Structure
Article

Predictive Atomic Resolution Descriptions of Intrinsically Disordered hTau40 and α-Synuclein in Solution from NMR and Small Angle Scattering

Martin Schwalbe,[1,2,9] Valéry Ozenne,[3,4,5,9] Stefan Bibow,[1] Mariusz Jaremko,[1] Lukasz Jaremko,[1] Michal Gajda,[1] Malene Ringkjøbing Jensen,[3,4,5] Jacek Biernat,[6,7] Stefan Becker,[1] Eckhard Mandelkow,[6,7] Markus Zweckstetter,[1,2,8,*] and Martin Blackledge[3,4,5,*]

[1]Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany
[2]German Center for Neurodegenerative Diseases (DZNE), 37077 Göttingen, Germany
[3]University Grenoble Alpes, Protein Dynamics and Flexibility, Institut de Biologie Structurale, 38000 Grenoble, France
[4]CNRS, Protein Dynamics and Flexibility, Institut de Biologie Structurale, 38000 Grenoble, France
[5]CEA, DSV, Protein Dynamics and Flexibility, Institut de Biologie Structurale, 38000 Grenoble, France
[6]CEASAR Research Center, 53175 Bonn, Germany
[7]German Center for Neurodegenerative Diseases (DZNE), 53175 Bonn, Germany
[8]Center for the Molecular Physiology of the Brain, University Medical Center, 37073 Göttingen, Germany
[9]These authors contributed equally to this work
*Correspondence: mazw@nmr.mpibpc.mpg.de (M.Z.), martin.blackledge@ibs.fr (M.B.)
http://dx.doi.org/10.1016/j.str.2013.10.020

SUMMARY

The development of molecular descriptions of intrinsically disordered proteins (IDPs) is essential for elucidating conformational transitions that characterize common neurodegenerative disorders. We use nuclear magnetic resonance, small angle scattering, and molecular ensemble approaches to characterize the IDPs Tau and α-synuclein. Ensemble descriptions of IDPs are highly underdetermined due to the inherently large number of degrees of conformational freedom compared with available experimental measurements. Using extensive cross-validation we show that five different types of independent experimental parameters are predicted more accurately by selected ensembles than by statistical coil descriptions. The improvement increases in regions whose local sampling deviates from statistical coil, validating the derived conformational description. Using these approaches we identify enhanced polyproline II sampling in aggregation-nucleation sites, supporting suggestions that this region of conformational space is important for aggregation.

INTRODUCTION

During the last decade it has become widely recognized that intrinsic protein conformational disorder is common in eukaryotic proteomes and that this disorder is involved in numerous essential cellular processes such as signaling and regulation (Uversky, 2002; Tompa, 2002; Dyson and Wright, 2005). The lack of stable secondary and tertiary structure, either over the entire protein length or only in localized regions, imparts intrinsically disordered proteins (IDPs) with a structural plasticity that is essential for their cellular function and that is not achievable by globular, ordered proteins (Fink, 2005; Dunker et al., 2008; Tompa and Fuxreiter, 2008).

IDPs have also been the focus of research into the molecular basis of numerous amyloid-based conformational disorders, including major human neurodegenerative diseases such as Alzheimer and Parkinson, characterized by conformational transitions from intrinsic disorder in the soluble form to insoluble aggregates of the same proteins (Carrell and Lomas, 1997; Dobson, 1999; Bucciantini et al., 2002; Uversky and Fink, 2004). The molecular basis of these transitions remains poorly understood, not only because the trigger that initiates aggregation is inherently difficult to study but also because relatively few details of the conformational behavior of the soluble forms are known. In response to this challenge, there has been considerable interest in the conception and application of experimental and analytical techniques to describe intrinsic conformational disorder (Mittag and Forman-Kay, 2007; Tompa, 2011).

Nuclear magnetic resonance (NMR) spectroscopy has emerged as the most promising tool for the characterization of unfolded proteins at amino acid resolution, and even atomic resolution (Dyson and Wright, 2004). The interpretation of NMR observables in terms of ensembles has required the development of specific procedures that account for the particular averaging behavior of each independent parameter, providing information about both local and long-range conformational behavior (Meier et al., 2008; Mittag et al., 2010). Solution state small angle X-ray scattering (SAXS) provides complementary information concerning the volume space sampled by IDPs (Sibille and Bernadó, 2012). Recent methodological developments of representative ensembles have concentrated on restrained or replica-exchange molecular dynamics simulation, or sample-and-select approaches whereby subensembles are derived from a broader distribution (Allison et al., 2009; Fisher and Stultz, 2011; Jensen et al., 2013; Mittag and Forman-Kay, 2007).

One of these approaches exploits flexible-meccano (Bernadó et al., 2005b; Ozenne et al., 2012a), a statistical coil generator, to flood conformational space as broadly as possible, and uses the genetic algorithm ASTEROIDS to select equivalent ensembles in agreement with experimental data within the level of experimental uncertainty (Jensen et al., 2010; Nodet et al., 2009; Salmon et al., 2010). A recent study was used to determine the ability of combinations of NMR chemical shifts (CSs) and residual dipolar couplings (RDCs) to map local potential energy surfaces on an amino acid-specific basis with special emphasis on the ability to distinguish populations of β strand (βS), polyproline II (βP), and α-helical conformations (Ozenne et al., 2012b). In this study we apply these approaches, using extensive experimental data, including ^{13}C, ^{15}N, and ^{1}H CSs; paramagnetic relaxation enhancements (PREs); $^{1}D_{HN}$, $^{2}D_{CHN}$, and $^{4}D_{HNiHai-1}$ RDCs; and SAXS, to describe the conformational behavior of full-length Tau (hTau40) and α-synuclein (αSyn) in solution. Both of these proteins are characterized by conformational transitions from soluble, physiological proteins to insoluble forms of the same protein, and in both cases these transitions are associated with onset of neurodegenerative disease.

Although αSyn has been previously characterized using NMR (Bernadó et al., 2005a; Bertoncini et al., 2005b; Cho et al., 2009; Dedmon et al., 2005; Eliezer et al., 2001; Marsh et al., 2006; Salmon et al., 2010; Ullman et al., 2011; Wu et al., 2008), a recent ensemble description of the protein derived from NMR data in its soluble form was used to suggest a conformational equilibrium between higher order oligomers and monomers in solution (Gurry et al., 2013). Here we apply the ASTEROIDS approach to develop an ensemble description of αSyn from solution-state NMR and scattering data to determine local and long-range conformational sampling propensities, but we also investigate whether a monomeric ensemble can reproduce available experimental data. hTau40 is a 441-amino acid protein and therefore is the largest protein so far described at atomic resolution using ensemble approaches, and as such poses several methodological challenges that are addressed here.

It is widely understood that ensemble descriptions of IDPs are likely to be underdetermined, due to the large number of degrees of conformational freedom compared with independent experimental measurements. It is therefore essential to calibrate experimental procedures used to define representative ensembles. We have recently published a detailed study (Ozenne et al., 2012b) that provided guidelines describing how accurately different combinations of CSs and RDCs can be expected to map backbone dihedral angle distributions when used in combination with ASTEROIDS. The current study draws from these guidelines, using combinations of RDCs and CSs that, according to calibration, should be able to delineate the regions of Ramachandran space. We place particular emphasis on cross-validation of ensembles of IDPs derived from experimental NMR data, showing that five different types of independent experimental parameters are predicted more accurately by selected ensembles than by statistical coil descriptions. The improvement increases in regions whose local sampling deviates from statistical coil, validating the derived conformational description, and demonstrating the predictive capacity of the ASTEROIDS approach.

The ASTEROIDS ensembles verify that both proteins are highly disordered in solution with weak long-range contacts between distant regions of the peptide chain, in both cases driven by electrostatic interactions. Intriguingly, however, we find a higher propensity for sampling of the βP region of conformational space than is present in common statistical coil models, and in particular this sampling is localized in the vicinity of the aggregation-nucleation sites of the proteins. This observation supports the hypothesis that population of the βP region of conformational space represents a precursor for aggregation and formation of stable β sheets.

RESULTS

Combined SAXS/NMR Ensemble Description of αSyn

$^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, ^{15}N, and $^{1}H^{N}$ CSs; $^{1}D_{HN}$ RDCs; SAXS; and four PRE data sets were combined with flexible-meccano sampling and ASTEROIDS selection to define the conformational sampling of αSyn (Figure 1). αSyn serves as a well-studied IDP against which to test the ability of ensemble descriptions to simultaneously reproduce diverse data types while predicting independent data sets.

The combination of $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, ^{15}N, and $^{1}H^{N}$ CSs and $^{1}D_{HN}$ has been shown to unambiguously distinguish populations of βP, βS, and right- (αR) and left (αL)-handed α-helical regions (Ozenne et al., 2012b). Populations of the different regions of Ramachandran space determined from the analysis of αSyn are shown in Figure 2. Most noticeably, in comparison with the statistical coil sampling (black line), the population of the βP region is significantly and continuously elevated in numerous segments, in particular residues 76–81, which forms the central region of the nonamyloid-β component of the amyloid plaque region. The enhanced βP propensity is accompanied by a generally reduced propensity for βS compared with the statistical coil.

