JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

The Dynamic Structure of α-Synuclein Multimers

Thomas Gurry,[†,⊥] Orly Ullman,[‡,⊥] Charles K. Fisher,[§] Iva Perovic,[∥] Thomas Pochapsky,[⊗] and Collin M. Stultz*[,†,§,#]

[†]Computational and Systems Biology Initiative, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, United States

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

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

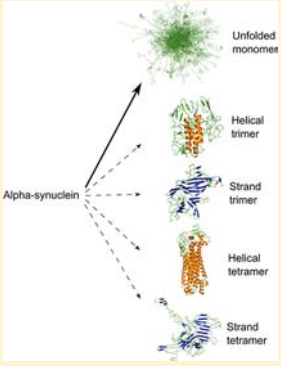
[∥]Department of Chemistry and Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02454, United States

[⊗]Department of Chemistry and Biochemistry and Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02454, United States

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

Supporting Information

ABSTRACT: α-Synuclein, a protein that forms ordered aggregates in the brains of patients with Parkinson's disease, is intrinsically disordered in the monomeric state. Several studies, however, suggest that it can form soluble multimers *in vivo* that have significant secondary structure content. A number of studies demonstrate that α-synuclein can form β-strand-rich oligomers that are neurotoxic, and recent observations argue for the existence of soluble helical tetrameric species *in cellulo* that do not form toxic aggregates. To gain further insight into the different types of multimeric states that this protein can adopt, we generated an ensemble for an α-synuclein construct that contains a 10-residue N-terminal extension, which forms multimers when isolated from *Escherichia coli*. Data from NMR chemical shifts and residual dipolar couplings were used to guide the construction of the ensemble. Our data suggest that the dominant state of this ensemble is a disordered monomer, complemented by a small fraction of helical trimers and tetramers. Interestingly, the ensemble also contains trimeric and tetrameric oligomers that are rich in β-strand content. These data help to reconcile seemingly contradictory observations that indicate the presence of a helical tetramer *in cellulo* on the one hand, and a disordered monomer on the other. Furthermore, our findings are consistent with the notion that the helical tetrameric state provides a mechanism for storing α-synuclein when the protein concentration is high, thereby preventing non-membrane-bound monomers from aggregating.



INTRODUCTION

α-Synuclein is a 140-residue protein that has been implicated in the pathogenesis of a number of neurodegenerative diseases, collectively known as synucleinopathies, the most well-known of which is Parkinson's disease.[1] The most notable pathological characteristic of these diseases is the aggregation of α-synuclein into amyloid fibrils, which have significant β-sheet secondary structure.[2,3] Although there is disagreement regarding whether the soluble oligomeric aggregates or insoluble aggregates are the most neurotoxic species, it is clear that α-synuclein self-association plays an integral role in neuronal dysfunction and death.[4−8] Given the importance of this protein in these neurodegenerative disorders, studies that help to elucidate its structure are of paramount importance.

However, the conformational landscape of α-synuclein is notoriously difficult to study, earning it the moniker of "chameleon" due to its tendency to adopt different conformations under different experimental conditions.[9,10] This has led to seemingly contradictory data about the dominant putative states in solution versus those under physiologic conditions.[11−13] While it is clear that monomeric α-synuclein is an intrinsically disordered protein[14] in solution, recent data suggests that it can adopt a tetrameric state that has a relatively high helical content under physiologic conditions.[11,13,15] By contrast, others have suggested that α-synuclein retains its monomeric disordered state *in cellulo*.[12,16]

Recently, NMR studies on an α-synuclein construct isolated from *Escherichia coli*, which contains a 10 residue N-terminal extension, suggested that the protein can exist as a "dynamic tetramer".[13] In short, these data are consistent with a model where the protein rapidly interconverts between different conformers, where some of these conformations are multimeric structures (trimers and tetramers) that contain significant

Received: October 24, 2012
Published: February 11, 2013

   dx.doi.org/10.1021/ja310518p | J. Am. Chem. Soc. 2013, 135, 3865−3872

helical content. To obtain a more comprehensive view of the types of structures that this particular α-synuclein construct can adopt, we generated an atomistic model for α-synuclein in its multimeric form. While we recognize that it is not possible to capture all possible monomeric and multimeric conformations that this protein can adopt in solution, our hope was to build a low-resolution description of the dominant states of the protein. More precisely, we define a conformational ensemble to consist of a structural library $S = \{\vec{s}_i\}_{i=1}^{n}$, where $\vec{s}_i$ is the Cartesian coordinates of structure i, and a corresponding set of weights $\vec{w} = \{w_i\}_{i=1}^{n}$, where w_i is the population weight of structure i. In this sense, the number of structures in the ensemble, n, is a function of the resolution with which one wishes to view the conformational landscape of the system.

As prior studies on this construct suggest that the purified protein contains primarily monomers, trimers and tetramers, we focused on these specific forms for our ensemble.[13] Since we had previously constructed an ensemble for monomeric α-synuclein using NMR chemical shifts, residual dipolar couplings (RDCs), and SAXS data,[17] we used these structures to represent the disordered, monomeric fraction. Using NMR chemical shifts and NH RDCs obtained on an α-synuclein construct, which contains a 10-residue N-terminal extension, we determine the relative fractions of different multimeric forms within the ensemble.

MATERIALS AND METHODS

Generation of Seed Structures. Our previous study on α-synuclein suggested that the monomeric, protein can sample amphipathic helices, which could in principle self-associate to form helical trimers and tetramers.[17]

All simulations used a model of α-synuclein that did not include the 10-residue N-terminal extension. An initial trimeric structure of the protein was generated by taking a monomer from the monomeric α-synuclein ensemble that has an amphipathic helix between residues 52 and 64 and threading the helix to a three-helix bundle from a crystal structure of myosin (PDB ID code 3GN4),[18] where the hydrophobic faces of the amphipathic helix were oriented such that they face inward. An initial tetrameric structure was generated by threading the same monomer to a four-helix bundle from a crystal structure of ferritin (PDB ID code 1FHA).[19,20] These structures were chosen from the Protein Data Bank such that the helix bundles in the structure used for threading the monomer were of sufficient length to accommodate the entire 12-residue helix in our monomer structure, while retaining a high enough resolution to be informative. A second initial helical tetrameric model was constructed using the available NMR data.[13] The model derived from the NMR data was obtained from a limited set of nuclear Overhauser effects (NOEs); i.e., we were not able to identify a sufficient number of sequential (Hα−HN i, i+3) NOEs in ^{15}N-edited NOESY spectra (see below). Consequently, the resulting model is not intended to represent a "high-resolution" structure of the helical tetramer. Instead, its only purpose is to serve as a structure (derived from limited experimental data) that is the starting point for additional simulations. More generally, each seed structure serves as a starting point from which to begin more extensive sampling.

