al. 2003),

charge-hydrophobicity plot etc have been developed that predict disordered proteins or regions. Currently, there are over 50 disorder tools (He, Wang et al. 2009; Rauscher and Pomes 2010) for the prediction of disordered proteins (regions).

Molecular dynamics simulations provide a powerful complement to the experimental studies of IDPs. While experimental observations measure average properties of the heterogeneous ensembles of disordered chains, computational simulations provide visualizations of individual conformers of the ensemble and further characterization of the sub-ensembles present. Atomistic simulations provide the most detailed description required for the characterization of conformational properties of disordered proteins (Click, Ganguly et al. 2010).

One approach to characterizing IDPs using molecular simulations involves the use of experimental observations as restraints in the simulations. These combined experimental and simulation approaches provide disordered ensembles that are consistent with experiment (Vendruscolo 2007). NMR chemical shift and J-coupling restraints have also been employed (Huang and Stultz 2008; Yoon, Venkatachalam et al. 2009), as also the RDCs (Esteban-Martin, Fenwick et al. 2010), to obtain ensembles consistent with these data. Lindoff-Larsen et al have also employed long-range PRE restraints over an ensemble of structures generated using the Monte Carlo Replica Sampling method (Lindorff-Larsen, Kristjansdottir et al. 2004). These restraints have been employed as an energy penalty over the ensemble averaged distances.

A number of studies have also been reported where a random ensemble of structures is weighted to create an ensemble fitting the experimental restraints (Choy and Forman-Kay 2001; Marsh, Neale et al. 2007; Marsh and Forman-Kay 2009). Similar approaches using

an Ensemble Optimization Method (EOM) have been used where ensembles generated using Flexibe-Meccano are used to fit SAXS measurements, and have been further refined by using other experimental measures from NMR, CD and FRET (Bernado, Mylonas et al. 2007). These ensembles constructed using the database approach have also been applied to α-synuclein where ensembles are constructed with specified regions of the chain in contact with each other, which are further compared to RDC measurements (Bernadó, Blanchard et al. 2005; Bernardó P 2005).

Another approach to characterizing IDPs is using de novo simulations, which are performed with no experimental restraints. The ensembles generated from these independent simulations can be used to predict the experimental measures. While simulations with explicit water provide the most realistic description of the solvent, inclusion of water molecules increases both the system size (Click, Ganguly et al. 2010) and simulation time. Implicit solvent models where solvent is represented as a continuous medium rather than explicit solvent molecules, provides a balance between computational cost and underlying physics principles.

In the subsequent chapters, molecular dynamics simulations and experimental NMR observables are integrated to provide the conformational description of the intrinsically disordered α-synuclein.

5.3 References:

Arawaka, S., Y. Saito, et al. (1998). "Lewy body in neurodegeneration with brian iron accumulation type 1 is immunoreactive for alpha-synuclein." Neurology **51**: 887-889.

Bax, A. (2003). "Weak alignment offers new NMR opportunities to study protein structure and dynamics." Protein Sci **12**(1): 1-16.

Bernadó, P., L. Blanchard, et al. (2005). "A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering." Proc. Natl. Acad. Sci. USA **102**: 17002-17007.

Bernado, P., E. Mylonas, et al. (2007). "Structural Characterization of Flexible Proteins using Small Angle X-ray Scattering." J. Am. Chem. Soc. **129**(17): 5656-5664.

Bernado, P. and D. I. Svergun (2012). "Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering." Molecular BioSystems **8**(1): 151-167.

Bernardó P, B. C., Griesinger C, Zweckstetter M, Blackledge M (2005). "Defining Long-range order and local disorder in native alpha-synuclein using Residual Dipolar Couplings." J. Am. Chem. Soc. **127**: 17968-17969.

Bertoncini, C. W., Y. S. Jung, et al. (2005). "Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein." Proc Natl Acad Sci U S A **102**(5): 1430-1435.

Bracken, C., L. M. Iakoucheva, et al. (2004). "Combining prediction, computation and experiment for the characterization of protein disorder." Curr Opin Struct Biol **14**(5): 570-576.

Chiti, F. and C. M. Dobson (2006). "Protein misfolding, functional amyloid, and human disease." Annu. Rev. Biochem. **75**: 333-366.

Choy, W. Y. and J. D. Forman-Kay (2001). "Calculation of ensembles of structures representing the unfolded state of an SH3 domain." J Mol Biol **308**(5): 1011-1032.

Click, T. H., D. Ganguly, et al. (2010). "Intrinsically disordered proteins in a physics-based world." International journal of molecular sciences **11**(12): 5292-5309.

Dancheck, B., A. C. Nairn, et al. (2008). "Detailed structural characterization of unbound protein phosphatase 1 inhibitors." Biochemistry **47**(47): 12346-12356.

Dedmon, M. M., J. Christodoulou, et al. (2005). "Heat shock protein 70 inhibits alpha-synuclein fibril formation via preferential binding to prefibrillar species." J Biol Chem **280**(15): 14733-14740.

Dobson, C. M. (1999). "Protein misfolding, evolution and disease." Trends Biochem Sci **24**(9): 329-332.

Dunker, A. K., C. J. Brown, et al. (2002). "Identification and functions of usefully disordered proteins." Adv Protein Chem **62**: 25-49.

Dunker, A. K., J. D. Lawson, et al. (2001). "Intrinsically disordered protein." J Mol Graph Model **19**(1): 26-59.

Dyson, H. and P. Wright (2005a). "Intrinsically Unstructured Proteins and their functions." Mol. Cell Biol. **6**: 197-208.

Dyson, H. J. and P. E. Wright (2001). "Nuclear Magnetic Resonance methods for elucidation of structure and dynamics in disordered states." Methods in Enzymology **339**: 259-270.

Dyson, H. J. and P. E. Wright (2002). "Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance." Adv Protein Chem **62**: 311-340.

Dyson, H. J. and P. E. Wright (2004). "Unfolded proteins and protein folding studied by NMR." Chem Rev **104**(8): 3607-3622.

Dyson, H. J. and P. E. Wright (2005b). "Intrinsically unstructured proteins and their functions." Nat Rev Mol Cell Biol **6**(3): 197-208.

Eliezer, D. (2009). "Biophysical characterization of intrinsically disordered proteins." Curr Opin Struct Biol **19**(1): 23-30.

Esteban-Martin, S., R. B. Fenwick, et al. (2010). "Refinement of ensembles describing unstructured proteins using NMR residual dipolar couplings." J Am Chem Soc **132**(13): 4626-4632.

He, B., K. Wang, et al. (2009). "Predicting intrinsic disorder in proteins: an overview." Cell research **19**(8): 929-949.

Huang, A. and C. M. Stultz (2008). "The effect of ΔK280 mutation on the unfolded state of a microtubule-binding repeat in Tau." PLoS Comput Biol **48**(8): e1000155.

Iakoucheva, L. M., C. J. Brown, et al. (2002). "Intrinsic disorder in cell-signaling and cancer-associated proteins." Journal of Molecular Biology **323**(3): 573-584.

Lansbury, P. T., Jr. (1999). "Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease." Proc Natl Acad Sci U S A **96**(7): 3342-3344.

Linding, R., L. J. Jensen, et al. (2003). "Protein disorder prediction: implications for structural proteomics." Structure **11**(11): 1453-1459.

Linding, R., R. B. Russell, et al. (2003). "GlobPlot: Exploring protein sequences for globularity and disorder." Nucleic acids research **31**(13): 3701-3708.

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

Marsh, J. A. and J. D. Forman-Kay (2009). "Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints." J Mol Biol **391**(2): 359-374.

Marsh, J. A., C. Neale, et al. (2007). "Improved structural characterization of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure." J. Mol. Biol. **367**: 1494-1510.

Merchant, K. A., R. B. Best, et al. (2007). "Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations." Proceedings of the National Academy of Sciences of the United States of America **104**(5): 1528-1533.

Mohana-Borges R, G. N., Kroon GJA, Dyson JH, Wright PE (2004). "Structural characterization of unfolded states of Apomyoglobin using Residual Dipolar Couplings." J. Mol. Biol **340**: 1131-1142.

Perutz, M. F. (1996). "Glutamine repeats and inherited neurodegenerative diseases: molecular aspects." Curr Opin Struct Biol **6**(6): 848-858.

Rauscher, S. and R. Pomes (2010). "Molecular simulations of protein disorder." Biochem Cell Biol **88**(2): 269-290.

Romero, Obradovic, et al. (1997). "Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family." Genome informatics. Workshop on Genome Informatics **8**: 110-124.

Romero, P., Z. Obradovic, et al. (2001). "Sequence complexity of disordered protein." PROTEINS **42**(1): 38-48.

Spillantini, M. G., R. A. Crowther, et al. (1998). "Filamentous alpha-synuclein inclusions link multiple system atrophy with Parkinson's disease and dementia with Lewy bodies." Neurosci Lett **251**(3): 205-208.

Tompa, P. (2002). "Intrinsically unstructured proteins." Trends Biochem Sci **27**(10): 527-533.

Tompa, P. (2009). "Structure and Function of Intrinsically Disordered Proteins." Taylor and Francis, Boca Raton, FL.

Uversky, V. N. (2002). "Natively unfolded proteins: A point where biology waits for

physics." <u>Protein Science</u> **11**: 739-756.

Uversky, V. N. (2003). "A protein chameleon: Conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders." <u>Journal of Biolmolecular Structure & Dynamics</u> **21**: 211-234.

Uversky, V. N., J. R. Gillespie, et al. (2000). "Why are "natively unfolded" proteins unstructured under physiologic conditions?" <u>Proteins</u> **41**(3): 415-427.

Uversky, V. N., C. J. Oldfield, et al. (2009). "Unfoldomics of human diseases: linking protein intrinsic disorder with diseases." <u>BMC genomics</u> **10 Suppl 1**: S7.

Vendruscolo, M. (2007). "Determination of conformationally heterogeneous states of proteins." <u>Curr Opin Struct Biol</u> **17**(1): 15-20.

Vise, P., B. Baral, et al. (2007). "Identifying long-range structure in the intrinsically unstructured transactivation domain of p53." <u>PROTEINS</u> **67**(3): 526-530.

Wright, P. E. and H. J. Dyson (1999). "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm." <u>J Mol Biol</u> **293**(2): 321-331.

Yoon, M. K., V. Venkatachalam, et al. (2009). "Residual Structure within the disordered C-terminal segment of p21 (Waf1/Cap1/Sdi1) and its implications for molecular recognition." <u>Prot. Sci.</u> **18**(2): 337-347.

Chapter 6

α-synuclein and Parkinson's Disease

6.1 Introduction

Parkinson's disease is the second most prevalent of the late onset neurodegenerative diseases and is characterized by the loss of dopaminergic neurons concurring with the formation of Lewy bodies. α-synuclein is the primary protein component of the Lewy body deposits that are the hallmark of Parkinson's (Spillantini, Schmidt et al. 1997). It is present in high concentrations in both soluble and membrane bound forms in the presynaptic terminals, accounting for about 1% of the total protein content in the cytosol. While the exact function of α-synuclein is unclear, a number of functions have been suggested. Functional repertoire of α-synuclein includes control of neuronal apoptotic response and regulation of enzymes and transporters (Dev, Hofele et al. 2003), vesicle transportation and neurotransmitter release (Norris, Giasson et al. 2005) and chaperone activity to promote protein complex assembly (Burre, Sharma et al. 2010).

The implication of α-synuclein in the pathogenesis of Parkinson's is evidenced from a number of observations from previous studies. A majority of the fibrillar materials deposited in the Lewy body and Lewy neurites were shown to be composed of α-synuclein (Spillantini, Schmidt et al. 1997; Spillantini, Crowther et al. 1998). Subsequent studies showed that point mutations in the primary sequence, namely A30P, E46K and A53T led to autosomally dominant early-onset Parkinson's disease (Polymeropoulos, Lavedan et al. 1997; Kruger, Kuhn et al. 1998; Zarranz, Alegre et al. 2004). It was also shown that triplication of the gene that encodes α-synuclein led to a doubling of the

protein concentration in the blood as well as the deposition of α-synuclein in the cerebral cortex where Lewy bodies are observed (Miller, Hague et al. 2004).

6.1.1 Sequence characteristics of α-synuclein

α-synuclein is best described as a highly conserved IDP (Rochet and Lansbury 2000) characterized by high net charge and low hydrophobicity and low sequence complexity. α-synuclein has been shown to exist in two primary forms – in the free form in the cytoplasm, and the membrane-bound form (Davidson, Jonas et al. 1998; Jo, McLaurin et al. 2000).

Figure 6.1. Primary sequence of α-synuclein. The N, NAC and C-terminal regions are displayed in blue, red and green respectively

α-synuclein is 140 residues long, with the sequence divided into 3 distinct domains, as shown in Figure 6.1. The N-terminal domains (1-60) contains four 11-residue repeats with the highly conserved hexamer motifs KTKEGV, which adopts α-helical conformation upon binding to membranes (Jo, McLaurin et al. 2000; Eliezer, Kutluay et al. 2001; Bussell and Eliezer 2003; Bussell, Ramlall et al. 2005). The Non-amyloid β-component (NAC) region (61-95) is the central hydrophobic region that has been proposed to be primarily responsible for aggregation (El-Agnaf, Bodles et al. 1998) and contains three of the 11-residue repeats. The C-terminal domain is highly acidic and proline-rich (Kim, Paik et al. 2002) and also contains the 3 conserved tyrosines (Uversky 2007).The sequence of α-synuclein shows an unbalanced distribution of charged

residues. Table shows the charge distribution in the different domains of α-synuclein.

pH	N-terminus			NAC			C-terminus			Whole chain		
	Net	Total	%	Net	Total	%	Net	Total	%	Net	Total	%
Neutral	+4	18	30.0	-1	3	8.6	-12	18	40	-9	39	27.8

Table 6.1 Charge distribution along the sequence of α-synuclein

6.1.2 Structural properties of α-synuclein

α-synuclein has been shown to adopt a variety of conformational states under different cellular conditions. It has been shown to adopt a helical conformational upon binding to membranes, with this form associated with the recruitment of dopamine in the presynapse and synaptic signal transmission, essential for physiological function (Yavich, Tanila et al. 2004). The monomeric form of α-synuclein has been extensively studied using a variety of biophysical techniques including NMR and is described as an intrinsically disordered ensemble with transient propensity for secondary and tertiary structures. The average size of α-synuclein under physiological conditions has been shown to be more compact than that expected for random coil conformation (Eliezer, Kutluay et al. 2001; Uversky, Li et al. 2001a; Bertoncini, Jung et al. 2005b; Dedmon, Lindorff-Larsen et al. 2005; Bertoncini, Rasia et al. 2007; Kim, Hiese et al. 2007; Sung and Eliezer 2007), while this size is considerably smaller than that expected for a random coil ensemble (Uversky, Li et al. 2001a).

NMR Chemical shift data suggest transient helical conformation for the N-terminal lipid-binding region, and an extended conformation for the acidic C-terminal region (Eliezer, Kutluay et al. 2001; Sung and Eliezer 2007). Transient long-range contacts have been

observed from PRE measurements suggesting long-range interactions between the N
and C-terminal regions and NAC and C-terminal regions (Eliezer, Kutluay et al. 2001;
Bernardó P 2005; Bertoncini, Jung et al. 2005b; Dedmon, Lindorff-Larsen et al. 2005;
Marsh, Singh et al. 2006; Sung and Eliezer 2007; Wu, Kim et al. 2008). These PRE
measurements have also been combined with molecular simulations to present an
ensemble in which the C-terminal region folds back to interact with the NAC (Bertoncini,
Jung et al. 2005b; Dedmon, Lindorff-Larsen et al. 2005). Interaction between N and C-
terminal regions have also been observed from a recent study that combines a bayesian
formalism with experimental data including chemical shifts, RDCs and SAXS
measurements (Ullman, Fisher et al. 2011).

6.2 Tetrameric conformation of α-synuclein

The native state of α-synuclein has long been described as a disordered monomeric
ensemble with little propensity for secondary structure. α-synuclein has been regarded as
an archetypal IDP - characterized by its high net charge and low hydrophobicity. The
monomeric state of α-synuclein was established from initial studies of the recombinant
protein expressed in bacterial systems and human cell lines (Weinreb, Zhen et al. 1996;
Bussell and Eliezer 2001). These findings were later confirmed in studies of other groups
(Bertoncini, Jung et al. 2005a; Dedmon, Lindorff-Larsen et al. 2005; Wu, Kim et al.
2008).

This view of α-synuclein as a disordered ensemble has been challenged by two studies,
which suggest that α-synuclein may adopt tetramer conformations with helical
propensities of the chains (Bartels, Choi et al. 2011; Wang, Perovic et al. 2011). The
Selkoe group has shown, using non-denaturing gel electrophoresis, that α-synuclein

obtained from human red blood cells and neuroblastoma cell lines, behaves as a 55-60 kDa protein, suggestive of a tetrameric conformation. Further, CD-spectra of native human RBC α-synuclein tetramer were observed to have two minima at 222 and 208 nm, suggesting an α-helical folded state. This structured, tetrameric form of α-synuclein has further been shown to be more resistant to aggregation than the monomeric form. These results have been confirmed by these two research groups using several independent experiments, indicating that the physiologically predominant species of α-synuclein is a helically folded, tetrameric form, although variable amounts of monomer and dimer have been observed in certain cell lines. The observations in previous studies of the disordered, monomeric form of α-synuclein have been attributed to the use of denaturing agents. Recent studies performed under conditions similar to that used by Selkoe et. al., have failed to support the above-mentioned observation of the tetrameric state (Fauvet, Mbefo et al. 2012) and suggest increasingly that α-synuclein adopts disordered monomeric conformations under cellular conditions (Fauvet, Mbefo et al. 2012; Kang, Moriarty et al. 2012; Maltsev, Ying et al. 2012; Trexler and Rhoades 2012).

6.3 α-synuclein aggregation and disease

α-synuclein has been shown to aggregate into insoluble fibrils and to form pathological structures including Lewy bodies and Lewy neurites. Aggregation of α-synuclein into fibrils has been suggested to play a causative role in the etiology of the disease. The conversion from the soluble monomeric state to amyloid fibrils has been proposed to arise from a nucleation-dependent mechanism in which the amyloidogenic proteins access unstable, partially ordered conformations prior to fibril formation (Lansbury 1997; Wood, Wypych et al. 1999; Rochet and Lansbury 2000).

A number of factors have been found to promote aggregation of this protein. Changes in the sequence, due to single point mutations are implicated in early-onset Parkinson's disease (Polymeropoulos, Lavedan et al. 1997; Kruger, Kuhn et al. 1998; Zarranz, Alegre et al. 2004). Truncation of the C-terminal domain was shown to result in accelerated aggregation of α-synuclein. Posttranslational modifications including nitration and phosphorylation have been shown to impact the aggregation and toxicity of α-synuclein (Giasson, Duda et al. 2000; Chen and Feany 2005). Environmental factors including pH, temperature, metal ions and molecular crowding were shown to promote aggregation of α-synuclein (Uversky, Li et al. 2001a; Uversky, Li et al. 2001b; Uversky, E et al. 2002). Understanding the aggregation of α-synuclein requires the identification of the conformational characteristics of the monomeric state, and correlation of the transient secondary and tertiary structural propensities of these states with fibrillation rate. The following chapters illustrate the conformational characteristics of the monomeric state of α-synuclein under physiological conditions and under conditions promoting faster aggregation of α-synuclein.

6.4 References:

Bartels, T., J. G. Choi, et al. (2011). "alpha-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation." Nature **477**(7362): 107-110.

Bernardó P, B. C., Griesinger C, Zweckstetter M, Blackledge M (2005). "Defining Long-range order and local disorder in native alpha-synuclein using Residual Dipolar Couplings." J. Am. Chem. Soc. **127**: 17968-17969.

Bertoncini, C., Y.-s. Jung, et al. (2005a). "Release of long-range teriary interactions potentiates aggregation of natively unstructured alpha-synuclein." Proceedings of the National Academy of Sciences **102**: 1430-1435.

Bertoncini, C., R. Rasia, et al. (2007). "Structural characterization of the intrinsically unfolded protein beta-synuclein, a natural negative regulator of alpha-synuclein regulation." J. Mol. Biol. **372** : 708-722.

Bertoncini, C. W., Y. S. Jung, et al. (2005b). "Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein." Proc Natl Acad Sci U S A **102**(5): 1430-1435.

Burre, J., M. Sharma, et al. (2010). "Alpha-synuclein promotes SNARE-complex assembly in vivo and in vitro." Science **329**(5999): 1663-1667.

Bussell, R., Jr. and D. Eliezer (2001). "Residual structure and dynamics in Parkinson's disease-associated mutants of alpha-synuclein." J Biol Chem **276**(49): 45996-46003.

Bussell, R., Jr. and D. Eliezer (2003). "A structural and functional role for 11-mer repeats in alpha-synuclein and other exchangeable lipid binding proteins." J Mol Biol **329**(4): 763-778.

Bussell, R., Jr., T. F. Ramlall, et al. (2005). "Helix periodicity, topology, and dynamics of membrane-associated alpha-synuclein." Protein Sci **14**(4): 862-872.

Chen, L. and M. B. Feany (2005). "Alpha-synuclein phosphorylation controls neurotoxicity and inclusion formation in a Drosophila model of Parkinson disease." Nature Neuroscience **8**(5): 657-663.

Davidson, W. S., A. Jonas, et al. (1998). "Stabilization of alpha-synuclein secondary structure upon binding to synthetic membranes." J Biol Chem **273**(16): 9443-9449.

Dedmon, M. M., K. Lindorff-Larsen, et al. (2005). "Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations." J. Am. Chem. Soc. **127**: 476-477.

Dev, K. K., K. Hofele, et al. (2003). "Part II: alpha-synuclein and its molecular

pathophysiological role in neurodegenerative disease." Neuropharmacology **45**(1): 14-44.

El-Agnaf, O. M., A. M. Bodles, et al. (1998). "The N-terminal region of non-A beta component of Alzheimer's disease amyloid is responsible for its tendency to assume beta-sheet and aggregate to form fibrils." Eur J Biochem **258**(1): 157-163.

Eliezer, D., E. Kutluay, et al. (2001). "Conformational properties of alpha-synuclein in its free and lipid-associated states." J Mol Biol **307**(4): 1061-1073.

Giasson, B. I., J. E. Duda, et al. (2000). "Oxidative damage linked to neurodegeneration by selective alpha-synuclein nitration in synucleinopathy lesions." Science (New York, N Y) **290**(5493): 985-989.

Jo, E., J. McLaurin, et al. (2000). "alpha-Synuclein membrane interactions and lipid specificity." J Biol Chem **275**(44): 34328-34334.

Kim, H.-Y., H. Hiese, et al. (2007). "Correlation of amyloid fibril beta-structure with the unfolded state of alpha-synuclein." Chembiochem **8**: 1671-1674.

Kim, T. D., S. R. Paik, et al. (2002). "Structural and functional implications of C-terminal regions of alpha-synuclein." Biochemistry **41**(46): 13782-13790.

Kruger, R., W. Kuhn, et al. (1998). "Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease." Nat Genet **18**(2): 106-108.

Lansbury, P. T., Jr. (1997). "Inhibition of amyloid formation: a strategy to delay the onset of Alzheimer's disease." Curr Opin Chem Biol **1**(2): 260-267.

Marsh, J. A., V. K. Singh, et al. (2006). "Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation." Prot. Sci. **15**(12): 2795-2804.

Miller, D. W., S. M. Hague, et al. (2004). "Alpha-synuclein in blood and brain from familial Parkinson disease with SNCA locus triplication." Neurology **62**(10): 1835-1838.

Norris, E. H., B. I. Giasson, et al. (2005). "Reversible inhibition of alpha-synuclein fibrillization by dopaminochrome-mediated conformational alterations." J Biol Chem **280**(22): 21212-21219.

Polymeropoulos, M. H., C. Lavedan, et al. (1997). "Mutation in the alpha-synuclein gene identified in families with Parkinson's disease." Science **276**(5321): 2045-2047.

Rochet, J. C. and P. T. Lansbury, Jr. (2000). "Amyloid fibrillogenesis: themes and variations." Curr Opin Struct Biol **10**(1): 60-68.

Spillantini, M. G., R. A. Crowther, et al. (1998). "alpha-Synuclein in filamentous

inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies." Proc Natl Acad Sci U S A **95**(11): 6469-6473.

Spillantini, M. G., M. L. Schmidt, et al. (1997). "Alpha-synuclein in Lewy bodies." Nature **388**(6645): 839-840.

Sung, Y. H. and D. Eliezer (2007). "Residual Structure, Backbone Dynamics, and Interactions within the Synuclein Family." J Mol Biol **372**(3): 689-707.

Ullman, O., C. K. Fisher, et al. (2011). "Explaining the Structural Plasticity of alpha-Synuclein." Journal of the American Chemical Society.

Uversky, V. N. (2007). "Neuropathology, biochemistry, and biophysics of alpha-synuclein aggregation." J Neurochem **103**(1): 17-37.

Uversky, V. N., M. C. E, et al. (2002). "Accelerated alpha-synuclein fibrillation in crowded milieu." FEBS Lett **515**(1-3): 99-103.

Uversky, V. N., J. Li, et al. (2001a). "Evidence for a partially folded intermediate in alpha-synuclein fibril formation." J Biol Chem **276**(14): 10737-10744.

Uversky, V. N., J. Li, et al. (2001b). "Metal-triggered structural transformations, aggregation, and fibrillation of human alpha-synuclein. A possible molecular NK between Parkinson's disease and heavy metal exposure." J Biol Chem **276**(47): 44284-44296.

Wang, W., I. Perovic, et al. (2011). "A soluble alpha-synuclein construct forms a dynamic tetramer." Proceedings of the National Academy of Sciences of the United States of America **108**(43): 17797-17802.

Weinreb, P. H., W. Zhen, et al. (1996). "NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded." Biochemistry **35**(43): 13709-13715.

Wood, S. J., J. Wypych, et al. (1999). "alpha-synuclein fibrillogenesis is nucleation-dependent. Implications for the pathogenesis of Parkinson's disease." J Biol Chem **274**(28): 19509-19512.

Wu, K. P., S. Kim, et al. (2008). "Characterization of conformational and dynamic properties of natively unfolded human and mouse alpha-synuclein ensembles by NMR: implication for aggregation." J Mol Biol **378**(5): 1104-1115.

Yavich, L., H. Tanila, et al. (2004). "Role of alpha-synuclein in presynaptic dopamine recruitment." J Neurosci **24**(49): 11165-11170.

Zarranz, J. J., J. Alegre, et al. (2004). "The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia." Ann Neurol **55**(2): 164-173.

Chapter 7

Polymer Theory for Intrinsically Disordered Proteins

7.1 Basic polymer models

There are significant advantages associated with the characterization of IDPs using polymer models (Bright, Woolf et al. 2001). Polymer theory, pioneered by Flory (Flory 1969), describe the statistical properties of chains. A number of different models have been adopted to characterize polymer chains.