The previously described dependence of RDCs on the long-range contact present between the C and N termini is again noticeable and is not commented on further here (Allison et al., 2009; Bernadó et al., 2005a; Bertoncini et al., 2005a). We note that the effective radius of gyration is higher when SAXS data are incorporated in the selection (Figure 2). This is to be expected, because none of the other data are unambiguously sensitive to the more extended conformations that contribute uniquely to SAXS.

The predictive nature of the flexible-meccano/ASTEROIDS approach was investigated by removing data from the analysis, in this case $^{1}D_{HN}$ RDCs, and calculating these values from the conformational sampling derived from the chemical shifts alone. As shown in Figure 1 and Table 1, these independent data are well reproduced by the local and long-range conformational ensemble determined from CSs, SAXS, and PRE data (nearly a factor of two better than the statistical coil; Table 1). To test whether this improvement derives from regions whose conformational sampling is identifiably different from the starting pool, we have compared the overall reproduction of subsets of residues for which Δ_{Rama}, a metric of the difference in local sampling between the selected ensemble and the statistical coil pool (see Experimental Procedures), is higher than 0.05 (Table 1). Although the χ^2 falls in both cases, the effect is more significant for the residues for which the local conformational sampling changes the most compared to the statistical coil.

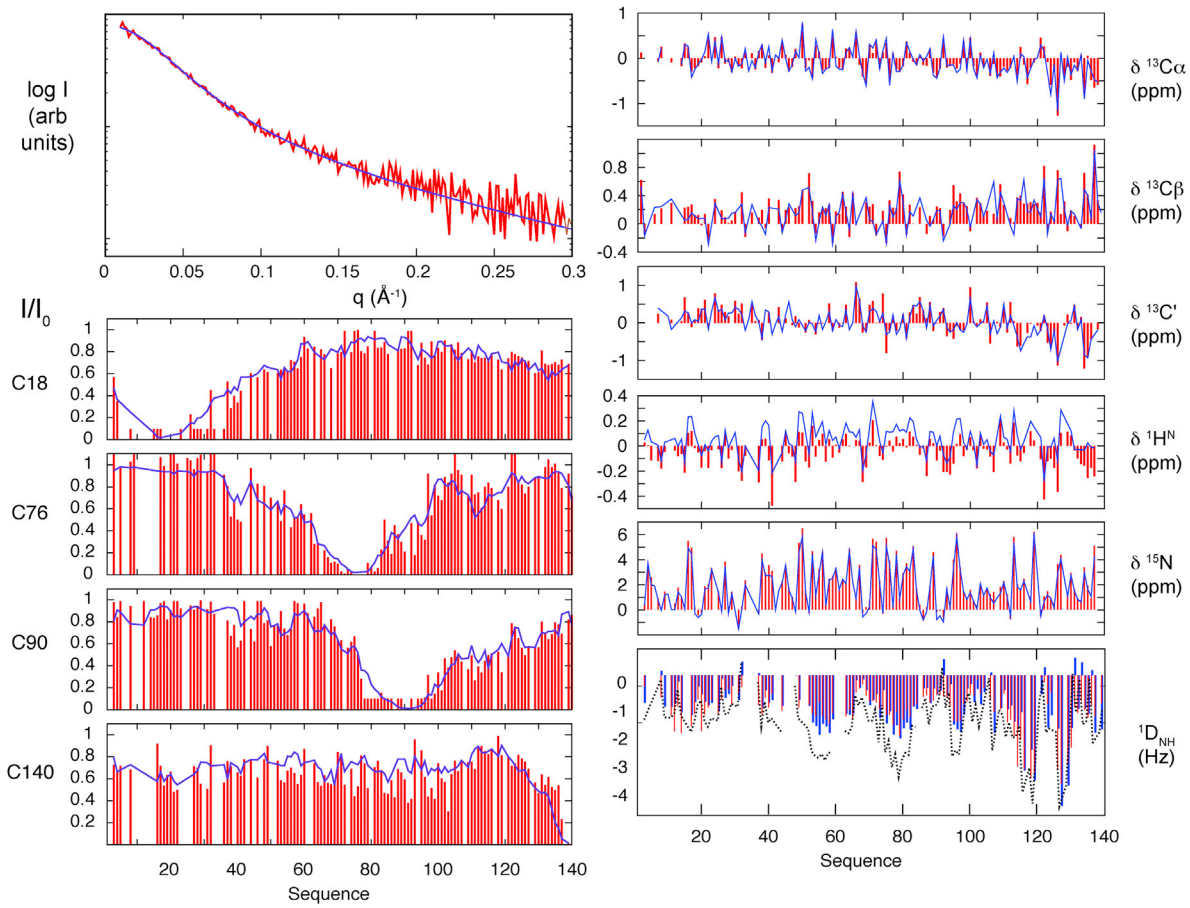

Figure 1. ASTEROIDS Reproduction of Experimental Data from αSyn

Top left: experimental (red) and calculated (blue) SAXS data.

Bottom left: experimental and calculated PREs.

Top right: experimental and calculated CSs.

Bottom right: experimental RDCs (red) and RDCs calculated from ASTEROIDS selection using SAXS, PREs, and CSs in the absence of RDCs (blue) and RDCs calculated using the statistical coil model (no selection) to describe the conformational sampling (dotted line).

ASTEROIDS Description of the K32 Construct of Tau Protein

Next, we applied the flexible-meccano/ASTEROIDS analysis to $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, ^{15}N, and $^{1}H^{N}$ CSs as well as $^{1}D_{HN}$, $^{2}D_{CHN}$, and $^{4}D_{HNiHai-1}$ RDCs of the 198-residue Tau fragment K32 (Figure 3). K32 encompasses the four repeat sequences that include the microtubule binding and aggregation-nucleation sites of Tau, as well as the proline-rich region (residues 198–240) (Mukrasch et al., 2007b). The structural propensities in the proline-rich region are particularly interesting because it is targeted by several kinases and is essential for efficient Tau-driven assembly of tubulin into microtubuli (Preuss et al., 1997).

Access to $^{1}D_{HN}$, $^{2}D_{CHN}$, and $^{4}D_{HNiHai-1}$ RDCs and five different types of CS allows for a more rigorous validation of the ensembles. Again following the recent calibration study identifying combinations of experimental data that accurately delineate sampling of Ramachandran space (Ozenne et al., 2012b), $^{2}D_{CHN}$ and $^{4}D_{HNiHai-1}$ RDCs and ^{15}N and $^{1}H^{N}$ CSs were separately removed from the selection procedure and back-calculated from the selected ensembles (Figure 4 and Table 1). The reproduction of both types of RDCs and CSs is better in comparison with the statistical coil (Table 1), with $^{1}H^{N}$ chemical shifts and $^{4}D_{HNiHai-1}$ RDCs showing the most significant improvements. When only residues for which Δ_{Rama} is higher than 0.05 are included, this differential is maintained or improved (Table 1). The analysis again demonstrates the generally predictive capacity of the flexible-meccano/ASTEROIDS ensemble selection of IDPs.

Amino acid-specific conformational sampling is presented in Figure 5 in terms of populations of four broad regions of Ramachandran space. In agreement with studies of a shorter construct (K18), where chemical shifts and $^{1}D_{HN}$ were used to define the conformational sampling (Mukrasch et al., 2007a; Ozenne et al., 2012b), the four glycine-rich sequences in the repeat domain, as well as four type I β-turns, have significantly increased αR population (Figure 5). Interestingly, higher populations of βP are observed in between residues (275–282), (307–313) and (338–346); the two former regions comprise hexapeptides that have been identified as aggregation-nucleation sites. In line with the high abundance of prolines within residues

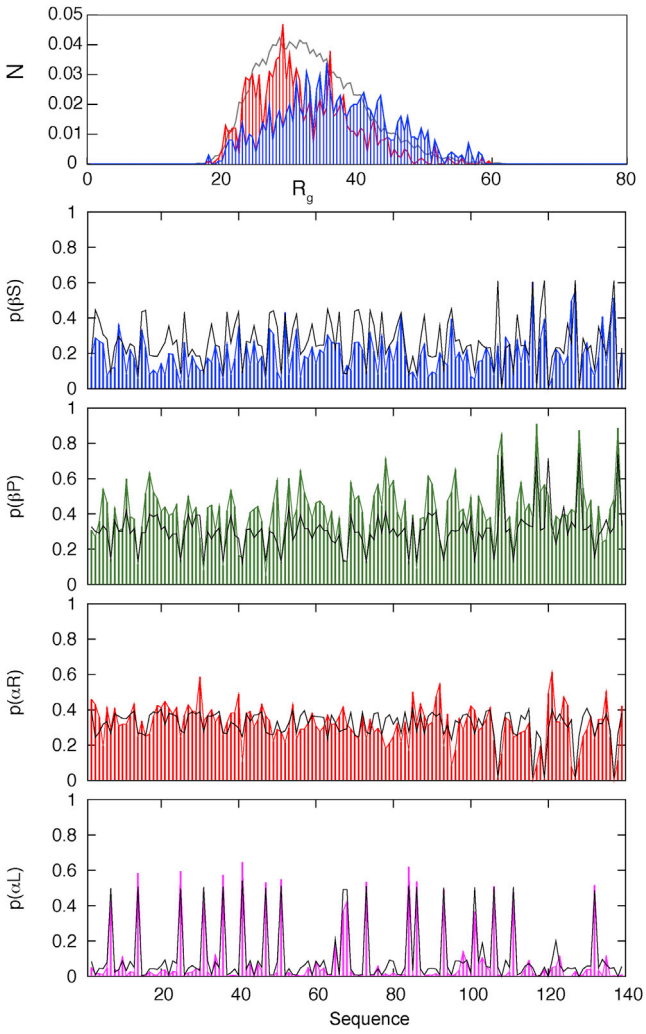

Figure 2. Conformational Propensities in αSyn Derived from RDCs, CSs, PREs, and SAXS

Top: distribution of radius of gyration (Å) in statistical coil pool (black), in the SAXS+RDC+PRE+CS-selected ensemble (blue), and in the PRE-only selected ensemble (red).