Generation of α-Synuclein Structural Library. The conformational space of α-synuclein was sampled by subjecting the initial seed structures to replica exchange molecular dynamics (REMD) simulations.[21] Each initial structure underwent REMD with the EEF1[22] implicit solvent model as implemented in the CHARMM[23] force field. Sixteen replicas were used, with temperatures equally spaced in 5 K increments over the 293−368 K range. Prior studies of IDPs with this implicit solvent model have yielded useful insights.[17,24,25] Initially, higher temperature replicas were explored, along with quenched molecular dynamics simulations at higher temperatures, but we found that these led to dissociation of multimers into monomers free of intermolecular contacts. We therefore limited the highest temperature to 368 K, the highest temperature at which intermolecular contacts were retained in oligomers for the duration of the trajectory. Each replica was run for 20 ns, and structures were collected at each picosecond. A total of 20 000 conformations per REMD simulation were collected, all from the 298 K window, making a total of 60 000 conformations for the trimeric and tetrameric structures.

The set of 60 000 structures was pruned down by enforcing a minimum pairwise root-mean-square deviation of 9 Å to ensure that the resulting library would span a range of heterogeneous conformations. The resulting set contained 234 structures. These were then combined with 299 monomer structures from a previously constructed monomeric ensemble of α-synuclein[17] to yield our structural library $S = \{\vec{s}_i\}_{i=1}^{533}$ of 533 conformers.

Generation of the Ensemble and Calculation of Confidence Intervals. To obtain the set of weights associated with each conformer in our structural library, we employ the variational Bayesian weighting (VBW) algorithm previously described,[26] which is a variational approximation to a Bayesian weighting formalism used in the past.[17,24] This algorithm generates a posterior distribution $f_{\vec{W}|\vec{M},S}(\vec{w}|\vec{m},S)$ for the weights, conditioned on the set of 533 structures, and the provided experimental measurements. The form of the posterior distribution is dictated by Bayes's rule:

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

where the term $f_{\vec{W}|S}(\vec{w}|S)$ is the prior distribution and $f_{\vec{M}|\vec{W},S}(\vec{m}|\vec{w},S)$ is the likelihood function for the experimental observations $\vec{m}$, whose full descriptions can be found in the original publication of the method.[26] Experimental observables, specifically Cα, Cβ, N, H, and Hα chemical shifts from a previous work[13] in combination with backbone NH RDCs, were used (Supporting Information Table S1). Predicted measurements for each conformer were generated using SHIFTX[27] for chemical shifts and PALES[28] for RDCs. RDCs were uniformly scaled to account for uncertainty in the magnitude of the alignment tensor. Similarly, like-atom chemical shifts were uniformly offset to account for uncertainty in chemical shift referencing. To increase computational efficiency and analytical tractability, an approximation from variational Bayesian inference was applied by choosing a simpler probability density function (PDF),[26] which approximates the full posterior distribution, calculated from eq 1. For a vector of weights, a natural choice is the Dirichlet distribution with parameters $\{\alpha_i > 0\}_{i=1}^{N}$. This results in an approximate PDF for the weights:[26]

$$g(\vec{w}|\vec{\alpha},\,S) = \frac{\Gamma(\alpha_0)}{\sum_{i=1}^{n}\Gamma(\alpha_i)}\prod_{i=1}^{N} w_i^{\alpha_i - 1} \tag{2}$$

where α_i is the Dirichlet parameter associated with weight i and $\alpha_0 = \sum_i \alpha_i$. The Kullback−Leibler distance (i.e., the KL divergence) between $g(\vec{w}|\,\vec{\alpha},S)$ and $f_{\vec{W}|\vec{M},S}(\vec{w}|\vec{m},S)$ is then minimized to find the optimal set of Dirichlet parameters, $\vec{\alpha}' = \{\alpha_i'\}_{i=1}^{N}$, which provides an approximation to the true posterior from which one can easily calculate quantities of interest.

We then compute the Bayes estimate for the weights $\vec{w}^B = \{w_i^B\}$, which is the expected value of the vector of weights over the new approximate posterior distribution:

$$\vec{w}^B = \int d\vec{w}\, g(\vec{w}|\vec{\alpha}',\,S)\vec{w} \tag{3}$$

The Bayes estimate can be calculated from the Dirichlet PDF according to

$$w_i^B = \frac{\alpha'_i}{\alpha'_0} \tag{4}$$

where $\alpha'_0 = \sum_i \alpha'_i$. The uncertainty parameter $\sigma_{\vec{w}^B}$, called the posterior expected divergence, corresponds to the average distance from the Bayes weights over the entire space of weights:

$$\sigma_{\vec{w}^B} = \sqrt{\int d\vec{w}\Omega^2(\vec{w}^B, \vec{w})g(\vec{w}|\hat{\alpha}', S)} \quad (5)$$

where $\Omega^2(\vec{w}^B$ and $\vec{w})$ is the Jensen−Shannon divergence, a metric which quantifies the distance between the vectors $\vec{w}^B$ and $\vec{w}$.[24]

The covariance between the weights of conformers i and j can be calculated analytically from

$$\mathrm{cov}(w_i, w_j) = \frac{\alpha'_i\alpha'_0\delta_{ij} - \alpha'_i\alpha'_j}{\alpha'^2_0(\alpha'_0 + 1)} \quad (6)$$

where δ_{ij} is the Kronecker delta function. Any quantity D that can be calculated for a given conformer can then be assigned a variance across the ensemble according to

$$\mathrm{var}(D) = \sum_i \sum_j D_i D_j \mathrm{cov}(w_i, w_j) \quad (7)$$

95% confidence intervals can then be computed using a Gaussian approximation from CI = 1.54 × 1.96 × $(\mathrm{var}(D))^{1/2}$, where 1.54 is an empirical factor relating the variational approximation of the posterior distribution to the true posterior distribution under the complete BW formalism.[26]

A backward elimination procedure starting with our initial structural library of 533 conformers was used to ensure that the ensemble only contained essential structures. The procedure computed the VBW posterior distribution iteratively. After each iteration, all non-essential structures were identified by finding the largest set I such that the joint probability that each weight of the structures in I fell below a cutoff exceeded a chosen confidence level; i.e., $\Pi_{i\in I}P(w_i \leq c) \geq 1 - \theta$, where $P(\cdot)$ denotes the cumulative distribution function of the weights. The cutoff (c) and confidence level (θ) were set to 0.005 and 0.05 (95%), respectively. Each of the non-essential structures in I was removed and the weighting procedure repeated. This process was iterated until convergence, i.e., until the cardinality of I was zero.

