JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

Explaining the Structural Plasticity of α-Synuclein

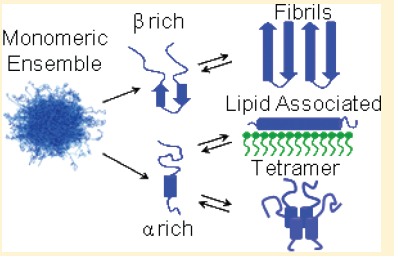
Orly Ullman,[†] Charles K. Fisher,[‡] and Collin M. Stultz*[,†,‡,§]

[†]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, United States

[‡]Committee on Higher Degrees in Biophysics, Harvard University Cambridge, Massachusetts 02139-4307, United States

[§]Harvard-MIT Division of Health Sciences and Technology, Department of Electrical Engineering and Computer Science, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, United States

ABSTRACT: Given that α-synuclein has been implicated in the pathogenesis of several neurodegenerative disorders, deciphering the structure of this protein is of particular importance. While monomeric α-synuclein is disordered in solution, it can form aggregates rich in cross-β structure, relatively long helical segments when bound to micelles or lipid vesicles, and a relatively ordered helical tetramer within the native cell environment. To understand the physical basis underlying this structural plasticity, we generated an ensemble for monomeric α-synuclein using a Bayesian formalism that combines data from NMR chemical shifts, RDCs, and SAXS with molecular simulations. An analysis of the resulting ensemble suggests that a non-negligible fraction of the ensemble (0.08, 95% confidence interval 0.03−0.12) places the minimal toxic aggregation-prone segment in α-synuclein, NAC(8−18), in a solvent exposed and extended conformation that can form cross-β structure. Our data also suggest that a sizable fraction of structures in the ensemble (0.14, 95% confidence interval 0.04−0.23) contains long-range contacts between the N- and C-termini. Moreover, a significant fraction of structures that contain these long-range contacts also place the NAC(8−18) segment in a solvent exposed orientation, a finding in contrast to the theory that such long-range contacts help to prevent aggregation. Lastly, our data suggest that α-synuclein samples structures with amphipathic helices that can self-associate via hydrophobic contacts to form tetrameric structures. Overall, these observations represent a comprehensive view of the unfolded ensemble of monomeric α-synuclein and explain how different conformations can arise from the monomeric protein.



INTRODUCTION

α-Synuclein is a 140 amino acid protein that has been implicated in several neurodegenerative diseases, often referred to as synucleopathies, such as Parkinson's disease (PD), Dementia with Lewy bodies (DLB), and Multiple System Atrophy (MSA).[1−3] PD, in particular, is neuropathologically characterized by α-synuclein aggregates and the loss of dopaminergic neurons within the substantia nigra.[4,5] While a number of theories have been advanced to explain how α-synuclein self-association is related to neuronal dysfunction, the precise relationship between α-synuclein aggregation and cell death remains unclear.[6] Consequently, understanding the structural basis of α-synuclein self-association is of particular importance.

Although monomeric α-synuclein is intrinsically disordered in aqueous solution and is therefore considered an intrinsically disordered protein (IDP), it cannot be simply described as a random coil.[7−9] For example, the average radius of gyration of a random coil that is 140 amino acids long is larger than the measured average radius of gyration for α-synuclein obtained via small-angle X-ray scattering (SAXS) experiments.[10] This suggests that α-synuclein is, on average, more compact than the classic random coil.

In addition, α-synuclein can form ordered structures under different experimental conditions. The amino acid sequence of α-synuclein contains 11-residue imperfect repeats that are distributed among the highly basic N-terminal region of the protein (residues 1−60), and the hydrophobic NAC region (Non-Aβ Component of α-synuclein, residues 61−95). These repeats were proposed to form amphipathic α-helices capable of interacting with different types of lipid structures.[11,12] It was found that when α-synuclein is bound to micelles, two helices can form.[13,14] The first helix encompasses residues 3−37 and is therefore contained within the N-terminal region and the second helix is formed between residues 45 and 92, a region that begins in the N-terminal region and extends into the NAC region. The two helices are aligned antiparallel to one another.[15,16] Other studies suggest that α-synuclein can also form a continuous helix that begins in the N terminal and continues through to the NAC region and that the precise form of the helical segment depends on the precise experimental conditions.[17−20] For example, in a recent study it was found, using pulsed dipolar ESR spectroscopy, that depending on the relative protein-to-detergent concentrations, α-synuclein can adopt either a single extended helix form or the broken helix form, similar to the one previously described.[21] Moreover, recent data further suggest that under physiologic conditions, α-synuclein exists in a tetrameric form that has considerable helical content.[22]

By contrast, α-synuclein aggregation is characterized by an increase in β-sheet content. Atomic force microscopy and Raman spectroscopy demonstrated that soluble α-synuclein oligomers have reduced α-helical content relative to protofilaments, and

Received: September 14, 2011
Published: October 26, 2011

that the β sheet content is relatively increased in protofilaments and filaments.[23] Fiber diffraction of α-synuclein fibrils further demonstrated the presence of cross-β structure which is characteristic of amyloid fibrils.[24] Consequently, the available experimental evidence suggests that α-synuclein can adopt helical structures or extended structures depending on the binding partner and experimental conditions.

A number of studies have constructed α-synuclein ensembles, using a combination of computational methods and experiments, to better understand the nature of the unfolded state.[25−29] Some of these studies combine data obtained from NMR, PRE, and conformational sampling to construct an appropriate ensemble.[25−28] While these studies have provided insights into the accessible states of α-synuclein in solution, there are still many unanswered questions regarding the unfolded state of this protein, including the precise role of secondary structure in the unfolded ensemble and the presence of long-range contacts, in particular. In addition, the recent observation that α-synuclein can also form ordered helical tetramers in the native cell environment has not been addressed in the previous studies.

In this work, we use a recently developed Bayesian Weighting (BW) algorithm to construct an ensemble for wild-type (WT) α-synuclein.[30] Data from NMR chemical shifts,[31] RDCs,[32] and SAXS[33] experiments are used to guide the construction of the ensemble. An analysis of the ensemble (1) helps to clarify the role of secondary structure propensity and the different binding characteristics of α-synuclein, (2) identifies potential aggregation prone structures within the ensemble, (3) clarifies the relationship between long-range contacts and aggregation propensity, and (4) provides insights into how the disordered monomeric protein can form tetrameric helical structures.

RESULTS AND DISCUSSION

Construction of an α-Synuclein Ensemble. We began by generating a relatively large structural library of energetically favorable conformations and then used a Bayesian weighting (BW) algorithm[30] to assign weights (or relative stabilities) for each conformer in the library. Hence, an 'ensemble' is defined as a set of structures, $\{S_i\}$ and a corresponding set of weights $\vec{w} = \{w_i\}$ where w_i is the weight (or probability) of structure S_i and $\Sigma_i w_i = 1$.

For a given structural library, there are many possible ways to weight the different structures within the structural library, and each possible weighting scheme represents a different ensemble. The BW method assigns a probability to every possible weighting scheme, and hence every possible ensemble, that can be constructed from the structural library. Parenthetically we note that, since some of the w_i can be 0, finding a correct weighting scheme also enables us to exclude structures from the structural library if they consistently lead to ensembles that are inconsistent with the experimental data.

The probability of a given ensemble is calculated using methods from Bayesian statistics as described in our previous work[30] and as reviewed in the Methods. Overall the probability of an ensemble is related to the agreement between the data predicted by the ensemble and the experimental data. In the Bayesian formalism, we compute a probability distribution (which we refer to as the posterior density) over all possible ensembles, and this distribution is used to make statements about the conformational properties of α-synuclein. Since the posterior density is a multidimensional function, we summarize its properties in two ways.

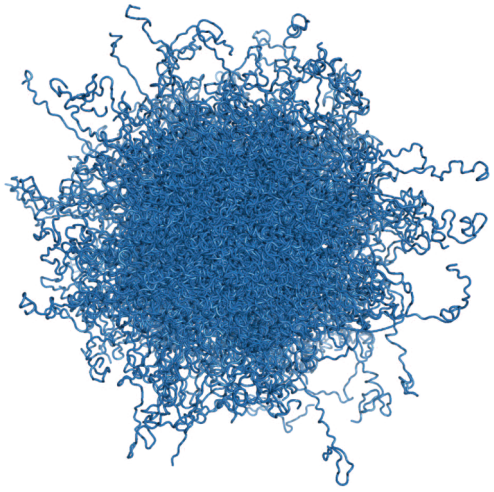


Figure 1. An alignment of all structures within the α-synuclein ensemble.