Bottom four panels: local conformational propensities. Black lines indicate proportion of conformers in the statistical coil pool. Colored bars indicate the population level in ensembles selected on the basis of the experimental data. See Experimental Procedures for definitions of different regions.

212–232, the levels of βP sampling were up to 20% higher than expected from the statistical coil. In proline-containing segments the population of βP even exceeded 60% (Figure 5).

Elevated Population of βP Conformations in 441-Residue Tau Protein

Backbone resonances of the 441-residue isoform of the Tau protein, hTau40, the longest Tau isoform found in the human brain, have been fully assigned, providing the basis for a detailed analysis of its conformational properties (Mukrasch et al., 2009). Here we incorporated different backbone and $^{13}C^{\beta}$ CSs in combination with $^{1}D_{HN}$ RDCs, SAXS data, and PREs from 12 different cysteine mutants into a single flexible-meccano/ASTEROIDS ensemble selection of hTau40 (Figure 6). Selection of representative ensembles against such a large molecule, comprising 441-amino acids constrained by more than 2,500 PRE-derived peak intensity ratios, more than 2,000 chemical shifts, as well as RDCs and SAXS data, challenges the conformational ensemble selection by the ASTEROIDS algorithm. The optimal number of structures in the selection was estimated from cross-validation procedures to be approximately 400, so that each selection requires the averaging of nearly three million simulated observables. To achieve convergence, 100,000 iterations of the genetic algorithm were required (3×10^{11} operations).

Figure 7 depicts conformational sampling as a function of the four regions of Ramachandran space. Within the common 198 amino acids between K32 and full-length Tau, the local conformational sampling described above is reproduced almost identically in the full-length protein (chemical shifts are superimposable in the two constructs except for three to five residues in the immediate vicinity of the N and C termini). For the remainder of the protein, two regions with significant helical propensity are clearly visible (II and IX in Figure 7), as well as two shorter turn regions around residues 138 and 146, containing a high concentration of charged residues. Apart from these helix/turn motifs, the most marked observation concerns the elevated population of βP conformations at specific regions in the protein (I, IV, V, and VI). Although the presence of proline residues throughout the protein is clearly correlated with seeding of elevated βP populations, there are notable exceptions, where elevated βP exists in the absence of vicinal prolines, such as regions III, VI, and VIII. In line with results obtained for K18 and K32, both microtubule binding and aggregation-nucleation hexapeptides (VI and VII) present elevated βP populations, and they are followed by β-turns. Besides the description of local structural propensities, the flexible-meccano/ASTEROIDS ensemble description identifies the transient long-range contacts that characterize the Tau protein (Figure 8). Analysis of the long-range contacts in the ensembles not only reveals the long-range contact between the N terminus and the proline-rich region (Bibow et al., 2011; Mukrasch et al., 2009) but also points to a more extended nature of the repeat domain (240–370). Again comparison of distribution of radii of gyration over ensembles selected with and without SAXS data underlines the complementary nature of SAXS and PRE data (Figure 8).

To validate the ensemble of hTau40 conformers, $^{1}D_{NH}$ RDCs were again left out of an entire analysis and compared with prediction using the ensemble selected by the remaining experimental constraints (Table 1; Figure 7). Reproduction of experimental RDCs from the selected ensemble was more than a factor two better than those predicted from the statistical coil. Gradual removal of residues that have similar conformational sampling to the statistical coil again significantly increases the differential between back-calculated χ^2 for ASTEROIDS selections compared with the statistical coil (Table 1; Figure S2 available online). This is further supported by Figure 7, where the differences between statistical coil and ASTEROIDS selection derived from CS, SAXS, and PREs are shown to coincide with regions for which RDCs are better predicted by the selection than the statistical coil (for the regions I, IV, V, and VI showing a higher level of βP ASTEROIDS-predicted RDCs are closer to the experimental values). This provides strong, independent

Table 1. Testing Predictive Nature of ASTEROIDS Ensembles by Prediction of Independent Observables

	Active[a]	Passive[b]	$\chi^2_{CV,AST}$[c]	$\chi^2_{CV,COIL}$[d]	$\chi^2_{CV,AST,\Delta Rama}$[e]	$\chi^2_{CV,COIL,\Delta Rama}$[f]
hTau40	$\Delta(^{13}C^{\alpha}, ^{13}C^{\beta}, ^{13}C', ^{15}N, ^{1}H)$, SAXS, PRE	$^{1}D_{HN}$	1.05 (228)[g]	2.02 (228)	1.23 (171)[h]	3.63 (171)
					1.32 (59)	4.45 (59)
					0.35 (20)	6.4 (20)
K18	$\Delta(^{13}C^{\alpha}, ^{13}C^{\beta}, ^{13}C', ^{15}N, ^{1}H)$	$^{1}D_{HN}$	1.68 (130)	3.20 (130)	0.98 (34)	6.15 (34)
K32	$\Delta(^{13}C^{\alpha}, ^{13}C^{\beta}, ^{13}C', ^{15}N, ^{1}H)$, $^{1}D_{HN}$	$^{2}D_{CHN}$	1.42 (168)	1.81 (168)	1.45 (47)	2.28 (47)
K32	$\Delta(^{13}C^{\alpha}, ^{13}C^{\beta}, ^{13}C', ^{15}N, ^{1}H)$, $^{1}D_{HN}$	$^{4}D_{HNHa}$	1.93 (139)	3.00 (139)	2.70 (37)	3.98 (37)
K32	$\Delta(^{13}C^{\alpha}, ^{13}C^{\beta}, ^{13}C')$, $^{1}D_{HN}$	$\Delta(^{1}H)$[i]	1.87 (167)	2.49 (167)	1.66 (44)	2.49 (44)
K32	$\Delta(^{13}C^{\alpha}, ^{13}C^{\beta}, ^{13}C')$, $^{1}D_{HN}$	$\Delta(^{15}N)$	2.00 (167)	2.25 (167)	2.04 (44)	2.18 (44)
K32	$\Delta(^{13}C^{\alpha}, ^{13}C^{\beta}, ^{13}C')$, $^{1}D_{HN}$, $^{2}D_{CHN}$	$\Delta(^{1}H)$	1.73 (167)	2.49 (167)	1.21 (44)	2.49 (44)
K32	$\Delta(^{13}C^{\alpha}, ^{13}C^{\beta}, ^{13}C')$, $^{1}D_{HN}$, $^{2}D_{CHN}$	$\Delta(^{15}N)$	1.99 (167)	2.25 (167)	1.97 (44)	2.18 (44)
αSyn	$\Delta(^{13}C^{\alpha}, ^{13}C^{\beta}, ^{13}C', ^{15}N, ^{1}H)$, SAXS, PRE	$^{1}D_{HN}$	0.44 (104)	0.91 (104)	0.29 (41)	0.80 (41)

See also Figure S2.

[a]Experimental data used in the ASTEROIDS ensemble selection.

[b]Experimental data against which the validity of the ensemble is tested (not included in the selection).

[c]Reduced χ^2 of passive data set compared with the ASTEROIDS selection.

[d]Reduced χ^2 of passive data set compared with the statistical coil description of the protein.

[e]Reduced χ^2 of passive data set compared with the ASTEROIDS selection. Amino acids are only included if the local potential energy function of the residue is different from statistical coil ($\Delta_{Rama} > \Delta_{Rama,threshhold}$).

[f]Reduced χ^2 of passive data set compared with the statistical coil description of the protein. Amino acid selection as in footnote e.

[g]Number in parentheses refers to number of residues included in selection.

[h]Number in parentheses refers to number of residues included in selection; this falls as Δ_{Rama} is more selective.

[i]χ^2 calculated with uncertainty of 0.1 ppm for ^{1}H CSs and 1.0 ppm for ^{15}N CSs.

evidence that the conformational description is meaningful, and predictive.

Finally, the agreement between experiment and prediction deteriorates slightly if the underlying modulation of locally calculated RDCs, due to the long-range contact between the N terminus of Tau and the proline-rich region (212–232), is not included in the analysis, reinforcing previous observations about the importance of long-range interactions in experimentally observed RDC profiles (Salmon et al., 2010; Schneider et al., 2012).

