Domain Conformation of Tau Protein Studied by Solution Small-Angle X-ray Scattering[†]

Efstratios Mylonas,[‡] Antje Hascher,[§,||] Pau Bernadó,[‡,⊥] Martin Blackledge,[@] Eckhard Mandelkow,[§] and Dmitri I. Svergun[‡,#,*]

European Molecular Biology Laboratory, Hamburg Outstation, Notkestrasse 85, 22603 Hamburg, Germany, Max-Planck-Unit for Structural Molecular Biology, Notkestrasse 85, 22607 Hamburg, Germany, Institut de Biologie Structurale Jean-Pierre Ebel, CEA-CNRS-UJF, 41 Rue Jules Horowitz, 38027 Grenoble, France, and Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia

Received May 14, 2008; Revised Manuscript Received August 5, 2008

ABSTRACT: Tau is one of the two main proteins involved in the pathology of Alzheimer's disease via formation of β-sheet rich intracellular aggregates named paired helical filaments (PHFs). Given that tau is a natively unfolded protein with no folded core (even upon binding to physiological partners such as microtubules), its structural analysis by high-resolution techniques has been difficult. In this study, employing solution small-angle X-ray scattering from the full length isoforms and from a variety of deletion and point mutants the conformation of tau in solution is structurally characterized. A recently developed ensemble optimization method was employed to generate pools of random models and to select ensembles of coexisting conformations, which fitted simultaneously the scattering data from the full length protein and deletion mutants. The analysis of the structural properties of these selected ensembles allowed us to extract information about residual structure in different domains of the native protein. The short deletion mutants containing the repeat domain (considered the core constituent of the PHFs) are significantly more extended than random coils, suggesting an extended conformation of the repeat domain. The longer tau constructs are comparable in size with the random coils, pointing to long-range contacts between the N- and C-termini compensating for the extension of the repeat domain. Moreover, most of the aggregation-promoting mutants did not show major differences in structure from their wild-type counterparts, indicating that their increased pathological effect is triggered only after an aggregation core has been formed.

Alzheimer's disease is the most common form of dementia among older people (*1*). Common histological findings among patients include the formation of amyloid plaques and neurofibrillary tangles in neurons rich in β-amyloid and tau protein, respectively. The neurofibrillary tangles are formed from β-sheet rich (*2*) paired helical filaments (PHFs)[1] (*3*) which in turn are aggregates of hyperphosphorylated tau protein (*4*). Physiologically, tau is a microtubule-associated protein (*5, 6*) occurring mainly in the axons of neurons. It stabilizes microtubules (*7*) and promotes neurite outgrowth. In the human central nervous system (CNS), it is found in six alternatively spliced isoforms ranging from 352 to 441 residues resulting from the presence or absence of exons 2, 3, and 10 (*8, 9*). In fetal brain, the smallest isoform (ht23, lacking exons 2, 3, and 10) is the predominant one, whereas in adult brain, all isoforms can be found in roughly equal amounts. Tau contains four semiconserved sequences of 31 or 32 residues, so-called "repeats". The second repeat is exon 10 and may be absent in some of the isoforms. The repeat domain is essential both for the binding to microtubules and for the aggregation of tau into PHFs (*10*). Mutations of the tau gene can cause frontotemporal dementias with parkinsonism (FTDP-17) and are mostly found in the repeat domain (*11*). The pathological aggregation of tau appears to be nucleated by the hexapeptide motifs PHF6*=275-VQIINK-280 (in R2) and PHF6=306-VQIVYK-311 (in R3) probably because of the propensity of these residues to form β structure (*12, 13*).

Overall, tau has a very low content of secondary structure as shown by sequence analysis (very high content of polar residues) and circular dichroism (CD) experiments (*14*−*16*). It is considered a natively unfolded protein (NUP) (*17*), as shown for example by circular dichroism spectra (*12*), and retains its disordered state even upon binding to microtubules, without adopting well-defined periodic binding sites in

[†] This work was funded in part by a grant from the Deutsche Forschungsgemeinschaft (to E. Mandelkow).