First, we calculate the average weight of each structure in the structural library using the posterior density function. The ensemble consisting of the structures $\{S_i\}$ and these average weights, $\vec{w}^B = \{w_i^B\}$ is called the Bayes ensemble; that is, it is the Bayesian analogue of a 'best fit' ensemble. Of course, the average of a distribution may not be very informative if the standard deviation, a measure of uncertainty, is large. To reflect this, we use the distribution over ensembles (the posterior density) to calculate confidence intervals for conformational characteristics of α-synuclein as a way of quantifying statistical uncertainty. Note that the confidence intervals do not refer to a specific ensemble, but rather to the distribution over all possible ensembles that could be constructed from the structural library.

An advantage of the BW formalism is that it provides a built in estimate of the uncertainty in the Bayes ensemble. Since agreement with experiment alone does not ensure that an ensemble is correct, such quantitative measures of uncertainty are important.[30] A further advantage of the method is that, even when the uncertainty in the Bayes ensemble is relatively large, we can calculate error bars to quantify the uncertainty in any observable quantity that is calculated from the ensemble.

First, we constructed a structural library of 100 000 energetically stable structures by breaking the protein into overlapping 8-residue segments and exhaustively sampling the conformational space of each segment using Replica Exchange Molecular Dynamics (REMD).[34] Segments were then joined to form a structure of the full 140 residue protein. To reduce the number of conformations to a more manageable size, the structural library was pruned using a coarse clustering method to generate a set of 299 structures that largely preserves the structural heterogeneity that was present in the original structural library (Figure 1).

Application of the BW algorithm to obtain the Bayes weight for each structure yielded a Bayes ensemble that agrees with measured NMR chemical shifts[31] (Figure 2A) and RDCs (Figure 2B)[32] as well as SAXS derived radius of gyration[33] (ensemble average value 41 $\pm$ 1 Å vs experimentally determined value of 40 $\pm$ 2 Å). These data demonstrate that the BW algorithm accomplishes its goal of generating ensembles that agree with the input experimental data.

As discussed above, the BW method provides a built-in metric, called the uncertainty parameter, that quantifies our uncertainty in the Bayes ensemble, and is analogous to the standard deviation

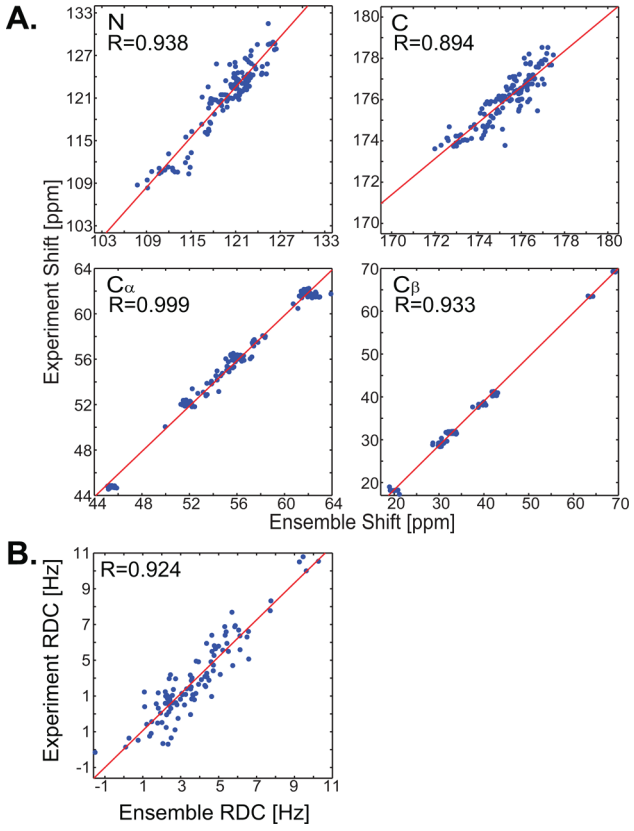


Figure 2. Ensemble agreement with experimental data. Comparison of experimental results with the corresponding calculated (A) chemical shifts and (B) RDCs. Correlation coefficients are explicitly shown. Calculated root mean square error (RMSE) for the chemical shifts was found to be within accuracy provided by SHIFTX[66].

of a Gaussian distribution.[30] If it is likely that the Bayes ensemble is correct, the uncertainty parameter approaches 0. Conversely, if it is unlikely that the Bayes ensemble is correct, then the uncertainty parameter approaches 1. In other words, as the uncertainty parameter approaches 1, we cannot say with any certainty that the constructed ensemble is correct. In the present case, the uncertainty parameter is 0.4. In this scenario, we can further quantify our uncertainty by computing confidence intervals for specific conformational characteristics.

An analysis of the Bayes ensemble provides additional information about the relative distribution of different conformer sizes that are accessible to the protein. As shown in Figure 3, the ensemble itself contains structures with radii of gyration that range from approximately 20 to 60 Å. To put these values into perspective, we note that the average radius of gyration for a globular folded protein containing 140 aa is approximately 15 Å, while the average radius of gyration for a random coil with the same amino acid length is approximately 52 Å.[35] The fraction of the ensemble with a radius of gyration near 20 Å is 0.09 (95% confidence interval 0.05−0.19), while the fraction that has a radius of gyration greater than would be expected based on the random coil calculation is 0.17 (95% confidence interval 0.13−0.22). This suggests that α-synuclein samples structures that are nearly as compact as a globular protein of the same size in addition to structures that are more extended than that of the average random coil value.

Residual Secondary Structure in α-Synuclein. To assess the secondary structure content in the Bayes ensemble, we used the

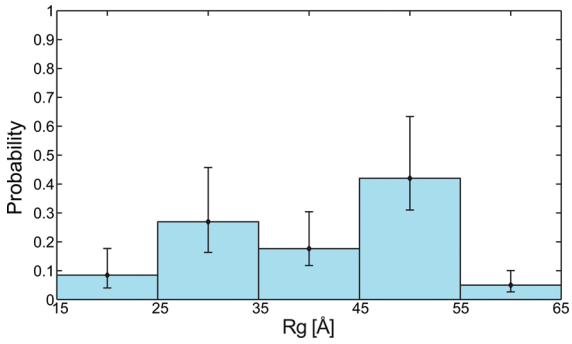


Figure 3. Distribution of radii of gyration in the calculated ensemble. The 95% confidence intervals show that the bar heights are significantly different from zero.

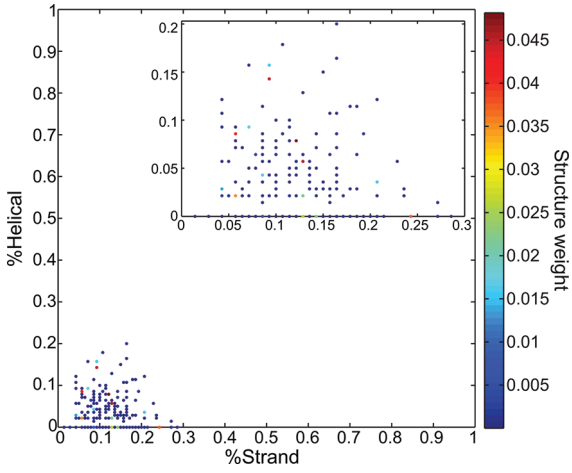


Figure 4. Helical and strand (or extended) content for each structure in the ensemble. The inset is an expanded view of the data.