The simplest model for polymer chains used for the description of polypeptide chains is the freely jointed chain model in which the chain is divided into N statistical segments or virtual bonds with a constant bond length l (Flory 1969). The polymer chain has a constant bond length with no correlations between the directions of the different bond vectors (Rubinstein and Colby 2003). The radius of gyration (Rg) for the freely jointed chain represented as follows:

$$\left\langle Rg^2 \right\rangle = \frac{1}{6Nl^2}$$

The worm-like chain model, also called the Kratky-Porod model (Kratky O 1949; Porod 1949) is especially useful for the description of stiff polymer chains like DNA. The average dimensions of the chain are dependent on the Persistence length (l_p), corresponding to the length over which the correlations between bond vectors are lost. The average Rg for the worm-like chain is calculated as:

$$\langle Rg^2 \rangle = \frac{1}{3}l_p l_c - l_p^2 + \frac{2l_p^3}{l_c}\left[1 - \frac{2l_p}{l_c}\left(1 - \exp\left(-\frac{l_c}{l_p}\right)\right)\right]$$

Here, l_p and l_c are the persistence and contour lengths respectively of the polypeptide chain.

7.2 Excluded volume and the effect of solvent quality on size of polymer chains

Real chains are characterized by interaction of monomeric units with each other and the solvent in which they are present. The interaction energy between two monomers is dependent on the attractive interactions between the monomers, relative to that of the solvent, and the repulsive barrier due to steric repulsion between overlapping monomers. The statistical properties of polymer chains are dependent on the quality of the solvent in which they are present. The excluded volume (v) is related to the kuhn length (b) and the diameter of the monomeric units (d).

For chains in good solvents, the chain is more expanded with the net attraction between monomers being weaker than the hard-core repulsion resulting in a positive excluded volume given by: $0 < v < b^2d$. The Rg of the chains in good solvent is dependent on the chain length (N) as

$$Rg = R_o N^{0.588}$$

where, R_0 is a constant dependent on the persistence length of the chain.

Chains in poor solvent are more collapsed with the net attraction between monomers being stronger than the hard-core repulsion resulting in a negative excluded volume given

by: $-b^2d < v < 0$. The Rg of these chains is related to the chain length as follows:

$$Rg = R_o N^{0.33}$$

Under special solvent conditions called theta solvent, the attractive and hard-core repulsive contributions to the excluded volume cancel each other resulting in a net zer excluded volume. Under these conditions, the Rg scales with chain length as

$$Rg = R_o N^{0.5}$$

It is important to note that Flory theory of polymer chains gives rise to the universal power law dependence of the Rg and the chain length as Rg   N , where   is the scaling exponent.

7.3 Charged polypeptide chains:

Heteropolymer chains with an even distribution of positively and negatively charged groups are classified as polyampholytes. Polypeptide chains like that of α-synuclein are characterized by the presence of charged amino acids, which possess a net charge under physiological conditions. Theory of polyampholytes is therefore very relevant to the study of charged polypeptide chains (Kantor, Kardar et al. 1998).

Charge balanced polyampholytes, with equal numbers of positive and negative charges, tend to form collapsed globules in solution due to fluctuation-induced attraction between opposite charges (Kantor and Karder 1994; Kantor and Karder 1995; Kantor, Kardar et al. 1998; Dobrynin, Colby et al. 2004). However, uneven distribution of charged residues along the chain have been shown to affect the shapes of disordered polypeptide chains (Khokhlov 1980; Kantor and Karder 1994; Dobrynin, Rubinstein et al. 1996; Bright,

Woolf et al. 2001). Polymer chains carrying net charges tend to expand, with the shape of the polymer conformations largely determined by the distribution of charges within the chain sequence. Charge imbalances satisfying the relation

$$N\left|f_+ - f_-\right| > \sqrt{fN}$$

where N is the number of residues, f_+ and f_- are the fraction of positively and negatively charged residues respectively, and $f = f_+ + f_-$, lead to polyampholytes forming a necklace globule conformation. The necklace is formed of beads with a net neutral charge connected by thin, charged strings. Simulation studies of polyampholyte chains (Kantor, Kardar et al. 1998; Dobrynin, Colby et al. 2004) have also shown that chains with very unbalanced charge distributions can form globules with one or more extended, charged fingers.

7.4 Persistence length of polymer chains

Local stiffness of polymer chains is quantified in terms of the persistence length, which corresponds to the length scale over which the chain behaves like a rigid rod (Grosberg and Khoklov 1994; Tran and Pappu 2006). Persistence length (l_p) is most commonly calculated as follows. Here, $l_{i(j)}$ is the length of bond vector i (j) and N if the chain length.

$$l_p = \left\langle \frac{1}{N} \sum_{i=1}^{N} \sum_{j=i}^{N} \frac{l_i \cdot l_j}{|l_i|} \right\rangle = \left\langle \sum_{i=1}^{N} \frac{l_i \cdot l_N}{|l_i|} \right\rangle$$

Persistence length is also estimated from a plot of the $\langle Cos\theta_{ij}\rangle$ as a function of the sequence separation ($|j - i|$) between bond vectors i and j, where l_p is calculated as the sequence separation at which $\langle Cos\theta_{ij}\rangle = e^{-1}$ (Thirumalai and Ha 1998).

A number of other approximate procedures are used to calculate the persistence length of

neutral polymer chains (Cifra 2004). Different estimates of the calculated persistence lengths are obtained from these different methods.

Polymer chains with a predominance of either positive or negatively charged residues, classified as polyelectrolytes, are characterized by greater chain stiffness. The extension of the worm-like chain model for stiff polyelectrolyte chains was introduced independently by Skolnick and Fixman, and Odijk (Odijk 1977; Skolnik J 1977). The total persistence length (P_T) for polyelectrolyte chain is calculated as the sum of the persistence length in the absence of electrostatic forces (P_o) and that of the electrostatic contribution to the persistence length (P_e):

$$\left\langle R_e^2 \right\rangle = 2l_c \left(l_0 + l_e \right), l_e = \frac{l_B f^2}{4 \left(kb \right)^2}$$

Here, f is the fraction of charged monomers, l_B the Bjerrum length (corresponding to the distance between the monomers at which the interaction energy becomes equal to the thermal energy), 1/k the Debye screening length and b the bond length.

Ha and Thirumalai (Ha and Thirumalai 1995) determined the persistence length for weak polyelectrolyte chains using a variational theory referred to as the uniform expansion method and have extended their theory to account for the crossover from stiff to flexible chains.

Persistence lengths have also been measured experimentally using a variety of techniques. These include force extension measurements from Atomic Force Microscopy (AFM), Small Angle X-ray Scattering (SAXS), Optical tweezers, NMR relaxation measurements, Forster Resonance Energy Transfer, quenching methods (Gast K 1995; Kellermayer MSZ 1997; Schwalbe H 1997; Rief M 1998; Rief M 1999; Yang G 2000; Schwarzinger, Wright et al. 2002; Hoffmann A 2007). Force extensions corresponding to

the mechanical unfolding of proteins studied using single-molecule AFM studies have been used to determine the persistence length l_p of proteins using the following interpolation formula (Bustamante C 1994; Marko JF 1995).

$$\frac{l_p F}{K_B T} = z/l_c + 1/4 \left(1 - z/l_c\right)^2 - 1/4$$

where T is the absolute temperature , K_B the Boltzmann constant, F the mechanical force applied and z the extension along the direction of the force.

The persistence lengths determined from AFM measurements range between 4 and 8 Å (Kellermayer MSZ 1997; Rief M 1998; Rief M 1999). The range of forces applied to unfold the protein has been shown to affect the magnitude of the persistence length (Rief M 1999). It has been suggested that the observation of different magnitudes of persistence lengths for different applied forces maybe attributed to the effect on the excluded volume interactions under these conditions (Zhou 2004).

Similar estimates of the persistence lengths have been obtained using other methods (Rief M 1999; Yang G 2000; Lapidus LJ 2002). Hoffmann et al have determined the persistence length of the collapsed, unfolded state ensemble of the cold shock protein CspTm to range from about 4 to 11 Å with change in the denaturant concentration (Hoffmann A 2007), calculated using the worm-like chain model. Buscaglia et al, using quenching experiments, have shown that in the presence of denaturant (8M Urea and 6M GdmCl), the persistence length of polypeptide chain reduces from 6 to 4 Å (Buscaglia M 2006). The presence of proline residues that leads to chain stiffening has been shown to increase the persistence length of the polypeptide chain (Schuler B 2005; Wells M 2008). Persistence length calculated from NMR relaxation measurements for a number of unfolded proteins has been shown to be about seven residues (Klein-Seetharaman J 2002;

Schwarzinger, Wright et al. 2002), similar to that estimated from SAXS measurements of about 10 residues (Gast K 1995). It has been proposed that the Residual Dipolar Couplings (RDCs) in unfolded proteins originate from the transient alignment of statistical segments, as long as the persistence lengths of these chains, against the orienting medium (Louhivuori M 2003; Mohana-Borges R 2004). More recently, it has been shown that the RDCs calculated based on the alignment of short, uncoupled segments of the chain – the Local Alignment Windows (LAWs) provides a better fit to experimental RDCs with a smaller number of conformers (Marsh JA 2008; Nodet, Salmon et al. 2009). The choice of the length of these LAWs has been correlated with the stiffness of the polypeptide chain, corresponding to the persistence length of the protein (Nodet, Salmon et al. 2009).

7.5 References

Bright, J. N., T. B. Woolf, et al. (2001). "Predicting properties of intrinsically unstructured proteins." Prog Biophys Mol Biol **76**(3): 131-173.

Buscaglia M, L. L., Eaton WA, Hofrichter J (2006). "Effects of denaturants on the dynamics of loop formation in polyepeptides." Biophys. J **91**: 276-288.

Bustamante C, M. J., Siggia E, Smith S (1994). "Entropic elasticity of lambda-phage DNA." Science **265**: 1599-1600.

Cifra, P. (2004). "Differences and limits in estimates of persistence length for semi-flexible macromolecules." Polymer **45**: 5995-6002.

Dobrynin, A. V., R. H. Colby, et al. (2004). "Polyampholytes." Journal of Polymer Science **42**: 3513-3538.

Dobrynin, A. V., M. Rubinstein, et al. (1996). "Cascade of transitions of polyelectrolytes in poor solvents." Macromolecules **29**: 2974-2979.

Flory, P. J. (1969). "Statistical Mechanics of Chain Molecules." InterScience.

Gast K, D. H., Schulze-Forster K, Maurer HR, Muller-Frohne M, Zirwer D, Czarnecki J, Damaschun G (1995). "Prothymosin α - A biologically Active Protein with Random Coil Conformation." Biochemistry **34**: 13211-13218.

Grosberg, A. and A. Khoklov (1994). "Statistical Physics of Macromolecules." AIP Press, New York.

Ha, B.-Y. and D. Thirumalai (1995). "Electrostatic persistence length of a polyelectrolyte chain." Macromolecules **28**: 577-581.

Hoffmann A, K. A., Nettels D, Hertzog DE, Baumgartel P, Lengefeld J, Reichardt G, Horsley DA, Seckler R, Bakajin O, Schuler B (2007). "Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy." Proc. Natl. Acad. Sci. USA **104**: 105-110.

Kantor, Y., M. Kardar, et al. (1998). "Necklace model of randomly charged polymers." Physica A **249**: 301-306.

Kantor, Y. and M. Karder (1994). "Excess Charge in Polyampholytes." Europhysics Letters **27**(9): 643-648.

Kantor, Y. and M. Karder (1995). "Instabilities of Charge Polyampholytes." Physical Review E **51**(2): 1299-1312.

Kellermayer MSZ, S. S., Granzier HL, Bustamante C (1997). "Folding-Unfolding Transitions in single titin molecules characterized with laser tweezers." Science **276**: 1112-1116.

Khokhlov, A. R. (1980). "Collapse of weakly charged poly-electrolytes." J. Phys. A **13**: 979-987.

Klein-Seetharaman J, O. M., Wirmer J, Ducardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002). "Long-range interactions within a non-native protein." Science **295**: 1719-1722.

Kratky O, P. G. (1949). "Röntgenuntersuchung gelöster Fadenmoleküle." Rec. Trav. Chim. **68**: 1106-1123.

Lapidus LJ, S. P., Eaton WA, Szabo A, Hofrichter J (2002). "Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: Testing a 1-Dimensional diffusion model for peptide dynamics." J. Phys. Chem. B 2002 **106**: 11628-11640.

Louhivuori M, P. K., Fredriksson K, Permi P, Lounila J, Annila A (2003). "On the origin of Residual Dipolar Couplings from Denatured Proteins." J. Am. Chem. Soc. **125**: 15647-15650.

Marko JF, S. E. (1995). "Stretching DNA." Macromolecules **28**: 8759-8770.

Marsh JA, B. J., Tollinger M, Forman-Kay JD (2008). "Calculation of Residual Dipolar Couplings from Disordered State Ensembles using Local Alignment." J. Am. Chem. Soc. **130**: 7804-7805.

Mohana-Borges R, G. N., Kroon GJA, Dyson JH, Wright PE (2004). "Structural characterization of unfolded states of Apomyoglobin using Residual Dipolar Couplings." J. Mol. Biol **340**: 1131-1142.

Nodet, G., L. Salmon, et al. (2009). "Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Dipolar Couplings." J. Am. Chem. Soc. **131**(49): 17908-17918.

Odijk, T. (1977). "Polyelectrolytes near rod limit." J. Polym. Sci. B Polym. Phys. **15**: 477-483.

Porod, G. (1949). "Zusammenhang zwischen mittlerem endpunktsabstand und kettenlange bei fadenmolekulen." Monatsh. Chem. **80**: 251-255.

Rief M, G. M., Schemmel A, Gaub HE (1998). "The Mechanical Stability of Immunoglobulin and Fibronectin III Domains in the Muscle Protein Titin Measured by Atomic Force Microscopy." Biophys. J **75**: 3008-3014.

Rief M, P. J., Saraste M, Gaub HE (1999). "Single Molecule Force Spectroscopy of Spectrin Repeats: Low Unfolding Forces in Helix Bundles." J. Mol. Biol **286**: 553-561.

Rubinstein, M. and R. H. Colby (2003). "Polymer Physics."

Schuler B, L. E., Steinbach PJ, Kumke M, Eaton WA (2005). "Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence." Proc. Natl. Acad. Sci. USA **102**(8): 2754-2759.

Schwalbe H, F. K., Buck M, Jones JA, Grimshaw SB, Spencer A, Glaser SJ, Smith LJ, Dobson CM (1997). "Structural and Dynamical Properties of a Denatured Protein. Heteronuclear 3D NMR Experiments and Theoretical Simulations of Lysozyme in 8M Urea." Biochemistry **36**: 8977-8991.

Schwarzinger, S., P. Wright, et al. (2002). "Molecular Hinges in Protein Folding: the Urea-Denatured State of Apomyoglobin." Biochemistry **41**: 12681-12686.

Skolnik J, F. M. (1977). "Electrostatic persistence length of a wormlike polyelectrolyte." Macromolecules **10**: 944-948.

Thirumalai, D. and B.-Y. Ha (1998). "Statistical Mechanics of semi-flexible chains: A mean field variational approach. A. Grosberg, Editor." Theoretical and Mathematical Models in Polymer Research, Academic Press, Boston: 1-35.

Tran, H. T. and R. V. Pappu (2006). "Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions." Biophys J **91**(5): 1868-1886.

Wells M, T. H., Rutherford TJ, Markwick P, Jensen MR, Mylonas E, Svergun DI, Blackledge M, Fersht AR (2008). "Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain." Proc. Natl. Acad. Sci. USA **105**(15): 5762-5767.

Yang G, C. C., Baase WA, Vetter IR, Breyer WA, Haack JA, Matthews BW, Dahlquist FW, Bustamante C (2000). "Solid-State Synthesis and Mechanical Unfolding of T4 Lysoyme." Proc. Natl. Acad. Sci. USA **97**(1): 139-144.

Zhou, H.-X. (2004). "Polymer Models of Protein Stability, Folding, and Interactions." Biochemistry **43**(8): 2141-2154.

Chapter 8

Investigation of the Polymeric Properties of α-Synuclein and Comparison with NMR Experiments: A Replica Exchange Molecular Dynamics Simulations

8.1 Introduction

Intrinsically Disordered Proteins like α-synuclein exist as rapidly interconverting ensemble of conformations in solution. Polymer theory is well suited for the description of the statistical properties of IDPs (Bright, Woolf et al. 2001; Vitalis, Wang et al. 2007). The nature of solvent influences the chain-chain and chain-solvent interactions, with chain-chain interactions favored under poor solvent conditions resulting in collapsed conformations, and chain-solvent interactions are preferred under good solvent conditions resulting in the swelling of polymer chains.

The average size, determined as the radius of gyration of polymer chains scales with chain length as $<R_g> \sim N^n$. The average size of homopolymeric chains scale as $N^{0.33}$ and $N^{0.59}$ under poor and good solvent conditions respectively. Proteins are heteropolymeric chains composed of up to 20 different monomeric units. Intrinsically disordered proteins are also characterized by the unbalanced distribution of charged residues resulting influencing the polyampholyte and/or polyelectrolyte behavior of chains. One focus of this work is to determine if the scaling laws of homopolymeric chains can be applied to heteropolymeric chains like α-synuclein. This is done by comparing the statistical

properties of α-synuclein under different solvent conditions, using temperature to modulate solvent quality, with that of the predictions for homopolymeric chains.

Persistence length, corresponding to the distance over which the memory of the direction of the chain persists, describes the average local stiffness of the polymer chain. It has been proposed, for experimental residual dipolar couplings (RDCs) arising due to the transient alignment of the chain, that the RDC patterns persist over length scales corresponding to the persistence length of the chain (Annila and Permi 2004; Mohana-Borges R 2004). Experimental measurements of the persistence length for unfolded proteins are reported to be in the range of 4 to 8 Å based on atomic force microscopy measurements (Zhou 2004). A number of approximate procedures are used to estimate the persistence length of polymer chains (Cifra 2004). In this study, persistence length of α-synuclein is determined for ensembles generated from replica exchange simulations; and the relationship between persistence length and the length scale over which angular correlations within polypeptide chains are determined.

Conformational characterization of the monomeric form of α-synuclein has been performed using a number of biophysical techniques and computational approaches (Eliezer 2009). Nuclear magnetic resonance (NMR) experiments hace been used extensively to probe the conformational characteristics of α-synuclein (Bussell and Eliezer 2001; Eliezer, Kutluay et al. 2001; Bertoncini, Jung et al. 2005a; Sung and Eliezer 2007; Wu, Kim et al. 2008; Wu, Weinstock et al. 2009). Residual dipolar couplings and paramagnetic relaxation enhancements (PREs) are commonly used to probe the structural properties of IDPs (Bernardó P 2005; Bertoncini, Jung et al. 2005b; Dedmon, Lindorff-

Larsen et al. 2005; Sung and Eliezer 2007; Ullman, Fisher et al. 2011). Conformational characterization of the monomeric state of α-synuclein using NMR suggests the presence of residual structure in solution (Bussell and Eliezer 2001; Bertoncini, Jung et al. 2005b; Dedmon, Lindorff-Larsen et al. 2005; Sung and Eliezer 2007; Wu, Kim et al. 2008), with the average size of the protein more compact than that expected for a random coil (Uversky, Li et al. 2001; Uversky 2002; Kim, Heise et al. 2007). The conformational characteristics for the α-synuclein ensemble reported on the basis of fits to PREs are inconsistent, with some studies suggesting conformations involving interactions between the N- and C-terminal regions (Bertoncini, Jung et al. 2005a; Ullman, Fisher et al. 2011) and others suggesting predom- inantly extended conformations for the C-terminal region (Wu, Weinstock et al. 2009). These observations highlight the underdetermined nature of fitting ensembles of IDPs to any given experimental parameter. In this study, conformational properties of monomeric α-synuclein are determined by fitting to both the experimental RDCs and PREs.

Statistical properties of α-synuclein determined as a function of solvent quality shows that at extremes of solvent quality, corresponding to the poor and good solvent conditions, α-synuclein scales as expected for homopolymer chains under these conditions. At intermediate temperature, the identity of the monomeric units and the charge distribution influence the polymer characteristics of α-synuclein. The conformational characteristics of α-synuclein determined by fitting to

experimental parameters consequently fit the local and long-range conformational characteristics. Back calculation of experimental RDCs left out of the fitting procedure show good fit to the experiment, validating the fitting procedure (Figure 8.1).
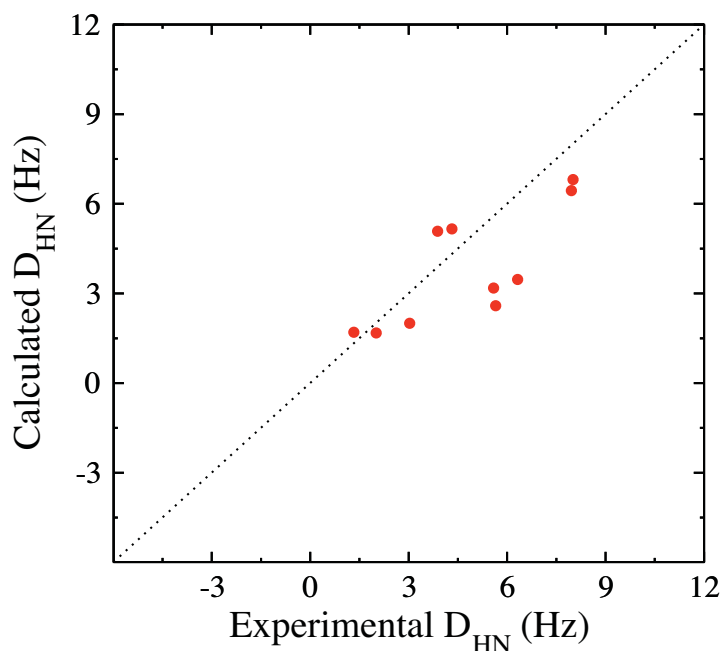

Figure 8.1: Comparison of the calculated RDCs (red dots) not used in the fitting procedure for the selection of α-synuclein ensembles fitting experiment.

8.2 Methods and Results

The conformational ensembles used in this study were generated using replica exchange molecular dynamics simulations over a series of twenty temperatures from 300-500K. To assess the convergence of simulations, standard errors of the means were calculated for the scaling of internal distances. Figure 8.2 shows the polymer scaling of internal distances plotted as a function of sequence separation for the low, intermediate and high

temperature ensembles. The error bars represent the calculated standard errors of the means. The low values of the calculated standard errors of the mean indicate that the simulations are well converged.
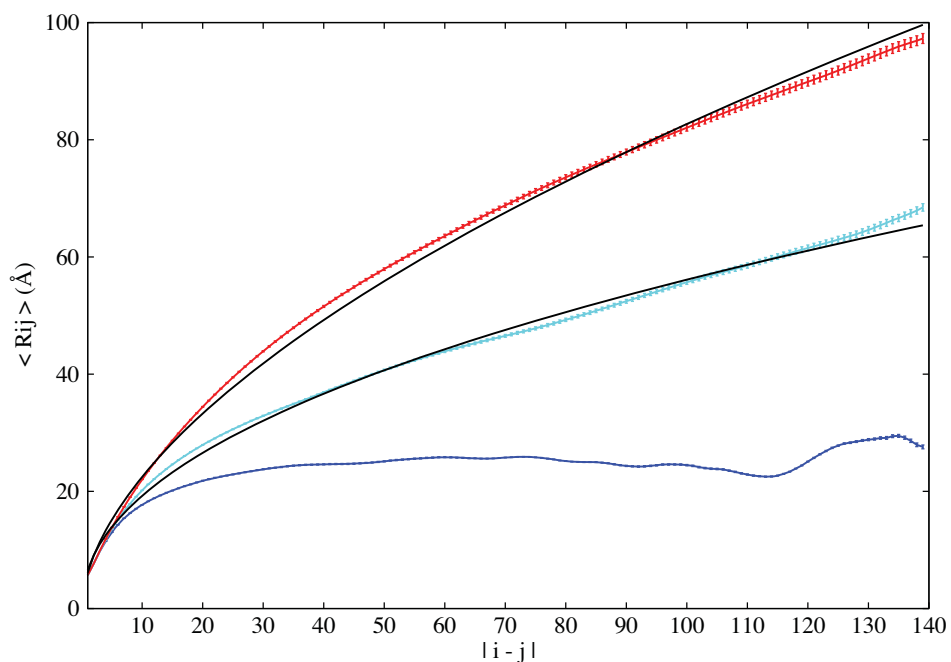

Figure 8.2: Scaling of internal distances as a function of sequence separation for the low (blue), intermediate (cyan) and high (red) temperature ensembles. Error bars represent the standard errors of the mean.

The procedure and results for this study are presented here as a reprint of the paper published in the *Journal of Chemical Theory and Computation, 2012*.

8.3 Publication Reprint

Investigation of the Polymeric Properties of α-Synuclein and Comparison with NMR Experiments: A Replica Exchange Molecular Dynamics Study

Chitra Narayanan,[†] Daniel S. Weinstock,[‡] Kuen-Phon Wu,[§] Jean Baum,[‡,§] and Ronald M. Levy[‡,§,*]

[†]Graduate Program in Biochemistry, Rutgers University, Piscataway New Jersey 08854, United States

[‡]BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway New Jersey 08854, United States

[§]Department of Chemistry and Chemical Biology, Rutgers University, 610 Taylor Road, Piscataway New Jersey 08854, United States

⑤ *Supporting Information*

ABSTRACT: Intrinsically disordered proteins (IDPs) have been shown to be involved in a number of cellular functions, in addition to their predominance in diseased states. α-Synuclein may be described as one such IDP, implicated in the pathology of Parkinson's disease. Understanding the conformational characteristics of the monomeric state of α-synuclein is necessary for understanding the role of the monomer conformation in aggregation. Polymer theories have been applied to investigate the statistical properties of homopolymeric IDPs. Here, we use Replica Exchange Molecular Dynamics (REMD) simulations using temperature as a proxy for solvent quality to examine how well these theories developed for homopolymeric chains describe heteropolymeric α-synuclein. Our results indicate that α-synuclein behaves like a homopolymer at the extremes of solvent quality, while in the intermediate solvent regime, the uneven distribution of charged residues along the sequence strongly influences the conformations adopted by the chain. We refine the ensemble extracted from the REMD simulations of α-synuclein, which shows the best qualitative agreement with experimental results, by fitting to the experimental NMR Residual Dipolar Couplings (RDCs) and Paramagnetic Relaxation Enhancements (PREs). Our results demonstrate that the detailed shapes of the RDC patterns are sensitive to the angular correlations that are local in sequence while longer range anticorrelations which arise from packing constraints affect the RDC magnitudes.

■ INTRODUCTION

Intrinsically Disordered Proteins (IDPs) have gained much attention in light of the finding that over 30% of the proteins encoded by the eukaryotic genome contain unstructured regions over 50 residues long.[1,2] The functional repertoire of IDPs includes signal transduction, transcription and translation, and protein complex assembly.[3] A number of neurodegenerative diseases including Alzheimer's, Parkinson's, and Huntington's diseases have been correlated with the aggregation of IDPs.[4] IDPs are best described as fluctuating ensembles of conformations in solution lacking a stable structure under physiological conditions[3] and are typically characterized by their low sequence complexity, high net charge, and low hydrophobicity.[5] Characterizing the conformational states of IDPs under physiological conditions is important for understanding their function, in addition to determining the driving forces that promote aggregation of these proteins in diseased states. Structural analysis of IDPs is challenging due to their highly flexible nature. NMR[6] and molecular simulations[7−10] can provide high-resolution data for the conformational characterization of disordered proteins.