* To whom correspondence should be addressed: EMBL Hamburg C/O DESY, Notkestrasse 85, 22603 Hamburg, Germany. Telephone: +49 40 89902 125. Fax: +49 40 89902 149. E-mail: Svergun@EMBL-Hamburg.de.

[‡] European Molecular Biology Laboratory.

[§] Max-Planck-Unit for Structural Molecular Biology.

[||] Present address: Department of Hematology and Oncology, Universitatsklinikum Münster, Domagkstrasse 3, 48149 Münster, Germany.

[⊥] Present address: Institut de Recerca Biomèdica, Parc Científic de Barcelona, Josep Samitier 1-5, 08028 Barcelona, Spain.

[@] CEA-CNRS-UJF.

[#] Russian Academy of Sciences.

[1] Abbreviations: PHF, paired helical filament; SAXS, small-angle X-ray scattering; FTDP-17, frontotemporal dementia with parkinsonism linked to chromosome 17; EOM, ensemble optimization method; NUP, natively unfolded protein.

10.1021/bi800900d CCC: $40.75 © 2008 American Chemical Society
Published on Web 09/05/2008

contrast to motor proteins (*18, 19*). In vitro tau aggregation can be induced efficiently only by incubation with polyanions (e.g., heparin) (*20*). Moreover, phosphorylation negatively regulates both tau-microtubules and tau−tau interactions, with some of the sites being responsible for both (*21*). Interestingly, in Alzheimer's disease PHFs, tau is hyperphosphorylated, a process that is poorly understood.

The family of natively unfolded proteins is neither small nor lacking biological and structural significance. An increasing number of proteins or protein parts is found to have very low content of secondary structure and thus no folded core. It is suggested that more than 30% of the eukaryotes genome's open reading frames consists of sequences, at least 50 residues long, that encode unfolded proteins (*22*−*24*). Their range of biological functions is also quite diverse (although they are mainly found in regulatory and functional roles rather than cellular structures). NUPs are involved in transcriptional and translation regulation, cell cycle control, modulation of the assembly of other proteins, and neuron development (as is the case for tau). Their unfoldedness appears to be a desirable feature for binding to several partners, even simultaneously. Tau is not the only NUP involved in pathologic amyloid fibril formation; for example, α-synuclein forms amyloids in the brains of patients with Parkinson's disease.

The high flexibility of NUPs presents several challenges, and the concept of structure for NUPs is quite different from that of folded proteins. For a flexible macromolecule, the structure does not signify a fixed and rigid arrangement of the polypeptide in three dimensions, but it rather describes the freedom to explore the available conformational space. Individual protein molecules will have different overall structures, but some arrangements might be more frequent than others. Of course, this notion of structure does not permit the application of structural techniques such as X-ray crystallography and electron microscopy. Even if we disregard the technical difficulties (e.g., in growing the crystals), the crystal structures would probably not tell much about the actual conformations of the flexible protein. The palette of the structural methods is thus limited to those applicable in solution. Here, nuclear magnetic resonance (NMR) and small-angle X-ray scattering (SAXS) are the major structural techniques providing high- and low-resolution information, respectively.

In this study, two isoforms (the longest and the shortest) as well as a number of deletion and point mutants (including pseudophosphorylated constructs) were analyzed in solution using SAXS. The combination of measurements of the full length constructs and deletion mutants allowed for analysis of residual structure in different domains of the native protein. Finally, the comparison of the scattering patterns from point mutants resistant or prone to aggregation (termed "anti-aggregation" and "pro-aggregation" mutants) to those of the wild-type protein provided information about the pathological influence of these mutations.