Secondary Structure Assignments. Secondary structure was assigned using DSSP.[29] A residue was assigned to the class of "helix" if it was assigned as α-helix, π-helix, or 3−10 helix by DSSP. Similarly, a residue was assigned to the class of "strand" if it was assigned as a bridge or extended by DSSP. The remaining assignments were grouped into the class of "other". Structures appearing in the uppermost quartile of tetramers ranked by helical content were classified as helical tetramers, and structures in the uppermost quartile of tetramers ranked by strand content were classified as strand tetramers. Trimers were classified in the same manner.

Solvent Accessibility Calculations. Solvent-accessible surface area (SASA) was calculated for each conformation using CHARMM.[23] Since only the backbone atoms N, H, C, Cα, and O are involved in the formation of secondary structure, only SASA values for these atoms were considered. The solvent accessibility for the entire protein was computed by summing each atom's SASA value and normalized by dividing the result by the SASA of the α-synuclein backbone atoms when in a fully extended conformation.

NMR Studies. It is important to note that these NMR studies were insufficient to uniquely determine the structure of a helical tetrameric state (primarily due to an insufficient number of measured NOEs). Hence, the structure arising from these studies represents a model that only serves as the starting point for further simulations, as opposed to a well-defined structure for the helical tetramer.

Samples of ^{15}N- and ^{13}C-labeled αSyn for NMR spectroscopy were prepared using uniformly ^{13}C- and ^{15}N-labeled media (supplemented M9 media, ^{13}C source being glucose). NMR samples were typically prepared to a final concentration of ∼0.5 mM in 100 mM Tris-HCl pH 7.4, 100 mM NaCl, 0.1% BOG, 10% glycerol, 10% D_2O. All NMR spectroscopy was performed on a Bruker Avance 800 NMR spectrometer operating at 800.13 MHz (^{1}H), 81.08 MHz (^{15}N), and 201.19 MHz (^{13}C) and equipped with a TCI cryoprobe and pulsed field gradients. Experiments used for sequential resonance assignments include three-dimensional (3D) experiments HNCA, HNCACB, ^{15}N-HSQC TOCSY, and ^{15}N-HSQC NOESY. Quadrature detection was obtained in the ^{15}N dimension of 3D experiments using sensitivity-enhanced gradient coherence selection,[30] and in the ^{13}C dimension using States-TPPI, with coherence selection obtained by phase cycling. In all cases, spectral widths of 8802.82 Hz (^{1}H) and 2920.56 Hz (^{15}N) were used. For ^{13}C, spectral widths of 6451.61 Hz (HNCA) and 15105.74 Hz (HNCACB) were used. All experiments were performed at 298 K unless otherwise noted. NMR data were processed using TOPSPIN (Bruker Biospin Inc.), and data were analyzed using either TOPSPIN or SPARKY.[31]

^{1}H−^{15}N, ^{13}C′−^{15}N, and ^{13}C′−^{13}Cα RDCs were recorded for a ^{15}N- and ^{13}C-labeled wild-type αSyn oligomer sample in the presence and absence of alignment media using a standard IPAP-HSQC sequence or a variation of a standard HNCO pulse sequence. Sample alignment was accomplished using a 5% polyacrylamide stretched gel. We chose to use PA rather than bicelle or liquid crystalline phases for alignment because such phases contain long chain hydrocarbon moieties that might be expected to bind αSyn and could interfere with oligomer formation.

The stretched gel was prepared using a commercial apparatus (New Era, Vineland, NJ) according to the manufacturer's protocol and following guidelines reported by Bax.[32] After polymerization was complete, the gel was dialyzed against water overnight at room temperature, and then incubated with a 0.5 mM αSyn sample in standard NMR buffer for 48 h at 4 °C. The diameter of the gel was 6.0 mm before and 4.2 mm after stretching. Alignment was confirmed by observing the residual quadrupolar splitting (9.4 Hz) of the ^{2}H water signal.

We used solution NMR to localize the transient formation of α-helices in αSyn. Resonance assignments were made using standard methods (HNCO, HN(CO)CA, HNCA, HNCACB, ^{15}N-edited NOESY and TOCSY). Although a high degree of spectral overlap is present even in three-dimensional data sets, we were able to identify a number of sequential (Hα-HN i, i+3) NOEs in ^{15}N-edited NOESY spectra to confirm the transient existence of α-helical structure between residues Phe4-Thr43 and His50-Asn103. In many cases, these NOEs are quite weak, consistent with fractional occupancy, however, only the most reliable (strongest) experimental NOEs were used in model construction (Figure S1). Note that if long stretches of NOEs interrupted by several residue pairs without NOEs were observed, the missing pairs were included in the helical restraints applied in XPLOR-NIH. A total of 73 unique inter-residue NOEs per monomer were used to construct a model for the helical tetramer.

Given the relatively small number of NOEs any structure arising from these data merely represents a model (derived from limited experimental data) that serves as fodder for additional simulations, rather than a detailed high-resolution structure of the tetrameric state.

A combined torsional and Cartesian dynamics simulated annealing method was used to calculate an average tetramer structure using XPLOR-NIH v. 2.18.[33] Secondary structural restraints were applied to those regions of the polypeptide identified as forming α-helical structure from sequential NOEs. RDC restraints were applied for residues 1−103, and in some cases, non-crystallographic symmetry restraints were applied to residues 4−36, 47−85, and 89−98. Preliminary structures were crafted manually using PyMOL,[34] and initial values for alignment tensors determined by singular value decomposition (SVD) using the program PALES.[28] As refinement proceeded, best-fit structures were used to recalculate the alignment tensors via a combined SVD−least-squares fit which permits the rhombic terms to be fixed at zero. This was applied iteratively until no further improvements of fit were observed. PyMOL was also used for visualization of the structures generated by XPLOR-NIH. Proton chemical shifts were referenced directly to the water signal at 4.7 ppm, while ^{15}N and ^{13}C shifts were indirectly referenced.[35] NMR data are available in Table S1. Structural models for the multimeric state of α-synuclein will be freely available via http://www.rle.mit.edu/cbg.