Table 1. Ensemble Average Secondary Structure Content (with 95% Confidence Intervals) and Experimental Values Obtained from CD Spectroscopy (With Experimental Error Bounds)

	ensemble average	CD
Helix	0.03 (0.02−0.04)	0.02 ± 0.03
Strand	0.11 (0.10−0.13)	0.11 ± 0.07
Other	0.85 (0.84−0.87)	0.86 ± 0.22

STRIDE secondary structure assignment algorithm,[36] to calculate the propensity of each residue, in every structure in the ensemble, to adopt one of three mutually exclusive classes of secondary structure: helix, strand (also referred to as extended), and other (see Methods). Analysis of individual structures within the ensemble reveals that the highest helical content is 20% while the highest strand content is approximately 28%. Of note, the highest weighted structures in the Bayes ensemble have helical content less than 15% and strand content less than 25% (Figure 4). Nevertheless, the ensemble average secondary structure content is considerably less; that is, the overall strand content is less than 11% and the helical content less than 2% (Table 1). However, the ensemble average, which corresponds to the experimentally observed value, is in excellent agreement with estimated secondary

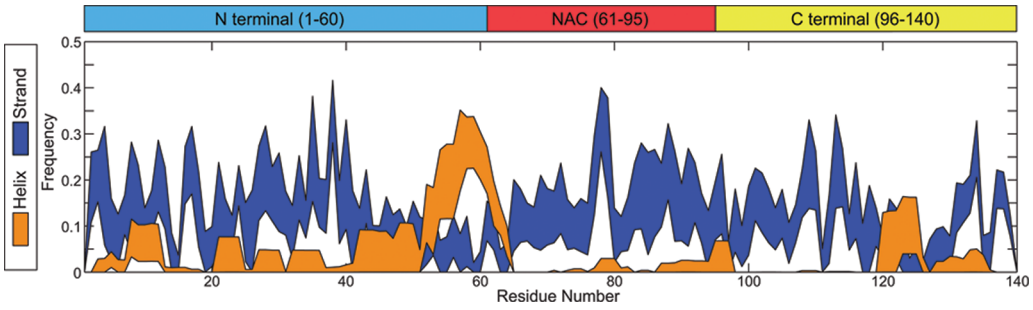


Figure 5. Ensemble average secondary structure propensity. For each residue, we present the probability to adopt a helical structure (orange area) vs a strand structure (blue area). The thickness of the lines corresponds to the 95% confidence interval.

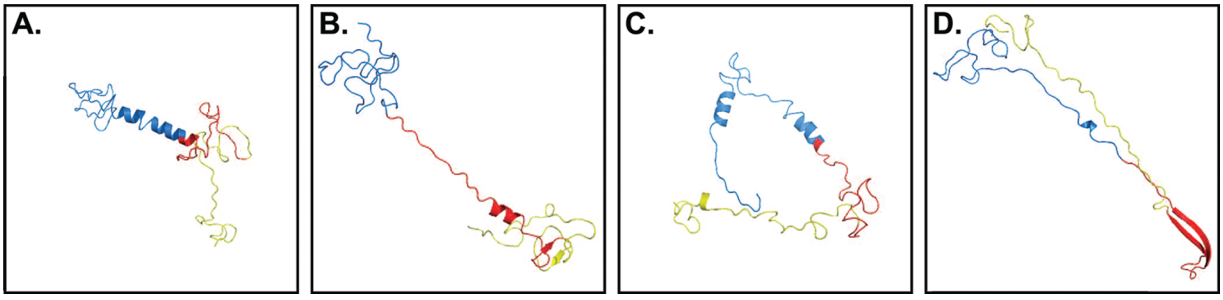


Figure 6. Representative ensemble conformations. A sample of structures from the ensemble of α-synuclein. In all structures, blue denotes the N-terminal region (residues 1−60), red denotes the NAC region (residues 61−95), and yellow denotes C-terminal region (residues 96−140).

structure content obtained from CD spectroscopy[37] (Table 1). Although the experimental error bounds and the confidence intervals from the BW algorithm are relatively large for the helical and strand content, both BW and CD spectroscopy agree that the protein has minimal helical and strand content.

In addition to the overall secondary content of the protein, we also computed the expected (or ensemble average) relative helix and strand propensities for each residue in the protein with their corresponding 95% confidence intervals (Figure 5). On average, most of the helical propensity resides in residues 52−64. This region contains a highly conserved hexamer motif within the fifth 11-mer imperfect repeat, which is proposed to form amphipathic α-helices.[11,12] Additionally, within the NAC segment, the strand propensity is peaked in the immediate vicinity of residue 78. Interestingly, this region, NAC(8−18), has been experimentally determined to be the minimal toxic aggregate forming segment in α-synuclein in vitro.[38]

To further demonstrate that α-synculein samples structures with varying amounts of secondary structure, we explicitly show four conformations in Figure 6. The N-terminal region is marked in blue, the NAC region in red and the C-terminal region in yellow. In Figure 6A−C, three structures are shown that contain varying degrees of helical content in the N-terminal and NAC region. Figure 6A shows a structure containing a helix between residues 42 and 64; this helical conformation is in agreement with the contiguous helix model. Figure 6B shows a structure that contains a helix in residues 74−82 in the middle of the NAC region and Figure 6C presents two helices, one in the range 52−62 and the other in the range 15−24. Figure 6D shows a structure with significant strand content in the NAC region, in particular residues 68−94.

Long Range Contacts in α-Synuclein. A number of studies have used Paramagnetic Relaxation Enhancement (PRE) experiments to detect long-range contacts in α-synuclein.[26−29,39,40] These experiments allow for the detection of interactions between a paramagnetic group and nuclear spins of residues at a distance up to 25 Å away.[41] Some of these studies argue that long-range contacts, especially involving the N-terminal (residues 1−60) and C-termini (residues 96−140), can be found in the unfolded ensemble of α-synuclein. To determine whether our data are consistent with these observations, we computed the distribution of such long-range contacts from the Bayes ensemble. For these calculations, we define a long-range contact between the N- and C-terminal regions to occur when the center of mass of the N-terminal region (residues 1−60) and the center of mass of the C-terminal region (residues 96−140) are within 25 Å. We computed these center of mass distances for each structure and used these data to compute the distribution of such distances along with the associated confidence intervals (Figure 7). Figure 7 demonstrates that structures in the Bayes ensemble span a wide range of N- to C-terminal distances, ranging from less than 25 Å to more than 125 Å. In addition, a significant fraction (0.14 with a 95% confidence interval of 0.04−0.23) of the ensemble has structures that place the center of masses of the N- and C-terminal regions within 25 Å of one another.

Since results from PRE experiments correspond to ensemble averages, we also computed a Residual Contact Map (which is a function of the ensemble average number of long-range contacts per residue) to better compare our results to the previous PRE data (Figure 8A). This pseudoenergy difference map represents the stability of a long-range contact between two residues in the Bayes ensemble compared to what one would expect from a random coil ensemble (see Methods).[25,26] These data suggest that there is a distinct preference for forming long-range contacts between the C-terminal (residues ~120−140) and the N-terminal region (residues 1−61) and also between the C-terminal and

the NAC region (residues ~61−70). In essence, residues 120−140 in the C-terminal region make contact with residues 1−70, which encompass both the N-terminal region and the beginning of the NAC segment. A less favorable contact forms between the NAC region (residues ~80−95) and the N-terminal region (residues ~1−30); that is, our data are in qualitative agreement with prior experimental observations made from PRE data.[25,26] In this regard, it is important to note again that our α-synuclein ensemble was generated without incorporating any data from prior PRE experiments, and therefore, no explicit distance constraints were used to construct the model.

In Figure 8B, we show the most stable contact for each residue, along with the corresponding 95% confidence interval. The relatively small error bars for residues 20−70 and residues 120−135 suggests that the model is relatively certain about these particular inter-residue contacts. However, the large error bars between residues 70−90 argues that the model is unsure about the inter-residue contacts in this region. These data complement the residual contact map in Figure 8A; for example, the residual contact map suggests that there is a relatively small preference for forming long-range contacts between the NAC (residues ~80−95) and the N terminal region (residues ~1−30); however, the uncertainty analysis (Figure 8B) suggests that the model is very uncertain about this particular observation.

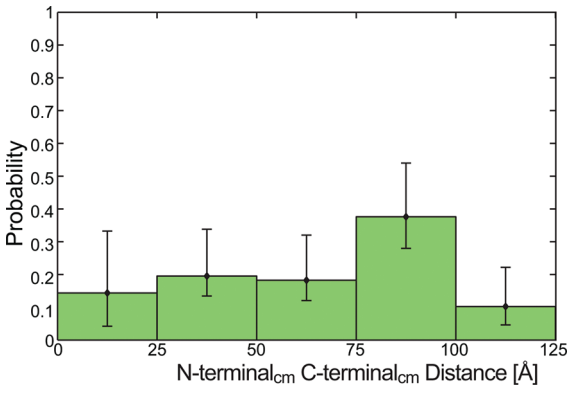


Figure 7. Distribution of N- to C-terminal distances in the calculated ensemble. The *x*-axis represents the distance between the center of mass of the N-terminal region and the center of mass of the C-terminal region. The 95% confidence intervals show that the bar heights are significantly different from zero. Two low probability structures had N- to C-terminal distances higher than 125 Å and we therefore excluded this information from the distribution.