DISCUSSION

The development of molecular descriptions of intrinsically disordered proteins on the basis of experimental data poses obvious questions of underdetermination and sampling bias. Our approach to the interpretation of experimental data from IDPs exploits two distinct algorithms. Flexible-meccano (Bernadó et al., 2005b; Ozenne et al., 2012a) samples the phase space available to a statistical coil model of the protein using amino acid-specific conformational energy potentials, combined with simple steric exclusion. The experimental data are not used at this stage, the aim being to sample the full energy landscape accessible for a statistical coil model for the specific primary sequence with negligible introduction of bias either from data or molecular dynamics force fields. The experimental data are then used in a second step, performed by the ASTEROIDS algorithm (Jensen et al., 2010; Nodet et al., 2009), to define conformational sampling of ensemble space that verifies experimental data.

We have combined NMR, SAXS, and the flexible-meccano/ASTEROIDS approach to characterize the local and long-range structure of two IDPs, demonstrating that both proteins are highly disordered in solution. In the case of αSyn, this result stands in contrast to recent descriptions of the protein, where the presence of large, multimeric conformational heteromers was proposed to constitute up to 25% of the ensemble of structures present in solution (Gurry et al., 2013; Wang et al., 2011). We find no evidence to support the presence of such large-molecular-weight assemblies, neither from NMR spectroscopy, where they would presumably give rise to negligible signal in comparison with flexible monomers, nor from SAXS where their presence should also be detectable.

The combination of SAXS and PREs in a single ensemble description reconciles two apparently contradictory observations regarding the presence of long-range contacts in the context of the generally extended nature of the Tau protein in solution. With respect to an entirely unfolded statistical coil description of Tau, the pairwise distribution of interresidue contacts derived from experiment demonstrates the presence of both compaction (due to the contact between the N terminus and the proline-rich region) and a more extended region containing the four repeats comprising the K18 construct, in line with previously published SAXS data (Mylonas et al., 2008).

Importantly, the analysis presented here exploits combinations of experimental constraints that were recently calibrated against simulation to accurately determine local conformational sampling (Ozenne et al., 2012b). Although CSs and RDCs predominantly report on local order and PREs and SAXS mainly report on transient intrachain contacts and sampling of the volume space by the unfolded chain, all parameters are to some extent dependent on both local and long-range effects. It is therefore preferable to combine all experimentally measurable constraints in a single ensemble, if only to reduce the ambiguity

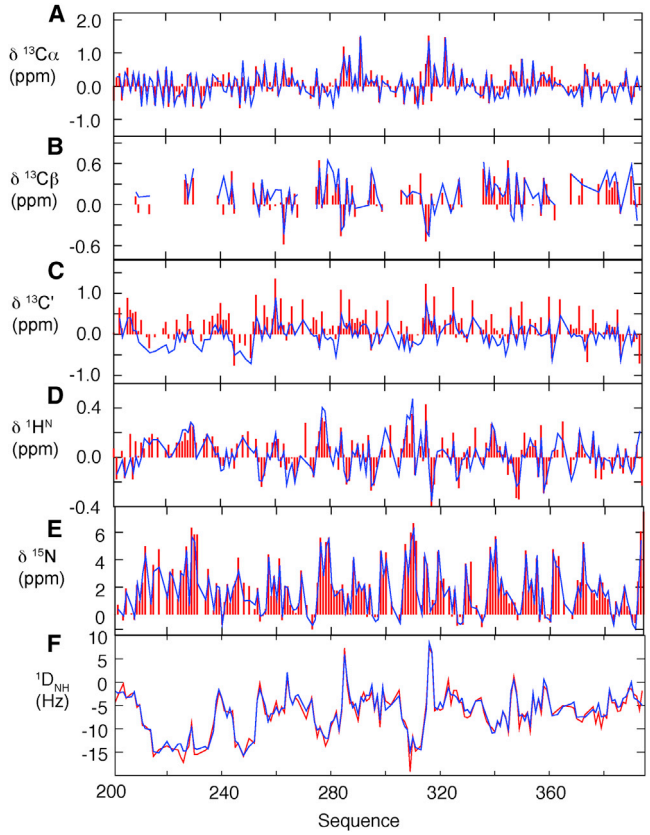

Figure 3. Reproduction of Experimental Data from K32 when Included as Active Data in the ASTEROIDS Target Selection Function
(A–E) Experimental (red bars) and calculated (blue lines) CS data.
(F) Experimental RDCs (red) and values calculated from ASTEROIDS selection (blue).

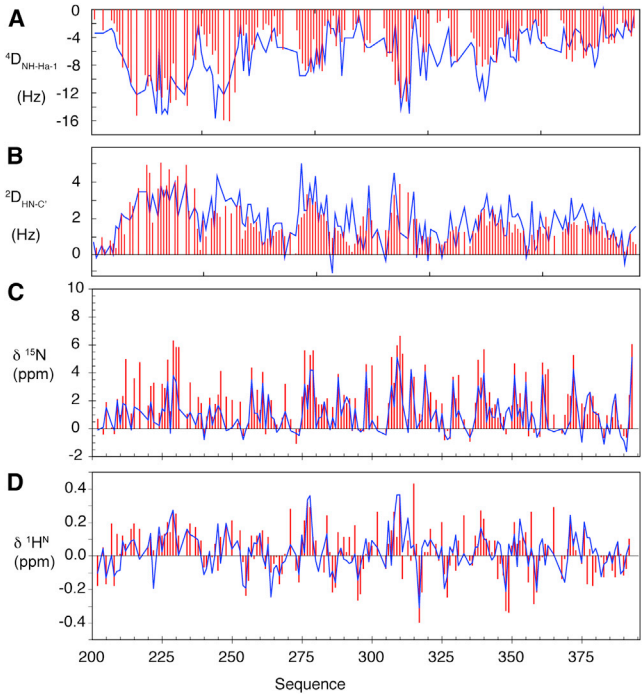

Figure 4. Cross-Validation of K32 Ensemble
Reproduction of experimental (red) data from K32 using ASTEROIDS ensemble selection (blue) when not included as active data in the ASTEROIDS target selection function.
(A) Experimental and calculated $^{4}D_{HNiHai-1}$ RDCs from selection using $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, ^{15}N, and $^{1}H^{N}$ CSs and $^{1}D_{HN}$ RDCs.
(B) Experimental and calculated $^{2}D_{HNC'}$ RDCs from selection using $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, ^{15}N, and $^{1}H^{N}$ CSs and $^{1}D_{HN}$ RDCs.
(C) Experimental and calculated ^{15}N CSs from selection using $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ CSs and $^{1}D_{HN}$ and $^{2}D_{HNC'}$ RDCs.
(D) Experimental and calculated $^{1}H^{N}$ CSs from selection using $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ CSs and $^{1}D_{HN}$ and $^{2}D_{HNC'}$ RDCs.

inherent in the interpretation of the available experimental data. The resulting ensemble descriptions can be checked for consistency by removing data from the selection procedure to analyze the predictive nature of the approach.

A total of nine independent sets of five different types of experimental parameter ($^{1}D_{HN}$, $^{2}D_{CHN}$, and $^{4}D_{HNiHai-1}$ RDCs and $^{1}H^{N}$ and ^{15}N CSs from K18, K32, hTau40, and αSyn) were predicted from ASTEROIDS selections, using combinations of active experimental data that have been identified to accurately map Ramachandran space of each amino acid. For the 198-amino acid construct of K32, this allows for an extensive level of cross-validation that supports previous simulation-based prediction (Ozenne et al., 2012b) concerning the ability of ASTEROIDS to probe the free-energy landscape of IDPs. Critically, we find that for K18, K32, full-length Tau, and αSyn, the different RDCs and CSs are better predicted by the ASTEROIDS selection than by the general statistical coil model (Table 1 and Figures 4 and 7). This is already a significant result, because in all cases the majority of the protein does not exhibit a significantly different behavior from statistical coil. If the higher accuracy of the prediction using ASTEROIDS derives from a real improvement in the conformational description, then this difference should increase when amino acids showing statistical coil-like conformational distributions are removed from the selection. This is indeed the case (Table 1), most notably for hTau40, K18 $^{1}D_{HN}$, K32 $^{2}D_{CHN}$, and $^{4}D_{HNiHai-1}$ RDCs and K32 ^{1}H CSs where the increase in the reduced χ^2 is remarkable when focusing on regions exhibiting significantly different behavior compared to statistical coil. These cross-validation results provide quantifiable evidence that the identified local conformational details reported by the ASTEROIDS ensemble are genuine and accurate and therefore unlikely to derive from noise or poor conformational sampling. The results also stand in contrast to molecular dynamics-based approaches, where RDCs appear to be difficult to reproduce from trajectories of IDPs unless they are explicitly included in a restrained molecular dynamics target function (Ball et al., 2011; Sgourakis et al., 2011).