■ RESULTS AND DISCUSSION

To generate a set of energetically favorable multimers for the ensemble, we began with a set of "seed" structures that served as starting points from which a diverse library of multimeric

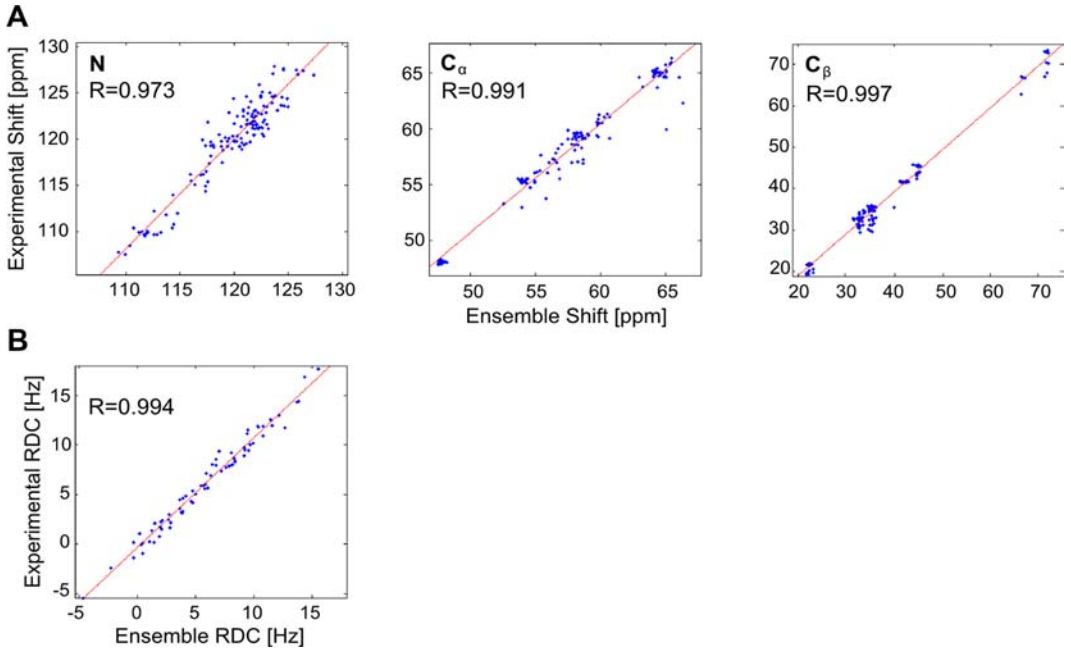


Figure 1. Calculated ensemble averages vs experimental measurements: (A) N, Cα, and Cβ chemical shifts; (B) N−H residual dipolar couplings. Correlation coefficients for each plot are explicitly shown.

structures could be built. Our previous study on α-synuclein suggested that the monomeric protein can sample amphipathic helices, which could in principle self-associate to form higher order structures.[17] Hence, we constructed trimeric and tetrameric structures using amphipathic helices from the monomeric ensemble. Structures for both the trimeric and tetrameric species were obtained by threading these amphipathic helices onto three- and four-helix bundles, respectively, from the Protein Data Bank such that the hydrophobic faces of these helices form the contact-interface (see Materials and Methods). A second helical tetrameric model was constructed using the available NMR data.[13] The model derived from the NMR data was obtained from a limited set of NOEs because a high degree of spectral overlap is present even in three-dimensional data sets. Consequently, the resulting model is not intended to represent a "high-resolution" structure of the helical tetramer. Instead, it is a model, constructed from limited experimental data, which serves as a starting point for additional simulations. Indeed, all seed structures represent initial structures (derived from experimental data and from prior studies on the monomeric state) from which to begin sampling, rather than high-resolution structures for trimeric and tetrameric structures.

Each seed structure was subjected to REMD[21] (16 replicas, each replica run for 20 ns). Structures from the 298 K window were output every picosecond and added to the structural library. In total, the structural library contained 60 000 structures (monomers, trimers, and tetramers). All of these structures were then clustered using a crude pruning algorithm to ensure that the final set of structures largely retained the structural heterogeneity present in the original 60 000. The final set of structures, including monomers, trimers, and tetramers, contained 533 conformers.

We note that each of the replica exchange simulations began with a predominantly helical seed structure because several studies suggest that α-synuclein multimers had significant helical content.[11,13,15] However, many of the helical multimers rearranged to form strand-rich conformers during the course of the simulations. Hence the final set of 533 structures constitutes a heterogeneous set of conformers that have a range of both helical and strand content.

The final step in our ensemble construction procedure was to assign population weights to each of the 533 structures. One approach to accomplish this is to obtain a single set of weights, $\vec{w} = \{w_i\}_{i=1}^{n}$, such that calculated observables from the final ensemble agree with the corresponding experimentally determined values. However, as we have previously shown, agreement with experiment alone is insufficient to ensure that the constructed ensemble is correct.[24,36] This is because the construction of ensembles for disordered systems is an inherently degenerate problem; i.e., the number of experimental constraints pales in comparison to the number of degrees of freedom for the system. To overcome this limitation, we used a previously developed formalism, grounded in Bayesian statistics, to compute the population weights. This BW algorithm computes the full posterior distribution over all possible ways of weighting structures in the structural library. From this posterior distribution we can compute an uncertainty measure, $0 \leq \sigma_{\vec{w}^B} \leq 1$, which describes the spread of the posterior distribution—a metric that is akin to the standard deviation of a Gaussian distribution.[24,26] Our prior work suggests that the numeric value of $\sigma_{\vec{w}^B}$ is correlated with model correctness. When $\sigma_{\vec{w}^B} = 0$, we can be relatively certain that the model is correct. By contrast when $\sigma_{\vec{w}^B} = 1$, it is likely that the ensemble is far from the truth. Nevertheless, when $\sigma_{\vec{w}^B} \neq 0$, we can construct rigorous confidence intervals for quantities of interest that are calculated from the ensemble. The ability to calculate rigorous confidence intervals enables us to perform rigorous hypothesis tests and therefore determine what conclusions we can make from the ensemble with statistical significance.