Potential Aggregation Prone Conformers in α-Synuclein. Given that a relatively small segment of α-synuclein, consisting of residues 68−78, which is found in the NAC region (i.e., NAC-(8−18)), was experimentally determined to be the minimal toxic aggregate-prone segment in α-synuclein in vitro,[38] we explored the conformational preferences of this segment. Structures that place this segment in a solvent exposed and extended orientation may be more likely to form toxic aggregates containing cross-β structure.

Figure 9 shows the normalized solvent accessible surface area (SASA) of the NAC(8−18) region versus the number of residues in that segment that are in an extended conformation, as identified by STRIDE.[36] Calculations of the SASA only included the atoms

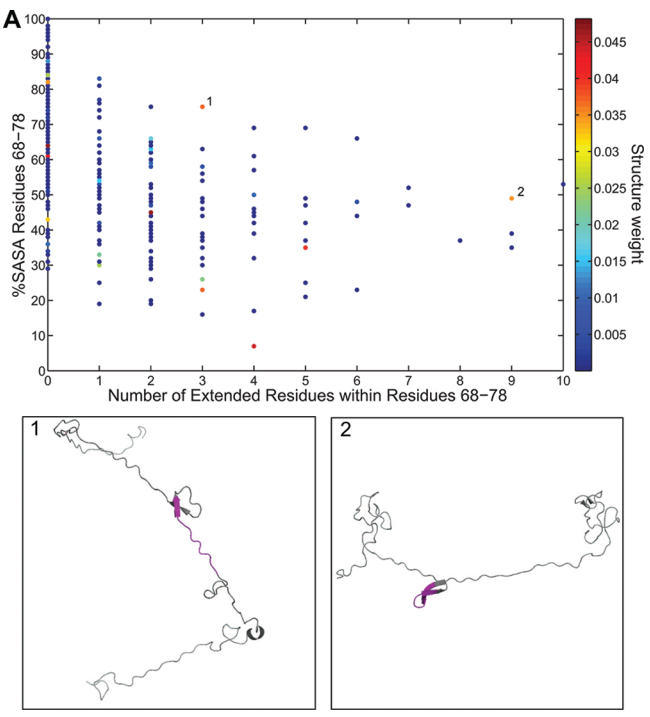


Figure 9. The SASA vs number of residues in an Extended orientation for the NAC(8−18) region. Two relatively high weighted structures with an exposed and a significant Extended content (more than 3 residues) for the NAC(8−18) segments are explicitly shown.

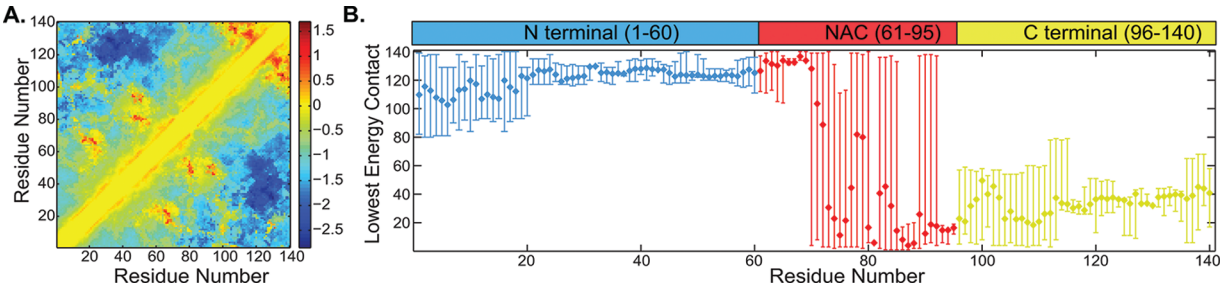


Figure 8. Stabilized long-range contacts. (A) Pseudoenergy values are calculated for each contact as $-\ln(p_{ij}^{\text{Ensemble}}/p_{ij}^{\text{RC}})$. A contact is defined between residues where the Cα atoms are less than 25 Å apart. A large negative pseudoenergy (in blue) represents contacts that are energetically favorable in the ensemble compared to the random coil ensemble (in units of kT). Positive values (color range yellow to red) represent relatively unfavorable contacts. (B) For each residue, we calculated the contact that is associated with the lowest pseudoenergy along with the 95% confidence interval. The *x*-axis is the residue number and the *y*-axis is the position of the residue that forms the lowest energy contact. We used the following color code to depict the different regions of the protein: blue, N-terminal region (residues 1−60); red, NAC region (residues 61−95); and yellow, C-terminal region (residues 96−140).

dx.doi.org/10.1021/ja208657z |J. Am. Chem. Soc. 2011, 133, 19536–19546

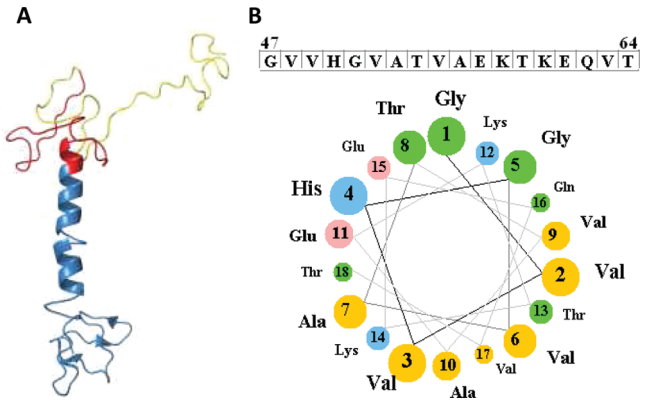


Figure 10. (A) Structure of the conformation within the ensemble that has the longest helical segment (expanded view of structure shown in Figure 6A). As before, blue denotes the N-terminal region (residues 1−60); red denotes the NAC region (residues 61−95), and yellow denotes C-terminal region (residues 96−140). (B) Associated helical wheel: orange, nonpolar residues; green, polar uncharged residues; blue, basic; pink, acidic.

H−N−Cα−C−O as this ensures that large SASA values identify structures that can form the intermolecular hydrogen bonds that are needed for cross-β structure formation. The Bayes ensemble contains several structures that place the aggregation prone segment, NAC(8−18), in a relatively extended and solvent exposed orientation. We define a residue as solvent exposed when it has a normalized SASA > 40%, as this cutoff has been used in previous studies and useful results were obtained.[42] In total, the fraction of structures that have the NAC(8−18) segment in a relatively extended and solvent exposed orientation is 0.08 with a 95% confidence interval 0.03−0.12.

It has been postulated that the formation of long-range contacts in α-synuclein may provide a mechanism to shield regions of the NAC segment.[26] Burying regions of the NAC segment could potentially hinder the formation of cross-β structure and the formation of toxic aggregates. To investigate the relationship between solvent exposure of the NAC(8−18) segment and long-range contacts, we computed the SASA of the structures that have the center of mass of the N-terminal and C-terminal regions within 25 Å. We find that the majority of structures that have the aforementioned long-range contacts also place the NAC(8−18) segment in a solvent exposed orientation; that is, 65% of structures (28%-100%, confidence interval) with long-range contacts have the NAC(8−18) segment with a SASA> 40%.

A Potential Mechanism for Helical Self-Association. In a recent study, α-synuclein was isolated from human RBCs in a tetrameric form and the CD spectrum of this tetramer was quite distinct from that of recombinant α-synuclein obtained from *Escherichia coli*.[22] Indeed, the spectrum of the tetramer had minima at 208 and 222 suggesting that, on average, the tetrameric structure had considerable helical structure.