The analysis developed here uses PREs, SAXS, and a total of eight types of CS or RDC that each have unique conformational dependences. Nevertheless, there is clear potential for improvement of this kind of ensemble description with other solution approaches. Although we have not directly cross-validated our ensembles against other data, comparison of our results on αSyn and Tau with descriptions derived from Förster resonance energy transfer (Jeganathan et al., 2006; Nath et al.,

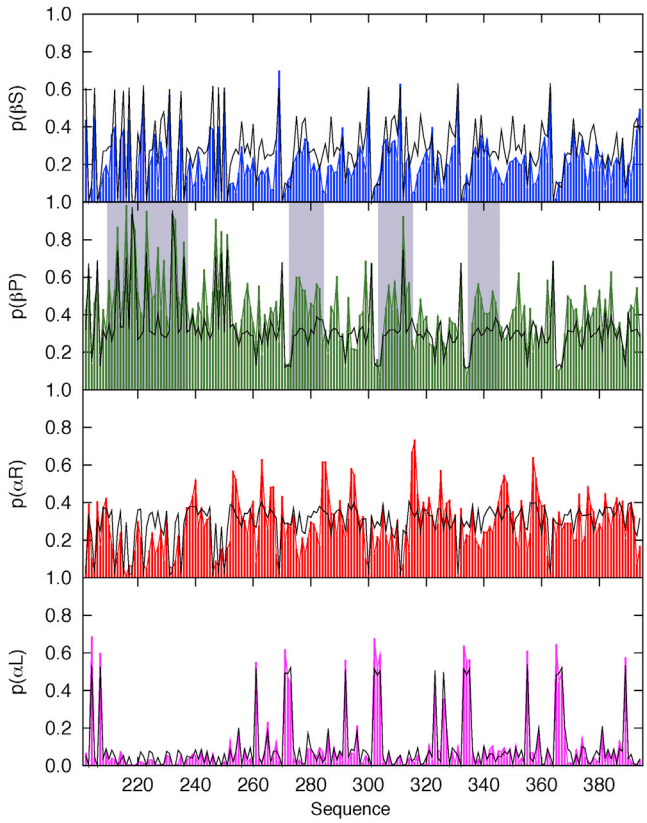

Figure 5. Conformational Propensities in K32 Derived from RDCs and CSs

Color scheme as for Figure 2. The proline-rich region (212–232) and regions (275–282), (307–313), and (338–346) exhibit elevated βP propensity (shading).

2012), electron paramagnetic resonance (Jeganathan et al., 2006), and mass spectrometry (Bernstein et al., 2004) are in broad general agreement in terms of long-range structure. Combination of SAXS and PRE data provides clear complementarity, defining both local close contacts and pairwise distance distributions, both of which are essential for an accurate description of long-range behavior.

Local conformational sampling is nonuniform along the sequence. A key finding of our study is that both IDPs exhibit enhanced propensity to sample the βP region of Ramachandran space compared with the statistical coil. We note that the description that is used here closely resembles alternative statistical coil libraries (Fitzkee et al., 2005; Jha et al., 2005; Smith et al., 1996; Swindells et al., 1995), bearing the same local sampling features. The raised propensities observed in αSyn and Tau are often found to be locally continuous, even though the ensemble selection treats each amino acid independently. Our analysis therefore suggests a "seeding" mechanism whereby the probability of βP is energetically enhanced by cooperative effects. We also found segments that do not contain proline residues that exhibit enhanced βP propensities, including regions that are involved in the initial steps of aggregation.

There is a long history of the implication of βP in denatured or unstructured peptides. Almost 40 years ago Tiffany and Krimm (1969) proposed that denatured proteins are not ensembles of random coils, but rather consist of short βP helix-like stretches interspersed with bends. This idea was corroborated by various groups demonstrating the propensity of unfolded/disordered proteins to adopt βP helix-like conformations (Creamer and Campbell, 2002; Pappu and Rose, 2002; Shi et al., 2006; Woody, 2009; Zagrovic et al., 2005). The propensity of unfolded proteins to adopt a βP helix-like conformation has been rationalized not only by a stabilization of the βP helix through favorable interactions of the backbone with the solvent but also with the minimization of electrostatic repulsion or the formation of side chain-to-backbone hydrogen bonds (Shi et al., 2002). It was also previously proposed that βP represents a "killer conformation" in amyloidogenic proteins, facilitating the transition to βS (Blanch et al., 2000), and eventually amyloid fibril formation.

Until recently, βP conformations in IDPs could only be identified through general conformational methods such as ultraviolet circular dichroism (UVCD) (Eker et al., 2003), vibrational circular dichroism (VCD), Raman optical activity (ROA) (McColl et al., 2004; Syme et al., 2002), or Fourier transform infrared spectroscopy (Mästle et al., 1995). In particular CD and related techniques designated the proteins Tau and αSyn (Schweers et al., 1994; Weinreb et al., 1996) as having predominantly random coil structure. Although quantitative analysis is possible from peptides and model systems (Keiderling and Xu, 2002; Krimm and Tiffany, 1974; Schweitzer-Stenner, 2013), we are not aware of quantification of the polyproline content in αSyn or Tau by VCD and UVCD. A detailed ROA analysis of αSyn proposed a global mixture of α helix, β sheet, and βP with obvious parallels to the site-specific analysis described here (Maiti et al., 2004). Homonuclear NMR, using nuclear Overhauser effect spectroscopy or ^{3}J coupling constants (Shi et al., 2002), have been used to invoke the general presence of βP in unfolded peptides, but evidence was not entirely unambiguous (Makowska et al., 2006). The recently developed approaches used here, based on CSs and RDCs, seem to provide the promise of a more quantitative analysis of the population of βP in IDPs.

Conclusions

In conclusion, we have presented amino acid-specific ensemble descriptions of the intrinsically disordered amyloidogenic polypeptides involved in human neurodegenerative disease, αSyn and Tau. Both long-range and short-range conformational sampling was described on the basis of CSs, RDCs, PREs, and SAXS. The ensemble descriptions were systematically tested for consistency with independent experimental data that were not incorporated into the analysis, providing a quantitative measure of the predictive capacity of conformational ensembles. We believe that the systematic observation of improvement in the accuracy of different types of back-calculated RDCs relative to statistical coil prediction provides an unprecedented test of the predictive nature of an ensemble-based approach to IDP description.

We provide ensemble descriptions of Tau and αSyn that account for local conformational propensity and transient long-range contacts and that also describe the volume space sampled by the entire unfolded chain. The analysis revealed that the region of Ramachandran space corresponding to polyproline II is locally favored in both proteins, in particular in

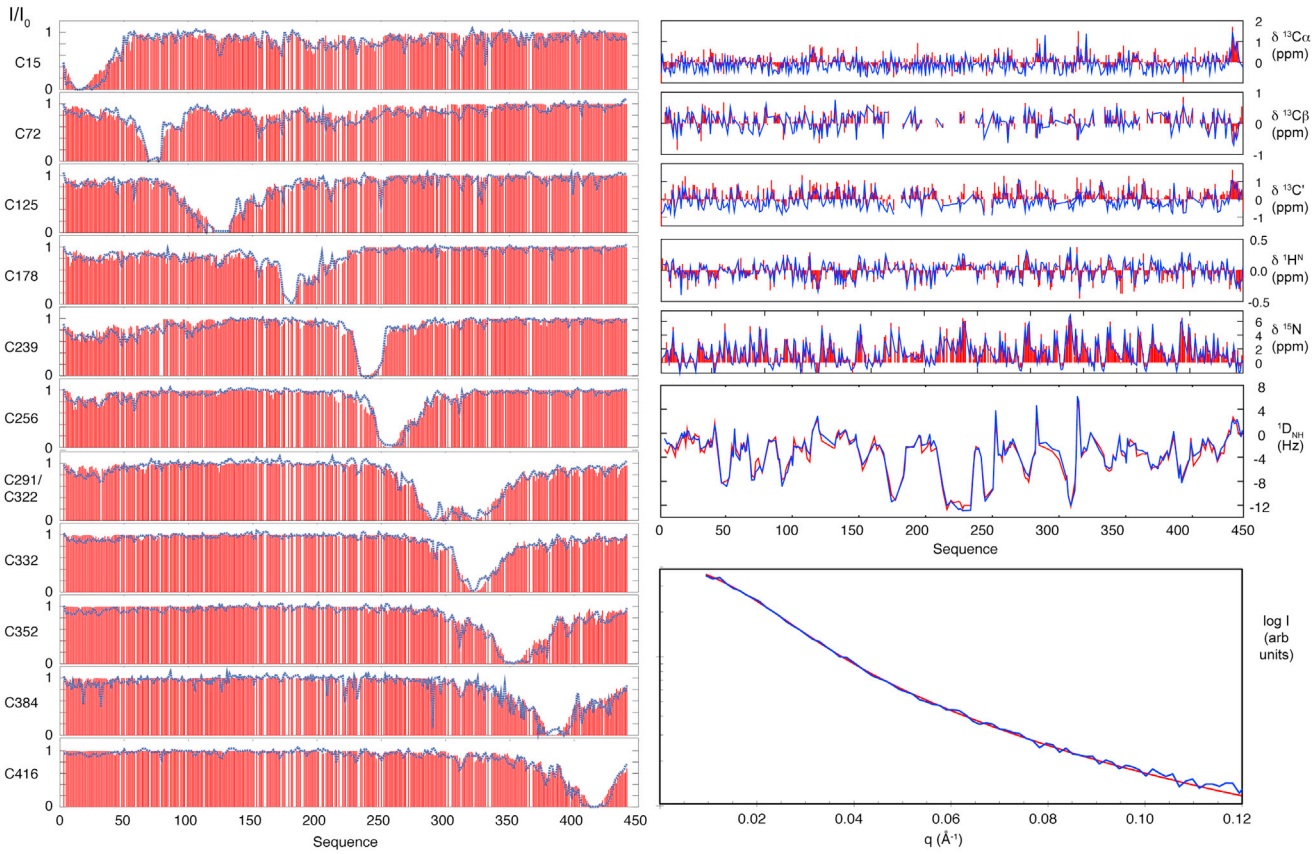

Figure 6. Reproduction of Experimental Data from 441-Residue Tau Using ASTEROIDS
Left: experimental (red) and calculated (blue) PRE data.
Top right: experimental (red) and calculated (blue) CS data.
Mid-right: experimental (red) and calculated (blue) RDCs.
Bottom right: experimental (red) and calculated (blue) SAXS data.
See also Figure S1.

aggregation-prone regions, providing potential precursors for pathogenic βS formation.