The final Bayes's ensemble consists of a set of weights, $\vec{w}^B = \{w_i^B\}$, which corresponds to the expected value of the weights calculated from the posterior distribution, and the structural

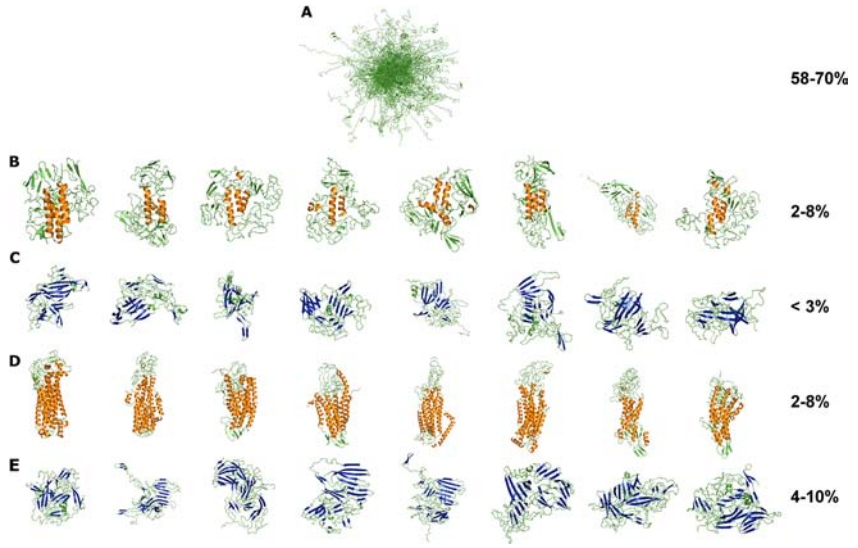


Figure 2. Types of α-synuclein structures in our ensemble. Monomers are aligned to each other (A) to demonstrate that they form a structurally heterogeneous set. For the multimeric species, the top eight structures from each category in terms of secondary structure content are shown: (B) helical-rich trimers, (C) strand-rich trimers, (D) helical-rich tetramers, and (E) strand-rich tetramers.

library $S = \{\vec{s}_i\}_{i=1}^{n}$. The algorithm also ensures that we restrict our analysis to the most important conformers. More precisely, the ith structure is excluded from the ensemble when we can say with 95% confidence that $w_i \leq c$. In the end, a total of 311 structures survived this criterion. While the resulting Bayes's ensemble achieves a good fit to the NMR experimental data (Figure 1), the corresponding uncertainty parameter is nonzero: $\sigma_{\vec{w}^B} = 0.47$. Consequently, we express ensemble average values along with their corresponding 95% confidence intervals.

The ensemble is composed mostly of monomeric species (64.1% ± 6.4%), with tetrameric species making up the next most common species (28.2% ± 6%), and trimeric structures making up only 7.7% ± 3.6%. Since we have already reported on the types of structures that are sampled in the monomeric protein,[17] here we focus on the types of multimeric structures that appear in the ensemble. Both trimeric and tetrameric structures mainly come in two forms, either predominantly helical, or predominately strand. A small fraction of multimeric structures contain so little secondary structure that they fall into neither category. Representative structures from each species are shown in Figure 2.

To determine how each of these multimers may influence α-synuclein self-association, we focus on the position and conformation of the subsequence NAC(8−18), which corresponds to the minimal segment of α-synuclein that can initiate the formation of toxic β-strand-rich aggregates *in vitro*.[37] This is of particular interest because toxic soluble oligomers of α-synuclein and other related IDPs contain significant β-structure.[38,39] Of all the multimeric species in the ensemble, the normalized solvent accessibility of the NAC(8−18) region in helical tetramers is significantly lower than for other types of structures, with an expected value of only 30.6% ± 1.0% (Figure 3). For comparison, the solvent exposure of the NAC(8−18) region in the monomeric fraction is 58.6% ± 4.2%. Consequently, helical tetrameric species bury the NAC(8−18) segment relative to the monomeric state. Our findings are consistent with a model where the NAC(8−18) segment initiates the formation of β-rich structures, which then progress to form higher order aggregates. In the β-rich conformers, the NAC(8−18) segment has already been incorporated into β sheet and therefore it is not surprising that their solvent accessibility is reduced. In the helical tetramer the NAC(8−18) segment is hidden in a nonamyloidogenic conformation and is therefore not available to initiate the formation of β-strand-rich multimers.

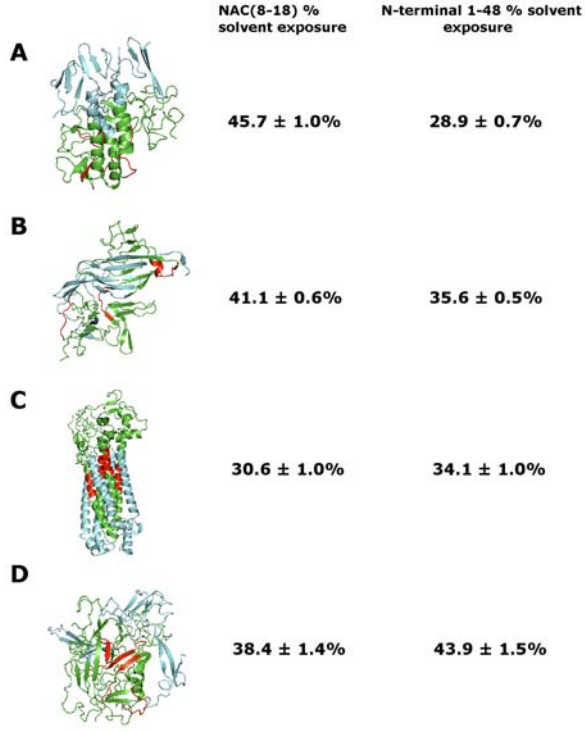


Figure 3. Normalized solvent accessibility (±95% confidence intervals) for the NAC(8−18) region and N-terminal residues 1−48 for (A) helical-rich trimers, (B) strand-rich trimers, (C) helical-rich tetramers, and (D) strand-rich tetramers. Representative structures are shown on the left. The N-terminal residues are shown in cyan, the NAC(8−18) in red, and the remaining residues in green.