MATERIALS AND METHODS

Protein Preparation. Figure 1summarizes the constructs of tau protein used in this study, including different tau isoforms or combinations of domains and different mutations observed in FTDP-17. The constructs include ht40 (largest isoform in human CNS, 441 residues), and its deletion mutants K32, K16, and K18, and ht23 (the smallest isoform, 352 residues), and its deletion mutants K27, K17, K19, K44, and K10. Deletion mutants containing no repeats were also measured, the N-terminal fragment K25 and the N+C-terminal construct K23. Several point mutants were also used, such as "E mutants" mimicking the phosphorylation at Ser or Thr residues and named after the phospho-epitopes of well-known antibodies against Alzheimer tau: K32-AT8*(E)-AT100(E) (K32 containing mutations S199E, S202E, T205E, T212E, and S214E), ht23-S214E, ht23-AT8*(E)-AT100(E) (ht23 containing mutations S199E, S202E, T205E, T212E, and S214E), K18-P301L and K18-ΔK280 (K18 with two natural FTDP-17 mutations), and K18-ΔK280-I277P-I308P (K18 with the natural FTDP-17 mutation ΔK280 and two proline mutations within two hexapeptide motifs breaking their β-sheet structure). Proteins were prepared as described previously (*16*). Briefly, the tau isoforms, deletion constructs, and point mutants were obtained using PCR amplification and subcloned into expression vector pNG2 [a derivative of pET-3a (Merck-Novagen, Darmstadt, Germany)]. All recombinant proteins were expressed in the *Escherichia coli* BL21(DE3) strain (Merck-Novagen). The expressed proteins were purified from bacterial extracts by making use of the heat stability of the tau protein and by FPLC SP-Sepharose (GE Healthcare, Freiburg, Germany). The cell pellet was resuspended in the boiling extraction buffer [50 mM MES, 500 mM NaCl, 1 mM $MgSO_4$, 1 mM EGTA, and 5 mM DTT (pH 6.8)] complemented with protease inhibitor cocktail. The cells were disrupted with a French pressure cell and subsequently boiled for 20 min. The soluble extract was isolated by centrifugation, and the supernatant was dialyzed against two changes of cation exchange chro-

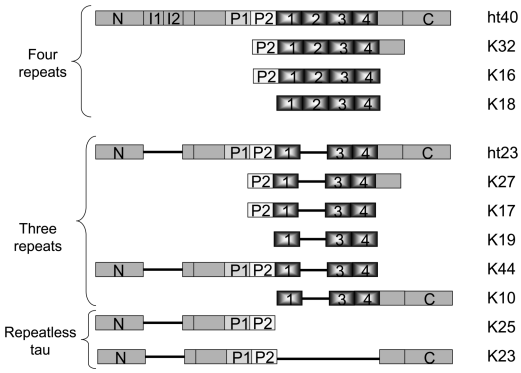

FIGURE 1: Bar diagrams of isoforms and mutants of tau protein. All residue numbers refer to the sequence of ht40, the longest of human tau isoforms in the central nervous system [441 residues (*9*)]. Constructs with four repeats: ht40 (441 amino acids); K32, residues (M)S198−Y394; K16, residues (M)S198−E372; and K18, residues (M)Q244−E372. Constructs with three repeats: ht23 (352 amino acids); K27, residues (M)S198−Y394 without the second repeat; K17, residues (M)S198−E372 without the second repeat; K19, three repeats where the second repeat is missing (M)Q244−E372; K44, residues M1−E372 without the second repeat and where two N-terminal inserts (E45−T102) are missing; K10, residues (M)Q244−L441 without the second repeat. For repeat-less tau, K25 contains the amino-terminal domain of ht23 and consists of residues M1−L243 (residues E45−T102, representing the amino-terminal inserts in ht40, are missing) and K23 represents the ht23 molecule without repeats (residues Q244−N368 are missing).