α-Synuclein. The monomeric conformation of α-synuclein is described as a 140 residue polypeptide with classic IDP characteristics—low sequence complexity, low overall hydrophobicity with hydrophobic patches, and a high net charge.[5] The physiological function of α-synuclein has been attributed to acting as a chaperone to promote the assembly of large protein complexes,[11] vesicle transportation, and neurotransmit-

ter release.[12] The sequence of α-synuclein (Figure 1A) has an uneven distribution of charged residues along the chain and is divided into three regions—the N-terminal domain (residues 1−60) with a balanced distribution of positive and negatively charged residues corresponding to a polyampholyte chain, the Non Amyloid-beta Component of Alzheimer's disease (NAC, residues 61−95), which forms the hydrophobic core of the protein having minimal charged residues and the highly acidic C-terminal domain (residues 96−140) with a predominance of negative charges characteristic of a polyelectrolyte chain. α-Synuclein can adopt different conformations under various conditions. The N-terminal region of α-synuclein has been shown to form amphipathic helices upon binding to lipid membranes and micelles.[13] A conformational change to the monomeric form of α-synuclein leading to its aggregation into fibrils has been shown to play a key role in the pathology of Parkinson's disease.[14−16] Characterizing the monomeric form is important for understanding the conformational changes leading to the aggregated state.

The monomeric form of α-synuclein, long thought to be the dominant physiological species, has been studied using a variety of experimental and simulation techniques.[7] NMR experiments have been used extensively to probe the conformational

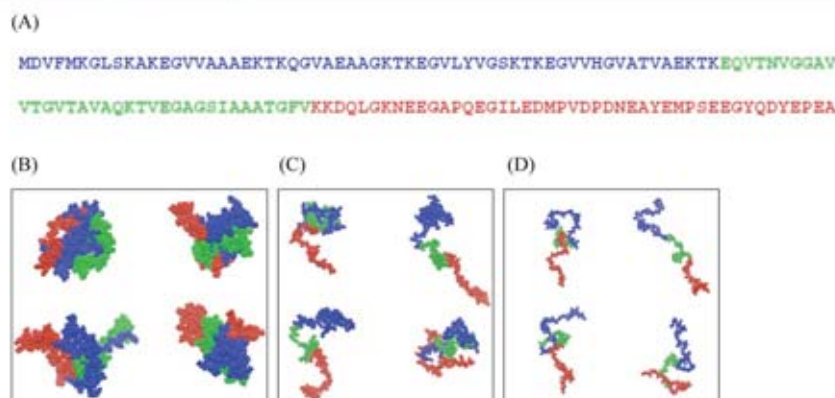**ACS Publications** © XXXX American Chemical Society

A

dx.doi.org/10.1021/ct300241t | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(A)

MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTKEGVVHGVATVAEKTKEQVTNVGGAV

VTGVTAVAQKTVEGAGSIAAATGFVKKDQLGKNEEGAPQEGILEDMPVDPDNEAYEMPSEEGYQDYEPEA

(B) (C) (D)

Figure 1. (A) Primary sequence of human α-synuclein. The N, NAC, and C-terminal regions are represented in blue, green, and red, respectively. (B–D) Representative conformations of α-synuclein selected for the low (B), intermediate (C), and high temperature (D) ensembles. The representative structures were chosen on the basis of the top four clusters, determined using the hierarchical clustering method.[101] The color scheme used here is the same as that in A.

propensities of α-synuclein.[10,17–21] Residual Dipolar Couplings (RDCs) and Paramagnetic Relaxation Enhancements (PREs) are the most commonly used measurements for probing the structural properties of this IDP.[19,20,22–24] NMR PRE measurements have been used as constraints in molecular dynamics simulations to generate ensembles consistent with experimental data.[23] A number of simulation studies have also reported the generation of ensembles based on the use of combinations of one or more experimental parameters including RDCs, PREs, chemical shifts, diffusion, and SAXS measurements.[10,20,22–25]

NMR studies of the monomeric state of α-synuclein suggest the presence of a residual structure in solution,[17,19–21,23] with the N-terminal region having a propensity to adopt transient helical conformations.[18,21] The average size of α-synuclein is more compact than would be expected for a random coil conformation,[2,26,27] arising as a result of transient long-range contacts between different domains.[19–21,23] The conformational characteristics of α-synuclein leading to its aggregation reported in the literature are disparate, with some studies suggesting the requirement of partially folded intermediate conformations for fibril formation while others hint that the release of long-range interactions promotes aggregation.[10,28] A more recent study suggests that exposure of the aggregation-prone region of the NAC (residues 8–18 in this region) in a non-negligible fraction of the ensemble could lead to the formation of a cross-beta structure.[24] The conformational characteristics for the α-synuclein ensemble reported on the basis of fits to PREs are inconsistent, with some studies suggesting conformations involving interactions between the N- and C-terminal regions[20,24] and others suggesting predominantly extended conformations for the C-terminal region.[10] These studies underscore the underdetermined nature of the problem; structural characteristics of IDPs are not uniquely determined by fitting any given experimental parameter. Addressing this degeneracy is necessary in order to provide a proper representation for the conformational characteristics of α-synuclein.

Polymer Properties of IDPs. The conformational characteristics of IDPs have been described in terms of their statistical properties using well-established concepts from polymer physics.[29,30] These statistical properties include chain descriptors such as the ensemble-averaged radius of gyration (R_g) and the persistence length (L_p).[31–33] The radius of gyration of homopolymer chains scales with chain length N as $\langle R_g \rangle = R_o N^\nu$, where R_o is a constant and the scaling exponent ν is dependent on the nature of the solvent.[34,35] In good and poor solvents, the scaling factors have been determined to be 0.59 and 0.33, respectively, for chains with excluded volume interactions between monomers. Proteins however are heteropolymeric chains composed of amino acids with varying side-chain chemical characteristics. IDPs characterized by their high net charge are more expanded due to charge repulsion.[9,36] How is the scaling affected for the heteropolymeric chain of α-synuclein, with an unbalanced distribution of charged residues resulting in regions with polyampholyte and polyelectrolyte characteristics? This question forms one focus of the current work.

Persistence length describes the average local stiffness of polymer chains, which corresponds to the distance over which the memory of the direction of the chain persists. The characteristics of the monomeric units making up the polypeptide chain can be expected to influence the magnitude of the persistence length. Experimental measurements of persistence lengths of unfolded proteins are reported to be within the range of 4 to 8 Å[37] based on atomic force microscopy measurements. A number of approximate procedures are used to calculate the persistence length of polymer chains.[38] For NMR residual dipolar couplings arising due to the transient alignment of the chain, it has been proposed that these patterns persist over length scales corresponding to the persistence length of the chain.[39,40] We have used REMD simulations to estimate the persistence length of α-synuclein, and we discuss the relationship between the persistence length of α-synuclein estimated from the simulations and the length scale over which angular correlations of the polypeptide chain affect the RDCs.

In this study, we focus on the following problems: (1) We compare the polymer theory of homopolymer chains to the statistical properties of heteropolymeric α-synuclein under different solvent conditions using temperature as a proxy for solvent quality. (2) We generate conformational ensembles of α-synuclein that fit the experimental RDC and PRE measure-

Article
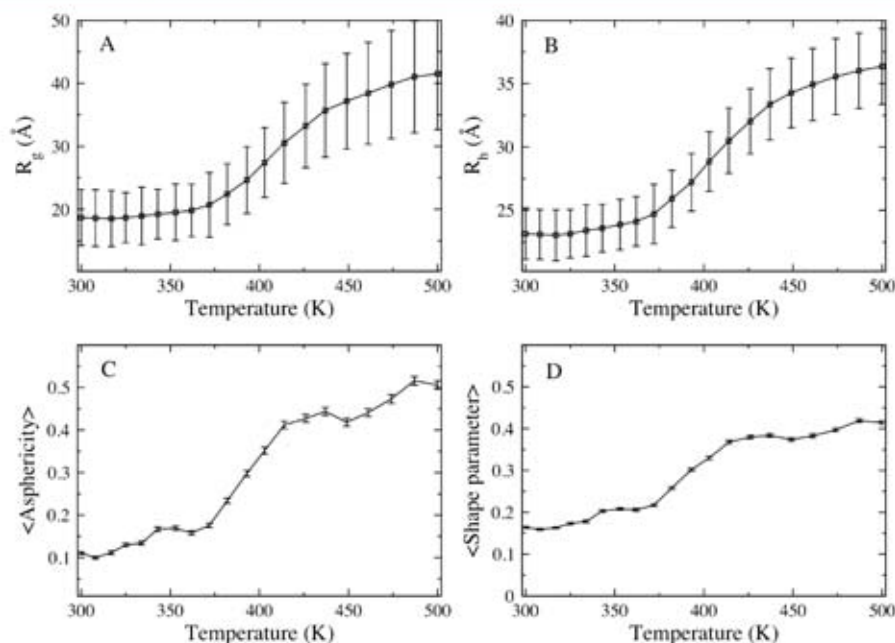
Figure 2. Global shape and size descriptors for α-synuclein. Average radius of gyration, R_g (A); hydrodynamic radius, R_h (B); asphericity, δ (C); and shape parameter, S (D) plotted as a function of simulation temperature. Standard deviations within ensembles are represented as error bars.

ments. (3) We analyze the relationship between the persistence length of α-synuclein and the length of the transiently aligning segments used in the calculation of RDCs from models of the α-synuclein conformational ensemble generated from REMD simulations. Our results demonstrate that at the extremes of solvent quality, α-synuclein scales as expected for a homopolymer chain under poor and good solvent conditions, while at intermediate values, the heterogeneity of the charge distribution and the identity of the monomeric units significantly influence the polymeric characteristics of the chain. We construct α-synuclein conformational ensembles which fit both the experimental RDCs and PREs, which consequently fit both the local and long-range conformational characteristics of the monomeric form of α-synuclein. Our results suggest that RDCs are sensitive to the angular correlations of the chain that are local in sequence while longer-range correlations act as a scaling factor which affects the magnitudes of the RDCs but has little effect on the RDC pattern of positive and negative RDCs.

■ RESULTS

Polymer Chain Characteristics of α-Synuclein. *Scaling Behavior of α-Synuclein.* REMD simulations were performed to generate the conformational ensembles of α-synuclein at neutral pH over a range of temperatures between 300 K and 500 K. Temperature has been used previously[41,42] to alter the conformational equilibrium of polypeptides in a way that mimics the effects of changing solvent conditions, with low temperatures corresponding to poor solvent conditions while higher temperatures correspond to good solvent conditions. In this work, we vary the temperature as a surrogate for changing solvent conditions. We note that the temperature scale over which structural changes occur is characteristic of implicit

solvent models like that used in this work.[43−45] Figure 2 shows the global size, represented by the radius of gyration (R_g) and hydrodynamic radius (R_h), and shape descriptors including the asphericity (δ) and shape parameter (S) as a function of the simulation temperature. Expansion of the α-synuclein ensemble is observed as the temperature is raised from 300 to 500 K. Here, we elaborate on the observations for the low, intermediate, and high temperatures represented by the 300, 414, and 500 K ensembles, respectively. At low temperatures, the ensemble adopts a collapsed conformation, with an average size of 17 Å (and 23.15 Å) for the R_g (and R_h), slightly more expanded than that expected for a well folded protein that is 140 residues long (R_g, ~15 Å and R_h, 20.5 Å)[46] with dimensions similar to those of a molten globule.[47] The asphericity (δ) and shape parameter (S) are around 0.1 and 0.15, respectively, values consistent with an approximately spherical chain shape (Figure 1B). At high temperatures, the ensemble adopts an extended conformation, with an average R_g of 41.5 Å (R_h of 36.36 Å), consistent with previous reports for the average size for the random coil conformation of α-synuclein[48] and that expected from scaling laws for unfolded conformations.[46,49] The asphericity and shape parameters for the high temperature ensemble are 0.5 and 0.42, respectively, consistent with a prolate ellipsoid or a cigar-shaped conformation. At intermediate temperatures, the conformational ensemble consists of a heterogeneous set of structures that span a range of sizes with both the average R_g and R_h of 30.5 Å at the midpoint. The δ and S values for the intermediate temperature ensemble are 0.41 and 0.37, respectively. We note that while the intermediate temperature ensemble adopts a smaller average size compared to high temperature, the shape characteristics (δ and S) at both temperatures are similar, with both ensembles adopting prolate ellipsoid conformations.
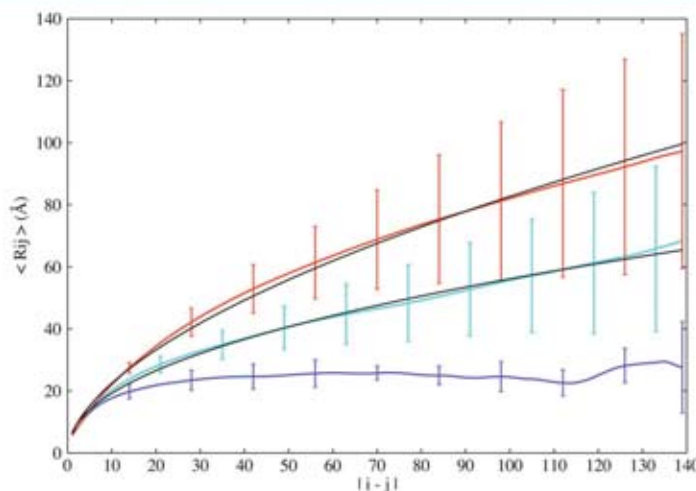
Figure 3. Scaling of average internal distances (R_{ij}) plotted as a function of sequence separation $|i - j|$ for the low (blue), intermediate (cyan), and high (red) temperature ensembles. The fits to the intermediate and high temperature ensembles, shown as solid lines, have scaling exponents of 0.46 and 0.57, respectively. Error bars represent standard deviations displayed here only for select residues along the sequence for clarity.
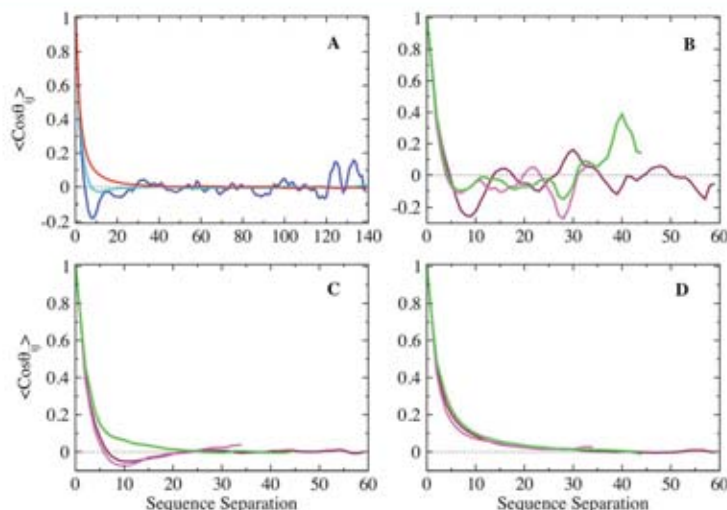
Figure 4. Ensemble averaged angular correlation function plotted as a function of sequence separation for the (A) low (blue), intermediate (cyan), and high (red) temperature ensembles, and the N (maroon), NAC (magenta), and C (green) terminal domains of the (B) low, (C) intermediate, and (D) high temperature ensembles.

Representative conformations for the low, intermediate, and high temperature ensembles are presented in Figure 1B–D.

The scaling of internal distances (R_{ij}) is another polymer parameter that is used to characterize the polymer chain under different solvent conditions. This parameter, similar to the average size (R_g and R_h), for homopolymeric chains follows the Flory scaling laws under different solvent conditions. R_{ij} is calculated as follows:

$$\langle R_{ij} \rangle = \left\langle \frac{1}{Z_{ij}} \sum_{m \in i} \sum_{n \in j} |\mathbf{r}_m^i - \mathbf{r}_n^j| \right\rangle \tag{1}$$

In eq 1, i and j are the amino acid indices, while m and n denote the atoms corresponding to residues i and j, respectively, and Z_{ij} is the total number of distances between the two residues. Figure 3 shows the plot of R_{ij} as a function of sequence separation. At low temperatures, R_{ij} plateaus; this is a signature of a collapsed chain, consistent with that expected for polymer chains in a poor solvent. At high temperatures, R_{ij} scales with sequence separation as $|i - j|^{0.57}$, approaching the theoretical scaling exponent of 0.59 expected for polymer chains in good solvent.[50] At intermediate temperatures, the polymer chain has a scaling exponent of 0.46.

The Angular Correlation Function. The angular correlation function, which provides insight into the topology of the chain

D
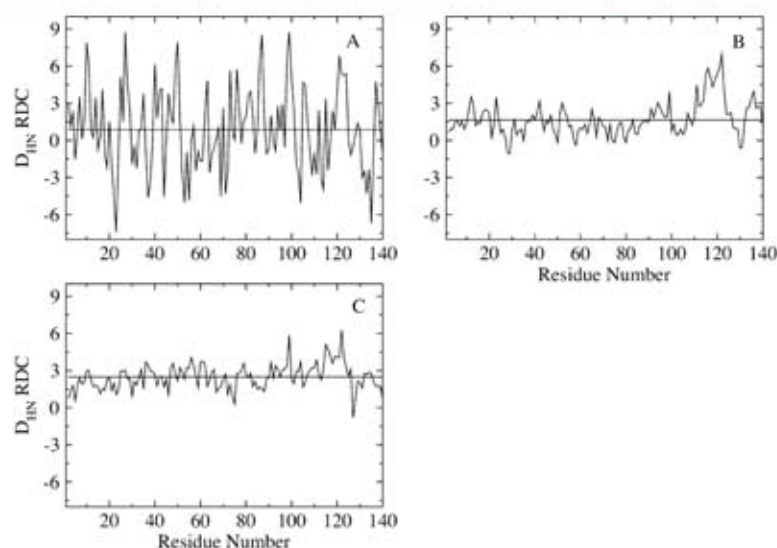
Figure 5. Residual dipolar couplings calculated on the basis of global alignment for the (A) low, (B) intermediate, and (C) high temperature ensembles. The horizontal lines represent the calculated RDCs averaged over the chains, with values of 0.86, 1.65, and 2.5 Hz for the low, intermediate, and high temperature ensembles, respectively.

under different solvent conditions,[30] is calculated as a function of sequence separation as follows:

$$\langle \cos \Theta_{ij} \rangle = \left\langle \frac{\mathbf{l}_i \cdot \mathbf{l}_j}{l^2} \right\rangle \tag{2}$$

where $\mathbf{l}_i$ ($\mathbf{l}_j$) represents the vector between the backbone nitrogen and carbonyl carbon of residue i (j) and l is the length of this bond vector. Figure 4A shows the ensemble averaged angular correlation function plotted as a function of sequence separation for the low, intermediate, and high temperature ensembles. The low temperature ensemble shows significant anticorrelations corresponding to chains with collapsed conformations, which on average reverse chain direction due to excluded volume effects. Strong correlations between residues separated by over 120 residues correspond to stable long-range interactions between residues in the N and C terminal regions. In contrast, the angular correlation function decays exponentially for the high temperature ensemble. The decay of the angular correlation function for the intermediate temperature ensemble is not exponential; it also shows anticorrelations similar to that observed for transient collapsed globular conformations, but suppressed in magnitude. Figure 4B−D show the angular correlations corresponding to the N, NAC, and C-terminal regions for the low, intermediate, and high temperature ensembles, respectively. Significant anticorrelations are observed for the three regions at low temperatures, while all three regions show exponential decay at high temperatures, consistent with random coil distributions of conformations. However, the N, NAC, and C-terminal regions show distinctly different topological characteristics at intermediate temperatures (Figure 4C). The N and NAC regions show nonexponential decay of the angular correlation function, with anticorrelations corresponding to chains under a packing restraint. The C-terminal region however exhibits more complete angular averaging, decaying exponentially with sequence separation at intermediate temperatures.

Persistence Length. Persistence length provides information about the local intrinsic stiffness of polymer chains. We have calculated the persistence lengths for the low, intermediate, and high temperature ensembles. The persistence lengths for the low and high temperature ensembles are ~3 Å and 12 Å, respectively, while the persistence length for the intermediate temperature ensemble is about 7 Å. To probe the effect of the heteropolymeric nature of α-synuclein on the persistence length, we calculated the persistence lengths separately for the N, NAC, and C-terminal domains of the low, intermediate, and high temperature ensembles. A comparison of the N, NAC, and C-terminal regions shows no significant difference in the persistence lengths observed for the three regions under the extreme (low and high temperature) conditions. However, at intermediate temperatures, the N-terminal and NAC regions have a persistence length of around 6 Å, while the C-terminal region has a persistence length of 11 Å, which also shows single exponential decay for the angular correlation function, corresponding to a more extended random coil conformation in this region. The larger persistence length for the C-terminal region at intermediate temperature suggests that this region of the chain is significantly stiffer than the N and NAC regions, consistent with the high negative charge density in the C-terminal domain.

Residual Dipolar Couplings. Figure 5 shows the residual dipolar coupling calculated using global alignment for the low (A), intermediate (B), and high temperature (C) ensembles. The RDCs are described below in terms of both (1) the range of values (difference between the magnitudes of the largest and smallest RDCs) and (2) the average values of the RDCs when averaged over the chain. The range of the RDCs is larger (~9 Hz) at low temperatures, with both positive and negative values, while the RDCs averaged over the chain are close to zero (~0.86 Hz). At intermediate and high temperatures, the RDCs have a much smaller range compared to the low temperature values and are almost all the same sign (positive).
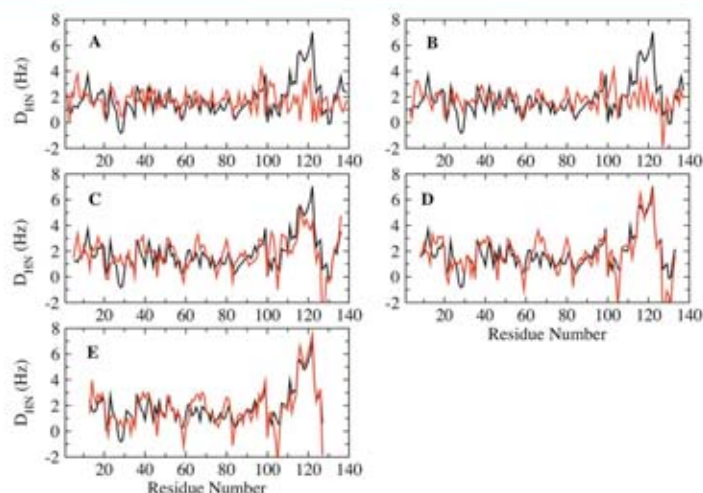
Figure 6. Correlation between the RDCs calculated from global (black) and local (red) alignments using LAWs lengths 3 (A), 5 (B), 9 (C), 15 (D), and 25 (E).
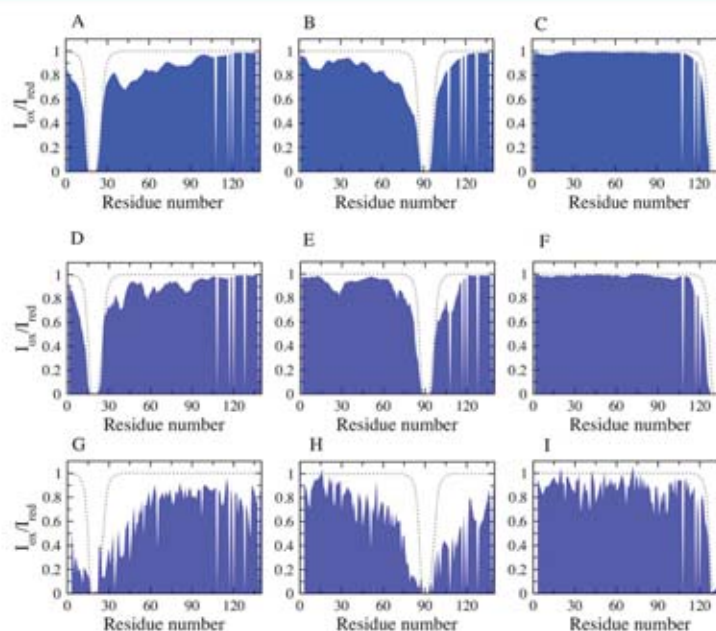

Figure 7. Conformational characteristics of the simulation ensembles correlating with experimental results: PREs for the REMD ensemble (A, B, C), reconstructed ensemble (D, E, F), and NMR experimental data (G, H, I) for spin label at positions A19, A90, and G132 respectively. The dotted lines represent the theoretical PRE values calculated for α-synuclein with no long-range contacts, determined as described previously.[21]

The RDCs at high temperatures (Figure 5C) are the most uniform along the chain with higher magnitudes in the middle of the chain; the end segments have smaller values, consistent with those calculated for random flight chain models.[51] The RDC averaged over the chain is also larger, ~2.5 Hz, at high temperatures. At intermediate temperatures, the C-terminal region exhibits larger RDCs, while the N-terminal and NAC regions show relatively smaller values (Figure 5B).

The concept of the Local Alignment Window (LAW) has been proposed previously, according to which RDCs are calculated by aligning short segments of the chain, instead of the whole chain, against the orienting medium.[52,53] This approach has been proposed to provide a good representation of the disordered state of polypeptide chains using fewer structures. We have analyzed the intermediate temperature ensemble to determine the LAW length best suited for calculating RDCs, by comparing the ability of LAWs of

F
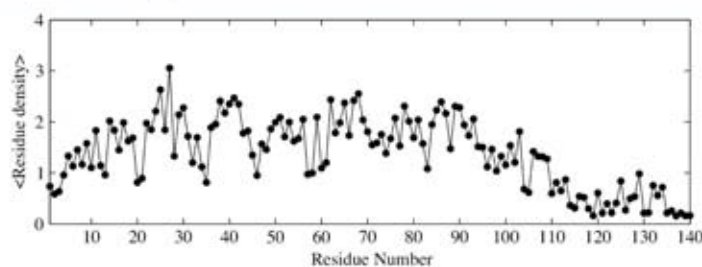
Figure 8. Residue density along the sequence for the intermediate temperature ensemble of α-synuclein. The residue density is calculated as the count of the average number of residues within 7 Å of the side chains of any residue along the sequence.
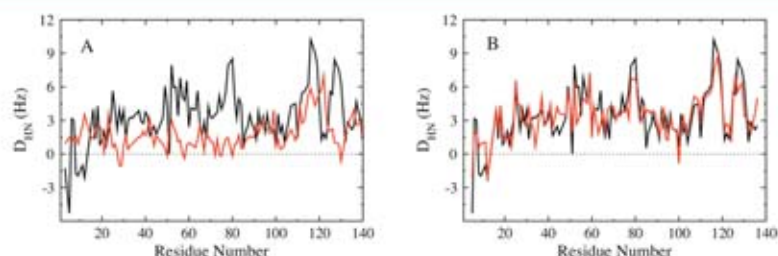
Figure 9. Comparison of the experimental HN RDCs (black) with (A) the average RDC determined from global alignment of the intermediate temperature REMD ensemble (red) and (B) RDCs calculated using local alignment for the weighted subset of the reconstructed ensemble (red).

different lengths to reproduce the global alignment average. LAW lengths of 3, 5, 9, 15, and 25 residues were used for the calculation of the RDCs. The average RDCs of the different window lengths were scaled to fit the global alignment average. Figure 6A−E shows the accuracy of the RDCs calculated using the different LAWs compared to the global alignment average. These results show that LAW lengths of nine residues and longer reproduce the global alignment average reasonably well for the entire length of the polypeptide chain, especially in the C-terminal region, which is not well fit by a smaller alignment window.