dx.doi.org/10.1021/ja310518p | *J. Am. Chem. Soc.* 2013, 135, 3865−3872

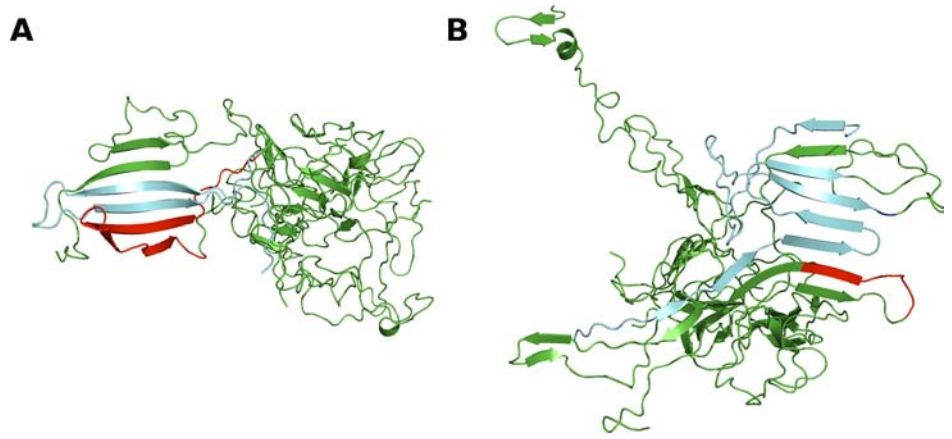


Figure 4. Two representative structures of strand-rich tetramers. The N-terminal residues 1−48 of the monomers participating in sheets are shown in cyan. NAC 8−18 residues participating in sheets are shown in red.

Several studies also suggest that the N-terminal region of α-synuclein may act as an initiation site for the formation of strand-rich oligomeric aggregates. The observation that aggregation-inhibiting small molecules bind preferentially to the N-terminal region of human α-synuclein is consistent with this notion.[40] More importantly, ^{15}N relaxation experiments performed on monomeric mouse α-synuclein (which has faster aggregation kinetics than the human homologue) suggest that the N-terminal region of the protein has decreased backbone flexibility as compared to both a random coil model as well as measurements on human α-synuclein—a finding suggesting that secondary structure formation is more prevalent in the mouse form of the protein.[41] It has further been proposed that KTK(E/Q)GV, which are mainly found within the first 48 residues of the protein, can serve as initiation sites for aggregation in mouse α-synuclein.[41] Therefore, we computed the average solvent accessibility of the N-terminal 48 residues in each multimeric state to explore the conformation of the N-terminal region of α-synuclein in each of these multimeric states, as shown in Figure 3. Helical trimers and tetramers preferentially place the N-terminal region of α-synuclein in positions that are hidden from solvent; i.e., the solvent exposure of these regions is 28.9% ± 0.7% and 34.1% ± 1.0% for helical trimers and tetramers, respectively. We note that several studies suggest that the N-terminal region of α-synuclein plays a critical role in the formation of helical structures,[42−44] hence this region may be important for assembly of the helical tetramer. By contrast, the solvent exposure for the monomeric state is 52.5% ± 3.6%. Figure 4 shows two structures that involve the N-terminal residues in β-sheet formation, highlighting the β-strand propensity of these residues.

Interestingly, however, β-strand-rich trimers and tetramers, preferentially have the N-terminal residues 1−48 involved in a sheet that contains the NAC(8−18) segment; i.e., the segment that can initiate α-synuclein aggregation *in vitro* (Figure 4). Although it is not clear whether the NAC component or the N-terminal region provides the primary impetus behind the oligomerization propensity of α-synuclein, our data are consistent with a model whereby the initial stages in toxic oligomer formation is the formation of an N-terminal β-strand-rich region that contains the NAC(8−18) segment. In this regard, it is interesting that the helical tetrameric species sequesters both of these regions from the surrounding solvent by involving them in the formation of helices, as shown in Figure 3, supporting the notion that this structure acts as a nontoxic storage mechanism.

■ CONCLUSIONS

In this study we constructed an ensemble for the multimeric state of α-synuclein. Our data reveal a number of important insights into the types of structures that multimeric forms of the protein can adopt. Given that generating a comprehensive list of the thermally accessible states of both the monomeric and multimeric protein is not tractable, our goal was to generate a low-resolution description of the dominant states that are available to the protein. However, even with this proviso additional assumptions are needed to make the calculations feasible. In this regard we restricted our sampling of multimeric states to trimers and tetramers; i.e., the primary multimeric states that have been observed when α-synuclein constructs are isolated from *E. coli*, red blood cells, and human neuroblastoma cell lines.[11,13] Replica exchange molecular dynamics (REMD) simulations were used to generate a representative set of heterogeneous set of energetically favorable conformers that served as the template from which a structural ensemble could be built. Given that earlier studies had described the existence of helical trimers and tetramers of α synuclein, the REMD simulations began using a predefined set of seed structures that were intended to capture conformations that were observed in earlier experiments on α-synuclein multimers. Given that our previous study suggested that the monomeric α-synuclein can sample amphipathic helices, we generated a model for helical trimers and tetramers assuming that multimeric structures were formed from self-association of these amphipathic helices. A second model seed structure was derived from limited NMR data on α-synuclein at high concentrations. Given the limited number of NOEs obtained, it was not possible to uniquely determine the structure of any tetrameric state; therefore the resulting seed structure serves as fodder for additional simulations, rather than a detailed high-resolution structure of the tetrameric state. Although the REMD simulations began with these seed structures, the resulting trajectories sample a wide region of conformational space leading to the generation of some structures that are very different from the initial seeds (Figures S2 and S3). The BW method is then used to construct a probability density over all possible ways of assigning population weights to structures arising from the trajectories.[24]

These data are then used to calculate ensemble average properties with their corresponding confidence intervals.