EXPERIMENTAL PROCEDURES

Experimental NMR

NMR spectra were acquired on Bruker spectrometers with proton frequencies of 600, 700, 800, and 900 MHz equipped with cryogenic probes and Agilent spectrometers with proton frequencies of 600 and 800 MHz equipped with cryogenic probes. NMR data were processed with NMRPipe (Delaglio et al., 1995) and analyzed using Sparky (T.D. Goddard and D.G. Kneller, SPARKY3, University of California, San Francisco). CSs for the longest 441-residue splice-isoform of hTau40 were determined previously at pH 6.8 and 25°C (Narayanan et al., 2010). $^{1}D_{HN}$ RDCs in hTau40 were obtained using an in-phase-antiphase (IPAP)-heteronuclear single quantum correlation (HSQC). $^{1}D_{HN}$ values were calculated as the difference between the splitting in the isotropic and weakly aligned state. Weak alignment of hTau40 was achieved in a nonionic liquid crystalline medium of 5% pentaethyleneglycolmonooctylether/*n*-octanol (Rückert and Otting, 2000). Otherwise, the NMR buffer contained 50 mM NaH_2PO_4/Na_2HPO_4 (pH 6.8), 10% D_2O. IPAP-HSQC spectra were acquired at 5°C.

Experimental CSs of the K32 construct of Tau were obtained at 5°C as described previously (Mukrasch et al., 2007b). To avoid any bias we calculated the CSs of K32 corresponding to 25°C by comparing the 5°C assignment of K32 with the 25°C assignment of full-length Tau (Mukrasch et al., 2009), and we applied a uniform shift to each nucleus type independently. Different types of RDCs were measured in stretched polyacrylamide gels: $^{1}D_{HN}$, $^{2}D_{CHN}$, and $^{4}D_{HNiHai-1}$. The RDCs were measured using band-selective excitation short-transient-type three-dimensional triple resonance [HNCO or HN(CO)CA] experiments allowing for coupling evolutions (Rasia et al., 2011). Spectra were acquired at 5°C and an ^{1}H resonance frequency of 800 MHz. In the case of αSyn, previously determined CSs and RDCs were used (Bertoncini et al., 2005b; Cho et al., 2009).

SAXS

SAXS data were collected at X33 at the European Molecular Biology Laboratory on DORIS III (DESY) at a wavelength of 1.5 Å at 25°C using a Pilatus 1M photon counting detector. Samples were exposed to eight frames of 15 s. Tau samples contained either 1.0, 2.5, 4.0, or 10 mg/ml hTau40 in 50 mM phosphate buffer (pH 6.8) and 1 mM dithiothreitol. In the case of αSyn, protein concentrations of 1.6, 3.3, and 5.0 mg/ml in 10 mM phosphate buffer (pH 7.4) were measured. For both proteins, comparison of SAXS data at different protein concentrations excluded any concentration effect. Data at 5.0 mg/ml αSyn and 4.0 mg/ml hTau40 were used for further analysis using ATSAS 2.3 (Konarev et al., 2006).

Flexible-Meccano/ASTEROIDS

The conformational sampling and ensemble selection algorithms flexible-meccano (Bernadó et al., 2005b; Ozenne et al., 2012a) and ASTEROIDS

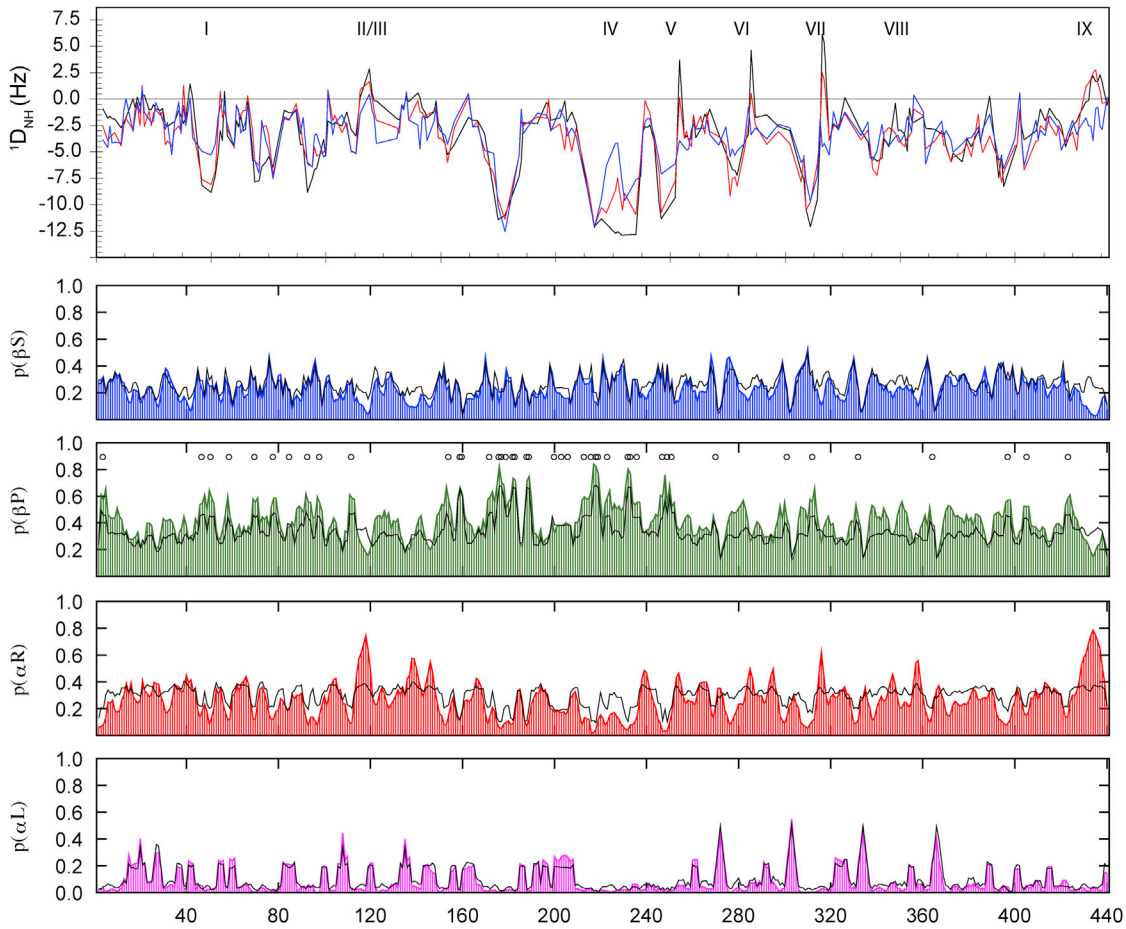

Figure 7. Conformational Propensities in 441-Residue Tau Derived from RDCs, CSs, PREs, and SAXS

Color scheme as for Figure 2. The position of proline residues in Tau are shown as open circles in the second panel. Top panel shows experimental $^1D_{HN}$ RDCs (black) compared with values predicted using the statistical coil model (blue) and values predicted using the ASTEROIDS selection based on CSs, PREs, and SAXS data only (red). The numbered regions are described in the text and show either differences in conformational sampling (calculated using Equation 1) or clear improvement in prediction based on the selected ensemble compared with the statistical coil model.

(Nodet et al., 2009) were used as recently described. In brief, a pool of statistical coil conformers was generated using flexible-meccano, and experimental observables were calculated for each conformer: CSs using SPARTA (Shen and Bax, 2007), RDCs using an in-house algorithm, SAXS using Crysol (Bernadó et al., 2005b; Svergun et al., 1995; Wells et al., 2008), and PREs as previously described (Salmon et al., 2010). Expected average values were calculated over each subensemble, and the genetic algorithm ASTEROIDS was used to select subsets of conformers reproducing experimental data (Ozenne et al., 2012b).

Local alignment windows of 15 amino acids in length were used to calculate averaged RDCs (Marsh et al., 2008). These were then combined with a generic baseline to account for long-range effects (Nodet et al., 2009; Salmon et al., 2010). Random coil values for calculation of secondary shifts were taken from RefDB (Zhang et al., 2003).