Comparison between Simulations and NMR Experiments. The presence of residual structure and long-range contacts involving the different regions of α-synuclein at neutral pH have been observed experimentally using NMR PREs.[10,19−21] The choice of the REMD ensemble for comparison with NMR experiments was based on the average hydrodynamic radius calculated for the simulation ensemble which best matches the experiments.[10,21] The intermediate temperature neutral pH REMD ensemble has an average R_H of ~30.5 Å, consistent with the size determined experimentally from PFG NMR diffusion measurements.[21] Figure 7 shows the comparison of the back-calculated PREs (Figure 7A−C) with that of the NMR data (Figure 7G−I). The calculated PREs are in reasonably good agreement with the experimental observations.[10] This ensemble is characterized by a heterogeneous set of structures, with transient local and long-range interactions involving the residues in the N and NAC domains. The C-terminal residues (121−140 in particular), containing eight negatively charged residues, show very few interactions with the rest of the protein chain.

The lack of any significant transient interactions involving the last 20 residues of α-synuclein is also observed in the residue density plots shown in Figure 8. Residue densities are obtained by calculating the average number of residues whose side chains

are within 7 Å of any residue along the sequence. We note that the contributions from neighboring residues (up to five residues on either side of the given residue) were ignored in the calculation of the average densities. The residue density is significantly greater along the sequence for the N and NAC regions and up to residue 110 in the C-terminus. The regions with the highest average densities correspond to short stretches of hydrophobic residues. The residue density for the residues in the C-terminus is minimal, indicating few contacts of residues in this region with the rest of the chain, consistent with the PRE data.

The Conformational Ensemble of α-Synuclein That Best Fits the Experimental RDCs and PREs. The REMD ensemble at intermediate temperature shows qualitative agreement with the average size and residue contacts (PREs) observed experimentally. However, the calculated average RDCs for this ensemble, shown in Figure 9A, are not in good agreement with the experimental data. The quality of the fit is assessed using the Q-factor,[54] calculated as the ratio of Root Mean Square (RMS) deviation between the experiment and calculated ensemble to that of the RMS average of the experimental RDCs. Q = 0.78 for the REMD ensemble. Smaller values of the Q-factor correspond to a better fit, with a value between 0.1 and 0.3 observed for well ordered proteins.[55,56] To construct an ensemble of α-synuclein conformations with better agreement with experimental RDCs, we refine the conformational ensemble of α-synuclein using a reweighting approach elaborated in the methods section. Figure 9B shows that the average RDCs calculated for the reweighted ensemble are in good agreement with the experimental data. The reconstructed ensemble has a Q-factor of 0.35, showing a substantial improvement in the fit to experiment when compared with the fit from the original ensemble. The back-calculated RDCs left out of the fitting procedure are also in good agreement with experimental results (Figure S1),

indicating that the data are not overfit. The Q-factors for the comparison of the REMD ensemble with experimental results for the three spin label positions (A19C, A90C, and G132C) are 0.37, 0.3, and 0.18, while the Q-factors for the comparison between the reconstructed ensemble and experimental results are 0.24, 0.4, and 0.16, respectively. While the RDC values have changed, the calculated persistence length and the global shape characteristics of the reconstructed ensemble still closely resemble those of the original REMD ensemble. The persistence length for the reconstructed ensemble is ~6 Å, while the global shape characteristics reported as the average R_g (R_h), asphericity, and shape parameter values are 30.7 Å (30.9 Å), 0.35, and 0.39, respectively. The long-range conformational characteristics of the reconstructed ensemble, estimated from the back-calculated PREs, are similar to that of the experimental PREs (Figure 7D−F).

■ DISCUSSION

Polymer Theory Provides a Statistical Description of Conformational Properties of α-Synuclein. We have used changes in temperature as a proxy for solvent quality to explore the polymeric properties of α-synuclein under different conditions which induce varying degrees of compaction. We report the average size in terms of the radius of gyration and the hydrodynamic radius, for comparison with measurements from SAXS and NMR diffusion experiments, respectively. The relationship between R_g and R_h is dependent on the nature of the solute−solvent interactions. In poor solvents, R_g is smaller than R_h, while the inverse is true in a good solvent. Limiting ratios are provided by native, globular proteins for which R_g/R_h = 0.775[57] and excluded volume chains where R_g/R_h = 1.5.[58] However, the ratio under strongly denaturing conditions, determined experimentally, has been reported to be 1.06,[49] smaller than 1.5, predicted from the theoretical Zimm relationship.[58] For the low and high temperature ensembles for α-synuclein, we calculate R_g/R_h ratios of 0.808 and 1.14, respectively, values which are very close to those reported previously.[49,57]

The limiting values for the average size of heteropolymeric α-synuclein at the lowest and highest temperatures agree with predictions from the Flory theory of homopolymeric chains (Table 1). This is consistent with experimental observations that globular proteins in their native states behave like homopolymers in poor solvents, while denatured proteins scale like homopolymers in good solvents.[49] The R_g calculated for the intermediate temperature ensemble (30.5 Å) is different from that reported from SAXS measurements (~ 40 Å),[59] while the R_h value closely matches the NMR measurements. It

appears that the SAXS measurement of the size is not consistent with the estimates of size from NMR measurements. The larger value of R_g reported from SAXS measurements is in contrast to values expected for polypeptide random coil models (41.5 Å) and close to a previous report of random coil simulations of α-synuclein[23] where the attractive dispersion interaction between residues was turned off.

Although the scaling of chain size is insensitive to the sequence at the extremes of solvent quality, the effects of chain heterogeneity are significant at intermediate solvent quality, represented here by the intermediate temperature ensemble of α-synuclein. This ensemble is heterogeneous with the N, NAC, and C-terminal regions exhibiting distinctly different polymeric properties. Note that the internal distances at intermediate and high temperatures are overlapping (Figure 3). The C-terminal region of the protein has a predominance of acidic residues, with a net charge of −8 at neutral pH. It is mostly extended with very few contacts with other regions of the chain. The extended conformation is consistent with the expected polyelectrolyte behavior of a chain with a high net charge.[60,61] In contrast, the N-terminal region of the chain with a relatively large total charge density but small net charge shows transient contacts with other parts of the chain. Polyampholyte chains with unbalanced charges have also been shown to form globules in the charge-balanced regions, and one or more charged fingers corresponding to the charged regions,[62] similar to conformations observed for the N-terminal domain of α-synuclein in the intermediate temperature ensemble.

The observation of transient contacts between the N-terminal and NAC regions for the intermediate temperature ensemble suggests that these interactions may be influenced by favorable electrostatic interactions through the formation of salt bridges. The calculation of distances between all charged side chains within a distance of 4.5 Å however shows salt bridge interactions in fewer than 10% of the population, for the intermediate temperature neutral pH ensemble. These ion-pair interactions are observed mostly between residues local in sequence rather than between distant residues, suggesting locally collapsed regions, as shown for polyampholyte chains.[63]

The effect of the charge-balanced state on the collapsed conformation of the N-terminal region of α-synuclein at neutral pH becomes more evident when compared to the conformations of this region at low pH (see Supporting Information). The low pH ensembles were generated for a study published previously[10] and are used here to highlight the effect of charged residues on the conformational characteristics of α-synuclein. With a shift in pH from low to neutral pH, the residue density in the N-terminal region increases with a corresponding increase in the charge density (but lower net charge), while the C-terminal shows a decrease in residue density with an increased charge density (and higher net charge; Figure S2). Schuler et al.[36] showed the collapsed conformations of charge-balanced polypeptides due to attractive interactions between the opposite charges, consistent with our observations for the N and NAC regions, also consistent with observations from the theory of polyampholyte chains. In contrast, the expanded conformation of the C-terminal region at neutral pH with an increase in charge density and the corresponding net charge arises due to charge repulsion, a result supported by similar observations reported by Pappu et al.[9]

Correlating the Polymer Properties of α-Synuclein with the Calculated Residual Dipolar Couplings. By

Table 1. Comparison between the Expected R_h (R_g) and the Calculated R_h (R_g) from REMD Simulations[a]

REMD ensemble	expected R_h (R_g) Å	REMD simulation R_h (R_g) Å
low temperature	20.1 (15.1)	23.15 (17.0)
intermediate temperature	30.8	30.5
high temperature	35.1 (40.7)	36.36 (40.5)

[a]The expected values for R_h are determined using empirical equations which are based on a comparison of the hydrodynamic radii for a number of globular, disordered proteins and proteins under strongly denaturing conditions.[46,49] The R_g is calculated from the relationship $R_g = R_0 N^\nu$. The values for R_0 are used from previous reports of this constant value for globular[99] and unfolded[100] proteins.

changing the solvent quality using temperature as a proxy, α-synuclein adopts a variety of conformations—near-spherical conformations at low temperatures, while, at intermediate and high temperatures, akin to good solvent conditions, the chain adopts prolate ellipsoid conformations. The RDC averaged over the chain is close to zero (0.86 Hz) at low temperatures; this average increases continually with increasing temperature to ~2.5 Hz at the highest temperature (Figure 5). These observations highlight the strong influence of the shape of the chain on the averaged RDCs, with the average values increasing with increasing asphericity of the chain. This effect of the shape of the chain on average RDCs is also made clearer by comparisons of the angular correlations under these conditions (Figure 4A). Significant (anti)correlations observed at low temperatures, corresponding to collapsed conformations, correlate with RDCs with both large positive and negative values, whose average is close to zero. In contrast, the RDC patterns at intermediate and high temperatures exhibit patterns similar to random flight chains,[51,64,65] where the RDCs are observed to all have the same sign, larger in the middle of the chain, with smaller RDCs at the ends. The corresponding angular correlation functions under these conditions show exponential decay. The average RDC at intermediate temperature within the N, NAC, and C-terminal regions (Figure 4C) can also be related to the characteristic shapes within the three regions. The N and NAC regions, which show significant anticorrelations, have smaller average RDCs, while the C-terminal region, which shows a single exponential decay of the angular correlation function for segments within this region, corresponds to more extended conformations and greater asphericity in this region, leading to higher average RDCs (2.47 Hz). The range of the RDCs in the C-terminal region is also larger than those of the values of the N and NAC regions.

We note that while the averaged RDCs increase with increasing temperature, the range of RDCs sampled (both positive and negative) by individual residues is significantly larger at low temperatures (up to ~9 Hz). The large range of the calculated RDCs suggests that there is local structural regularity within chains at low temperatures. In contrast, at intermediate and high temperatures, where structures with more diverse conformational characteristics are sampled, the RDC pattern is closer to that of a random flight chain. The high temperature RDC pattern also shows a flat distribution of RDCs along the middle of the chain, while terminal residues have smaller RDCs, consistent with the results from models of unfolded proteins described using random flight chains.[51,64]

The persistence length reflects the local chain stiffness and is influenced by a number of factors, including the nature of the monomeric units making up the polypeptide chain and the temperature or solvent quality. For a heteropolymeric chain like α-synuclein, the stiffness varies along the chain due to the variation in the side chain groups and the resulting interactions between different segments of the chain with each other and with the solvent.[29,66] Our results show that the persistence length of α-synuclein increases from 3 to 12 Å going from low to high temperatures, corresponding to a shift from poor to good solvent conditions. While the heteropolymeric nature of the chain has a negligible effect on the persistence length observed for the N, NAC, and C-terminal domains at the lowest and highest temperatures (data not shown), significant differences are observed at intermediate temperatures. For this ensemble, the N and NAC regions have a persistence length of about 6 Å, while the C-terminal one shows a much larger

persistence length of about 11 Å. This difference in the persistence lengths of the different regions can be explained on the basis of the charge pattern, which is best described for the N-terminal region as a polyampholyte chain, with an unbalanced distribution of charges, while the C-terminal region has a predominance of acidic residues. The C-terminal region of the protein is characterized by the presence of five proline residues, which also increases the chain stiffness.[67]

Residual dipolar couplings in intrinsically disordered proteins have been proposed to originate due to transient alignment of short segments of the chain, the length of which corresponds to the persistence length of the polymer chain.[39,64] Polymer persistence length has also been correlated with the size of the Local Alignment Window (LAW), a recently introduced concept,[52,53,68] which allows one to align individual fragments corresponding to the windows local in sequence separately to the ordering frame, rather than aligning the entire polypeptide chain. Analysis using the intermediate temperature ensemble shows that local alignments with LAW lengths of nine residues and higher reproduce the RDCs calculated from the global alignment data well (Figure 6). However, average angular correlations for α-synuclein at intermediate temperatures extend beyond nine residues. Our results show that for the intermediate temperature ensemble, the RDC pattern (the signs of the RDCs and the range) is sensitive to angular correlations that decay on a short length scale (over which the angular correlation function decays to $1/e$ of it is initial value) but that the longer range anticorrelations between HN groups separated by 10 or more residues have little effect on the RDC pattern. The longer-range anticorrelations, however, are much smaller compared to that observed for the low temperature ensemble, indicating the lack of structural regularity of chains under these conditions. The results imply that the analysis of RDCs using local alignment windows is appropriate when the polypeptide chain obeys chain statistics that do not deviate markedly from random flight chain statistics.

Fitting to Both the Experimental RDCs and PREs Provides a Better Representation for the Conformational Ensemble of α-Synuclein.

The structural characterization of α-synuclein has been carried out using a variety of experimental and computational methods.[23,28,69−71] We have previously shown that intermediate temperature ensembles obtained from REMD simulations fit experimental NMR measurements for model peptides.[45] Qualitative agreement of the α-synuclein conformational ensemble at intermediate temperatures with experimental PREs[10] was also shown in a previous study where we noted the underdetermined nature of the problem of fitting conformational ensembles to experimental PREs, resulting in different representative ensembles fitting the same PREs.[10] A common problem encountered while constructing conformational ensembles for IDPs like α-synuclein is that many different ensembles with varying conformational propensities can fit any given experimental parameter. A number of approaches have been devised to deal with this problem of degeneracy of conformational ensembles,[72−74] but care must be taken to prevent overfitting the experimental data.[53,74] It has been suggested that the degeneracy problem encountered when fitting RDCs can be removed by fitting PREs as well.[75]

In our approach, ensembles were generated repeatedly by selecting sets of 50 structures from a large pool of conformers, with the subsequent introduction of a set of weights for each conformer that are adjusted to fit the RDC data. Cross-
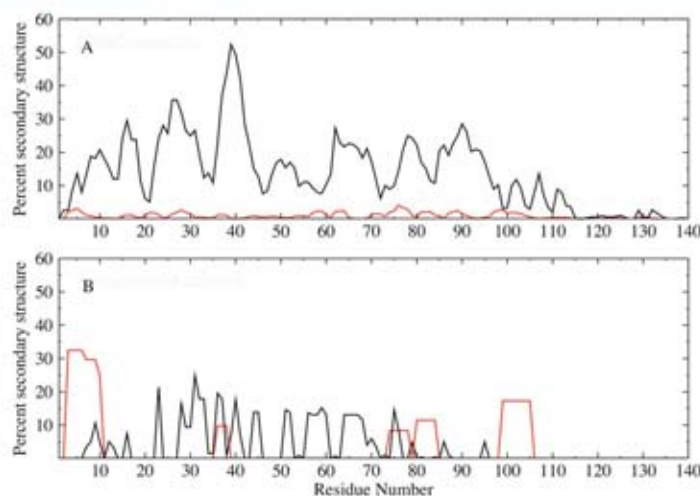
Figure 10. Ensemble averaged helix (red) and beta strand (black) propensities for the original REMD (A) and reconstructed (B) α-synuclein ensembles. Helix propensies were calculated by combining the alpha pi and 3−10 conformations, while the strand propensities were obtained by combining the extended and bridge conformations.[77]

validation has been used to avoid overfitting the data (Figure S1). From the many ensembles that can be constructed in this way, we calculate the fit of each ensemble to the long-range PREs and then choose as the most representative ensemble the one with the best fit to both the short-range RDC data and the long-range PRE data. We find that ensembles fitting only the RDCs do not always reproduce the global shape and size characteristics. Previous studies of α-synuclein, based on fits to experimental RDCs and PREs, suggest that RDCs agree better with experimental results when the long-range information from PREs is explicitly included to fits obtained from local alignment windows.[25,76] We calculated RDCs based on local alignments and further screened ensembles fitting the long-range PREs. The reconstructed ensemble fitting both of these parameters consistently retains the global shape and size properties observed for the REMD ensemble, in contrast to ensembles fitting either just the RDCs or the PREs alone.

A comparison of the conformational characteristics between the reconstructed ensemble fitting both RDC and PRE measurements and the original REMD ensemble shows a significant change in the secondary structural properties, assigned using STRIDE[77] (Figure 10). The helical propensity of the reconstructed ensemble is enhanced upon fitting to the RDCs, with helicities being prominent over short stretches along the sequence including residues 1−13, which has been shown to form helices under crystallization conditions.[78] The strand propensity of the reconstructed ensemble is marginally diminished compared to the REMD ensemble and is prominent in the N-terminal and NAC domains, encompassing regions predicted to have strand propensity in previous studies.[79]

This study focuses on the conformational characteristics of the monomeric disordered state of α-synuclein, which was long thought to be the physiological form of this protein. Two reports have challenged this view by suggesting that the physiological form of α-synuclein might actually be a stable helical tetramer,[80,81] while the most recent studies under similar conditions have been unable to confirm the report that α-synuclein forms a stable helical structure in mammalian cells.

Instead, evidence is presented that that α-synuclein remains a disordered monomer under cellular conditions.[82] It remains to be seen if the helical tetramer is indeed the physiological form of α-synuclein. While this is still debated, it is essential to note the importance of the monomeric disordered state of α-synuclein. Changes in the environment change subpopulations of conformations in the monomeric state,[10] which is presumed to eventually lead to aggregation. The monomeric conformations of α-synuclein have also been shown to associate with each other to form transient dimers[83] which can proceed to form higher order aggregates and fibrils.

Implications for Association and Aggregation. Conformational characterization of the monomeric state of α-synuclein by experiment and computation has been performed in order to determine structural characteristics that might be associated with the aggregation propensity of α-synuclein. A variety of factors including secondary structure propensity[19] and long-range interactions[10,20,23,24,48] have been linked to the aggregation properties of α-synuclein. Our intermediate temperature ensemble, reconstructed by fitting both the experimental RDCs and PREs, shows the secondary structural propensity to be predominantly turn and coil-like, and a small average propensity for α-helical and β-strand conformations (Figure 10B). The propensity for helical conformations over short stretches along the chain suggests that these helical stretches could potentially act as seeds and accelerate the formation of a long contiguous helix under suitable conditions, including the lipid bound state[13] and the recently suggested tetrameric helical state.[80,81]

The observation of nonexponential decay of the angular correlations (including a sign change) in the N and NAC domains points to a more collapsed conformation of these regions consistent with the transient long-range interactions observed between the N and NAC domains, while the C-terminal domain is largely extended. It seems likely that the collapsed conformation of the N and NAC regions, through transient contacts within these regions, will retard intermolecular chain interactions. Coarse-grained modeling studies of

Article

polypeptide chains have suggested that stabilizing aggregation-prone conformations of chains can result in the formation of ordered fibrils with on-pathway oligomeric intermediates.[84] The propensity of α-synuclein to populate helical conformations in the N-terminal region is suggestive of a helix-mediated association between chains which can facilitate favorable intermolecular interactions between the aggregation-prone, hydrophobic NAC region, consistent with ideas proposed recently for natively unfolded chains[85] and shown for α-synuclein in the form of helical intermediates[86] under some conditions. It appears that in it is monomeric state, α-synuclein shows a propensity for adopting a variety of conformations, which can facilitate either the formation of ordered structures or association to form higher order aggregates under suitable cellular conditions.

■ MATERIALS AND METHODS

Setup of Replica Exchange Molecular Dynamics Simulations. We have used Replica Exchange Molecular Dynamics (REMD)[87,88] simulations to generate the conformational ensembles of α-synuclein. In this approach, a number of replicas are run in parallel over a specified temperature range. Adjacent replicas (T_i and T_j) are allowed to exchange periodically, with an acceptance criterion based on the following Metropolis transition probability.

$$W\{T_i, T_j\} \rightarrow \{T_j, T_i\} = \min(1, \exp[-(\beta_i - \beta_j)(E_i - E_j)]) \quad (3)$$

where, $\beta_{i(j)} = 1/KT_{i(j)}$ and $E_{i(j)}$ is the potential energy of the ith (jth) replica. This method generates canonical probability distributions for the ensembles over the specified temperature range. The REMD method has been implemented in the IMPACT simulation package.[89] Simulations were performed using the AGBNP implicit solvent model[90] and the OPLS-AA force field.[91]

All simulations were initiated with a fully extended conformation of the α-synuclein molecule. The simulations start with a short minimization using the conjugate gradient method followed by a production run for a total of 25 ns each over 20 replicas at the following temperatures: 300, 308, 317, 325, 334, 343, 353, 362, 372, 382, 393, 403, 414, 426, 437, 449, 461, 474, 487, and 500 K. The molecular simulation time step was 1.5 fs, and exchanges were attempted every 1 ps. The cumulative simulation time, with a total of 25 ns for each of the 20 replicas, corresponds to a total of 500 ns.

Analysis of the Global Shape Characteristics. The global shape characteristics of a chain are determined using the inertia tensor[92,93] defined in eq 4 as

$$T_{\alpha\beta} = \frac{1}{2N^2} \sum_{i,j=1}^{N} (r_{i\alpha} - r_{j\alpha})(r_{i\beta} - r_{j\beta}) \quad (4)$$

Here, N is the total number of atoms in the molecule, $r_{i\alpha}$ ($r_{j\beta}$) is the αth (βth) component of the position of atom i (j), and α,β = x, y, z. The radius of gyration (R_g), asphericity (δ), and shape parameter (S) can be derived from the eigenvalues of $T_{\alpha\beta}$ represented as λ_1, λ_2 and λ_3, as follows:

$$R_g = \sqrt{\lambda_1 + \lambda_2 + \lambda_3} \quad (5)$$

$$\delta = 1 - 3\left(\frac{\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3}{(\lambda_1 + \lambda_2 + \lambda_3)^2}\right) \quad (6)$$

$$S = 27\frac{(\prod_{i=1}^{3}(\lambda_i - \bar{\lambda}))}{(\lambda_1 + \lambda_2 + \lambda_3)^3}, \quad \bar{\lambda} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (7)$$

The asphericity values range from 0 for a sphere to 1 for a rod with intermediate values corresponding to ellipsoidal conformations. The range of the shape parameter is between −0.25 and 2. Negative values of S represent oblate conformations while positive values correspond to prolate conformations. The hydrodynamic radii of the structures were calculated using Hydropro.[94] Hydropro calculations were performed using the following values for the input parameters: A hydrodynamic model of each chain was obtained for non-hydrogen atoms using spherical elements of radii 3.1 Å. The resulting structure with overlapping spheres is used to obtain the shell model as described in the original reference.[94] The minimum and maximum radii of the beads in the shell were set to 1.5 and 2 Å, respectively.

Error bars representing the standard deviations were calculated as the square root of the variance of the simulation data for the size and shape parameters.

Estimation of the Persistence Length of Polymer Chains. Persistence length is calculated as the average projection of the end-to-end vector onto every bond vector (l_i) along the sequence. Since the persistence length is calculated as an average over all possible sections along the chain of any given sequence separation, this method makes it easier to identify the effect of the varying stiffness along the sequence arising due to the heteropolymeric nature of the chain.

$$l_p = \left\langle \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{l_i \cdot l_j}{|l_i|} \right\rangle = \left\langle \sum_{i=1}^{N} \frac{l_i \cdot l_N}{|l_i|} \right\rangle \quad (8)$$

Ensemble Reconstruction by Fragment Assembly and Ensemble Selection by Fitting to Experimental Parameters. To provide a coherent representation for the conformational ensemble of α-synuclein, we have designed an approach to fit both the experimental RDCs and PREs, which have been used most commonly for determining the structural properties of IDPs. The residual dipolar couplings between two nuclei (HN) are determined using eq 9 where the dipolar couplings are dependent on the angle θ between the internuclear vector and the magnetic field.[95] Here, γ_H (γ_N) corresponds to the gyromagnetic ratio of nuclei H (N) and r_{HN} is the internuclear distance.

$$D = \frac{\mu_0 h \gamma_H \gamma_N}{4\pi r_{HN}^3} \left\langle \frac{3\cos^2 \theta - 1}{2} \right\rangle \quad (9)$$

Ensembles of weighted structures were selected from a large pool of structures, based on fits of the calculated average RDCs to the experimental RDCs. The pool of structures was generated by reconstructing the intermediate temperature REMD simulation ensemble. RDCs calculated using sliding windows over short overlapping segments of polypeptide chains have been used previously for structural motif determination of folded proteins.[96,97] In our approach, ensemble reconstruction was performed by first cleaving the structures generated from the REMD simulations into short fragments, each 14 residues long, which are subsequently spliced together to generate a new pool of structures. The RDCs for each of these new structures were calculated using a

sliding window centered on a given residue, using the alignment tensor for this short segment. Fifty weighted structures were chosen at random from the pool of structures. The fits of this ensemble to the experimental RDCs were determined using the following selection criterion:

$$\chi^2 = \sum_i^n \left(\sum_j^N D_{ij}^{calc} w_j - D_i^{exp} \right)^2 \tag{10}$$

where w_j is the weight associated with structure j, i corresponds to the residue number, N is the number of structures, n is the number of RDCs, D_{ij}^{calc} is the calculated HN couplings of residue i in structure j, and D_{ij}^{exp} is the corresponding experimental HN couplings. The following constraints were applied to the weights assigned to chosen structures: (a) The sum of the weights should be equal to 1. (b) The weight associated with any structure should be greater than zero. The RDCs have been calculated on the basis of the alignment of individual structures using PALES.[98] Ensembles with the smallest χ^2 have been chosen for further analysis. Ensemble selection using a genetic algorithm for fitting to experimental RDCs, with selection criteria similar to that mentioned above, have been reported previously.[53]

This selection procedure was performed iteratively to obtain weighted ensembles best fitting to the RDCs. Cross-validation of RDC data not employed in the fitting was predicted and compared with experimental results to evaluate the validity of the fitting procedure. The RDCs were calculated using a segment length determined on the basis of extensive analysis of different segment lengths and their ability to reproduce the average global alignment of the RDCs.

The paramagnetic intensity ratios were back-calculated for ensembles fitting the experimental RDCs, corresponding to the three spin-label sites used in the experiment. The distances were converted to intensity ratios using the following two relations.

The distance data (r) obtained from the coordinates of structures in the conformational ensemble is first converted to PREs (Γ_2) using the following relation, where K is a constant (1.23×10^{-32} cm^6 S^{-2}), τ_c is the correlation time (4 ns), and ω_H is the proton larmor frequency.

$$r = \left[\frac{K}{\Gamma_2} \left(4\tau_c + \frac{3\tau_c}{1 + \omega^2 \tau_c^2} \right) \right]^{1/6} \tag{11}$$

The PREs are then converted to intensity ratios using the following relation, where I_{ox}/I_{red} is the intensity ratio and t is the total relaxation time. The average intrinsic relaxation rate is set to 20 Hz, taken from experimental measurements.[83]

$$\frac{I_{ox}}{I_{red}} = \frac{R_2 \exp(-\Gamma_2 t)}{R_2 + \Gamma_2} \tag{12}$$

Of the many ensembles that can be obtained using the above approach, the ensemble best fitting the experimental PREs was chosen as the representative ensemble. We note that multiple ensembles generated independently on the basis of fits to both the RDCs and PREs have similar shape properties as described for the reconstructed ensemble presented here (data not shown).