Given that construction of an ensemble for an intrinsically disordered protein is an inherently degenerate problem, it is important to provide estimates of one's uncertainty in the resulting ensemble.[24,36] One advantage of the BW formalism is that it has a built in measure of uncertainty, $0 \leq \sigma_{\vec{w}^B} \leq 1$, that is correlated with model correctness.[24] When $\sigma_{\vec{w}^B} = 0$, we can be relatively certain that the model is correct. By contrast when $\sigma_{\vec{w}^B} = 1$, it is likely that the ensemble is far from the truth. In the present case, this uncertainty parameter is non-zero: $\sigma_{\vec{w}^B} = 0.47$. However, even when the uncertainty parameter is non-zero,nonzero, one can still quantify the uncertainty in calculated ensemble average quantities via the use of confidence intervals. In this work, we present ensemble averages ±95% confidence intervals. Confidence intervals comprise a standard statistical method to quantify uncertainty in an underlying model. The meaning of the confidence interval for the ensemble average $\langle M \rangle$ is that, if one calculated $\langle M \rangle$ from many different ensembles (that also fit the experimental data), then those values would fall within the 95% confidence intervals approximately 95% of the time. The 95% confidence interval therefore provides a quantitative measure for the range of values one would see if they constructed many different ensembles. Overall we find that helical tetramers represent a relatively small fraction (5.1% ± 2.9%) of an otherwise predominantly disordered, monomeric, ensemble. These findings are consistent with recent bacterial in-cell experiments that suggest that α synuclein is predominantly disordered within the crowded intracellular environment.[16]

Our data suggest that the multimeric ensemble contains tetrameric states that have significant helical content. However, while some groups have been able to isolate helical tetramers by using gentle purification protocols, the isolation of such structures by other groups has remained elusive.[11,12,45] These latter experiments have led some to conclude that α-synuclein predominantly exists as a disordered monomer under physiologic conditions.[12] We believe our data help to reconcile these seemingly contradictory observations. Our findings argue that helical tetramers are present within the unfolded ensemble, albeit at very low concentrations. Successful isolation of helical tetramers would therefore require additional measures to increase the relative population weight of these states. Indeed, it has been shown that the tetrameric species elute from purification columns in a concentration-dependent manner when the protein is acetylated at its N-terminus.[15] This suggests that the relative abundance of this species is a function, in part, of the post-translational state of the protein, the purification protocol, and the protein concentration. These observations are consistent with the notion that the helical tetramer provides a mechanism for *in cellulo* α-synuclein storage when the protein concentration is high. Formation of aggregation resistant helical tetramers may provide a method to sequester non-membrane-bound monomers in a form that both prevents them from aggregating and preserves them in a conformation amenable to lipid binding upon dissociation.

To understand why helical states are aggregation resistant, we focus on the minimal segment, NAC(8−18), needed to initiate α-synuclein aggregation *in vitro*.[37] Of all the multimeric states in our ensemble, the solvent exposure of the NAC(8−18) is the lowest for the helical tetramer. Burying the NAC(8−18) segment ensures that is not available to initiate the formation of β-strand-rich oligomers. In the β-rich tetramer conformers, the NAC(8−18) segment has already been subsumed in a central β sheet and therefore it is not surprising that its solvent accessibility is reduced relative to the monomeric state. Our findings are consistent with a model where the NAC(8−18) segment initiates the formation of β-rich tetramer structures, which then progress to form higher order aggregates.

The appearance of strand-rich states in our ensemble is somewhat surprising given that previously published CD spectra of multimeric α-synuclein suggested that the protein had considerable helical content on average.[11,13] Although the reported CD spectra have distinct minima at 208 and 222 nm—a finding indicative of considerable helical content—estimating the precise helical content from CD spectra alone is problematic.[46,47] For example, we used several different algorithms to quantify the helical content from the published CD spectrum of α-synuclein isolated from human red blood cells,[11] and depending on the algorithm used, the amount of helix varied from 10% to 80%. Hence, while the CD spectrum suggests that the helical content of the tetrameric species is higher than that of the monomeric protein, quantifying the amount of helicity from the CD spectrum alone is a non-trivial exercise. In addition, the multimeric ensemble was generated using data from NMR experiments that were performed at a concentration (0.5 mM) that was at least an order of magnitude greater than the concentration used for the CD experiments (~0.02 mM). This is important because the concentration of α-synuclein *in vitro* can influence its secondary structure propensity and the precise effect may vary on the post-translational state of the protein.[15,48,49] Therefore it is not clear whether the published CD spectrum reflects the structure of α-synuclein under the conditions used for the NMR experiments.

We also note that it is likely that a variety of factors, such as the ionic strength of the medium and presence of divalent metal cations,[50] would affect the relative stabilities of different conformations: it has been shown, for instance, that the abundance of β-rich monomeric structures increases in the presence of high ionic strength, as well as upon inclusion of Cu^{2+}.[51] The effect of these ions on the relative stability of multimeric forms is unknown and will be explored in future work.

Lastly, we note that a limitation of our study is that the NMR data were obtained on an α-synuclein construct that contains a 10-residue N-terminal extension relative to the wild-type protein. While the experimental data provided useful constraints that could be fruitfully applied to generate an ensemble, α-synuclein isolated from human neuroblastoma and red blood cell lines does not have an N-terminal extension and instead is acetylated at the N-terminus.[11] Nevertheless, our construct shares important characteristics with the N-acetylated protein. First, the monomeric form of the construct bearing a 10-residue N-terminal extension has a CD spectrum that is similar to that of the monomeric N-terminal acetylated form of α-synuclein[12] and both constructs form tetrameric structures with increased α-helical content.[11,13,15] Lastly, monomeric forms of both constructs have similar aggregation profiles while the tetrameric forms of both constructs do not aggregate.[11,13] These similarities suggest that acetylation of the N-terminal and the 10-residues elongation of the N-terminal region in α-synuclein serve a similar purpose with regard to their effect on the α-synuclein, albeit N-terminal acetylation may have more dramatic effects on the conformational distribution of the protein relative to the N-terminal

extension. Nonetheless, since the sequence of this construct differs from the wild-type protein, we cannot exclude the possibility that wild-type α-synuclein isolated from other cell types, such as neurons or red blood cells, may not be well described by the ensemble presented here.

■ ASSOCIATED CONTENT

Supporting Information

Experimental data, representative multimeric structures from the ensemble, and additional information on the structures sampled by the REMD simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

cmstultz@mit.edu

Author Contributions

$^{\perp}$T.G. and O.U. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Bellucci, A.; Zaltieri, M.; Navarria, L.; Grigoletto, J.; Missale, C.; Spano, P. *Brain Res.* **2012**, *1476*, 183.

(2) Spillantini, M. G.; S., M. L.; Lee, V.M.-Y.; Trojanowski, J. Q.; Jakes, R.; Goedert, M. *Nature* **1997**, *388*, 839.