Our data suggest that monomeric α-synuclein samples structures that have at most 20% helical content (Figure 4). The structure with the highest helical content is shown in Figure 10A. In general, the associated helix has a hydrophobic patch on one side (Figure 10B), that is akin to hydrophobic faces that have been observed in other proteins that form helical bundles.[43−45] These data are consistent with a model where helical segments within structures in the unfolded ensemble associate via hydrophobic interactions to form a tetrameric structure. If self-association of preformed helical structures was the dominant mechanism underlying the formation of tetrameric structures, then the expected helical content of the α-synuclein tetramers would be at most 20%. Interestingly, using the CD spectrum of the tetrameric structure (kindly provided by Tim Bartels and Dennis Selkoe), we obtain a predicted helical content of 29%, with an error of approximately 10%, using the program K2d.[46,47]

More recently, Wang et al.[48] were able to obtain NMR data on a tetrameric form of α-synuclein that was purified from *E. coli*. Weak (i,i + 3) Nuclear Overhauser Enhancements (NOEs) and secondary chemical shifts indicated helical propensity in residues 4−103. Intermolecular PREs, obtained using mixtures of α-synuclein with and without the spin label, suggested that the tetramer forms by the association of amphipathic helices formed within the region consisting of residues 50−103. It is important to note, however, that the NMR data are not consistent with a fully folded helix. Instead, they suggest transient helical formation with an overall helical content of approximately 20% (Tom Pochapsky, personal communication).[48] These data are qualitatively consistent with our model presented in Figure 10, which includes an amphipathic helix consisting of residues 47−64, and with our finding of α-helical propensity throughout the N-terminal region.

CONCLUSION

Although monomeric α-synuclein is intrinsically disordered in solution, it can adopt conformations that have varying secondary structure content depending on its environment.[9] A number of spectroscopic studies suggested that WT-α-synuclein in the presence of membranes can form helical structures, and two types of helical configurations have been observed: a continuous extended helix and two antiparallel helices separated by a short linker.[13−21] By contrast, α-synuclein was shown to acquire significant β-sheet content when it self-associates in vitro.[23] The most ordered form of these aggregates are fibrils which were found to contain cross-β-sheet structures.[24] Most recently, an α-synuclein tetramer has been isolated from both human red blood cells and *E. coli* and it has been argued that this structure is the dominant form under physiological conditions.[22,48] While initial reports suggested this structure was a 'folded tetramer' rather than an 'unfolded monomer', it is important to note that the data from recent NMR studies suggest that the actual amount of structural order is fairly low in the tetramer and that there is only fractional helix formation.[48] These observations highlight the need to obtain an accurate structural ensemble that describes the accessible states of the protein.

While a number of studies have generated ensembles for the unfolded state of α-synuclein,[25−29] the majority of these ensembles has not provided a detailed analysis of residual secondary content within the ensemble and has not addressed recent data on what has been described as a physiologically dominant helical tetramer.[22] More importantly, existing studies have led to contradictory observations. Although one prior study suggested a small preference for residues 6−37 (in the N-terminus) to form helices compared to residues 103−140 (in the C-terminus),[25] another study did not find significant helical structure within the ensemble.[27] Therefore, to further explore the nature of the unfolded state of α-synuclein and to understand the differential binding characteristics of the protein, we constructed and analyzed an ensemble that represents the unfolded state of monomeric α-synuclein in solution.

As we have previously noted, the construction of models that adequately represent the unfolded state of a protein is inherently difficult for a number of reasons. First, there is the conformational sampling problem; that is, sampling all possible conformations of even a modestly sized protein is intractable. Nevertheless, the form of the underlying energy surfaces helps because the space of energetically favorable conformations is likely far less than the space of all possible conformations. Some studies have shown that straightforward Boltzmann sampling for some IDPs yield calculated observables that are in reasonable agreement with experiment, thereby suggesting that extensive sampling, by itself, is a plausible approach for generating representative ensembles for IDPs.[49,50] However, while it is reasonable to apply such a direct sampling approach to relatively small proteins, the prospect of extensively sampling the relevant conformational space of a protein that is 140 residues in length (at 300K) is daunting. In this regard, a number of approaches that do not rely on direct Boltzmann sampling of large proteins have been developed and useful insights have been obtained using these methods.[51−54]

In the present study, we use a fragment based approach to sample energetically favorable conformations of the entire 140 residue protein. The motivation for this approach arose from a prior study that demonstrated that sampling the conformational states of fragments from folded proteins may reproduce the backbone structure of that peptide's structure in the context of the entire protein.[55] In our method, the protein was divided into eight residue long overlapping segments and REMD was used to sample the conformational space of each peptide. Eight residue segments were chosen because this length corresponds, roughly, to the average persistence length of a polypeptide.[52] Structures for α-synuclein were generated by combining these overlapping segments, and subsequent energy minimization of each reconstructed conformer ensures that each structure corresponds to local energy minima on the potential energy surface of the protein. While the approach is computationally efficient, we recognize that focusing on sampling small peptides may limit the formation of long-range interactions within the final ensemble. To mitigate this, once the peptide fragments have been combined to generate the α-synuclein sequence, the entire protein is energy minimized thereby allowing the different peptides to "see" one another. Using the resulting structural library with the BW algorithm, we arrive at a Bayes ensemble that: (1) contains conformations that correspond to local energy minima on the potential energy surface and (2) agrees with the available experimental data. Nevertheless, while our approach is computationally very efficient, we recognize that other sampling approaches for the initial structural library (e.g., using different lengths for the peptide segments) may lead to different structural libraries and this fact introduces some uncertainty in our analysis.

One additional source of uncertainty is the inherent degeneracy of the problem of constructing a good ensemble, even after the precise structural library has been specified. Given that the number of degrees of freedom (i.e., the number of energetically favorable conformations) is typically much greater than the number of independent experimental observables, the problem of choosing, or weighting, a set of structures is inherently degenerate; that is, there are many possible ways of weighting the structures that will agree with any given set of experimental observations.[30]

To deal with these sources of uncertainty, the BW method calculates a probability distribution over the space of ensembles. This posterior density function naturally leads to a new metric, the uncertainty parameter, which quantifies our uncertainty in the Bayes ensemble. This uncertainty parameter is akin to the standard deviation in a Gaussian distribution and reflects the overall spread of the calculated probability distribution. The uncertainty parameter varies between 0 and 1 where a value of 0 suggests that the Bayes ensemble is correct. By contrast, when the uncertainty parameter is nonzero, one cannot be certain that the Bayes ensemble is correct. However, in this latter instance, one can express values with the appropriate confidence intervals.[30] Therefore, the method provides a rigorous means to quantify the overall uncertainty in the final results.

The BW algorithm yields a Bayes ensemble that is in agreement with data obtained using NMR chemical shifts,[31] RDCs,[32] and the radius of gyration as determined by SAXS experiments.[33] Surprisingly, we find that some conformers in the ensemble are nearly as compact as a folded globular protein with the same amino acid length. In addition, the Bayes ensemble contains structures that have a radius of gyration that is larger than the average radius of gyration that would be expected from a 140 residue random coil. This highlights the fact that a single experimental value for the radius of gyration provides little insight into the full range of conformations that the protein can adopt in solution. Two recent experimental studies suggested the existence of distinct classes of conformers describing α-synuclein equilibrium.[56,57] Both studies argue that α-synuclein contains a range of conformations, where some are quite compact and others are quite extended and random-coil like. Our data are in agreement with these observations and quantify the range of radii of gyration within the ensemble. Moreover, these studies highlight the fact that while the overall secondary structure content of the ensemble is negligible (about 7% of the population was suggested to contain "β-like" conformation), there are subpopulations of structures that have more significant helical and strand content, a finding in agreement with our observations.

The resulting structural ensemble provides additional insight into the secondary structure propensities within different regions of the protein. We find that the Bayes ensemble contains several structures that have helical segments of varying length in the N-terminal region, extending into the NAC segment. In total, the helical regions span residues 1−92; that is, the segment that has been shown to adopt either a continuous helix or a broken helix in the presence of lipid membranes.[19−21] These data are consistent with a model of lipid binding where interaction with the membrane stabilizes these helical segments leading to the formation of either a continuous or a broken helix, depending on the precise experimental conditions. In this sense, the presence of relatively short helical segments may serve as intermediates that enable fast and efficient binding to lipid membranes. These data are of particular importance because interactions between α-synuclein, in its helical form, and membranes may play a role in cellular dysfunction in patients with Parkinson's disease.[58]

By contrast, on average, significant probability for extended structure is found throughout the α-synuclein sequence. These data are in qualitative agreement with previous Raman spectroscopic studies that suggest that the protein adopts an ensemble of rapidly interconverting secondary structural elements.[59] Of particular interest is the region spanning residues 68−78 in the NAC segment because this has been shown to be the minimal toxic peptide that can also initiate α-synuclein aggregation in vitro.[38] The probability of extended structure is relatively peaked around this region, thereby suggesting that this segment has an intrinsic predisposition to form extended structure that may

initiate the formation of β-sheet rich aggregates. However, in order to form intermolecular hydrogen bonds with other α-synuclein molecules, this segment must be exposed to solvent. Therefore, structures that place this segment in a solvent exposed and extended conformation may be more prone to form toxic aggregates. Our analysis suggests that approximately 8%, with a 95% confidence interval of 3−12%, of the structures in the ensemble have the NAC(8−18) segment in an extended and solvent exposed orientation. This suggests that the unfolded ensemble of α-synuclein contains preformed conformations that can readily form β-sheet rich toxic aggregates.