Secondary structural assignments for the α-synuclein ensembles were determined using STRIDE.[77]

ASSOCIATED CONTENT

Supporting Information
Figures S1–S3. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: ronlevy@lutece.rutgers.edu.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

DEDICATION

It is a pleasure to contribute to this special issue of JCTC in honor of Wilfred van Gunsteren. I have known Wilfred since we shared an office as postdocs in the Karplus group at Harvard in 1978. I very much respect Wilfred's commitment to clarity in his science and in his science writing, traits that I recall about Wilfred from our time together more than 30 years ago. Wilfred has used molecular dynamics simulations to connect with many areas of protein science, including those between MD simulations and NMR experiments designed to reveal the ensemble nature of protein structures, of which this manuscript is a recent illustration in the context of intrinsically disordered proteins.

REFERENCES

(1) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. J. Mol. Graphics Modell. 2001, 19, 26.
(2) Uversky, V. N. Protein Sci. 2002, 11, 739.
(3) Dyson, H.; Wright, P. Mol. Cell. Biol. 2005, 6, 197.
(4) Chiti, F.; Dobson, C. M. Annu. Rev. Biochem. 2006, 75, 333.
(5) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Proteins 2000, 41, 415.
(6) Dyson, H. J.; Wright, P. E. Chem. Rev. 2004, 104, 3607.
(7) Eliezer, D. Curr. Opin. Struct. Biol. 2009, 19, 23.
(8) Tran, H. T.; Mao, A.; Pappu, R. V. J. Am. Chem. Soc. 2008, 130, 7380.
(9) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. Proc. Natl. Acad. Sci. U. S. A. 2010, 107, 8183.
(10) Wu, K.-P.; Weinstock, D. S.; Narayanan, C.; Levy, R. M.; Baum, J. J. Mol. Biol. 2009, 391, 784.
(11) Burre, J.; Sharma, M.; Tsetsenis, T.; Buchman, V.; Etherton, M. R.; Sudhof, T. C. Science 2010, 329, 1663.
(12) Norris, E. H.; Giasson, B. L.; Lee, V. M. Curr. Top. Dev. Biol. 2004, 60, 17.
(13) Davidson, W. S.; Jonas, A.; Clayton, D. F.; George, J. M. J. Biol. Chem. 1998, 273, 9443.
(14) Moore, D. J.; West, A. B.; Dawson, V. L.; Dawson, T. M. Annu. Rev. Neurosci. 2005, 28, 57.
(15) Spillantini, M. G.; Schmidt, M. L.; Lee, V. M.-Y.; Trojanowski, J. Q.; Jakes, R.; Goedert, M. Nature 1997, 388, 839.
(16) Goedert, M. Nature 2001, 2, 492.
(17) Bussell, R., Jr.; Eliezer, D. J. Biol. Chem. 2001, 276, 45996.
(18) Eliezer, D.; Kulhay, E.; Bussell, R., Jr.; Browne, G. J. Mol. Biol. 2001, 307, 1061.

(19) Sung, Y. H.; Eliezer, D. *J. Mol. Biol.* **2007**, *372*, 689.

(20) Bertoncini, C. W.; Jung, Y. S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 1430.

(21) Wu, K. P.; Kim, S.; Fela, D. A.; Baum, J. *J. Mol. Biol.* **2008**, *378*, 1104.

(22) Bernadó, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2005**, *127*, 17968.

(23) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 476.

(24) Ullman, O.; Fisher, C. K.; Stultz, C. M. *J. Am. Chem. Soc.* **2011**.

(25) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 8407.

(26) Uversky, V. N.; Li, J.; Fink, A. *J. Biol. Chem.* **2001**, *276*, 10737.

(27) Kim, H.-Y.; Hiese, H.; Fernandez, C.; Baldus, M.; Zweckstetter, M. *ChemBioChem* **2007**, *8*, 1671.

(28) Bertoncini, C.; Jung, Y.-s.; Fernandez, C.; Hoyer, W.; Griesinger, C.; Jovin, T.; Zweckstetter, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1430.

(29) Bright, J. N.; Woolf, T. B.; Hoh, J. H. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 131.

(30) Vitalis, A.; Wang, X.; Pappu, R. V. *Biophys. J.* **2007**, *93*, 1923.

(31) Tran, H. T.; Pappu, R. V. *Biophys. J.* **2006**, *91*, 1868.

(32) Gennes, P. D. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.

(33) Grosberg, A.; Khoklov, A. *Statistical Physics of Macromolecules*; AIP Press: New York, 1994.

(34) Flory, P. *Statistical Mechanics of Chain Molecules.* Cornell University Press: Ithaca, NY, 1953.

(35) Chan, H. S.; Dill, K. A. *Annu. Rev. Biophys. Biol.* **1991**, *20*, 447.

(36) Muller-Spath, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Ruegger, S.; Reymond, L.; Nettels, D.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 14609.

(37) Zhou, H.-X. *Biochemistry* **2004**, *43*, 2141.

(38) Cifra, P. *Polymer* **2004**, *45*, 5995.

(39) Mohana-Borges, R.; G., N.; Kroon, G. J. A.; Dyson, J. H.; Wright, P. E. *J. Mol. Biol.* **2004**, *340*, 1131.

(40) Annila, A.; Permi, P. *Concepts Magn. Reson., Part A* **2004**, *23A*, 22.

(41) Vitalis, A.; Lyle, N.; Pappu, R. V. *Biophys. J.* **2009**, *97*, 303.

(42) Vitalis, A.; Wang, X.; Pappu, R. V. *J. Mol. Biol.* **2008**, *384*, 279.

(43) Felts, A. K.; Gallicchio, E.; Chekmarev, D.; Paris, K. A.; Friesner, R. A.; Levy, R. M. *J. Chem. Theory Comput.* **2008**, *4*, 855.

(44) Gallicchio, E.; Paris, K.; Levy, R. M. *J. Chem. Theory Comput.* **2009**, *5*, 2544.

(45) Weinstock, D. S.; Narayanan, C.; Felts, A. K.; Andrec, M.; Levy, R. M.; Wu, K. P.; Baum, J. *J. Am. Chem. Soc.* **2007**, *129*, 4858.

(46) Marsh, J. A.; Forman-Kay, J. D. *Biophys. J.* **2010**, *98*, 2383.

(47) Kataoka, M.; Kuwajima, K.; Tokunaga, F.; Goto, Y. *Protein Sci.* **1997**, *6*, 422.

(48) Allison, J. R.; Varnai, P.; Dobson, C. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 18314.

(49) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. *Biochemistry* **1999**, *38*, 16424.

(50) Schäfer, L. *Excluded Volume Effects in Polymer Solutions As Explained by the Renormalization Group*; Springer: New York, 1999.

(51) Obolensky, O. I.; Schlepckow, K.; Schwalbe, H.; Solov'yov, A. V. *J. Biomol. NMR* **2007**, *39*, 1.

(52) Marsh, J. A.; Baker, J. M. R.; Tollinger, M.; Forman-Kay, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 7804.

(53) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 17908.

(54) Cornilescu, G.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 10143.

(55) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836.

(56) Wang, X.; Bansal, S.; Jiang, M.; Prestegard, J. H. *Protein Sci.* **2008**, *17*, 899.

(57) Choy, W. Y.; Mulder, F. A.; Crowhurst, K. A.; Muhandiram, D. R.; Millett, I. S.; Doniach, S.; Forman-Kay, J. D.; Kay, L. E. *J. Mol. Biol.* **2002**, *316*, 101.

(58) Doi, M.; Edwards, S. F. *The Theory of Polymer Dynamics*; Clarendon Press: Oxford, U. K.,1988.

(59) Binolfi, A.; Rasia, R. M.; Bertoncini, C. W.; Ceolin, M.; Zweckstetter, M.; Griesinger, C.; Jovin, T. M.; Fernandez, C. O. *J. Am. Chem. Soc.* **2006**, *128*, 9893.

(60) Dobrynin, A. V.; Rubinstein, M. *Prog. Polym. Sci.* **2005**, *30*, 1049.

(61) Dobrynin, A. V.; Rubenstein, M. *J. Phys. II* **1995**, *5*, 677.

(62) Dobrynin, A. V.; Colby, R. H.; Rubinstein, M. *J. Polym. Sci., Part B: Polym. Phys.* **2004**, *42*, 3513.

(63) Kantor, Y.; Kardar, M.; Ertas, D. *Physica A* **1998**, *249*, 301.

(64) Louhivuori, M.; P., K.; Fredriksson, K.; Permi, P.; Lounila, J.; Annila, A. *J. Am. Chem. Soc.* **2003**, *125*, 15647.

(65) Cubrovic, M.; Obolensky, O. I.; Solov'yov, A. V. *Eur. Phys. J. D* **2009**, *51*, 41.

(66) Jensen, M. R.; M., P.; Griesinger, C.; Zweckstetter, M.; Grzesiek, S.; Bernardo, P.; Blackledge, M. *Structure* **2009**, *17*, 1169.

(67) Schuler, B.; L., E.; Steinbach, P. J.; Kumke, M.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2754.

(68) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 8407.

(69) Eleizer, D.; Kutluay, E.; Bussel, R.; Browne, G. *J. Mol. Biol.* **2001**, *307*, 1061.

(70) Wu, K.-P.; Kim, S.; Fela, D.; Baum, J. *J. Mol. Biol.* **2008**, *378*, 1104.

(71) Sung, Y.; Eleizer, D. *J. Mol. Biol.* **2007**, *372*, 689.

(72) Vendruscolo, M. *Curr. Opin. Struct. Biol.* **2007**, *17*, 15.

(73) Fisher, C. K.; Huang, A.; Stultz, C. M. *J. Am. Chem. Soc.* **2010**, *132*, 14919.

(74) Marsh, J. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2009**, *391*, 359.

(75) Shi, L.; Traaseth, N. J.; Verardi, R.; Gustavsson, M.; Gao, J.; Veglia, G. *J. Am. Chem. Soc.* **2011**, *133*, 2232.

(76) Schneider, R.; Huang, J. R.; Yao, M.; Communie, G.; Ozenne, V.; Mollica, L.; Salmon, L.; Jensen, M. R.; Blackledge, M. *Mol. BioSyst.* **2012**, *8*, 58.

(77) Frishman, D.; Argos, P. *Proteins* **1995**, *23*, 566.

(78) Zhao, M.; Cascio, D.; Sawaya, M. R.; Eisenberg, D. *Protein Sci.* **2011**, *20*, 996.

(79) Zibaee, S.; Makin, O. S.; Goedert, M.; Serpell, L. C. *Protein Sci.* **2007**, *16*, 906.

(80) Bartels, T.; Choi, J. G.; Selkoe, D. J. *Nature* **2011**, *477*, 107.

(81) Wang, W.; Perovic, I.; Chittuluru, J.; Kaganovich, A.; Nguyen, L. T.; Liao, J.; Auclair, J. R.; Johnson, D.; Landeru, A.; Simorellis, A. K.; Ju, S.; Cookson, M. R.; Asturias, F. J.; Agar, J. N.; Webb, B. N.; Kang, C.; Ringe, D.; Petsko, G. A.; Pochapsky, T. C.; Hoang, Q. Q. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17797.

(82) Fauvet, B.; Mbefo, M. K.; Fares, M. B.; Desobry, C.; Michael, S.; Ardah, M. T.; Tsika, E.; Coune, P.; Prudent, M.; Lion, N.; Eliezer, D.; Moore, D. J.; Schneider, B.; Aebischer, P.; El-Agnaf, O. M.; Masliah, E.; Lashuel, H. A. *J. Biol. Chem.* **2012**.

(83) Wu, K. P.; Baum, J. *J. Am. Chem. Soc.* **2010**, *132*, 5546.

(84) Pellarin, R.; Caflisch, A. *J. Mol. Biol.* **2006**, *360*, 882.

(85) Abedini, A.; Raleigh, D. P. *Protein Eng. Des. Sel.* **2009**, *22*, 453.

(86) Anderson, V. L.; Ramlall, T. F.; Rospigliosi, C. C.; Webb, W. W.; Eliezer, D. *Proc. Natl. Acad. Sci. U. S. A.* **2010**.

(87) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.

(88) Felts, A.; Harano, Y.; Gallicchio, E.; Levy, R. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 310.

(89) Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2005**, *26*, 1752.

(90) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479.

(91) Kaminsky, G.; Friesner, R.; Tirado-Rives, J.; Jorgensen, W. *J. Phys. Chem. B* **2001**, *105*, 6474.

(92) Dima, R. I.; Thirumalai, D. *J. Phys. Chem. B* **2004**, *108*, 6564.

(93) Honeycutt, J.; Thirumalai, D. *J. Chem. Phys.* **1989**, *90*, 4542.

Article

(94) de la Torre, J. G.; Huertas, M. L.; Carrasco, B. *Biophys. J.* **2000**, *78*, 719.

(95) Ernst, R.; Bodenhausen, G.; A., W. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Oxford University Press: Oxford, U. K., 1987.

(96) Andrec, M.; Du, P.; Levy, R. M. *J. Am. Chem. Soc.* **2001**, *123*, 1222.

(97) Andrec, M.; Du, P.; Levy, R. M. *J. Biomol. NMR* **2001**, *21*, 335.

(98) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791.

(99) Gast, K.; Damaschun, H.; Eckert, K.; Schulze-Forster, K.; Maurer, H. R.; Muller-Frohne, M.; Zirwer, D.; Czarnecki, J.; Damaschun, G. *Biochemistry* **1995**, *34*, 13211.

(100) Fitzkee, N. C.; Rose, G. D. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12497.

(101) Johnson, S. *Psychometrica* **1967**, *32*, 241.
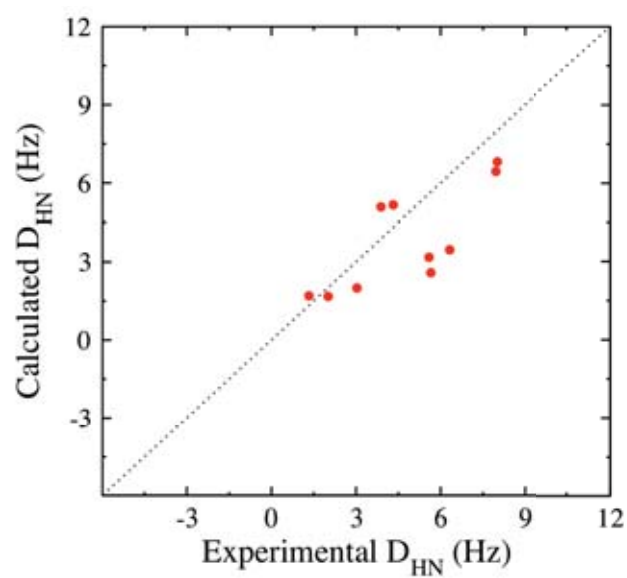
Figure S1

Comparison of the calculated HN RDCs not used in the fitting procedure. Red dots correspond to the back-calculated RDCs not used in the fitting procedure used for the selection of ensembles fitting the experiment.
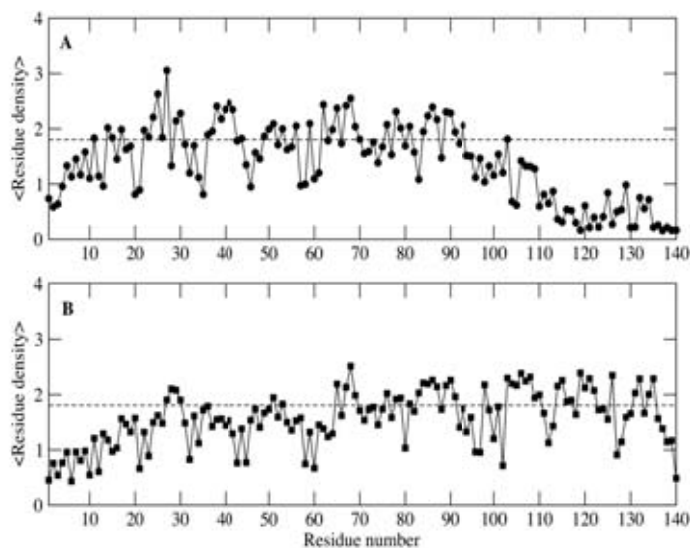
Figure S2

Comparison of residue densities of α-synuclein at neutral (A) and low pH (B) conditions.

   The low pH simulation ensembles were generated for a study published previously [1]. The N and C terminal regions of α-synuclein show significant changes in the residue density with change in pH. With shift in pH from low to neutral pH, the total charge in the N-terminus increases significantly while the net charge is low. This region adopts a more collapsed conformation with increased charge density, as predicted from theory of polyampholyte chains. The C-terminal region is observed to also show an increase in charge density with increasing pH, although net charge also increases significantly in this case, leading to charge repulsion resulting in an expanded conformation of this region with increase in pH. The behavior of the C-terminal region at neutral pH is consistent with that expected for a polyelectrolyte chain.

3

Convergence of Simulations:

To assess the convergence of simulations, we calculated the standard errors of the means for the scaling of internal distances. Fig. S3 shows the scaling of internal distances plotted as a function of sequence separation for the low, intermediate and high temperature ensembles. The error bars represent the calculated standard errors of the means. The low values of the calculated standard errors of the mean indicate that the simulations are well converged.
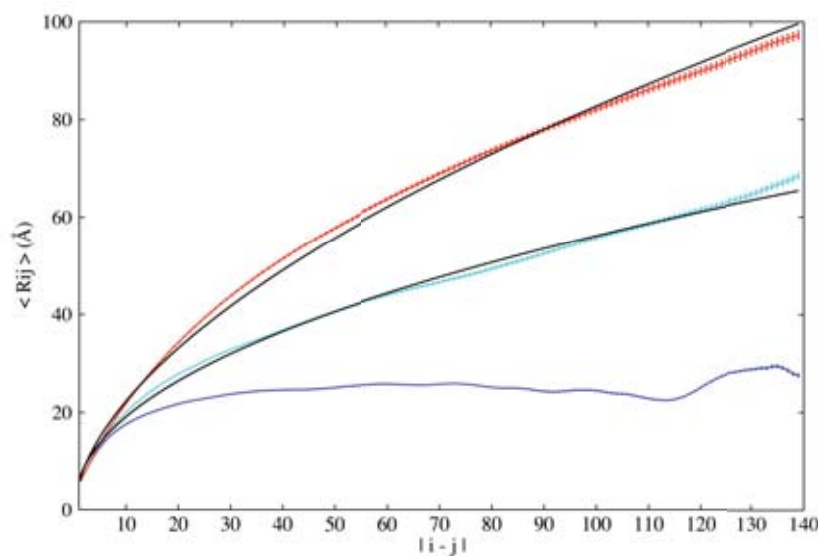

Figure S3

8.4 References:

Annila, A. and P. Permi (2004). "Weakly aligned biological macromolecules in dilute aqueous liquid crystals." Concepts in Magnetic Resonance Part A **23A**(1): 22-37.

Bernardó P, B. C., Griesinger C, Zweckstetter M, Blackledge M (2005). "Defining Long-range order and local disorder in native alpha-synuclein using Residual Dipolar Couplings." J. Am. Chem. Soc. **127**: 17968-17969.

Bertoncini, C., Y.-s. Jung, et al. (2005a). "Release of long-range teriary interactions potentiates aggregation of natively unstructured alpha-synuclein." Proceedings of the National Academy of Sciences **102**: 1430-1435.

Bertoncini, C. W., Y. S. Jung, et al. (2005b). "Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein." Proc Natl Acad Sci U S A **102**(5): 1430-1435.

Bright, J. N., T. B. Woolf, et al. (2001). "Predicting properties of intrinsically unstructured proteins." Prog Biophys Mol Biol **76**(3): 131-173.

Bussell, R., Jr. and D. Eliezer (2001). "Residual structure and dynamics in Parkinson's disease-associated mutants of alpha-synuclein." J Biol Chem **276**(49): 45996-46003.

Cifra, P. (2004). "Differences and limits in estimates of persistence length for semi-flexible macromolecules." Polymer **45**: 5995-6002.

Dedmon, M. M., K. Lindorff-Larsen, et al. (2005). "Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations." J. Am. Chem. Soc. **127**: 476-477.

Eliezer, D. (2009). "Biophysical characterization of intrinsically disordered proteins." Curr Opin Struct Biol **19**(1): 23-30.

Eliezer, D., E. Kutluay, et al. (2001). "Conformational properties of alpha-synuclein in its free and lipid-associated states." J Mol Biol **307**(4): 1061-1073.

Kim, H. Y., H. Heise, et al. (2007). "Correlation of Amyloid Fibril beta-Structure with the Unfolded State of alpha-Synuclein." Chembiochem **8**(14): 1671-1674.

Mohana-Borges R, G. N., Kroon GJA, Dyson JH, Wright PE (2004). "Structural characterization of unfolded states of Apomyoglobin using Residual Dipolar Couplings." J. Mol. Biol **340**: 1131-1142.

Sung, Y. H. and D. Eliezer (2007). "Residual Structure, Backbone Dynamics, and Interactions within the Synuclein Family." J Mol Biol **372**(3): 689-707.

Ullman, O., C. K. Fisher, et al. (2011). "Explaining the Structural Plasticity of alpha-Synuclein." Journal of the American Chemical Society.

Uversky, V. N. (2002). "Natively unfolded proteins: A point where biology waits for physics." Protein Science **11**: 739-756.

Uversky, V. N., J. Li, et al. (2001). "Evidence for a partially folded intermediate in alpha-synuclein fibril formation." J Biol Chem **276**(14): 10737-10744.

Vitalis, A., X. Wang, et al. (2007). "Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories." Biophys J **93**(6): 1923-1937.

Wu, K.-P., D. S. Weinstock, et al. (2009). "Structural Reorganization of alpha-Synuclein at Low pH Observed by NMR and REMD Simulations." J Mol Biol **391**(4): 784-796.

Wu, K. P., S. Kim, et al. (2008). "Characterization of conformational and dynamic properties of natively unfolded human and mouse alpha-synuclein ensembles by NMR: implication for aggregation." J Mol Biol **378**(5): 1104-1115.

Zhou, H.-X. (2004). "Polymer Models of Protein Stability, Folding, and Interactions." Biochemistry **43**(8): 2141-2154.

Chapter 9

Structural Reorganization of α–synuclein at low pH observed by NMR and REMD simulations

9.1 Introduction

Aggregation rate of α-synuclein is influenced by changes in sequence resulting from mutations and environmental changes. α-synuclein has been shown to aggregate much faster at low pH compared to neutral pH with aggregation taking days rather than weeks (Uversky, Li et al. 2001). The disordered conformation of α-synuclein at neutral pH has been attributed to the high net charge and low hydrophobicity of the chain (Uversky, Gillespie et al. 2000). At neutral pH, α-synuclein has 15 positively and 24 negatively charged side-chains, resulting in a net charge of -9. Under low pH conditions, the acidic residues along the chain are protonated resulting in a net charge of +15 on the protein. Uversky et. al., showed that α-synuclein, under low pH conditions, adopts a partially folded conformation, resulting from a decrease in net charge and increased hydrophobicity. They showed an increase in beta-sheet content at low pH which was correlated with the increase in aggregation rate resulting from a shorter lag phase.

With the shift from neutral to low pH, the charge distribution of α-synuclein is primarily affected in the N and C-terminal regions of the protein. Under these conditions, the net charge of the NAC region remains the same while the C-terminal region becomes neutral and N-terminal becomes positively charged. In this study, molecular dynamics

simulations were used in combination with NMR to understand the influence of altered charge distributions on the conformational properties of α-synuclein and their relation to the aggregation rate.

Polymeric properties of α-synuclein at low pH, described using global size (R_g, R_h) and shape descriptors (asphericity and shape parameter) are shown in Figure 9.1. The R_g and R_h of 16.96 Å and 21.95 Å respectively for α-synuclein at the lowest temperature, are similar to that expected for a chain in poor solvent (15.1 Å and 20.1 Å) (Wilkins, Grimshaw et al. 1999; Marsh and Forman-Kay 2010). The highest temperature ensemble, which behaves like a chain in a good solvent has an average size of 38.6 Å and 34.7 Å for the R_g and R_h respectively. These values are slightly smaller than expected from theory (40.7 and 35.1 Å) (Wilkins, Grimshaw et al. 1999; Marsh and Forman-Kay 2010).
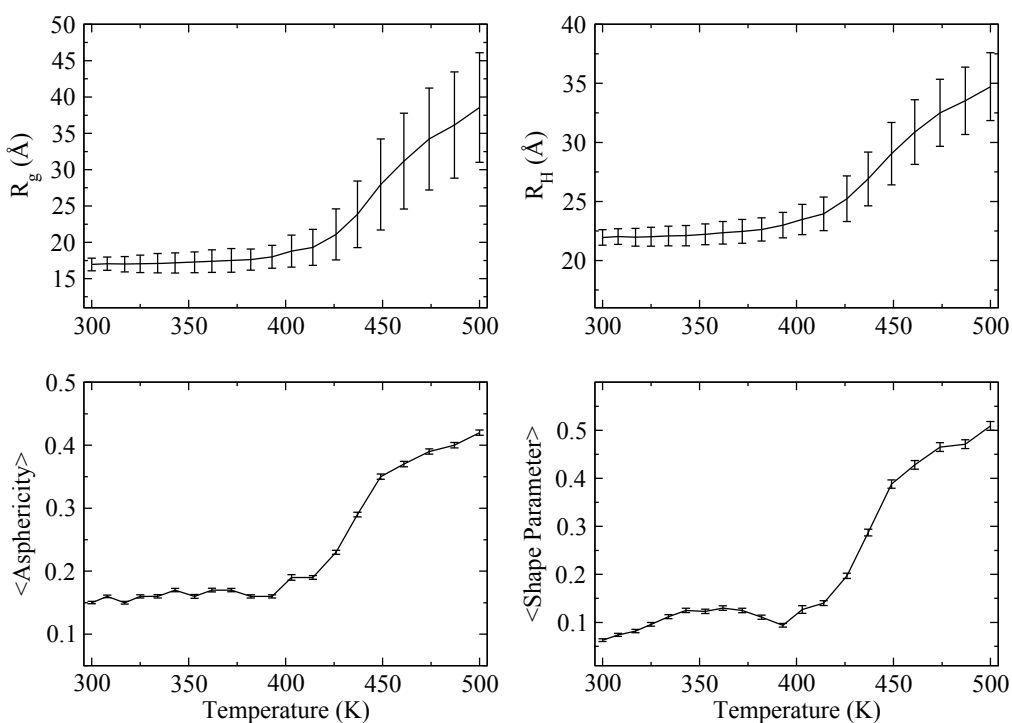
Figure 9.1 Global shape and size descriptors of α-synuclein at low pH. Average radius of gyration, (A) hydrodynamic radius, (B) Asphericity, (C) and shape parameter, (D) plotted as a function of simulation temperature. Standard deviations within ensembles are represented as error bars.

The average R_h of 29.1 Å at intermediate temperature, is slightly smaller than that expected for a typical IDP under physiological conditions (30.7 Å) (Marsh and Forman-Kay 2010). The average R_h and long-range conformational characteristics of this ensemble are also in good agreement with NMR R_h (30.1 Å) and PREs respectively. The intermediate temperature ensemble is used in the study presented here to characterize the conformations of α-synuclein in the aggregation-prone low pH state. The conformational characteristics of the low pH state reported in this study are consistent with those described in other studies (Cho, Nodet et al. 2009; McClendon, Rospigliosi et al. 2009; Trexler and Rhoades 2010).