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

(4) Conway, K. A.; Lee, S.-J.; Rochet, J.-C.; Ding, T. T.; Williamson, R. E.; Lansbury, P. T. *Proc. Natl. Acad. Sci.* **2000**, *97*, 571.

(5) Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdo, J.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M. *Nature* **2002**, *416*, 507.

(6) Kayed, R.; Head, E.; Thompson, J. L.; McIntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabe, C. G. *Science* **2003**, *300*, 486.

(7) Danzer, K. M.; Haasen, D.; Karow, A. R.; Moussaud, S.; Habeck, M.; Giese, A.; Kretzschmar, H.; Hengerer, B.; Kostka, M. *J. Neurosci.* **2007**, *27*, 9220.

(8) Winner, B.; Jappelli, R.; Maji, S. K.; Desplats, P. A.; Boyer, L.; Aigner, S.; Hetzer, C.; Loher, T.; Vilar, M. a.; Campioni, S.; Tzitzilonis, C.; Soragni, A.; Jessberger, S.; Mira, H.; Consiglio, A.; Pham, E.; Masliah, E.; Gage, F. H.; Riek, R. *Proc. Natl. Acad. Sci.* **2011**, *108*, 4194.

(9) Uversky, V. N. *J. Biomol. Struct. Dyn.* **2003**, *21*, 159.

(10) Drescher, M.; Huber, M.; Subramaniam, V. *ChemBioChem* **2012**, *13*, 761.

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

(12) 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**, *287*, 15345.

(13) Wang, W.; Perovic, I.; Chittuluru, J.; Kaganovich, A.; Nguyen, L. T. 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.* **2011**, *108*, 17797.

(14) Weinreb, P. H.; Zhen, W.; Poon, A. W.; Conway, K. A.; Lansbury, P. T. *Biochemistry* **1996**, *35*, 13709.

(15) Trexler, A. J.; Rhoades, E. *Protein Sci.* **2012**, *21*, 601.

(16) Binolfi, A.; Theillet, F. X.; Selenko, P. *Biochem. Soc. Trans.* **2012**, *40*, 950.

(17) Ullman, O.; Fisher, C. K.; Stultz, C. M. *J. Am. Chem. Soc.* **2011**, *133*, 19536.

(18) Mukherjea, M.; Llinas, P.; Kim, H.; Travaglia, M.; Safer, D.; Ménétrey, J.; Franzini-Armstrong, C.; Selvin, P. R.; Houdusse, A.; Sweeney, H. L. *Mol. Cell* **2009**, *35*, 305.

(19) Lawson, D. M.; Artymiuk, P. J.; Yewdall, S. J.; Smith, J. M. A.; Livingstone, J. C.; Treffry, A.; Luzzago, A.; Levi, S.; Arosio, P.; Cesareni, G.; Thomas, C. D.; Shaw, W. V.; Harrison, P. M. *Nature* **1991**, *349*, 541.

(20) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.

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

(22) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct. Bioinf.* **1999**, *35*, 133.

(23) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

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

(25) Huang, A.; Stultz, C. M. *PLoS Comput. Biol.* **2008**, *4*, e1000155.

(26) Fisher, C. K.; Ullman, O.; Stultz, C. M. *Pac. Symp. Biocomputing* **2012**, *17*, 82.

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

(28) Zweckstetter, M. *Nat. Protoc.* **2008**, *3*, 679.

(29) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.

(30) Kay, L.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663.

(31) Goddard, T. D.; Kneller, D. G. *Sparky 3*, Mac Intel version; University of California: San Francisco, CA, 2008; http://www.cgl.ucsf.edu/home/sparky/.

(32) Bax, A. *Protein Sci.* **2003**, *12*, 1.

(33) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Marius Clore, G. *J. Magn. Reson.* **2003**, *160*, 65.

(34) *PyMOL*; Schrodinger, LLC: 2010.

(35) Wishart, D.; Bigam, C.; Holm, A.; Hodges, R.; Sykes, B. *J. Biomol. NMR* **1995**, *5*, 67.

(36) Fisher, C. K.; Stultz, C. M. *Curr. Opin. Struct. Biol.* **2011**, *21*, 426.

(37) el-Agnaf, O. M.; Irvine, G. B. *Biochem. Soc. Trans.* **2002**, *30*, 559.

(38) Volles, M. J.; Lee, S. J.; Rochet, J. C.; Shtilerman, M. D.; Ding, T. T.; Kessler, J. C.; Lansbury, P. T., Jr. *Biochemistry* **2001**, *40*, 7812.

(39) Laganowsky, A.; Liu, C.; Sawaya, M. R.; Whitelegge, J. P.; Park, J.; Zhao, M. L.; Pensalfini, A.; Soriaga, A. B.; Landau, M.; Teng, P. K.; Cascio, D.; Glabe, C.; Eisenberg, D. *Science* **2012**, *335*, 1228.

(40) 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.

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

(42) Bartels, T.; Ahlstrom, L. S.; Leftin, A.; Kamp, F.; Haass, C.; Brown, M. F.; Beyer, K. *Biophys. J.* **2010**, *99*, 2116.

(43) Bodner, C. R.; Dobson, C. M.; Bax, A. *J. Mol. Biol.* **2009**, *390*, 775.

(44) Vamvaca, K.; Volles, M. J.; Lansbury, P. T., Jr. *J. Mol. Biol.* **2009**, *389*, 413.

(45) Kang, L.; Moriarty, G. M.; Woods, L. A.; Ashcroft, A. E.; Radford, S. E.; Baum, J. *Protein Sci.* **2012**, *21*, 911.

(46) Manavalan, P.; Johnson, W. C. *J. Biosci.* **1985**, *8*, 141.

(47) Greenfield, N. J. *Nat. Protoc.* **2006**, *1*, 2876.

(48) Iwai, A.; Yoshimoto, M.; Masliah, E.; Saitoh, T. *Biochemistry* **1995**, *34*, 10139.

(49) Jarrett, J. T.; Lansbury, P. T. *Cell* **1993**, *73*, 1055.

(50) Dudzik, C. G.; Walter, E. D.; Millhauser, G. L. *Biochemistry* **2011**, *50*, 1771.

(51) Sandal, M.; Valle, F.; Tessari, I.; Mammi, S.; Bergantino, E.; Musiani, F.; Brucale, M.; Bubacco, L.; Samorì, B. *PLoS Biol.* **2008**, *6*, e6.