In addition to these insights, our data further clarify the role of long-range contacts in the protein. Previous studies that have constructed ensembles based on the results of PRE experiments have found conflicting findings, even though many of these experiments were performed under similar experimental conditions. One study suggested that long-range interactions occur between residues 85−95 of the NAC and the C-terminal region (specifically residues 110−130).[28] Other studies suggested the formation of long-range contacts between the highly charged C-terminus (residues 120−140) and the large hydrophobic center (residues 30−100) resulting in a hydrodynamic radius significantly smaller than that expected for a random coil structure.[25,26] In another study, done under similar conditions, it was suggested that long-range contacts form between the N-terminus and the NAC region in contrast to the previously mentioned studies.[27] Therefore, while these data have provided new insights into the nature of the unfolded state of α-synuclein in solution, they leave the precise role of any long-range interactions in the protein unclear.

Our data suggest that, on average, there are long-range contacts between the N- and C-termini of the molecule. Interestingly, the Bayesian estimates allow us to say with confidence that the N-terminal region and the first nine residues from the N-terminal portion of the NAC make, on average, contacts with the C-terminal region of the protein, a result that is in qualitative agreement with prior studies.[25,26] It has been suggested that these long-range interactions provide a mechanism that effectively shields the aggregation prone region, and thereby minimizes the extent of aggregation.[25,26,60] However, a more detailed look at the actual distribution of structures within the ensemble (as opposed to an analysis of the ensemble average data) finds that most of the structures in our ensemble that contain long-range contacts between the N- and C-termini also place the NAC(8−18) segment in a solvent exposed conformation. An example of one such structure is shown in Figure 6D. Consequently, it is not clear that separation of the N- and C-termini is required to expose the most aggregation prone regions of the sequence. This claim is supported by recent PRE experiments comparing WT α-synuclein and A30P, E46K, and A53T naturally occurring mutants, which were all shown to have a higher aggregation rate in vitro. Results of this study suggest that A30P and A53T mutants did not have a significant decrease in the N- and C-termini contacts. Moreover, E46K presented an increase in these long-range contacts.[39] These data bring into question whether long-range contacts play a key role in regulating aggregation of α-synuclein.

We note that our model did not use any PRE derived distance restraints. While PRE-derived data have provided valuable information into the presence or absence of long-range contacts in several IDPs, it requires introducing a paramagnetic probe into the protein.[61] However, it may be that such probes alter the accessible states of the unmodified protein. In light of these observations, and the fact that some of the PRE-derived results are contradictory, we did not explicitly use PRE-derived data when building our ensemble. Nevertheless, we obtain results that corroborate and clarify many aspects of the prior PRE studies.

In addition, we recognize that there may be additional contacts between the N-terminal, NAC, and C-terminal regions, but we cannot make statements about these interactions with confidence given the very wide error bars associated with residues in the more central region of the NAC segment (Figure 8B). Interestingly, our uncertainty in the precise contacts that involve the entire NAC region is also reflected in the literature as the NAC region is suggested to interact with the C-terminus in some studies while other studies suggest that it interacts with the N-terminus instead.[26,27,39,40,60,62] In short, our model is unable to distinguish between these two possibilities with certainty.

Recent data suggest that α-synuclein forms helical tetramers under physiological conditions.[22,48] Our data suggest that monomeric α-synuclein samples structures that have at most 20% helical content and that some of these helices are amphipathic in character. These data are consistent with a model where tetrameric structures are formed via the interaction of hydrophobic patches on these amphipathic helices. Indeed there are many examples in the literature of such four-helical bundle structures composed of amphipathic helices.[43−45] Wang et al. independently proposed a model for the tetrameric state of α-synuclein on the basis of NMR data in which transiently formed amphipathic helices interact in just such a manner.[48]

Our results argue that the unfolded state of α-synuclein contains a heterogeneous set of conformations of both highly compact and extended structures, and that while the overall secondary structure content of these structures is low, there are regions that have a relatively high propensity for helical and extended structure. Regions with a significant propensity for either helical or strand content may facilitate the formation of lipid-associated helical structures, helical tetrameric structures, and aggregates that are rich in β-sheets. Our results also provide quantitative estimates for the percentage of structures that are compact, have long-range contacts between the N- and C-termini, and that have the minimal toxic aggregation fragment of α-synuclein that is in a position that is poised to make intermolecular beta strands. In sum, these data provide a comprehensive view of the unfolded ensemble of monomeric α-synuclein in solution and explains how different ordered structures conformers can arise from this disordered protein.

METHODS

Generation of an α-Synuclein Structural Library. The sequence of α-synuclein was divided into eight residue long segments resulting in 28 segments in total (the C-terminal segment was five residues long). Each segment had three residues overlap with the adjacent segments. A similar protocol was used to describe K18, an intrinsically disordered protein of comparable size, 130 amino acids long.[30] The size of the segments was chosen based on the average persistence length of a polypeptide.[52] Conformations for segments of this length were shown to be successfully sampled using a replica exchange molecular dynamics (REMD[34]) procedure.[63]

Each segment underwent REMD with the EEF1[64] implicit solvent model using CHARMM.[65] A total of 16 replicas, each at a different temperature, were used. Temperatures were spaced exponentially in the range 280−700 K . Segments were run for 10 ns, and structures were collected from the last 5 ns of the 298 K heat bath, allowing 5 ns of

equilibration period. A total of 5000 conformations per segment were collected.

Full length α-synuclein conformations were generated by piecing together the segments one at a time, starting with the N-terminal segment. Each segment was clustered according to the three overlapping residues at its ends. The segment to be added to the growing polypeptide chains was chosen from the cluster that had the most structural similarity in the overlapping region. The first residue coordinates of the overlapping segments were taken from the C-terminal region of the one segment and the two others from the N-terminal region of the adjoining segment. At the end of the procedure, the full length structure was subjected to 1000 steps of steepest descent minimization followed by 10 000 steps of adopted basis Newton−Raphson minimization to relieve any bad contacts in the molecule. Only structures with a negative energy were chosen for the structural library. Following the process, we found the structural library generated was composed of structures that were mainly compact when comparing their radius of gyration (Rg) to the one obtained by SAXS experiments. Therefore, the combined pre-energy minimization structures were used in additional energy minimization using an Rg restraint. Rg restraints varied from 27 to 75 Å. This process ensures that a wide range of conformations were generated. At the end of the process, ~100 000 structures were generated.

The structural library was reduced in size to 299 structures using our previously described pruning algorithm.[30] This number of structures was shown to be able to provide a good model for the K18 tau segment of comparable size (130 residues).[30]

Generation of an α-Synuclein Ensemble from the Pruned Structural Library and the Calculation of Confidence Intervals. To obtain the sets of weights for the pruned structural library, we employed the BW algorithm as previously described.[30] In this method, one generates a posterior distribution which represents the probability of all possible weighting schemes over the 299 structures, given the available experimental data. Experimental measurements used were C, Cα, Cβ, and N chemical shifts,[31] N−H RDCs,[32] and radius of gyration.[33] The carbonyl chemical shift value for residue 140 from the set of experimental data points was not used, as it was an extreme outlier from the other data.[31] To implement the BW method, we first need to calculate the corresponding chemical shifts for each atom, along with RDCs and the radius of gyration, in each structure. Chemical shifts were calculated with SHIFTX,[66] and the radii of gyration were calculated with CHARMM.[65] The RDCs of each individual conformer in the ensemble was calculated with PALES[67] based on a 'global alignment' model, that is, using the entire protein structure. This is in contrast to a 'local alignment' model, in which the RDCs are calculated from short segments of the protein. It has been suggested that one can reproduce experimental RDCs with a smaller number of conformers when using a local alignment model as compared to a global alignment model;[68,69] nevertheless, we were able to obtain good agreement with the experimental RDCs using the global alignment method with a relatively small number of highly populated conformers.