A comparison of the global size of the protein between neutral and low pH conditions shows that the size remains unchanged even though major topological reorganization is observed in these 2 states. Earlier studies of α-synuclein under physiological (neutral pH) conditions showed that it forms long-range contacts between the C-terminal region and the rest of the chain resulting in ring-like conformations. In these studies, it was also proposed that these long-range contacts protect the hydrophobic NAC region resulting in slower aggregation under physiological conditions (Bertoncini, Jung et al. 2005; Dedmon, Lindorff-Larsen et al. 2005). The results presented here show that C-terminal region is mostly extended with few contacts with other parts of the chain while the N and

NAC regions adopt collapsed conformations. Moreover, these findings have recently been confirmed by Irback and colleagues in a simulation study of α-synuclein where the C-terminal region was shown to have minimal interactions with the rest of the chain (Jonsson, Mohanty et al. 2012). It is not clear why different long-range conformational characteristics have been described in these studies based on a good agreement with PRE data. It is important to note that this reflects the underdetermined nature of determining ensembles from fitting to PRE data.

The intermediate temperature low pH ensemble shows topological characteristics very different from those under neutral pH conditions (Figure 9.2). At low pH, the C-terminal region shows regular contacts with itself, the NAC region and transient interactions with the N-terminal region near residue 40. At the same time, the first forty residues show fewer interactions with the rest of the chain.
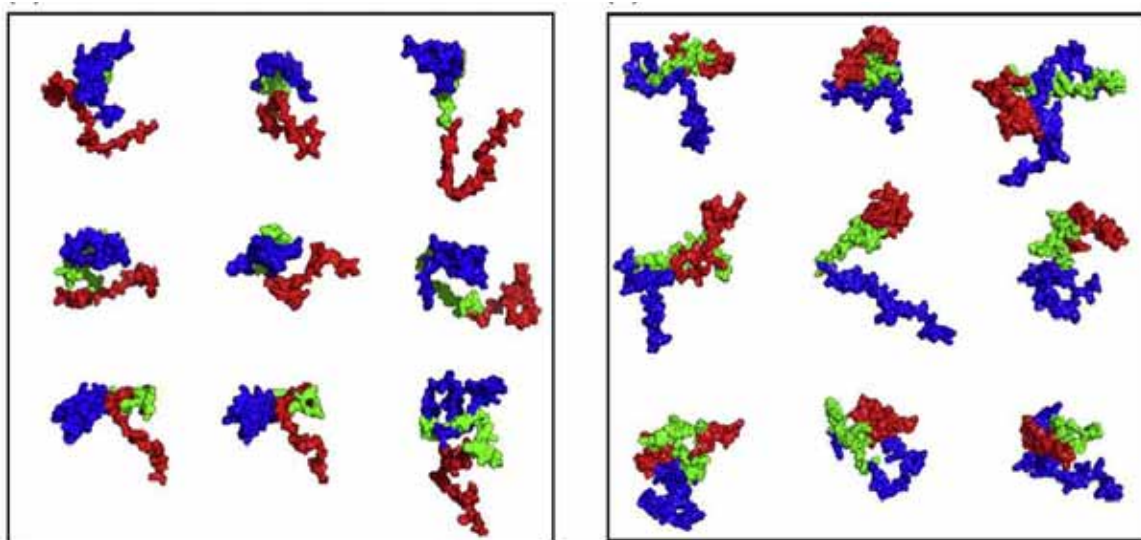

Figure 9.2 Comparison between the neutral (left) and low pH (right) simulation ensembles that exhibit good agreement with experiment.

The different topological characteristics under the two pHs are also observed in the calculated solvent accessible surface areas of the N, NAC and C-terminal regions. With the decrease in pH, the solvent exposure of the N-terminal region increases while the C-terminal solvent exposure decreases. The solvent exposure of the NAC region remains the same under both pH conditions.

The major structural reorganization of α-synuclein with decrease in pH can be explained based on the change in charge distribution along the sequence – the charge density in the C-terminal decreases significantly (from 40% to ~7%) leading to a collapse of this region at low pH. The N-terminal region, on the other hand, shows a more subtle change in charge density (30% to ~ 18%) with decreasing pH. At low pH, this region shows fewer contacts with itself and the rest of the chain compared to neutral pH conditions.

It is interesting to note that although the neutral and low pH ensembles show very different topological characteristics, the solvent accessible surface area of the NAC region does not change significantly. This suggests that the faster aggregation at low pH does not result from the exposure of NAC region. The reduction of net charge in the C-terminus and the corresponding increase in hydrophobicity, which leads to its collapsed conformation, may be involved in the early stages of aggregation at low pH.

9.2 Methods and Results

The procedures and results for this study are presented as a reprint of the paper published in the *Journal of Molecular Biology*, 2009, 391: 784-796.

9.3 Publication Reprint

Structural Reorganization of α-Synuclein at Low pH Observed by NMR and REMD Simulations

Kuen-Phon Wu[1], Daniel S. Weinstock[2], Chitra Narayanan[3], Ronald M. Levy[1,2]* and Jean Baum[1,2]*

[1]*Department of Chemistry and Chemical Biology, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854, USA*

[2]*BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway, NJ 08854, USA*

[3]*Graduate Program in Biochemistry, Rutgers University, Piscataway, NJ 08854, USA*

α-Synuclein is an intrinsically disordered protein that appears in aggregated forms in the brains of patients with Parkinson's disease. The conversion from monomer to aggregate is complex, and aggregation rates are sensitive to changes in amino acid sequence and environmental conditions. It has previously been observed that α-synuclein aggregates faster at low pH than at neutral pH. Here, we combine NMR spectroscopy and molecular simulations to characterize α-synuclein conformational ensembles at both neutral and low pH in order to understand how the altered charge distribution at low pH changes the structural properties of these ensembles and leads to an increase in aggregation rate. The N-terminus, which has a small positive charge at neutral pH due to a balance of positively and negatively charged amino acid residues, is very positively charged at low pH. Conversely, the acidic C-terminus is highly negatively charged at neutral pH and becomes essentially neutral and hydrophobic at low pH. Our NMR experiments and replica exchange molecular dynamics simulations indicate that there is a significant structural reorganization within the low-pH ensemble relative to that at neutral pH in terms of long-range contacts, hydrodynamic radius, and the amount of heterogeneity within the conformational ensembles. At neutral pH, there is a very heterogeneous ensemble with transient contacts between the N-terminus and the non-amyloid β component (NAC); however, at low pH, there is a more homogeneous ensemble that exhibits strong contacts between the NAC and the C-terminus. At both pH values, transient contacts between the N- and C-termini are observed, the NAC region shows similar exposure to solvent, and the entire protein shows similar propensities to secondary structure. Based on the comparison of the neutral- and low-pH conformational ensembles, we propose that exposure of the NAC region to solvent and the secondary-structure propensity are not factors that account for differences in propensity to aggregate in this context. Instead, the comparison of the neutral- and low-pH ensembles suggests that the change in long-range interactions between the low- and neutral-pH ensembles, the compaction of the C-terminal region at low pH, and the uneven distribution of charges across the sequence are key to faster aggregation.

Published by Elsevier Ltd.

Keywords: Parkinson's disease; α-synuclein; NMR; replica exchange molecular dynamics

Edited by P. Wright

*Corresponding authors.* E-mail addresses: ronlevy@lutece.rutgers.edu; jean.baum@rutgers.edu.

Introduction

Many human diseases are associated with proteins that convert from their normally soluble form to aggregates that accumulate in the affected organs. The final forms of the aggregates include fibrillar plaques known as amyloid. This disease class includes Alzheimer's disease, Parkinson's disease (PD), Huntington's disease, prion disease, and type II diabetes.[1-4] PD is the second most prevalent of the late-onset neurodegenerative diseases.[5] α-Synuclein is the primary protein component of the Lewy body deposits that are the diagnostic hallmark of PD and is expressed in high levels in the brain.[6] α-Synuclein can adopt two forms, the free form found in the cytoplasm and a highly helical membrane-bound form.[7,8] The membrane-bound α-synuclein may be crucial for physiological functions[9-12] and has been shown to be associated with the recruitment of dopamine in the presynapse and with the synaptic signal transmission.[13] The aggregation of α-synuclein has been shown to occur via a nucleation-dependent mechanism during which there is an initial lag phase followed by a growth phase.[14] The rate of fibrillization is highly dependent on the sequence identity of the protein as well as the solution conditions, and the details of the mechanism remain unclear.

One approach to understanding the progression of fibril formation is to study human α-synuclein and variants under different conditions in order to define the distribution of conformational states in the monomeric ensemble and how these may be related to the aggregation propensities. Human α-synuclein in the free form has been shown to be an intrinsically disordered protein (IDP)[15,16] and, similar to other IDPs, is characterized by a low sequence complexity, low overall hydrophobicity, and high net charge.[17] It consists of three regions: an N-terminal region (residues 1–60) with a highly conserved hexamer motif KTKEGV forms α-helices in association with membranes.[8,18,19] A central hydrophobic region, known as the "non-amyloid β component" (NAC) (residues 61–95), is proposed to be primarily responsible for aggregation.[20] The C-terminal region (residues 96–140) is acidic, proline-rich,[21] and also contains three highly conserved tyrosine residues. The charged residues are unevenly distributed within the sequence (Fig. 1) and result in a net charge of −9 at neutral pH (Table 1). The N-terminus has a high charge density at neutral pH (18 charged residues out of 60), but the balance between positive and negative charges leads to a net charge of +4, while the NAC, with a net charge of −1, only has three charged residues. The C-terminus at neutral pH has an even higher percentage of charged residues than the N-terminus (18 out of 45 residues), and a preponderance of negatively charged Asp and Glu residues results in a net charge of −12. The uneven distribution of charges within the three regions implies that the charge profile of α-synuclein is strongly dependent on pH.
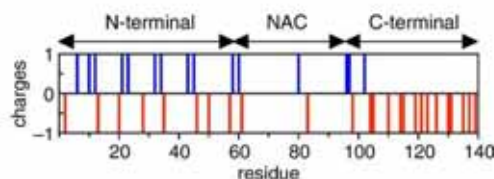
Fig. 1. Distribution of charged residues of α-synuclein. The distribution of positive (blue) and negative (red) charged residues of α-synuclein at pH 7.4 is shown.

The charge distribution of the NAC, which has very few charged residues, is not greatly affected by the change in pH. However, the N-terminus region that contains the highest fraction of positively charged residues becomes the region with the highest charge density at low pH, while the C-terminus, at low pH, becomes more like the NAC, with both a low charge density and a low net charge.

Developing a better understanding of how sequence-dependent effects alter the conformational propensities of monomeric α-synuclein, and correlating these effects with changes in aggregation rates, can provide insights into the mechanistic basis for abnormal folding and human disease. The conformational states of the synuclein family, consisting of human α-synuclein,[15] β-synuclein,[22,23] and γ-synuclein,[24] as well as the familial mutations of synuclein A53T,[25,26] A30P,[25,26] and E46K[27,28] and mouse α-synuclein,[29] which contains seven mutations relative to human α-synuclein, have been studied extensively using NMR and other spectroscopic approaches. Human α-synuclein at neutral pH has a relatively more compact conformation than is expected for a random-coil conformation, and transient α-helical structure has been detected in the N-terminal region of the protein.[22,23,29-31] Transient long-range interactions between the C- and N-terminal regions and also between the C-terminal and NAC regions have been proposed based on paramagnetic relaxation enhancement (PRE) measurements.[23,29,30,32,33] Observation of these long-range interactions has led to the proposal that the contacts between the NAC and the C-terminal regions shield the hydrophobic NAC region and therefore delay the aggregation of α-synuclein under these conditions.[23,30,32] Destabilization of these long-range contacts in mouse α-synuclein[29] and in A53T and A30P mutations[25] involved in early-onset disease has been correlated with faster aggregation, supporting the notion that protection of the NAC region may be important for delaying aggregation.[25] An alternative view is that there is little difference in long-range contacts between the wild type and mutant forms of α-synuclein and that changes in aggregation rate can be attributed to local changes in the physicochemical properties of individual residues.[28]

At low pH, α-synuclein has been shown to aggregate faster than at neutral pH, with fibrils taking days to form rather than several weeks.[31] Uversky

Table 1. Distribution of charges in the sequence of α-synuclein

pH	N			NAC			C			α-Synuclein		
	Net	Total	%	Net	Total	%	Net	Total	%	Net	Total	%
Neutral	+4	18	30.0	−1	3	8.6	−12	18	40.0	−9	39	27.8
Low	+11	11	18.3	+1	1	2.9	+3	3	6.7	+15	15	10.7

The table shows the net charge, total charge (total number of charged residues), and percentage of charged residues in the N, NAC, and C regions and the full α-synuclein sequence at neutral pH and low pH.

et al. have observed, via CD, Fourier transform infrared spectroscopy, and small-angle X-ray scattering measurements, a 1.3-fold compaction in the radius of gyration of α-synuclein at low pH as well as an increase in β structure.[31] They have suggested that the compaction of the protein into a partially folded conformation is responsible for the increased rate of aggregation at low pH.

NMR spectroscopy[34–37] and computer simulations[38–40] have been developed in recent years to characterize conformational ensembles of IDPs. Here, we use NMR and computational approaches to compare the conformational ensembles of α-synuclein at neutral and low pH. These studies help identify the role of electrostatic interactions in the pH-dependent structural changes of α-synuclein and relate these to the different aggregation propensities as a function of pH. The NMR measurements, when combined with structures generated by computer simulations, provide the first visualization of the conformational ensemble of α-synuclein at low pH. The results suggest that there is a significant structural reorganization relative to the neutral-pH ensemble in terms of long-range intra-

chain contacts and hydrodynamic radius. At neutral pH, the highly charged C-terminal tail makes contact with the other regions of the protein in only a small population of structures. In contrast, at low pH, the C-terminus is locally collapsed. The secondary-structure propensity and exposure of the NAC region to solvent are not affected by the pH change. The structural changes observed at low pH highlight the effects that the very different charge characteristics in the N-terminal, NAC, and C-terminal chain segments have on the pH-dependent structural reorganization of α-synuclein.

Results

Acid-induced changes in structure and dynamics probed by NMR

The ^{1}H–^{15}N heteronuclear single quantum coherence (HSQC) spectrum at pH 2.5 and 15 °C exhibits similar spectral features to the HSQC spectrum at pH 7.4;[15] the narrow ^{1}H^{N} chemical shift range and poor dispersion suggest that α-synuclein is intrinsi-
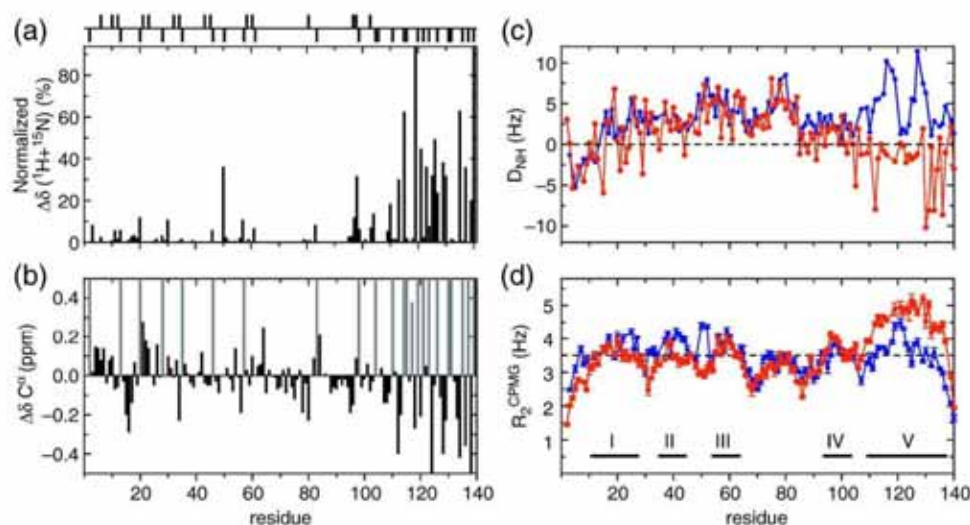
Fig. 2. NMR parameters for α-synuclein at neutral and low pH. Distribution of charged residues shown in Fig. 1 is displayed at the top of the figure. Comparison of chemical shifts (a and b), RDC (c), and R_2^{CPMG} (d) data of α-synuclein at acidic and neutral pH. (a) Average differences of H^{N} and ^{15}N chemical shifts at pH 2.5 and 7.4 with the formula $\Delta\delta = \sqrt{\{[(\Delta H^{N})^2 + 0.2(\Delta^{15}N)^2]\}/2}$ and the values are normalized. (b) C$^{\alpha}$ chemical shift differences between pH 7.4 and 2.5 ($\Delta\delta C^{\alpha} = \delta C^{\alpha}_{7.4} - \delta C^{\alpha}_{2.5}$) are shown. Asp and Glu residues are colored light gray. (c) ^{1}H–H^{N} RDCs at pH 2.5 and 7.4 are shown in red and blue lines with dots, respectively. (d) R_2^{CPMG} relaxation rates at pH 2.5 and 6.1 are displayed with red and blue lines with dots, respectively. Average value (3.5 Hz) of R_2^{CPMG} at pH 2.5 is indicated. Regions of five defined segments are shown in bars and numbered.

cally disordered at pH 2.5. Backbone assignments were carried out based on ^{15}N–^{15}N connections using the 3D HNN triple-resonance experiments.[41] Out of 140 residues, 134 (no Met[1] and 5 prolines) were assigned in the ^1H–^{15}N HSQC spectrum and most of the ^1H and ^{15}N chemical shift changes arise in the C-terminal region of the protein (Fig. 2a) where there are 3 Asp and 12 Glu residues. The major changes in C^α shifts ($\Delta\delta C^\alpha$, Fig. 2b) from pH 7.4 to pH 2.5 can be directly tied to the protonation of Asp and Glu residues at low pH. Once those changes are accounted for, the remaining small $\Delta\delta C^\alpha$'s in the N-terminus and NAC regions alternate between positive and negative values, indicating that the secondary-structure propensities are essentially the same at low and neutral pH. There is, however, a stretch of small negative values in the C-terminus region, which represents a small decrease in the β propensity in this region of the protein at low pH.

Residual dipolar coupling (RDC) constants have been employed to probe the local structural ordering[42,43] as well as long-range contacts[30,44] of IDPs. We have measured the D_{NH} values of α-synuclein at pH 7.4 and pH 2.5 using an n-alkyl-poly(ethylene glycol)/n-alkyl alcohol mixture [pentaethylene glycol monooctyl ether (C8E5)/1-octanol] as the alignment medium. Significant differences can be seen in the RDC profiles at pH 7.4 and pH 2.5, particularly at the C-terminus and the leading portion of the N-terminus (Fig. 2c). At pH 2.5, the D_{NH} values are close to zero or negative at residues 100–127 and even more negative (~−8 to −10 Hz) at the end of the C-terminus. The D_{NH} values in the NAC region and the N-terminal region are similar to the D_{NH} values measured at pH 7.4, indicating that these

regions are relatively insensitive to the environmental changes.

^{15}N backbone transverse relaxation experiments using the CPMG pulse train (R_2^{CPMG}) indicate that the low-pH and near-neutral-pH forms of α-synuclein exhibit different local dynamics along the sequence (Fig. 2d). We have compared the R_2^{CPMG} experiments at pH 2.5 and pH 6.1 rather than pH 7.4 as we have shown (unpublished data) that the intrinsic R_2^{CPMG} values at pH 7.4 are modulated by the fast hydrogen exchange rates that exist at this pH. A more accurate view of the intrinsic R_2^{CPMG} values can be obtained at pH 6.1 where the hydrogen exchange rates are slower. The R_2^{CPMG} values can be classified into five segments (10–25, 32–42, 52–62, 93–105, and 109–136) using $R_2^{CPMG}=3.0$ Hz as cutoff. The overall features of the R_2^{CPMG} data are similar for the two pH values with the exception of the C-terminal end from residues 105 to the end of the protein. Here, the R_2^{CPMG} values are uniformly greater at low pH than at neutral pH, suggesting that the C-terminal end of the protein is experiencing restricted motion and is more rigid at low pH. In the NAC region, for both pH values, the R_2^{CPMG} values are the smallest, indicating that this region is the most flexible.

The C-terminus is collapsed at pH 2.5

Site-directed spin labeling and PRE at three positions, A19, A90 and G132, are used to obtain information about long-range interactions within the unfolded protein. The PRE effects are compared at pH 7.4 and pH 2.5 and show differences, especially in terms of the interactions observed at the C-terminal end of the protein. At pH 7.4 (Fig. 3a), the
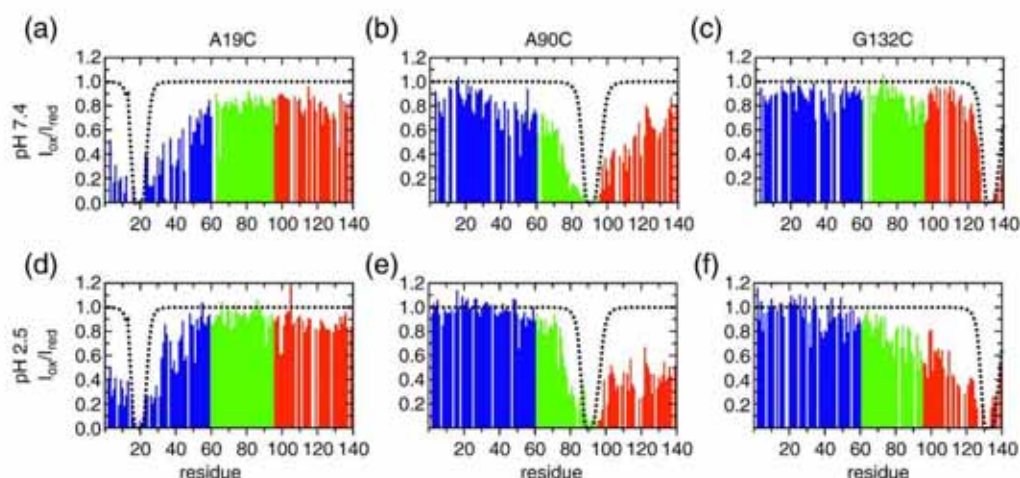
Fig. 3. PRE intensity ratios for spin-labeled α-synuclein. Paramagnetic relaxation experiments of amide protons in human α-synuclein at pH 7.4 (a–c) and pH 2.5 (d–f) using MTSL spin labels for cysteine mutants A19C, A90C, and G132C. Measured PRE intensity ratios (I_{ox}/I_{red}) are presented in the top (pH 7.4) and bottom (pH 2.5) panels. White gaps in each plot are unassignable residues including five prolines. N-terminal, NAC, and C-terminal regions are colored blue, green, and red, respectively. Broken lines are the theoretical PRE values of α-synuclein without any long-range contacts calculated as described previously.[29]

spin label at A19C results in diminished intensity throughout the entire sequence with a maximum observed value of $I_{ox}/I_{red} \sim 0.8$. At the N-terminus, there is a slow increase in I_{ox}/I_{red} up to residue 60, at which point the intensity plateaus at ~ 0.8. The same spin label at pH 2.5 (Fig. 3d) has a progressively faster increase in PRE intensity at the N-terminus and plateaus at a higher level close to 1.0 in the NAC region and C-terminal to it. There is a small reduction in intensity ($I_{ox}/I_{red} \sim 0.8$) at the very end of the C-terminus from residues 110 to 140. At neutral pH, there are numerous short-range contacts between residue 19 and the N-terminus as well as transient contacts between residue 19 and the NAC and C-terminal regions consistent with previous experimental results.[23,30] At low pH, the N-terminal region has fewer interactions with the N-terminal spin label and the NAC region is not interacting with the N-terminal spin label and is only making very local transient contact at a small region of the C-terminus (residues 115–130).

The PRE results at A90C in the NAC region (Fig. 3b and e) show distinct features at neutral and low pH. At pH 7.4, the PRE I_{ox}/I_{red} pattern is similar to the published data obtained by placing the PRE label at position A85[28] or A90.[30] The A90C spin label probe at pH 7.4 interacts primarily with the second half of the N-terminal region (40–60) as well as with the C-terminal region. In contrast, at pH 2.5, there are essentially no interactions with the N-terminal region and increased interactions with the C-terminal region ($I_{ox}/I_{red} = 0.4–0.6$) relative to neutral pH.

The spin label at position G132C (Fig. 3c and f) has a very different intensity profile at neutral and low pH in particular in the C-terminal region of the protein. At neutral pH, the spin label does not lead to strong signal attenuation anywhere in the sequence.

Three small dips in the intensity ratio ($I_{ox}/I_{red} = \sim 0.8$) are observed, one in the NAC region from residues 80 to 100 and two in the N-terminus, around residues 30–40 and near residue 5. This profile has similar trends to those seen by Bertoncini et al.[30] and Sung and Eliezer[23] who have studied the effects of spin labels at positions A140C and P120C, respectively; however, the decrease in the intensity ratio is somewhat smaller for the data observed here with the spin label at position 132. The PRE effects due to the G132C spin label at low pH (Fig. 3f) are quite different; there is no signal attenuation in the N-terminus, but there is strong signal attenuation in the C-terminus ($I_{ox}/I_{red} \sim 0.4$), which continues, gradually weakening (I_{ox}/I_{red} goes from 0.4 to 0.9), into the NAC region. The weak signal attenuation from G132C at neutral pH suggests that the protein is making infrequent transient C- to N-terminal contacts. However, at pH 2.5, the PRE profile shows a cluster of long-range contacts between the C-terminal spin label and the C-terminal end and NAC region.