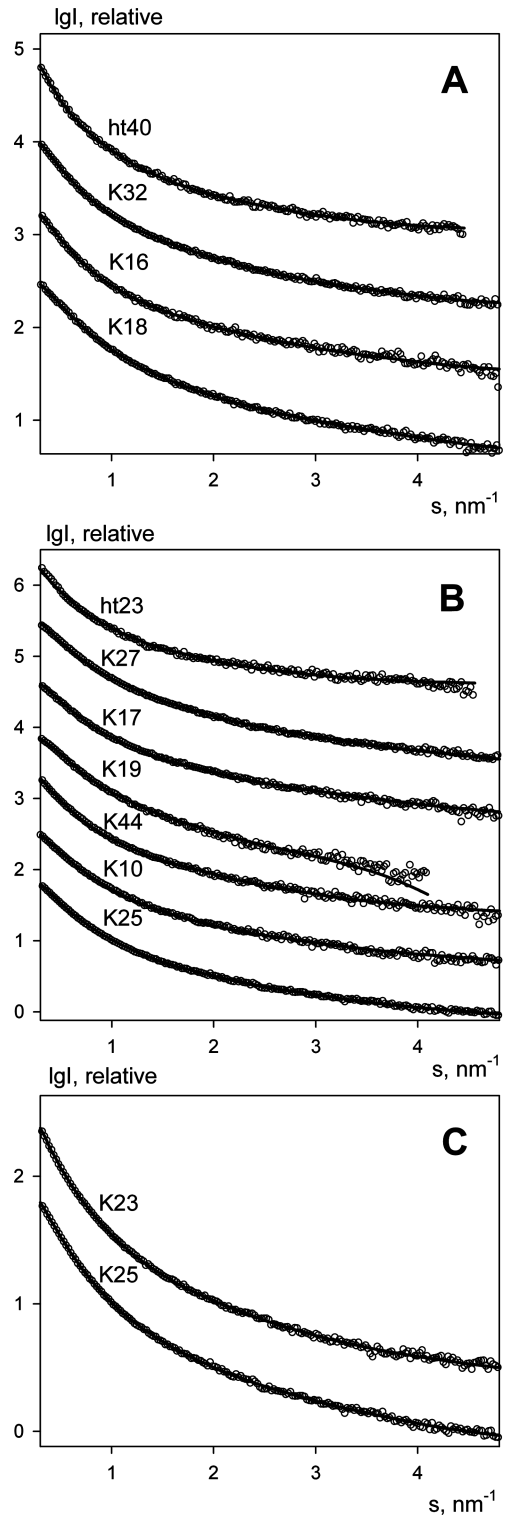

FIGURE 2: Experimental SAXS data (○) with the corresponding ensemble fit (—). Constructs were grouped in three categories: (A) ht40 (containing four repeats), (B) ht23 (containing three repeats with K25 which is the N-terminal part of ht23), and (C) K23 (repeatless tau equivalent to ht23 and the N-terminus of ht23).

matography buffer A [20 mM MES, 50 mM NaCl, 1 mM EGTA, 1 mM $MgSO_4$, 2 mM DTT, and 0.1 mM PMSF (pH 6.8)] and loaded on a FPLC SP-Sepharose column. The proteins were eluted with a linear gradient of cation exchange chromatography buffer B [20 mM MES, 1 M NaCl, 1 mM EGTA, 1 mM $MgSO_4$, 2 mM DTT, and 0.1 mM PMSF (pH 6.8)]. The purity of proteins was ascertained by SDS−PAGE. Where necessary, the breakdown products were removed by using the additional gel filtration column Superdex G75 or G200 with PBS buffer [137 mM NaCl, 3 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4 (pH 7.4), and 1 mM DTT]. Before SAXS data collection, protein samples were prepared by concentration on an Amicon Ultra-15 device (Millipore, Bedford, MA).

SAXS Data Collection and Processing. Synchrotron X-ray scattering data from solutions of the tau constructs were collected at beamline X33 of the EMBL (DESY, Hamburg, Germany) (*25*) using a MAR345 image plate detector. The scattering patterns were measured with a 3 min exposure time at 15 °C for multiple solute concentrations ranging from 2 to 10 mg/mL for the full length constructs and from 5 to 20 mg/mL for the smaller ones. To check for radiation damage, two 2 min exposures were compared; no radiation effects were observed. Using a sample−detector distance of 2.7 m, the range of momentum transfer 0.09 $nm^{-1} < s < 5\ nm^{-1}$ was covered [$s = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and λ (1.5 Å) is the X-ray wavelength].