The BW algorithm incorporates information from both the experimental errors and the errors associated with predictions for the experimental values of interest.[30] Experimental errors were taken to be 0.3 ppm (chemical shifts), 1 Hz (RDCs), and 2 Å (radius of gyration), respectively. As prediction errors for chemical shift values have been rather extensively studied, they were also included in the expression for the posterior distribution.[30,66]

Here, we provide a very brief review of the theoretical aspects of the BW framework; for a comprehensive description see Fisher et al.[30] Formally, the posterior probability distribution conditioned on the observed experimental data is obtained from Bayes' rule:

$$f_{\vec{W}|\vec{M}}(\vec{w}|\vec{m}) = \frac{f_{\vec{M}|\vec{W}}(\vec{m}|\vec{w}) f_{\vec{W}}(\vec{w})}{f_{\vec{M}}(\vec{m})} \quad (1)$$

where $f_{\vec{W}}(\vec{w})$ is a the prior probability distribution (for brevity, the specific form will not be reproduced here), and $f_{\vec{M}|\vec{W}}(\vec{m}|\vec{w})$ is the likelihood function for the vector of experimental observations, $\vec{m}$. We assume that the likelihood function can be decomposed as $f_{\vec{M}|\vec{W}}(\vec{m}|\vec{w}) = f^{R_G}_{\vec{M}|\vec{W}}(m^{R_G}|\vec{w})\, f^{RDC}_{\vec{M}|\vec{W}}(\vec{m}^{RDC}|\vec{w})\, f^{CS}_{\vec{M}|\vec{W}}(\vec{m}^{CS}|\vec{w})$ where each of the components is (multivariate)-Gaussian. Specifically, the likelihood functions are

$$f^{R_G}_{M|\vec{W}}(m|\vec{w}) = [2\pi\varepsilon^2_{R_G}]^{-1/2} \exp\left[-\frac{(m - E_{R_G}[m|\vec{w}])^2}{2\varepsilon^2_{R_G}}\right]$$

$$f^{RDC}_{\vec{M}|\vec{W},\Lambda}(\vec{m}|\vec{w}) \propto \int_{-\infty}^{\infty} \prod_{i=1}^{N_{RDC}} (2\pi\varepsilon^2_{i,RDC})^{-1/2} \exp\left[-\frac{(m_i - \lambda E_{RDC}[m_i|\vec{w}])^2}{2\varepsilon^2_{i,RDC}}\right] d\lambda$$

$$f^{CS}_{\vec{M}|\vec{W}}(\vec{m}|\vec{w}) = \prod_{i=1}^{N_{CS}} [2\pi(\varepsilon^2_{i,CS} + \alpha^2_{i,CS})]^{-1/2} \exp\left[-\frac{(m_i - E_{CS}[m_i|\vec{w}])^2}{2(\varepsilon^2_{i,CS} + \alpha^2_{i,CS})}\right] \quad (2)$$

Here, the letter ε denotes an experimental error, α denotes a prediction error, and λ is a factor for uniformly scaling the RDCs to account for uncertainty in the magnitude of alignment.

The Bayes estimate for the weight for each structure corresponds to the expected (or average) value of that structure's weight over the posterior distribution; that is, $w^B_j \equiv \langle w_j \rangle_{BW} = \int d\vec{w} w_j f_{\vec{W}|\vec{M}}(\vec{w}|\vec{m})$. The uncertainty parameter is the average distance from the Bayes weights, or $\sigma_{\vec{w}^B} \equiv [\int d\vec{w} \Omega^2(\vec{w}^B,\vec{w})\, f_{\vec{W}|\vec{M}}(\vec{w}|\vec{m})]^{1/2}$, where $\Omega^2(\vec{w}^B,\vec{w})$ is metric on the space of weight vectors called the Jensen-Shannon divergence.[30] To calculate these expected values, samples are taken from the posterior distribution using a Monte Carlo algorithm with Gibbs Sampling.[30] Each sample corresponds to a different weighting scheme over the 299 structures. A total of 100 million samples were generated as an equilibration period for the Markov Chain generated from the Monte Carlo algorithm. This was followed by an additional 1 billion samples, which constitutes the "production run". We followed the running average of the posterior divergence to ensure that convergence was reached. To calculate the Bayesian averages and the associated confidence intervals, we used 50 000 equally spaced samples from the 1 billion samples; this reduces the overall computation time. For a given quantity (e.g., the expected solvent exposure of a given residue), we computed this quantity using the chosen samples, yielding 50 000 estimates for the value of interest. The 95% confidence interval was obtained by finding the lower bound that excluded the bottom 2.5% of the estimates and the upper bound that excluded the top 2.5% of the estimates.

Random Coil Ensemble. The residual contact map shown in Figure 8A represents the stability of a long-range contact between two residues in our ensemble compared to what one would expect from a random coil ensemble. Therefore, to compute the contact map, we first need to generate a random coil ensemble for α-synuclein. We used the publicly available random coil ensemble posted in a web repository.[52] The ensemble contains 5000 structures; we therefore randomly selected 299 structures to ensure the two ensembles are of the same size. The selection process was repeated 20 times in order to reflect the full ensemble and each measurement of interest was averaged over this collection. The random coil model used to form this ensemble uses statistical potential and excluded volume constraints;[52] no α-synuclein experimental data was included in generating the ensemble.

Secondary Structure Assignments. We clustered STRIDE results of α-helix, π-helix, and 3-10 helix into a super class that we refer to as Helix. In addition, we cluster isolated bridge and extended results into a second super class we named Extended (or Strand). All other secondary structure assignments were combined into a single class denoted as Other. To have consistent definitions when comparing to the results obtained from CD spectroscopy (which assigned a helix, strand, turn, and unstructured),[37] we grouped the "turn" and the "unstructured" assignments into one category called "Other".

Solvent Accessible Surface Calculations. The solvent exposure surface area for each conformation was calculated using CHARMM.[65]

The SASA of the entire protein was computed, but only data from the solvent exposure of the backbone atoms N−H−C−Cα−O were used, since these represent atoms that are essential for the formation of cross-β-sheet interactions. The calculated SASA was normalized by dividing by solvent accessible surface of the backbone atoms when α-synuclein is in a fully extended conformation. A residue is said to be solvent exposed when its normalized SASA > 40%, as this cutoff has been used in previous studies and useful results were obtained.[42]

Calculating Distributions of Ensemble Properties. Two plots were generated presenting probabilities calculated from the posterior distribution. The radius of gyration for each structure in the ensemble is calculated in CHARMM[65] using the N−C−Cα atoms; structures are binned together in bins of 10 Å. Summation of the structures probabilities (their weights) in each bin comprises the probability of that bin. The 95% confidence intervals were then obtained using the 50 000 samples from the posterior distribution as outlined above. The histogram of N-terminal center-of-mass to C-terminal center-of-mass distances was generated in a similar fashion. Distances were calculated in CHARMM[65] using the center of mass of N−C−Cα backbone atoms for residues 1−60, the N-terminus, and the center of mass of N−C−Cα backbone atoms for residues 96−140, the C-terminus. Structures were binned in bins of 25 Å, corresponding to the maximal distance defined for formation of long-range contacts and again the 95% confidence intervals were then obtained from the 50 000 samples from the posterior distribution.

Helical Wheel Diagram. To generate the helical wheel, we used the freely available Helical Wheel program (http://cti.itc.virginia.edu/~cmg/Demo/wheel/wheelApp.html). Amino acid sequences taken from the conformation with the longest continuous helical structure were input to the Helical Wheel program to generate the associated diagram.

AUTHOR INFORMATION

Corresponding Author
cmstultz@mit.edu

ACKNOWLEDGMENT

The authors would like to thank Tim Bartels and Dennis Selkoe for kindly providing their CD data on the tetrameric α-synuclein structure isolated from red blood cells. We would also like to thank Tom Pochapsky for helpful discussions regarding the NMR data on the tetratmeric α-synuclein structure. Last, the authors thank Tomer Ullman for help with Figure 2. This work was supported in part by NIH 5R21NS063185-02.