Pulsed-field gradient NMR (PFG-NMR) translational diffusion experiments were performed at low and neutral pH in order to obtain the effective hydrodynamic radius (R_h) of the proteins. The R_h value at 15 °C and pH 2.5 ($R_h = 30.1$ Å) is smaller than that at pH 7.4 ($R_h = 31.7$ Å), suggesting that low pH induces a conformational compaction relative to the neutral pH value. Uversky et al. have also observed a compaction at low pH relative to neutral pH;[31] they saw a decrease in the radius of gyration from 40 Å at pH 7.5 to 30 Å at pH 3.0. The large gap in the radius of gyration values relative to the hydrodynamic radii is because the radius of gyration, which measures the average distance of each residue from the center of mass of the protein, is more sensitive to the extension of the chain than the
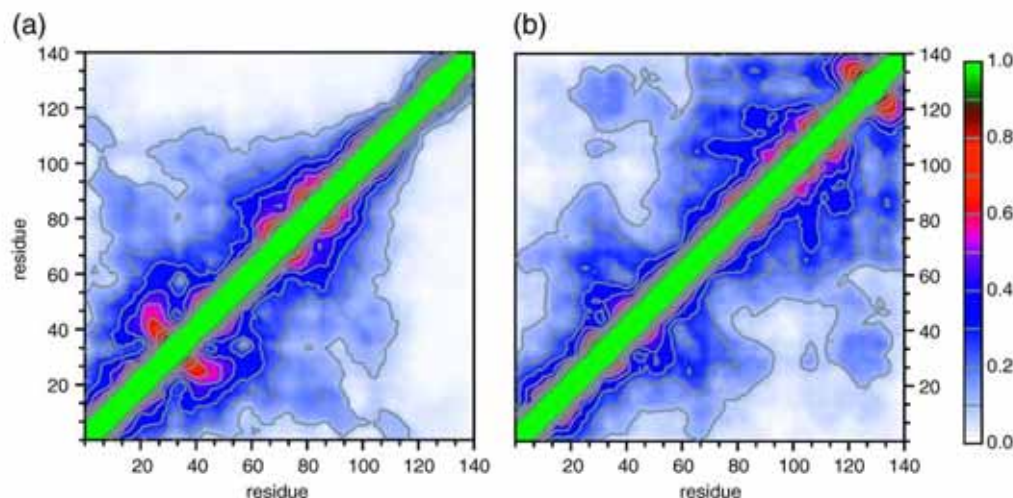

Fig. 4. REMD simulation results exhibit good agreement with experimental data. Contact maps of α-synuclein ensembles at neutral pH (a) and low pH (b) calculated using REMD. The color gradient from white, to blue, to red, to green represents percentages of the conformational ensembles that have inter-residue C^{α}–C^{α} distances within 20 Å.

hydrodynamic radius, which approximates the radius of a sphere whose diffusion constant equals that of the protein.[45]

Visualization of the conformational ensemble of α-synuclein by REMD at low pH

Replica exchange molecular dynamics (REMD) simulations were employed to generate conformational ensembles of α-synuclein at neutral and low pH. Long-range contacts within the neutral- and low-pH ensembles are consistent with the long-range contacts observed experimentally through PRE measurements. Figure 4 shows the percentage of the population within the low- and neutral-pH ensembles with inter-residue contacts (C^α–C^α distances) within 20 Å. A cutoff of 20 Å was chosen to facilitate comparison with experimental PRE measurements, which can detect contacts within 20–25 Å of the nitroxide spin label. At neutral pH (Fig. 4a), there are contacts observed throughout the N-terminus and NAC regions, while the C-terminal tail (residues 115–140) is largely extended, making contact with the other regions in only ~5% of the population. In the contact map for the low-pH ensemble (Fig. 4b), the region of the protein that shows many intramolecular contacts has switched relative to the neutral-pH ensemble. At low pH, the C-terminus is in regular contact with itself and the NAC region, as well as a small section of the N-terminus near residue 40, but only approaches the most N-terminal part of the protein in a small fraction of the population. The section of the N-terminus near residue 40 makes transient contact not only with the C-terminus but also with the N-terminal half of the NAC. However, the leading portion of the N-terminus, from residues 1 to 40, forms a separate cluster of long-range contacts and has fewer contacts with the rest of the protein. The switch from an extended C-terminus at neutral pH to a somewhat less extended N-terminus at low pH is also seen in calculations of the solvent-accessible surface area (SASA) (Fig. 5a). The C-terminus at neutral pH has the highest solvent accessibility (75.4%) of any region at either pH. However, at low pH, the N-terminus is the region most exposed to the solvent (68.6%), while the NAC and C-terminus have an equal amount of solvent accessibility (59%). Interestingly, the NAC solvent exposure is very similar at both neutral and low pH. It appears that the NAC region "swaps" intermolecular contacts with the N-terminal region at neutral pH for corresponding contacts with the C-terminal residues at low pH.

Simulated PRE intensity ratios for the neutral- and low-pH conformational ensembles (Fig. S1) portray the same trends seen in the contact maps. At neutral pH, the intensity ratio profiles show significant contacts between the N-terminus and NAC regions, while the C-terminal spin label site is only in contact with the rest of the sequence in a very small percentage of the population. Interestingly, the simulated ensemble does show a dip in the intensity ratio from
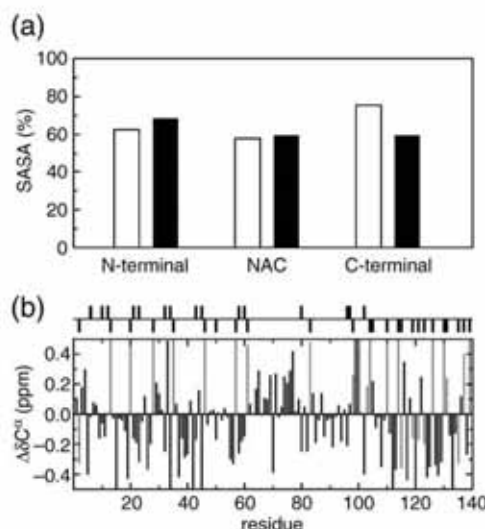
Fig. 5. SASAs at neutral and low pH. (a) Comparison of percentage of SASAs for the N-terminal, NAC, and C-terminal regions of the neutral (open bars) and low (filled bars) pH ensembles. A significant change in the solvent accessibility is observed in the C-terminal region, which is less exposed, supporting the idea of a collapsed conformation in this region. (b) Calculated differences in average chemical shifts between the low- and neutral-pH ensembles ($C^\alpha_{neutral} - C^\alpha_{low}$). Asp and Glu residues are colored light gray. The chemical shifts were calculated using ShiftX.[46]

G132C in the N-terminus near residue 10. At low pH, the calculated PRE profiles are consistent with fewer contacts between the N-terminus and the NAC but show very significant drops in the intensity ratio in the C-terminus from the A90C spin label and in the NAC from the G132C spin label. Overall, the calculated intensity ratio profiles, such as the contact maps, exhibit large changes in the C-terminus when going from neutral to low pH and smaller changes in the N-terminus.

The inter-residue contact maps (Fig. 4a and b) provide information about the long-range contacts of the ensemble. Local structural characteristics, particularly secondary-structure propensities, are often related to chemical shifts that are affected by the local environment of the nucleus being monitored.[24,47] We calculated the C^α chemical shifts for the low- and neutral-pH ensembles. There are large changes in C^α ($\Delta\delta C^\alpha$, Fig. 5b) that can be attributed to the change in the charge state of the Asp and Glu residues, just as with the experimental C^α measurements. However, even accounting for the protonation of the negatively charged residues, there are more differences in the C^α shifts of the simulation ensembles than are observed experimentally. The calculated $\Delta\delta C^\alpha$'s show that the biggest chemical shift difference is in the C-terminal region and that it not only corresponds to decreased β, similar to the experimental observation, but also features stretches

Table 2. Compaction ratios of REMD-simulated ensemble of α-synuclein

pH	N	NAC	C
Neutral	0.55	0.5	0.81
Low	0.63	0.56	0.52

The table shows the compaction ratios of the N, NAC, and C regions for the low-pH and neutral-pH α-synuclein simulation ensembles, calculated as the ratio of average end-to-end distance in the low (or neutral) pH ensemble to the same quantity for the denatured state ensemble.

of negative $\Delta\delta C^\alpha$ in the N-terminus and positive $\Delta\delta C^\alpha$ in the NAC region, indicative of decreased β propensity and decreased α propensity, respectively. The overall β structure of the ensembles, as calculated by STRIDE,[48] decreases as the pH is lowered, consistent with the trend predicted by the experimental $\Delta\delta C^\alpha$ measurement. There is no significant helical structure observed for either simulated ensemble. Though there is slightly more variation in the chemical shifts of the simulated ensembles compared to the experiments, in both cases, the largest differences are found in the C-terminus and the magnitude of the chemical shift differences throughout the sequence is very small. This suggests that there is little difference in secondary-structure propensity at neutral and low pH, with the biggest change being a slight decrease in β propensity in the C-terminus at low pH.

Relative packing of the N-terminal, NAC, and C-terminal regions of α-synuclein was evaluated by determining compaction ratios (see Materials and Methods). The values (Table 2) are consistent with the picture provided by the contact maps. At neutral pH, the N-terminus and NAC regions have similar compaction ratios, 0.55 and 0.5, respectively, but the C-terminus is much more expanded, with a compaction ratio of 0.81. In contrast, at low pH, the C-terminus has a compaction ratio similar to that of the NAC, 0.52 and 0.56, respectively, while the N-terminus is more expanded, with a compaction ratio of 0.63. Here, again, the region most affected by the change in pH is the C-terminus, whose compaction at low pH is reflected in the change of the compaction ratio from 0.81 to 0.52.

Molecular simulations allow for the visualization of individual structures within the conformational ensembles. We used a clustering algorithm[49] to select structures from the REMD simulations. The conformations from the neutral-pH ensemble shown in Fig. 6a were chosen from the top four clusters, representing more than 53% of the ensemble. The representative structures from the low-pH ensemble (Fig. 6b), chosen from the top three clusters and representing more than 60% of the total population, provide the first visualization of structures from an ensemble consistent with experimental observations at low pH. These figures highlight the structural changes at low pH relative to neutral pH previously observed in both the PRE measurements and the contact maps for the simulation ensembles, namely, the compaction of the C-terminal region at low pH coupled with an N-terminal region, which is less extended than the C-terminus at neutral pH.

Discussion

We have used two approaches, NMR spectroscopy and computer simulation, to characterize the conformational ensembles of the IDP α-synuclein at both neutral and low pH. By NMR, we have measured
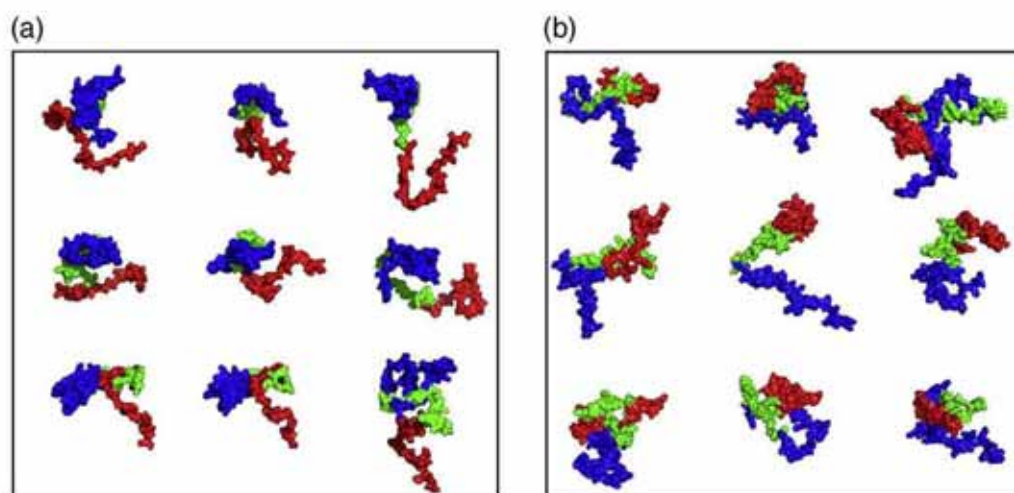
Fig. 6. Selected representative conformations of α-synuclein from the neutral-pH and low-pH ensembles. Representative structures chosen from (a) the top four clusters of the neutral-pH ensemble and (b) the top three clusters from the low-pH ensemble. These clusters account for more than 53% and 60% of the respective structural ensembles. The N-terminus is shown in blue, the NAC is shown in green, and the C-terminus is shown in red.

chemical shifts, RDCs, R_2^{CPMG}, and PREs of α-synuclein at pH 7.4 and 2.5 in order to compare the secondary structure and long-range interactions within the protein and to explore differences in dynamics. We have also used REMD with implicit solvent to simulate α-synuclein at neutral and low pH. The simulations allow us to visualize the conformational ensembles at low pH and compare these directly to the conformations populated at neutral pH.

α-Synuclein at neutral pH has been extensively studied by NMR.[15,29,30,32] Our experimental measurements at pH 7.4 reproduce the same general features observed in previous studies. The effect of the N-terminal spin label is seen throughout the remainder of the N-terminus and results in a small, uniform reduction in the intensity ratio in both the NAC and the C-terminus. We interpret the uniformly reduced intensity ratio throughout the NAC and C-terminus as a consequence of transient contacts between A19C and each residue from 60 to 140; it is a reflection of the heterogeneity of the structural ensemble since it implies that A19C is not involved in unique, long-range contacts with any particular residues in the C-terminal part of the protein. An alternative explanation for the plateau is that there is a scaling problem between the oxidized and reduced experiments that lowers the baseline but is not indicative of any actual contact between A19C and the residues of the NAC and C-terminal regions. However, the fact that a similar effect is seen in the PRE intensity ratios calculated from the REMD ensembles (Fig. S1) lends credibility to the observed results, since there can be no question of a scaling problem in the simulations. The neutral-pH conformational ensemble generated by the simulations qualitatively portrays the same long-range interactions observed by the NMR experiments. Intermittent contacts between the N-terminus and the other regions of the protein are seen both in the contact maps and in the calculated PRE intensity ratios. Moreover, in both the experimental and the calculated PRE plots, the C-terminal probe at G132C induces very little reduction in intensity from the line corresponding to the intensity ratio expected for a random coil, except for a small dip in the N-terminus near residue 20. This suggests that with the exception of a small population of conformations in which the N- and C-termini are in contact, the C-terminus is largely extended into solution.

While our analysis of the neutral-pH REMD ensemble suggests that it qualitatively captures the long-range interactions observed experimentally, it is quite different from the ensemble previously obtained by Bertoncini *et al.*[30] In their study, agreement with experiments was enforced through the use of PRE-derived distance constraints. The structures they show, representatives of seven clusters accounting for more than 50% of their conformational ensemble, are all shaped like rings, with the N- and C-termini in contact with each other. That two such different ensembles can be obtained, which both appear to agree with PRE data, is a reflection of

the fact that the determination of structural ensembles from PRE data is a very underdetermined problem.[50] PRE intensities are related to the distance between each residue and the nitroxide spin label via a $1/r^6$ average over the conformational ensemble. As a consequence of this $1/r^6$ distance dependence, PRE intensities are dominated by contributions from more compact conformations. Small changes in the population of conformations containing residues a short distance from the spin label will have a big effect on the $1/r^6$ average; this decreases the constraints on other conformers and can allow for the inclusion of large populations of expanded conformations in an ensemble matching a given PRE intensity ratio profile. This means that the distances derived from the spin labels attached to the protein in PRE experiments can be satisfied by many different conformational ensembles. It has recently been demonstrated that if it were experimentally feasible to attach many more spin labels, covering a large percentage of the protein, then there would be enough data generated to be able to uniquely identify the true conformational ensemble.[50]

NMR experiments and REMD simulations indicate that there is a significant structural reorganization within the low-pH ensemble relative to that at neutral pH in terms of long-range contacts, hydrodynamic radius, and the amount of heterogeneity within the conformational ensembles. The rearrangement of long-range contacts is observed by comparing the PRE measurements and REMD contact maps at low pH with the corresponding descriptions of the neutral-pH ensemble. Both the PREs and the contact maps show that the largest change takes place in the C-terminus. Whereas the C-terminal region in the neutral-pH ensemble is largely extended, only making contact with the other regions transiently, at low pH, there are numerous contacts between the C-terminus and the tailing portion of the NAC (Fig. 3e and f). The C-terminus at low pH is also more compact, has decreased flexibility based on ^{15}N relaxation data, and is less exposed to the solvent than at neutral pH. The response of the N-terminal region to the change in pH is less dramatic. While the N-terminus at low pH is both more extended and more exposed to solvent than at neutral pH, neither of these changes are as large as those observed in the C-terminus. These results highlight the change in the long-range interactions between the neutral- and low-pH ensembles; at neutral pH, there is a very heterogeneous ensemble with transient contacts between the N-terminus and the NAC; however, at low pH, there is a more homogeneous ensemble that exhibits strong contacts between the NAC and the C-terminus. Transient contacts between the N- and C-termini are observed at both pH values.

NMR experiments and REMD simulations show that the low-pH ensemble of α-synuclein has a smaller average hydrodynamic radius than the neutral-pH ensemble. From the experimental data and the visualization of the REMD structures, it is clear that there is no global collapse of the protein at

low pH but rather the compaction is local and exists primarily in the C-terminal domain and the C-terminal region of the NAC. Uversky *et al.* have proposed, based on CD and Fourier transform infrared spectroscopy measurements, that the low-pH form of α-synuclein has increased β content relative to the neutral-pH form, a relatively compact conformation, and can be thought of as a pre-molten globule state.[31] The detailed atomic description presented here indicates that the compaction is primarily restricted to the last 60 residues of the protein and that the N-terminal region is more exposed. In addition, there is no evidence at the individual residue level of an increase in β conformation in the low-pH ensemble relative to the neutral-pH ensemble.

The structural reorganization at low pH can be rationalized by considering the distribution of charges that exist in the low-pH form relative to the neutral-pH form. As seen in the charge density table (Table 1), the three regions of α-synuclein are impacted differently by lowering the pH. In particular, there is a large change in the total charge density in the C-terminus that decreases from 40% charged residues at neutral pH to only 6.7% charged residues at low pH. In contrast, the effect of low pH on the N-terminus is less pronounced with a shift from 30% charged residues at neutral pH to 18.3% charged residues at low pH and a total charge that is high at both pH values. Overall, the decrease in the total charge density of the C-terminus is considerably more than the change at the N-terminus and accounts for the fact that the reorganization of the C-terminus is the most pronounced and that of the N-terminus less so.

Uversky *et al.* have empirically observed that IDPs and globular proteins fall into two distinct regions as a function of their mean net hydrophobicity *versus* mean net charge.[51] They noted that in the plot of hydrophobicity *versus* charge, α-synuclein is one of the few examples of an IDP that is just over the line into the ordered region of the plot (Fig. 7). However, we consider the three functional regions of α-synuclein separately and note that the C-terminal region at neutral pH, with a net charge of −12, resides very definitively in the IDP region of the Uversky plot. At low pH, however, the C-terminus loses most of its net charge and becomes highly hydrophobic. The very dramatic movement of the C-terminus to the ordered region may be due to the high density of glutamates in the sequence that become very hydrophobic at low pH. The globular-like quality of the C-terminus that is predicted in the plot is reflected in the NMR experiments and REMD calculations where the C-terminal compaction is seen both in the experimental PRE measurements and the contact map of the simulated low-pH ensemble. The full α-synuclein sequence at low pH moves more fully into the ordered region of the charge *versus* hydrophobicity plot due to an increase in hydrophobicity. This shift is mirrored by the smaller hydrodynamic radius observed at low pH. Interestingly, because of the balance between
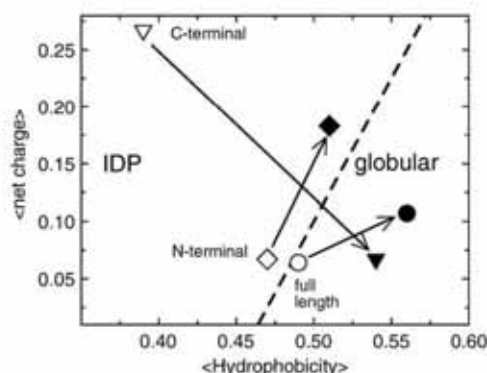
Fig. 7. Charge-hydrophobicity characteristics of α-synuclein. Charge-hydrophobicity characteristics of α-synuclein at low (filled symbols) and neutral (open symbols) pH. The hydrophobicity values for α-synuclein were calculated using the Cowan–Whittaker hydropathy indices[52] for the low- and neutral-pH conditions using a window size of five amino acids. The hydrophobicity values for the individual amino acids were normalized to a 0-to-1 scale. The mean net hydrophobicity (net charge) was calculated as the sum of normalized hydrophobicity (charge at specific pH) divided by the total number of residues in α-synuclein. The black line represents the boundary between globular proteins and IDPs determined using a list of these proteins, an analysis similar to that performed by Uversky *et al.*[51]

positively and negatively charged residues in the N-terminus at neutral pH, the N-terminus has both a higher net charge and a slightly higher net hydrophobicity at low pH and remains in the IDP region of the plot.

α-Synuclein has been shown to aggregate more rapidly at low pH, and a comparison of the structural and dynamical differences may aid in understanding which factors are critical to aggregation. Aggregation may be affected by a number of properties including propensity to form transient secondary structure,[23] the distribution of long-range interactions,[30,32] the hydrodynamic radius of the protein,[31] the flexibility of different regions and their exposure to solvent,[29,53] and the distribution of charges along the protein.[54] We find that the C^α chemical shifts of the neutral- and low-pH forms are quite similar, and this suggests that secondary-structure propensity is not a driving force to increase the aggregation rate at low pH. It has previously been suggested that the increased rates of aggregation observed for single point mutations[25] or upon binding of divalent metal ions[55] or polycations[30,56] are due to the release of the long-range N- to C-terminal contacts and subsequent exposure of the NAC. However, the results of the REMD simulations suggest that the NAC region has comparable exposure to solvent at both neutral and low pH; NMR R_2^{CPMG} measurements suggest that the NAC region has comparable flexibility in both pH forms. Similar exposure of the NAC region to solvent in both the slower, neutral-pH and faster, low-pH aggregating species suggests that protection or

exposure of the NAC region may not have a large impact on the aggregation propensity in this context.

At low pH, where the charge profiles of both the N- and C-termini are significantly different from their neutral-pH counterparts, it is possible that initiation of aggregation may arise through a different mechanism. Previously, we proposed for mouse α-synuclein at neutral pH that aggregation may be initiated through intermolecular N-terminal contacts. While the N-terminus has a balance of positively and negatively charged residues at neutral pH in both mouse and human α-synuclein, it carries a large positive charge at low pH. Without complementary negative charges anywhere in the sequence, it may be reasonable to assume that the N-terminus is unlikely to initiate intermolecular contacts at low pH. One scenario that is consistent with the structural changes and charge distribution at low pH is that the locally collapsed C-terminus, which loses its net charge and becomes highly hydrophobic under these conditions, is involved in the earliest stages of α-synuclein aggregation at low pH. These data indicate the importance of the charge distribution in directing both the mechanism and the rates of aggregation.[54]

Materials and Methods

Sample preparation

Expression and purification of human α-synuclein in *Escherichia coli* were followed by the described protocol.[29] Further purification using diethylaminoethyl or Q anion-exchange chromatography was applied to ensure high purity of α-synuclein for low-pH studies. α-Synuclein was eluted at 200–300 mM NaCl gradient. Solution containing pure α-synuclein was dialyzed against with 20 volumetric times of phosphate-buffered saline (PBS; pH 7.4) at 4 °C and repeated 3 times. Dialyzed and concentrated α-synuclein (200 µM) was transferred to 1.5-ml tubes and quickly frozen by cryogenic liquid nitrogen for −80 °C storage. For each NMR sample, the −80 °C stored proteins were unfrozen at 4 °C, polished using a 100-kDa filter (Millipore), and then dialyzed to pH 2.5 (10 mM phosphate and 140 mM NaCl) with the addition of 10% D$_2$O.

NMR experiments for assignment

NMR experiments including triple-resonance experiments and ^{15}N backbone dynamics experiments were performed on a Varian 800-MHz spectrometer equipped with triple gradient probe. PFG-NMR diffusion, PRE and RDC measurements were acquired on a Varian 600-MHz spectrometer with a cold probe or on a Bruker 700-MHz spectrometer. Acquired spectra were processed by NMRPipe[57] and analyzed by Sparky.[58]

^{13}C–^{15}N-labeled α-synuclein (650 µM) at pH 2.5 was used to perform triple-resonance experiments for backbone assignment at 15 °C. The 3D spectra were acquired via the conventional strategy[59] using HNCACB, CBCA (CO)NH, HNCO, and HN(CA)CO and a new approach using HNN[41] via ^{15}N–^{15}N connections. Data collection time for each 3D spectrum at 15 °C is shorter than 20 h (with 8 or 16 transient scans) to avoid aggregation during acquisition. HSQC spectra were collected to check the stability of α-synuclein before acquiring the next 3D spectrum.

Chemical shifts were calibrated using 2,2-dimethyl-2-silapentane-5-sulfonate dissolved at pH 2.5 as a reference.[60] Comparing the secondary-structure propensity using the C$^\alpha$ chemical shift deviation from the random-coil shifts is not appropriate since the random-coil C$^\alpha$ chemical shifts measured at pH 5.0[61] and 2.3[62,63] exhibit significant differences. Directly monitoring the changes of C$^\alpha$ chemical shifts ($\Delta\delta C^\alpha = \delta C^\alpha_{7.4} - \delta C^\alpha_{2.5}$) was used to inspect the secondary-structure propensities at pH 7.4 and 2.5.

Translational diffusion coefficients and hydrodynamic radii

PFG-NMR experiments incorporated with longitudinal eddy current pulse schemes and convection compensation were used to measure the translational diffusion coefficients (D_{trans}).[64] Samples containing 300 µM α-synuclein and 35 mM 1,4-dioxane were dissolved in PBS buffers at pH 7.4 and 2.5 (100% D$_2$O, pD uncorrected) for PFG-NMR experiments at 15 °C. Twenty-five 1D PFG-NMR spectra were acquired over a range of gradient strengths of 2 to 17 G/cm or 5 to 50 G/cm for 1,4-dioxane or α-synuclein, respectively. Integrated peak volumes of 1,4-dioxane and α-synuclein (methyl groups only) were fitted using VNMRJ 2.1 (Varian Inc., California) and used to calculate D_{trans}.[64] Since α-synuclein and dioxane were dissolved in one solution, the viscosity effect on D_{trans} can be ignored. Therefore, the Stokes–Einstein equation can be simplified to get the hydrodynamic radius of α-synuclein by the relationship:

$$R_h(\alpha - \text{synuclein}) \times D_{trans}(\alpha - \text{synuclein})$$
$$= R_h(\text{dioxane}) \times D_{trans}(\text{dioxane})$$

where R_h (dioxane) is 2.12 Å.

^{15}N relaxation

R_2^{CPMG} ^{15}N relaxation experiments were acquired using 250 µM ^{15}N-labeled α-synuclein at pH 2.5 and 6.1. Complex points for each spectrum were 1024×256 in the ^{1}H and ^{15}N dimensions, respectively. The relaxation times for R_2^{CPMG} were 0.01, 0.03, 0.05, 0.09, 0.13, 0.17, 0.19, 0.21, and 0.25 s. Recycle delays for each experiment were 2 s. Data were processed by NMRPipe and analyzed by Sparky.

Site-directed spin label

Methods for (1-oxy-2,2,5,5-tetra-methyl-3-pyrroline-3-methyl)-methanesulfonate (MTSL) (Toronto Research Chemicals, Ontario, Canada) spin-labeled ^{15}N-α-synuclein A19C, A90C, or G132C were done as described in our previous work.[29] In brief, 250 µM MTSL-labeled α-synuclein cysteine mutants were dialyzed to pH 2.5 or 7.4 and were divided to equal volume for experiments at paramagnetic (oxidized) and diamagnetic (reduced) states. Addition of 10 mM L-ascorbate (pH 2.5) or 10 mM dithiothreitol (DTT) (pH 7.4) was applied for generating the diamagnetic samples. It is not possible to use DTT or L-ascorbate as a reducing reagent at both pH values. DTT cannot be used at low pH as the disulfide bond is not efficiently broken at pH 2.5 because the thiol groups are

highly protonated. At neutral pH, we have observed significant signal enhancements and changes of ^{1}H and ^{15}N chemical shifts in the presence of L-ascorbate. PRE ratios (I_{ox}/I_{red}) were calculated as the intensity ratios of the same residue in the absence (I_{ox}) or in the presence (I_{red}) of L-ascorbate or DTT. At low pH, all mutant samples were normalized against L-ascorbate, and at neutral pH, all samples were normalized against DTT.