The Ramachandran space was divided into four regions defined as follows: αL {$\phi > 0°$}, αR {$\phi < 0$, $-120° < \psi < 50°$}, βP {$-100° < \phi < 0°$, $\psi > 50°$ or $\psi < -120°$}, and βS {$-180° < \phi < -100°$, $\psi > 50°$ or $\psi < -120°$}. The populations of these quadrants are denoted as p(αL), p(αR), p(βP), and p(βS), respectively. Differences between statistical coil sampling and selected ensembles were characterized with the following metric for each residue:

$$\Delta_{i,Rama} = \sum_k \left(p_{k,SC}(k) - p_{k,AST}(k)\right)^2, \quad (1)$$

where k covers the four regions (αL, αR, βP, and βS). The dependence of cross-validated χ^2 values on identifiable differences in conformational sampling between selection and statistical coil was determined by creating subsets of data that are only included in the χ^2 when $\Delta_{Rama} > \Delta_{Rama,threshhold}$ for the residue of interest.

Deposition of Ensembles

Atomic coordinates of representative ensembles of both proteins are deposited in the pE-DB database (Varadi et al., 2013).

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures and can be found with this article online at http://dx.doi.org/10.1016/j.str.2013.10.020.

ACKNOWLEDGMENTS

We thank Ilka Lindner for preparation of Tau samples. This work was supported financially by ANR Protein Disorder (JCJC 2010 to M.R.J.), TAUSTRUCT (MALZ 2010 to M.B.), Deutsche Forschungsgemeinschaft (71/7-1 to M.Z.), START and Ventures Programmes of Foundation for Polish Science (Fundacja na rzecz Nauki Polskiej; http://www.fnp.org.pl/) operated

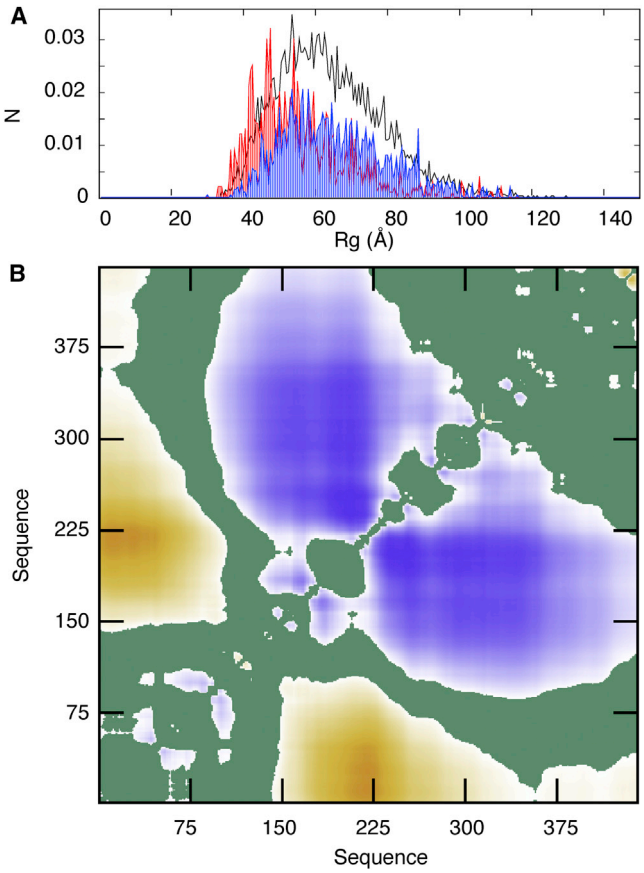

Figure 8. Overall Dimensions and Long-Range Contacts in hTau40

(A) Distribution of radius of gyration in statistical coil pool (gray), in the SAXS+RDC+PRE+CS-selected ensemble (blue), and in the PRE-only selected ensemble (red) for hTau40.

(B) C^{α} atom contact map of 441-residue Tau determined using CS, RDCs, SAXS, and PREs. Sampling represented in terms of $\delta_{ij} = \log(<d_{ij}>/< d_{ij,ref}>)$. Colors range from red (−0.3, regions in closer contact than pool) to green (0.0) to blue (0.3, regions farther apart than pool).

within the Innovative Economy Operational Programme (IE OP) 2007-2013 within European Regional Development Fund (to L.J. and M.J.), and the Iuventus Plus project IP2011 019471 from the Polish Ministry of Sciences and Higher Education (to M.J.).

Received: September 16, 2013
Revised: October 29, 2013
Accepted: October 30, 2013
Published: December 19, 2013

REFERENCES

Allison, J.R., Varnai, P., Dobson, C.M., and Vendruscolo, M. (2009). Determination of the free energy landscape of alpha-synuclein using spin label nuclear magnetic resonance measurements. J. Am. Chem. Soc. *131*, 18314–18326.

Ball, K.A., Phillips, A.H., Nerenberg, P.S., Fawzi, N.L., Wemmer, D.E., and Head-Gordon, T. (2011). Homogeneous and heterogeneous tertiary structure ensembles of amyloid-β peptides. Biochemistry *50*, 7612–7628.

Bernadó, P., Bertoncini, C.W., Griesinger, C., Zweckstetter, M., and Blackledge, M. (2005a). Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. J. Am. Chem. Soc. *127*, 17968–17969.

Bernadó, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R.W.H., and Blackledge, M. (2005b). A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc. Natl. Acad. Sci. USA *102*, 17002–17007.

Bernstein, S.L., Liu, D., Wyttenbach, T., Bowers, M.T., Lee, J.C., Gray, H.B., and Winkler, J.R. (2004). α-synuclein: stable compact and extended monomeric structures and pH dependence of dimer formation. J. Am. Soc. Mass Spectrom. *15*, 1435–1443.

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

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

Bibow, S., Ozenne, V., Biernat, J., Blackledge, M., Mandelkow, E., and Zweckstetter, M. (2011). Structural impact of proline-directed pseudophosphorylation at AT8, AT100, and PHF1 epitopes on 441-residue tau. J. Am. Chem. Soc. *133*, 15842–15845.

Blanch, E.W., Morozova-Roche, L.A., Cochran, D.A., Doig, A.J., Hecht, L., and Barron, L.D. (2000). Is polyproline II helix the killer conformation? A Raman optical activity study of the amyloidogenic prefibrillar intermediate of human lysozyme. J. Mol. Biol. *301*, 553–563.

Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C.M., and Stefani, M. (2002). Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature *416*, 507–511.

Carrell, R.W., and Lomas, D.A. (1997). Conformational disease. Lancet *350*, 134–138.

Cho, M.-K., Nodet, G., Kim, H.-Y., Jensen, M.R., Bernado, P., Fernandez, C.O., Becker, S., Blackledge, M., and Zweckstetter, M. (2009). Structural characterization of alpha-synuclein in an aggregation prone state. Protein Sci. *18*, 1840–1846.

Creamer, T.P., and Campbell, M.N. (2002). Determinants of the polyproline II helix from modeling studies. Adv. Protein Chem. *62*, 263–282.

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

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

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

Dunker, A.K., Silman, I., Uversky, V.N., and Sussman, J.L. (2008). Function and structure of inherently disordered proteins. Curr. Opin. Struct. Biol. *18*, 756–764.

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

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

Eker, F., Griebenow, K., and Schweitzer-Stenner, R. (2003). Stable conformations of tripeptides in aqueous solution studied by UV circular dichroism spectroscopy. J. Am. Chem. Soc. *125*, 8178–8185.

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

Fink, A.L. (2005). Natively unfolded proteins. Curr. Opin. Struct. Biol. *15*, 35–41.

Fisher, C.K., and Stultz, C.M. (2011). Constructing ensembles for intrinsically disordered proteins. Curr. Opin. Struct. Biol. *21*, 426–431.

Fitzkee, N.C., Fleming, P.J., and Rose, G.D. (2005). The Protein Coil Library: a structural database of nonhelix, nonstrand fragments derived from the PDB. Proteins *58*, 852–854.

Gurry, T., Ullman, O., Fisher, C.K., Perovic, I., Pochapsky, T., and Stultz, C.M. (2013). The dynamic structure of α-synuclein multimers. J. Am. Chem. Soc. *135*, 3865–3872.

Jeganathan, S., von Bergen, M., Brutlach, H., Steinhoff, H.-J., and Mandelkow, E. (2006). Global hairpin folding of tau in solution. Biochemistry *45*, 2283–2293.

Jensen, M.R., Salmon, L., Nodet, G., and Blackledge, M. (2010). Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. J. Am. Chem. Soc. *132*, 1270–1272.

Jensen, M.R., Ruigrok, R.W.H., and Blackledge, M. (2013). Describing intrinsically disordered proteins at atomic resolution by NMR. Curr. Opin. Struct. Biol. *23*, 426–435.

Jha, A.K., Colubri, A., Zaman, M.H., Koide, S., Sosnick, T.R., and Freed, K.F. (2005). Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. Biochemistry *44*, 9691–9702.

Keiderling, T.A., and Xu, Q. (2002). Unfolded peptides and proteins studied with infrared absorption and vibrational circular dichroism spectra. Adv. Protein Chem. *62*, 111–161.

Konarev, P.V., Petoukhov, M.V., Volkov, V.V., and Svergun, D.I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. J. Appl. Crystallogr. *39*, 277–286.

Krimm, S., and Tiffany, M. (1974). Circular-dichroism spectrum and structure of unordered polypeptides and proteins. Isr. J. Chem. *12*, 189–200.

Maiti, N.C., Apetri, M.M., Zagorski, M.G., Carey, P.R., and Anderson, V.E. (2004). Raman spectroscopic characterization of secondary structure in natively unfolded proteins: alpha-synuclein. J. Am. Chem. Soc. *126*, 2399–2408.