REFERENCES

(1) Galvin, J. E.; Lee, V. M. Y.; Trojanowski, J. Q. *Arch Neurol.* **2001**, *58*, 186.
(2) Spillantini, M. G.; Goedert, M. *Ann. N.Y. Acad. Sci.* **2000**, *920*, 16–27.
(3) Goedert, M. *Nat. Rev. Neurosci.* **2001**, *2*, 492–501.
(4) Forno, L. S. *J. Neuropathol. Exp. Neurol.* **1996**, *55*, 259–272.
(5) Spillantini, M. G.; Crowther, R. A.; Jakes, R.; Hasegawa, M.; Goedert, M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6469–6473.
(6) Maries, E.; Dass, B.; Collier, T. J.; Kordower, J. H.; Steece-Collier, K. *Nat. Rev. Neurosci.* **2003**, *4*, 727–38.
(7) Eliezer, D.; Kutluay, E.; Bussell, R.; Browne, G. *J. Mol. Biol.* **2001**, *307*, 1061–1073.
(8) Weinreb, P. H.; Zhen, W. G.; Poon, A. W.; Conway, K. A.; Lansbury, P. T. *Biochemistry* **1996**, *35*, 13709–13715.
(9) Uversky, V. N. *J. Biomol. Struct. Dyn.* **2003**, *21*, 211–234.
(10) Li, J.; Uversky, V. N.; Fink, A. L. *Biochemistry* **2001**, *40*, 11604–11613.
(11) Davidson, W. S.; Jonas, A.; Clayton, D. F.; George, J. M. *J. Biol. Chem.* **1998**, *273*, 9443–9449.
(12) George, J. M.; Jin, H.; Woods, W. S.; Clayton, D. F. *Neuron* **1995**, *15*, 361–372.
(13) Bussell, R.; Eliezer, D. *J. Mol. Biol.* **2003**, *329*, 763–778.
(14) Chandra, S.; Chen, X.; Rizo, J.; Jahn, R.; Sudhof, T. C. *J. Biol. Chem.* **2003**, *278*, 15313–15318.
(15) Ulmer, T. S.; Bax, A.; Cole, N. B.; Nussbaum, R. L. *J. Biol. Chem.* **2005**, *280*, 9595–9603.
(16) Borbat, P.; Ramlall, T. F.; Freed, J. H.; Eliezer, D. *J. Am. Chem. Soc.* **2006**, *128*, 10004–10005.
(17) Georgieva, E. R.; Ramlall, T. F.; Borbat, P. P.; Freed, J. H.; Eliezer, D. *J. Am. Chem. Soc.* **2008**, *130*, 12856–12857.
(18) Jao, C. C.; Hegde, B. G.; Chen, J.; Haworth, I. S.; Langen, R. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 19666.
(19) Trexler, A. J.; Rhoades, E. *Biochemistry* **2009**, *48*, 2304–2306.
(20) Ferreon, A. C. M.; Gambin, Y.; Lemke, E. A.; Deniz, A. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5645.
(21) Georgieva, E. R.; Ramlall, T. F.; Borbat, P. P.; Freed, J. H.; Eliezer, D. *J. Biol. Chem.* **2010**, *285*, 28261.
(22) Bartels, T.; Choi, J. G.; Selkoe, D. J. *Nature* **2011**, *477*, 107–10.
(23) Apetri, M. M.; Maiti, N. C.; Zagorski, M. G.; Carey, P. R.; Anderson, V. E. *J. Mol. Biol.* **2006**, *355*, 63–71.
(24) Serpell, L. C.; Berriman, J.; Jakes, R.; Goedert, M.; Crowther, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 4897.
(25) Allison, J. R.; Varnai, P.; Dobson, C. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 18314–18326.
(26) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 476–477.
(27) Wu, K.-P.; Weinstock, D. S.; Narayanan, C.; Levy, R. M.; Baum, J. *J. Mol. Biol.* **2009**, *391*, 784–96.
(28) Bernado, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2005**, *127*, 17968–17969.
(29) Koo, H. J.; Choi, M. Y.; Im, H. *Biochem. Biophys. Res. Commun.* **2009**, *386*, 165–169.
(30) Fisher, C. K.; Huang, A.; Stultz, C. M. *J. Am. Chem. Soc.* **2010**, *132*, 14919–14927.
(31) Rao, J. N.; Kim, Y. E.; Park, L. S.; Ulmer, T. S. *J. Mol. Biol.* **2009**, *390*, 516–529.
(32) Bertoncini, C. W.; Fernandez, C. O.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *J. Biol. Chem.* **2005**, *280*, 30649–30652.
(33) 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*, 9893–9901.
(34) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
(35) Uversky, V. N.; Li, J.; Fink, A. L. *J. Biol. Chem.* **2001**, *276*, 10737–10744.
(36) Heinig, M.; Frishman, D. *Nucleic Acids Res.* **2004**, *32*, W500–W502.
(37) Rekas, A.; Knott, R. B.; Sokolova, A.; Barnham, K. J.; Perez, K. A.; Masters, C. L.; Drew, S. C.; Cappai, R.; Curtain, C. C.; Pham, C. L. L. *Eur. Biophys. J. Biophys. Lett.* **2010**, *39*, 1407–1419.
(38) El-Agnaf, O. M. A.; Irvine, G. B. *Biochem. Soc. Trans.* **2002**, *30*, 559–565.
(39) Rospigliosi, C. C.; McClendon, S.; Schmid, A. W.; Ramlall, T. F.; Barre, P.; Lashuel, H. A.; Eliezer, D. *J. Mol. Biol.* **2009**, *388*, 1022–1032.
(40) Sung, Y.; Eliezer, D. *J. Mol. Biol.* **2007**, *372*, 689–707.
(41) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 158–169.
(42) Stultz, C. M.; White, J. V.; Smith, T. F. *Protein Sci.* **1993**, *2*, 305–314.
(43) Mathews, F. S.; Bethge, P. H.; Czerwinski, E. W. *J. Biol. Chem.* **1979**, *254*, 1699–706.
(44) Banner, D. W.; Kokkinidis, M.; Tsernoglou, D. *J. Mol. Biol.* **1987**, *196*, 657–75.
(45) Kamtekar, S.; Hecht, M. H. *FASEB J.* **1995**, *9*, 1013–22.
(46) Andrade, M. A.; Chacon, P.; Merelo, J. J.; Moran, F. *Protein Eng.* **1993**, *6*, 383–390.
(47) Merelo, J. J.; Andrade, M. A.; Prieto, A.; Moran, F. *Neurocomputing* **1994**, *6*, 443–454.

(48) 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*, 17797–17802.
(49) Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T. *J. Am. Chem. Soc.* **2008**, *130*, 6145–6158.
(50) Sgourakis, N. G.; Merced-Serrano, M.; Boutsidis, C.; Drineas, P.; Du, Z. M.; Wang, C. Y.; Garcia, A. E. *J. Mol. Biol.* **2011**, *405*, 570–583.
(51) Feldman, H. J.; Hogue, C. W. V. *Proteins: Struct., Funct., Bioinf.* **2000**, *39*, 112–131.
(52) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099–13104.
(53) Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002.
(54) Marsh, J. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2009**, *391*, 359–374.
(55) Ho, B. K.; Dill, K. A. *PLoS Comput. Biol* **2006**, *2*, e27.
(56) Frimpong, A. K.; Abzalimov, R. R.; Uversky, V. N.; Kaltashov, I. A. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 714–722.
(57) Sandal, M.; Valle, F.; Tessari, I.; Mammi, S.; Bergantino, E.; Musiani, F.; Brucale, M.; Bubacco, L.; Samorì, B. *PLoS Biol.* **2008**, *6*, 0099–0108.
(58) Auluck, P. K.; Caraveo, G.; Lindquist, S. *Annu. Rev. Cell Dev. Biol.* **2010**, *26*, 211–33.
(59) Frishman, D.; Argos, P. *Proteins: Struct., Funct., Bioinf.* **1995**, *23*, 566–579.
(60) 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.
(61) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.
(62) Cho, M. K.; Nodet, G.; Kim, H. Y.; Jensen, M. R.; Bernado, P.; Fernandez, C. O.; Becker, S.; Blackledge, M.; Zweckstetter, M. *Protein Sci.* **2009**, *18*, 1840–1846.
(63) Fink, A. L. *Acc. Chem. Res.* **2006**, *39*, 628–634.
(64) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133–152.
(65) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(66) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J. Biomol. NMR* **2003**, *26*, 215–240.
(67) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791–3792.
(68) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 17908–17918.
(69) Marsh, J. A.; Baker, J. M. R.; Tollinger, M.; Forman-Kay, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 7804–7805.