RDC measurement

n-Alkyl-poly(ethylene glycol)/*n*-alkyl alcohol mixture[65] was chosen to prepare the liquid crystalline medium for RDC experiments; 5% C8E5 and 1-octanol were mixed with PBS buffer at pH 2.5 or 7.4. The molar ratio of C8E5 and 1-octanol is 1.05. The quadrupolar deuterium splitting constants are 23.8 and 22.7 Hz at pH 7.4 and 2.5, respectively. High-resolution HSQC_IPAP spectra in the absence or in the presence of an alignment medium were collected at 15 °C with complex points of 2048 (t_2)×512 (t_1) and 16 transient scans.

REMD simulations

REMD[66,67] simulations were used to generate the conformational ensembles of α-synuclein at low and neutral pH. In this method, a number of simulations (replicas) are run in parallel over a specified temperature range. The adjacent replicas (T_i and T_j) are allowed to exchange periodically, with an acceptance criterion based on the following Metropolis transition probability:

$$W\{T_i, T_j\} \rightarrow \{T_j, T_i\} = \min\left(1, \exp\left[-\left(\beta_j - \beta_i\right)\left(E_j - E_i\right)\right]\right)$$

where $\beta_{i(j)} = 1/KT_{i(j)}$ and $E_{i(j)}$ is the potential energy of the *i*th (*j*th) replica. This method provides canonical probability distributions for the ensembles over the specified temperature range.[67] The REMD method has been implemented in the IMPACT simulation package.[68] The simulations were performed using the AGBNP implicit solvent[69] model and the OPLS-AA force field.[70]

The simulation was performed for a total of 10 ns over 20 temperatures (300, 308, 317, 325, 334, 343, 353, 362, 372, 382, 393, 403, 414, 426, 437, 449, 461, 474, 487, and 500 K), corresponding to a total simulation time of 200 ns. All simulations were started with a fully extended conformation of the α-synuclein molecule. The sequence of α-synuclein, for the low-pH condition, was modified to a state corresponding to that of the molecule under these conditions, by protonating the side-chain carbonyl groups of all of the acidic residues (aspartates and glutamates). α-Synuclein was also simulated with a purely repulsive potential, in which all interactions have been turned off except for the repulsive part of the Lennard–Jones potential, to obtain the ensemble of structures corresponding to the denatured or most expanded state. The 449 and 414 K ensembles were used for the structural characterization of the low- and neutral-pH ensembles. These ensembles were chosen because their hydrodynamic radii (29.1 and 30.5 Å, respectively) most closely matched the experimental measurements at 15 °C.

A number of analysis methods were used for the structural characterization of the low-pH ensembles. The average size of the ensembles was determined by calculating the hydrodynamic radii for individual structures using Hydropro.[71] The secondary structural assignments for the α-synuclein ensembles were made using STRIDE.[48] C$^\alpha$

chemical shifts were calculated using ShiftX.[46] The SASAs were calculated using the program Surfv.[72] The clustering of the neutral- and low-pH ensemble structures was performed using a hierarchical clustering method,[49] with the clustering based on a distance matrix of distances between the center of mass of the three regions of synuclein (N-terminus, NAC, and C-terminus). The compaction factors for the low- and neutral-pH conditions were calculated as the ratio of the average end-to-end distance between the low (neutral) pH and the denatured state of α-synuclein. The end-to-end distance of the denatured state was calculated from an ensemble generated using a purely repulsive potential, which gives the largest average dimensions possible for the protein.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.06.063

References

1. Lansbury, P. T., Jr (1999). Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease. *Proc. Natl Acad. Sci. USA*, **96**, 3342–3344.
2. Perutz, M. F. (1999). Glutamine repeats and neurodegenerative diseases: molecular aspects. *Trends Biochem. Sci.* **24**, 58–63.
3. Dobson, C. M. (1999). Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329–332.
4. Chiti, F. & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366.
5. Siderowf, A. & Stern, M. (2003). Update on Parkinson disease. *Ann. Intern. Med.* **138**, 651–658.
6. Rochet, J. C. & Lansbury, P. T., Jr (2000). Amyloid fibrillogenesis: themes and variations. *Curr. Opin. Struct. Biol.* **10**, 60–68.
7. Davidson, W. S., Jonas, A., Clayton, D. F. & George, J. M. (1998). Stabilization of alpha-synuclein secondary structure upon binding to synthetic membranes. *J. Biol. Chem.* **273**, 9443–9449.
8. Jo, E., McLaurin, J., Yip, C. M., St George-Hyslop, P. & Fraser, P. E. (2000). Alpha-synuclein membrane interactions and lipid specificity. *J. Biol. Chem.* **275**, 34328–34334.
9. Abeliovich, A., Schmitz, Y., Farinas, I., Choi-Lundberg, D., Ho, W.-H., Castillo, P. E. *et al.* (2000). Mice lacking alpha-synuclein display functional deficits in the nigrostriatal dopamine system. *Neuron*, **25**, 239–252.

10. Cooper, A. A., Gitler, A. D., Cashikar, A., Haynes, C. M., Hill, K. J., Bhullar, B. *et al.* (2006). Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. *Science,* **313,** 324–328.

11. Jenco, J. M., Rawlingson, A., Daniels, B. & Morris, A. J. (1998). Regulation of phospholipase D2: selective inhibition of mammalian phospholipase D isoenzymes by alpha- and beta-synucleins. *Biochemistry,* **37,** 4901–4909.

12. George, J. M., Jin, H., Woods, W. S. & Clayton, D. F. (1995). Characterization of a novel protein regulated during the critical period for song learning in the zebra finch. *Neuron,* **15,** 361–372.

13. Yavich, L., Tanila, H., Vepsalainen, S. & Jakala, P. (2004). Role of alpha-synuclein in presynaptic dopamine recruitment. *J. Neurosci.* **24,** 11165–11170.

14. Wood, S. J., Wypych, J., Steavenson, S., Louis, J. C., Citron, M. & Biere, A. L. (1999). Alpha-synuclein fibrillogenesis is nucleation-dependent. Implications for the pathogenesis of Parkinson's disease. *J. Biol. Chem.* **274,** 19509–19512.

15. Eliezer, D., Kutluay, E., Bussell, R., Jr & Browne, G. (2001). Conformational properties of alpha-synuclein in its free and lipid-associated states. *J. Mol. Biol.* **307,** 1061–1073.

16. Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A. & Lansbury, P. T., Jr (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry,* **35,** 13709–13715.

17. Uversky, V. N. (2003). A protein-chameleon: conformational plasticity of α-synuclein, a disordered protein involved in neurodegenerative disorders. *J. Biomol. Struct. Dyn.* **21,** 211–234.

18. Ulmer, T. S., Bax, A., Cole, N. B. & Nussbaum, R. L. (2005). Structure and dynamics of micelle-bound human alpha-synuclein. *J. Biol. Chem.* **280,** 9595–9603.

19. Bussell, R., Jr & Eliezer, D. (2003). A structural and functional role for 11-mer repeats in alpha-synuclein and other exchangeable lipid binding proteins. *J. Mol. Biol.* **329,** 763–778.

20. El-Agnaf, O. M., Bodles, A. M., Guthrie, D. J., Harriott, P. & Irvine, G. B. (1998). The N-terminal region of non-A beta component of Alzheimer's disease amyloid is responsible for its tendency to assume beta-sheet and aggregate to form fibrils. *Eur. J. Biochem.* **258,** 157–163.

21. Kim, T. D., Paik, S. R. & Yang, C. H. (2002). Structural and functional implications of C-terminal regions of alpha-synuclein. *Biochemistry,* **41,** 13782–13790.

22. Bertoncini, C. W., Rasia, R. M., Lamberto, G. R., Binolfi, A., Zweckstetter, M., Griesinger, C. & Fernandez, C. O. (2007). Structural characterization of the intrinsically unfolded protein beta-synuclein, a natural negative regulator of alpha-synuclein aggregation. *J. Mol. Biol.* **372,** 708–722.

23. Sung, Y. H. & Eliezer, D. (2007). Residual structure, backbone dynamics, and interactions within the synuclein family. *J. Mol. Biol.* **372,** 689–707.

24. Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. (2006). Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci.* **15,** 2795–2804.

25. Bertoncini, C. W., Fernandez, C. O., Griesinger, C., Jovin, T. M. & Zweckstetter, M. (2005). Familial mutants of alpha-synuclein with increased neurotoxicity have a destabilized conformation. *J. Biol. Chem.* **280,** 30649–30652.

26. Bussell, R., Jr & Eliezer, D. (2001). Residual structure and dynamics in Parkinson's disease-associated mutants of alpha-synuclein. *J. Biol. Chem.* **276,** 45996–46003.

27. Bussell, R., Jr & Eliezer, D. (2004). Effects of Parkinson's disease-linked mutations on the structure of lipid-associated alpha-synuclein. *Biochemistry,* **43,** 4810–4818.

28. Rospigliosi, C. C., McClendon, S., Schmid, A. W., Ramlall, T. F., Barrè, P., Lashuel, H. A. & Eliezer, D. (2009). E46K Parkinson's-linked mutation enhances C-terminal-to-N-terminal contacts in alpha-synuclein. *J. Mol. Biol.* **388,** 1022–1032.

29. Wu, K. P., Kim, S., Fela, D. A. & Baum, J. (2008). Characterization of conformational and dynamic properties of natively unfolded human and mouse alpha-synuclein ensembles by NMR: implication for aggregation. *J. Mol. Biol.* **378,** 1104–1115.

30. Bertoncini, C. W., Jung, Y. S., Fernandez, C. O., Hoyer, W., Griesinger, C., Jovin, T. M. & Zweckstetter, M. (2005). Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *Proc. Natl Acad. Sci. USA,* **102,** 1430–1435.

31. Uversky, V. N., Li, J. & Fink, A. L. (2001). Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J. Biol. Chem.* **276,** 10737–10744.

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

33. Bisaglia, M., Mammi, S. & Bubacco, L. (2009). Structural insights on physiological functions and pathological effects of alpha-synuclein. *FASEB J.* **23,** 329–340.

34. Mittag, T. & Forman-Kay, J. D. (2007). Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **17,** 3–14.

35. Eliezer, D. (2009). Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **19,** 23–30.

36. Bracken, C., Iakoucheva, L. M., Romero, P. R. & Dunker, A. K. (2004). Combining prediction, computation and experiment for the characterization of protein disorder. *Curr. Opin. Struct. Biol.* **14,** 570–576.

37. Wright, P. E. & Dyson, H. J. (2009). Linking folding and binding. *Curr. Opin. Struct. Biol.* **19,** 31–38.

38. Vitalis, A., Wang, X. & Pappu, R. V. (2007). Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories. *Biophys. J.* **93,** 1923–1937.

39. Khandogin, J. & Brooks, C. L., III (2007). Linking folding with aggregation in Alzheimer's β-amyloid peptides. *Proc. Natl Acad. Sci. USA,* **104,** 16880–16885.

40. Baumketner, A., Bernstein, S. L., Wyttenbach, T., Bitan, G., Teplow, D. B., Bowers, M. T. & Shea, J.-E. (2006). Amyloid β-protein monomer structure: a computational and experimental study. *Protein Sci.* **15,** 420–428.

41. Panchal, S. C., Bhavesh, N. S. & Hosur, R. V. (2001). Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and ^{15}N sequential correlations in (^{13}C, ^{15}N) labeled proteins: application to unfolded proteins. *J. Biomol. NMR,* **20,** 135–147.

42. Mohana-Borges, R., Goto, N. K., Kroon, G. J., Dyson, H. J. & Wright, P. E. (2004). Structural characterization

of unfolded states of apomyoglobin using residual dipolar couplings. *J. Mol. Biol.* **340**, 1131–1142.

43. Kristjansdottir, S., Lindorff-Larsen, K., Fieber, W., Dobson, C. M., Vendruscolo, M. & Poulsen, F. M. (2005). Formation of native and non-native interactions in ensembles of denatured ACBP molecules from paramagnetic relaxation enhancement studies. *J. Mol. Biol.* **347**, 1053–1062.

44. Meier, S., Blackledge, M. & Grzesiek, S. (2008). Conformational distributions of unfolded polypeptides from novel NMR techniques. *J. Chem. Phys.* **128**, 052204.

45. Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A. & Smith, L. J. (1999). Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry*, **38**, 16424–16431.

46. Neal, S., Nip, A. M., Zhang, H. & Wishart, D. S. (2003). Rapid and accurate calculation of protein ^{1}H, ^{13}C and ^{15}N chemical shifts. *J. Biomol. NMR*, **26**, 215–240.

47. Wishart, D. S. & Sykes, B. D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR*, **4**, 171–180.

48. Frishman, D. & Argos, P. (1995). Knowledge based protein secondary structure assignment. *Proteins*, **23**, 566–579.

49. Johnson, S. (1967). Hierarchical clustering schemes. *Psychometrika*, **32**, 241–254.

50. Ganguly, D. & Chen, J. Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states. *J. Mol. Biol.* **390**, 467–477.

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

52. Cowan, R. & Whittaker, R. (1990). Hydrophobicity indices for amino acid residues as determined by high-performance liquid chromatography. *Pept. Res.* **3**, 75–80.

53. Hoyer, W., Cherny, D., Subramaniam, V. & Jovin, T. M. (2004). Impact of the acidic C-terminal region comprising amino acids 109–140 on alpha-synuclein aggregation in vitro. *Biochemistry*, **43**, 16233–16242.

54. Rivers, R. C., Kumita, J. R., Tartaglia, G. G., Dedmon, M. M., Pawar, A., Vendruscolo, M. *et al.* (2008). Molecular determinants of the aggregation behavior of alpha- and beta-synuclein. *Protein Sci.* **17**, 887–898.

55. Binolfi, A., Rasia, R. M., Bertoncini, C. W., Ceolin, M., Zweckstetter, M., Griesinger, C. *et al.* (2006). Interaction of alpha-synuclein with divalent metal ions reveals key differences: a link between structure, binding specificity and fibrillation enhancement. *J. Am. Chem. Soc.* **128**, 9893–9901.

56. Fernandez, C. O., Hoyer, W., Zweckstetter, M., Jares-Erijman, E. A., Subramaniam, V., Griesinger, C. & Jovin, T. M. (2004). NMR of alpha-synuclein-polyamine complexes elucidates the mechanism and kinetics of induced aggregation. *EMBO J.* **23**, 2039–2046.

57. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.

58. Goddard, T. D. & Kneller, D. G. Sparky 3. University of California, San Francisco, CA.

59. Ferentz, A. E. & Wagner, G. (2000). NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q. Rev. Biophys.* **33**, 29–65.

60. Wishart, D. S. & Case, D. A. (2001). Use of chemical shifts in macromolecular structure determination. *Methods Enzymol.* **338**, 3–34.

61. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995).). ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR*, **5**, 67–81.

62. Schwarzinger, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E. & Dyson, H. J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* **123**, 2970–2978.

63. Schwarzinger, S., Kroon, G. J., Foss, T. R., Wright, P. E. & Dyson, H. J. (2000). Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J. Biomol. NMR*, **18**, 43–48.

64. Li, Y., Kim, S., Brodsky, B. & Baum, J. (2005). Identification of partially disordered peptide intermediates through residue-specific NMR diffusion measurements. *J. Am. Chem. Soc.* **127**, 10490–10491.

65. Ruckert, M. & Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122**, 7793–7797.

66. Felts, A., Harano, Y., Gallichio, E. & Levy, R. (2004). Free-energy surfaces of beta hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model. *Proteins: Struct., Funct., Bioinf.* **56**, 310–321.

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

68. Banks, J. L., Beard, H. S., Cao, Y., Cho, A. E., Damm, W., Farid, R. *et al.* (2005). Integrated modeling program, applied chemical theory (IMPACT). *J. Comput. Chem.* **26**, 1752–1780.

69. Gallicchio, E. & Levy, R. M. (2004). AGBNP: an analytic implicit solvent model suitable for molecular dynamics and high-resolution modeling. *J. Comput. Chem.* **25**, 479–499.

70. Kaminsky, G., Friesner, R., Tirado-Rives, J. & Jorgensen, W. (2001). 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.

71. Garcia de la Torre, J., Huertas, M. L. & Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78**, 719–730.

72. Sridharan, S., Nicholls, A. & Honig, B. (1992). A new vertex algorithm to calculate solvent accessible surface areas. *Biophys. J.* **61**, A194.

9.4 References

Bertoncini, C. W., Y. S. Jung, et al. (2005). "Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein." Proc Natl Acad Sci U S A **102**(5): 1430-1435.

Cho, M. K., G. Nodet, et al. (2009). "Structural characterization of alpha-synuclein in an aggregation prone state." Protein Sci **18**(9): 1840-1846.

Dedmon, M. M., K. Lindorff-Larsen, et al. (2005). "Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations." J. Am. Chem. Soc. **127**: 476-477.

Jonsson, S. A., S. Mohanty, et al. (2012). "Distinct phases of free alpha-synuclein-a monte carlo study." Proteins-Structure Function and Genetics.

Marsh, J. A. and J. D. Forman-Kay (2010). "Sequence determinants of compaction in intrinsically disordered proteins." Biophys J **98**(10): 2383-2390.

McClendon, S., C. C. Rospigliosi, et al. (2009). "Charge neutralization and collapse of the C-terminal tail of alpha-synuclein at low pH." Protein Sci **18**(7): 1531-1540.

Trexler, A. J. and E. Rhoades (2010). "Single Molecule Characterization of alpha-Synuclein in Aggregation-Prone States." Biophysical Journal **99**(9): 3048-3055.

Uversky, V. N., J. R. Gillespie, et al. (2000). "Why are "natively unfolded" proteins unstructured under physiologic conditions?" Proteins **41**(3): 415-427.

Uversky, V. N., J. Li, et al. (2001). "Evidence of partially folded intermediate in alpha-synuclein fibril formation." Journal of Biological Chemistry **276**: 10737-10744.

Wilkins, D. K., S. B. Grimshaw, et al. (1999). "Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques." Biochemistry **38**(50): 16424-16431.

Chapter 10

Summary

10.1 Conclusions, Implications and Future directions

We have studied the conformational characteristics of the intrinsically disordered protein α-synuclein using ensembles generated from replica exchange molecular dynamics simulations, and further comparison with experimental NMR observations. Partially stable minipeptides serve as good model systems for determining best approaches to combining computational simulations with experiments. By comparing conformational ensembles generated from simulations with experimental observations, the study of minipeptide GB1 provided models for visualizing the heterogeneity in the experimental ensemble, in addition to serving as a means for calibrating the effective potential. Studies on S-peptide system highlighted the inherent ambiguity of using backbone carbon chemical shifts in the determination of secondary structure by explicit distribution of structures than the average structure.

α-synuclein aggregates into fibrils in patients with Parkinson's disease – the mechanism by which the transition from the monomeric disordered state to final fibrillar state occurs is poorly understood. In this study, conformational characteristics of the intrinsically disordered α-synuclein are described under physiological and aggregation-prone states. The results of this study show that at extremes of solvent quality, the conformational properties of α-synuclein are insensitive to the presence of charged residues along the chain. However, at intermediate temperature, which has global characteristics similar to experiment, the presence of charged residues plays a key role in the observed

conformational properties of α-synuclein.

Weighted subensembles were selected based on fits to experimental parameters reflecting the local and long-range conformational properties of α-synuclein. As a next step to increasing the robustness of the ensemble characterization, experimental parameters could be incorporated as restraints in the form of energy functions (Esteban-Martin, Fenwick et al. 2010) to see if the inclusion of experimental data alters the properties of the ensembles or restricts the apparent heterogeneity.

Comparison of the conformational characteristics of α-synuclein under aggregation prone low pH state with that under physiological condition highlights the effect of charged residues on the conformations adopted under these conditions. With decrease in pH the net charge of and along the chain changes significantly. Consequently, a significant structural reorganization of the ensemble is observed with decrease in pH leading to the collapse of the C-terminal region, consistent with that expected for a chain with diminished charge and increased hydrophobicity.

The broader goal of these studies is to characterize the conformational characteristics of α-synuclein that lead to the formation of higher order aggregates and fibrils leading to the diseased state. A number of factors including increased temperatures and changes in salt result in accelerated aggregation of α-synuclein (Uversky, Li et al. 2001; Munishkina, Henriques et al. 2004). Aggregation proceeds from the disordered monomer to higher order aggregates via oligomeric species. A number of studies have emphasized the importance of on-pathway intermediate including oligomers in cellular toxicity (Volles and Lansbury 2003; Chiti and Dobson 2006; Roychaudhuri, Yang et al. 2009; Zraika, Hull et al. 2010). Understanding factors that contribute to the formation of these earliest

stages oligomers, including dimers, is essential to our basic understanding of the principles of protein misfolding.

The mechanisms by which monomers associate to form dimers can be studied using computational simulations. The REMD simulations used to characterize the monomeric state of α-synuclein could be used as starting conformations for dimer simulations. Experimental inter-chain PREs determined for transient dimer (Wu and Baum 2010) conformations could be used to guide simulations, and provide visualizations of the transient dimer conformations corresponding to experimental observations.

10.2 Implications of the tetramer conformation of α-synuclein

The physiological form of α-synuclein was long thought to be a disordered monomer with transient secondary and long-range conformational propensities. Two studies of α-synuclein obtained under non-denaturing purification conditions suggest that this protein exists as a tetramer with a tendency to adopt helical conformations (Bartels, Choi et al. 2011; Wang, Perovic et al. 2011). Other studies performed under similar conditions have failed to support this observation (Fauvet, Mbefo et al. 2012) and in turn suggest that α-synuclein adopts disordered monomeric conformations under cellular conditions (Fauvet, Mbefo et al. 2012; Kang, Moriarty et al. 2012; Maltsev, Ying et al. 2012; Trexler and Rhoades 2012).

The identification of ability of α-synuclein to adopt the tetrameric state, while still debated, necessitates a reassessment of our current understanding of the native state of α-synuclein and the key processes leading to the disease state. While this determination of α-synuclein's tetrameric conformation provides new pathways for developing therapeutic

drugs that stabilize this aggregation-resistant form, it is essential to confirm if this is indeed the physiological state in which α-synuclein exists in the cell. In the disease state, even the tetrameric form may be expected to destabilize and reassemble to form aggregated fibril forms (Bartels, Choi et al. 2011). It has been suggested that the tetramers are likely to undergo destabilization in the disease state, giving rise to monomers that would then aggregate to form fibrils (Bartels, Choi et al. 2011). This mechanism of conversion from the stable tetramers to fibrils via monomeric intermediates is analogous to the aggregation mechanism of transtheyretin (Quintas, Saraiva et al. 1999). The current knowledge of the conformational characteristics of the monomeric, disordered form can be applied to the monomeric intermediates of the pathway leading to the formation of oligomers and higher order aggregates and fibrils.

10.3 References

Bartels, T., J. G. Choi, et al. (2011). "alpha-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation." Nature **477**(7362): 107-110.

Chiti, F. and C. M. Dobson (2006). "Protein misfolding, functional amyloid, and human disease." Annu Rev Biochem **75**: 333-366.

Esteban-Martin, S., R. B. Fenwick, et al. (2010). "Refinement of ensembles describing unstructured proteins using NMR residual dipolar couplings." J Am Chem Soc **132**(13): 4626-4632.

Fauvet, B., M. K. Mbefo, et al. (2012). "Alpha-synuclein in the central nervous system and from erythrocytes, mammalian cells and E. coli exists predominantly as a disordered monomer." J Biol Chem.

Kang, L., G. M. Moriarty, et al. (2012). "N-terminal acetylation of alpha-synuclein induces increased transient helical propensity and decreased aggregation rates in the intrinsically disordered monomer." Protein science : a publication of the Protein Society **21**(7): 911-917.

Maltsev, A. S., J. Ying, et al. (2012). "Impact of N-Terminal Acetylation of alpha-Synuclein on Its Random Coil and Lipid Binding Properties." Biochemistry **51**(25): 5004-5013.

Munishkina, L. A., J. Henriques, et al. (2004). "Role of protein-water interactions and electrostatics in alpha-synuclein fibril formation." Biochemistry **43**(11): 3289-3300.

Quintas, A., M. J. Saraiva, et al. (1999). "The tetrameric protein transthyretin dissociates to a non-native monomer in solution. A novel model for amyloidogenesis." The Journal of biological chemistry **274**(46): 32943-32949.

Roychaudhuri, R., M. Yang, et al. (2009). "Amyloid beta-protein assembly and Alzheimer disease." J Biol Chem **284**(8): 4749-4753.

Trexler, A. J. and E. Rhoades (2012). "N-Terminal acetylation is critical for forming alpha-helical oligomer of alpha-synuclein." Protein science : a publication of the Protein Society **21**(5): 601-605.

Uversky, V. N., J. Li, et al. (2001). "Evidence of partially folded intermediate in alpha-synuclein fibril formation." Journal of Biological Chemistry **276**: 10737-10744.

Volles, M. J. and P. T. Lansbury, Jr. (2003). "Zeroing in on the pathogenic form of alpha-synuclein and its mechanism of neurotoxicity in Parkinson's disease." Biochemistry **42**(26): 7871-7878.

Wang, W., I. Perovic, et al. (2011). "A soluble alpha-synuclein construct forms a dynamic tetramer." Proceedings of the National Academy of Sciences of the United States of America **108**(43): 17797-17802.

Wu, K. P. and J. Baum (2010). "Detection of transient interchain interactions in the intrinsically disordered protein alpha-synuclein by NMR paramagnetic relaxation enhancement." J Am Chem Soc **132**(16): 5546-5547.

Zraika, S., R. L. Hull, et al. (2010). "Toxic oligomers and islet beta cell death: guilty by association or convicted by circumstantial evidence?" Diabetologia **53**(6): 1046-1056.

Curriculum Vitae

Chitra Narayanan

Education

1998-2001   B.Sc in Botany, S.D.N.B. Vaishnav College for Women, Chennai, India

2002-2004   M.Sc in Biotechnology, Indian Institute of Technology Roorkee, India

2004-2005   Research Assistant under the guidance of Prof. N. Srinivasan, Indian Institute of Science, Bangalore India

2005-2012   Ph.D in Biochemistry, Rutgers, The State University of New Jersey, Piscataway, NJ USA

Publications

1. Narayanan C, Weinstock DS, Wu K-P, Baum J, Levy RM. Investigation of Polymeric Properties of α-synuclein and Comparison with NMR Experiments: A Replica Exchange Molecular Dynamics Study. *J. Chem. Theory Comput*. Accepted.

2. Wu KP, Weinstock DS, Narayanan C, Levy RM, Baum J. 2009. Structural Reorganization of alpha-Synuclein at Low pH Observed by NMR and REMD Simulations. *J Mol Biol.* 391(4):784-96.

3. Weinstock DS, Narayanan C, Baum J, Levy RM. 2008. Correlation between 13Calpha chemical shifts and helix content of peptide ensembles. *Protein Sci*. 2008 May;17(5):950-4

4. Nanda V, Andrianarijaona A, Narayanan C. 2007. The role of protein homochirality in shaping the energy landscape of folding. *Protein Sci*. 16(8):1667-75.

5. Weinstock DS, Narayanan C, Felts AK, Andrec M, Levy RM, Wu KP, Baum J. 2007. Distinguishing among structural ensembles of the GB1 peptide: REMD simulations and NMR experiments. *J. Am. Chem. Soc*. 129(16):4858-9.

6. Talla V, Narayanan C, Srinivasan N, Balasubramanian D. 2006. Mutation causing self-aggregation in human gammaC-crystallin leading to congenital cataract. *Invest Ophthalmol Vis Sci*. 47(12):5212-7.