Makowska, J., Rodziewicz-Motowidło, S., Bagińska, K., Vila, J.A., Liwo, A., Chmurzyński, L., and Scheraga, H.A. (2006). Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. Proc. Natl. Acad. Sci. USA *103*, 1744–1749.

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

Marsh, J.A., Baker, J.M.R., Tollinger, M., and Forman-Kay, J.D. (2008). Calculation of residual dipolar couplings from disordered state ensembles using local alignment. J. Am. Chem. Soc. *130*, 7804–7805.

Mästle, W., Dukor, R.K., Yoder, G., and Keiderling, T.A. (1995). Conformational study of linear alternating and mixed D- and L-proline oligomers using electronic and vibrational CD and Fourier transform IR. Biopolymers *36*, 623–631.

McColl, I.H., Blanch, E.W., Hecht, L., Kallenbach, N.R., and Barron, L.D. (2004). Vibrational Raman optical activity characterization of poly(l-proline) II helix in alanine oligopeptides. J. Am. Chem. Soc. *126*, 5076–5077.

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

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

Mittag, T., Kay, L.E., and Forman-Kay, J.D. (2010). Protein dynamics and conformational disorder in molecular recognition. J. Mol. Recognit. *23*, 105–116.

Mukrasch, M.D., Markwick, P., Biernat, J., Bergen, Mv., Bernadó, P., Griesinger, C., Mandelkow, E., Zweckstetter, M., and Blackledge, M. (2007a). Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. J. Am. Chem. Soc. *129*, 5235–5243.

Mukrasch, M.D., von Bergen, M., Biernat, J., Fischer, D., Griesinger, C., Mandelkow, E., and Zweckstetter, M. (2007b). The "jaws" of the tau-microtubule interaction. J. Biol. Chem. *282*, 12230–12239.

Mukrasch, M.D., Bibow, S., Korukottu, J., Jeganathan, S., Biernat, J., Griesinger, C., Mandelkow, E., and Zweckstetter, M. (2009). Structural polymorphism of 441-residue tau at single residue resolution. PLoS Biol. *7*, e34.

Mylonas, E., Hascher, A., Bernadó, P., Blackledge, M., Mandelkow, E., and Svergun, D.I. (2008). Domain conformation of tau protein studied by solution small-angle X-ray scattering. Biochemistry *47*, 10345–10353.

Narayanan, R.L., Dürr, U.H.N., Bibow, S., Biernat, J., Mandelkow, E., and Zweckstetter, M. (2010). Automatic assignment of the intrinsically disordered protein Tau with 441-residues. J. Am. Chem. Soc. *132*, 11906–11907.

Nath, A., Sammalkorpi, M., DeWitt, D.C., Trexler, A.J., Elbaum-Garfinkle, S., O'Hern, C.S., and Rhoades, E. (2012). The conformational ensembles of α-synuclein and tau: combining single-molecule FRET and simulations. Biophys. J. *103*, 1940–1949.

Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M.R., and Blackledge, M. (2009). Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. J. Am. Chem. Soc. *131*, 17908–17918.

Ozenne, V., Bauer, F., Salmon, L., Huang, J.-R., Jensen, M.R., Segard, S., Bernadó, P., Charavay, C., and Blackledge, M. (2012a). Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics *28*, 1463–1470.

Ozenne, V., Schneider, R., Yao, M., Huang, J.-R., Salmon, L., Zweckstetter, M., Jensen, M.R., and Blackledge, M. (2012b). Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. J. Am. Chem. Soc. *134*, 15138–15148.

Pappu, R.V., and Rose, G.D. (2002). A simple model for polyproline II structure in unfolded states of alanine-based peptides. Protein Sci. *11*, 2437–2455.

Preuss, U., Biernat, J., Mandelkow, E.M., and Mandelkow, E. (1997). The 'jaws' model of tau-microtubule interaction examined in CHO cells. J. Cell Sci. *110*, 789–800.

Rasia, R.M., Lescop, E., Palatnik, J.F., Boisbouvier, J., and Brutscher, B. (2011). Rapid measurement of residual dipolar couplings for fast fold elucidation of proteins. J. Biomol. NMR *51*, 369–378.

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

Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M.R., Zweckstetter, M., and Blackledge, M. (2010). NMR characterization of long-range order in intrinsically disordered proteins. J. Am. Chem. Soc. *132*, 8407–8418.

Schneider, R., Huang, J.R., Yao, M., Communie, G., Ozenne, V., Mollica, L., Salmon, L., Jensen, M.R., and Blackledge, M. (2012). Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. Mol. Biosyst. *8*, 58–68.

Schweers, O., Schönbrunn-Hanebeck, E., Marx, A., and Mandelkow, E. (1994). Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. J. Biol. Chem. *269*, 24290–24297.

Schweitzer-Stenner, R. (2013). Different degrees of disorder in long disordered peptides can be discriminated by vibrational spectroscopy. J. Phys. Chem. B *117*, 6927–6936.

Sgourakis, N.G., Merced-Serrano, M., Boutsidis, C., Drineas, P., Du, Z., Wang, C., and Garcia, A.E. (2011). Atomic-level characterization of the ensemble of the Aβ(1-42) monomer in water using unbiased molecular dynamics simulations and spectral algorithms. J. Mol. Biol. *405*, 570–583.

Shen, Y., and Bax, A. (2007). Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J. Biomol. NMR *38*, 289–302.

Shi, Z., Olson, C.A., Rose, G.D., Baldwin, R.L., and Kallenbach, N.R. (2002). Polyproline II structure in a sequence of seven alanine residues. Proc. Natl. Acad. Sci. USA *99*, 9190–9195.

Shi, Z., Chen, K., Liu, Z., and Kallenbach, N.R. (2006). Conformation of the backbone in unfolded proteins. Chem. Rev. *106*, 1877–1897.

Sibille, N., and Bernadó, P. (2012). Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. Biochem. Soc. Trans. *40*, 955–962.

Smith, L.J., Bolin, K.A., Schwalbe, H., MacArthur, M.W., Thornton, J.M., and Dobson, C.M. (1996). Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. J. Mol. Biol. *255*, 494–506.

Svergun, D., Barberato, C., and Koch, M. (1995). CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. *28*, 768–773.

Swindells, M.B., MacArthur, M.W., and Thornton, J.M. (1995). Intrinsic phi, psi propensities of amino acids, derived from the coil regions of known structures. Nat. Struct. Biol. *2*, 596–603.

Syme, C.D., Blanch, E.W., Holt, C., Jakes, R., Goedert, M., Hecht, L., and Barron, L.D. (2002). A Raman optical activity study of rheomorphism in caseins, synucleins and tau. New insight into the structure and behaviour of natively unfolded proteins. Eur. J. Biochem. *269*, 148–156.

Tiffany, M., and Krimm, S. (1969). Circular dichroism of random polypeptide chain. Biopolymers *8*, 347.

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

Tompa, P. (2011). Unstructural biology coming of age. Curr. Opin. Struct. Biol. *21*, 419–425.

Tompa, P., and Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem. Sci. *33*, 2–8.

Ullman, O., Fisher, C.K., and Stultz, C.M. (2011). Explaining the structural plasticity of α-synuclein. J. Am. Chem. Soc. *133*, 19536–19546.

Uversky, V.N. (2002). Natively unfolded proteins: a point where biology waits for physics. Protein Sci. *11*, 739–756.

Uversky, V.N., and Fink, A.L. (2004). Conformational constraints for amyloid fibrillation: the importance of being unfolded. Biochim. Biophys. Acta (BBA)-Proteins Proteomics *1698*, 131–153.

Varadi, M., Kosol, S., Lebrun, P., Valentini, E., Blackledge, M., Dunker, A.K., Felli, I.C., Forman-Kay, J.D., Kriwacki, R.W., Pierattelli, R., et al. (2013). pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. Nucleic Acids Res. Published online October 29, 2013. http://dx.doi.org/10.1093/nar/gkt960.

Wang, W., Perovic, I., Chittuluru, J., Kaganovich, A., Nguyen, L.T.T., Liao, J., Auclair, J.R., Johnson, D., Landeru, A., Simorellis, A.K., et al. (2011). A soluble α-synuclein construct forms a dynamic tetramer. Proc. Natl. Acad. Sci. USA *108*, 17797–17802.

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

Wells, M., Tidow, H., Rutherford, T.J., Markwick, P., Jensen, M.R., Mylonas, E., Svergun, D.I., Blackledge, M., and Fersht, A.R. (2008). Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc. Natl. Acad. Sci. USA *105*, 5762–5767.

Woody, R.W. (2009). Circular dichroism spectrum of peptides in the poly(Pro)II conformation. J. Am. Chem. Soc. *131*, 8234–8245.

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

Zagrovic, B., Lipfert, J., Sorin, E.J., Millett, I.S., van Gunsteren, W.F., Doniach, S., and Pande, V.S. (2005). Unusual compactness of a polyproline type II structure. Proc. Natl. Acad. Sci. USA *102*, 11698–11703.

Zhang, H., Neal, S., and Wishart, D.S. (2003). RefDB: a database of uniformly referenced protein chemical shifts. J. Biomol. NMR *25*, 173–195.