The data were processed using standard procedures and extrapolated to infinite dilution with PRIMUS (*26*). The forward scattering $I(0)$ and the radii of gyration (R_g) were evaluated using the Guinier approximation (*27*) assuming that at very small angles ($s < 1.3/R_g$) the intensity is represented as $I(s) = I(0) \exp[-(sR_g)^2/3]$. For longer constructs, the experimentally available angular range contains a limited number of data points for accurate approximations (Supporting Information Figure 1). For this reason, the $I(0)$ and R_g as well as the maximum dimensions (D_{max}) and the interatomic distance distribution functions [$p(r)$] were computed using the indirect transform package GNOM (*28*). Because of the difficulty in determining the solute concentration (c) by optical absorption at 280 nm due to the scarcity of aromatic residues, especially tryptophans, molecular masses of the solutes could not be estimated from the forward scattering [$I(0)/c$] of the SAXS curves with sufficient accuracy. However, the scattering patterns collected at different concentrations of the same construct could be well scaled to each other (see examples in Supporting Information Figure 2), pointing to the absence of aggregation effects. The latter effects would have led to an increase in the apparent solute size [i.e., to a steeper slope of $I(s)$ at very small angles], which was not the case for all the samples measured except for the aggregating mutant K18-ΔK280 (the behavior of this construct will be discussed in detail below). If weak differences between the concentrations are observed, they rather point to repulsive interactions (the higher concentration has a less steep slope; see construct K32 in Supporting Information Figure 2), and a standard extrapolation to infinite dilution accounted for them. Furthermore, all samples came as single peaks from the gel filtration columns, indicating monodisperse preparations. Moreover, when more than one point mutants of the same construct were measured in the same experimental session, no significant differences were observed in their $I(0)$ values and thus in the molecular masses of the solutes.

For the calculation of the random coil R_g values, the equation $R_g = R_0 N^{\nu}$ was employed, where N is the number of residues, R_0 (0.1927 nm) is a constant that is a function of, among other things, the persistence length of the chain,

Table 1: Radii of Gyration

construct	no. of amino acids	experimental R_g (nm)	random coil R_g[a] (nm)	experimental R_g/RC R_g ratio	calculated R_g (PDB entry)[b]
ht40	441	6.5 ± 0.3	6.9	0.94	2.4 (1AQH)
K32	202	4.2 ± 0.3	4.4	0.96	1.8 (1AUN)
K16	174	3.9 ± 0.3	4.0	0.98	1.6 (1A33)
K18	130	3.8 ± 0.3	3.4	1.12	1.5 (8LYZ)
ht23	352	5.3 ± 0.3	6.1	0.88	2.1 (1AIR)
K27	171	3.7 ± 0.2	4.0	0.94	1.7 (1EUB)
K17	143	3.6 ± 0.2	3.6	1.02	1.8 (1J57)
K19	99	3.5 ± 0.1	2.9	1.21	1.4 (1AG6)
K44	283	5.2 ± 0.2	5.3	0.97	2.0 (1A9O)
K10	167	4.0 ± 0.1	3.9	1.02	1.6 (1B8Y)
K25	185	4.1 ± 0.2	4.1	1.00	1.6 (153L)
K23	254	4.9 ± 0.2	5.0	0.99	2.0 (1A8P)
mutants					
K32 AT8 AT100	202	4.1 ± 0.3	4.4	0.95	1.8 (1AUN)
ht23 S214E	352	5.4 ± 0.3	6.1	0.89	2.1 (1AIR)
ht23 AT8 AT100	352	5.2 ± 0.3	6.1	0.86	2.1 (1AIR)
K18 P301L	130	3.5 ± 0.2	3.4	1.05	1.5 (8LYZ)
K18 ΔK280	129	7.9 ± 1.0	3.4	2.34	1.5 (8LYZ)
K18 ΔK280 I277P I308P	129	3.5 ± 0.2	3.4	1.05	1.5 (8LYZ)

[a] Radii of gyration calculated according to ref *29*; see Materials and Methods. [b] Radii of gyration calculated from the PDB structures (entries in parentheses) of globular proteins having approximately the same number of residues as the given construct: α-amylase (1AQH, 448 residues), pathogenesis-related protein 5D (1AUN, 208 residues), peptidylprolyl isomerase (1A33, 174 residues), hen lysozyme (8LYZ, 129 residues), pectate lyase (1AIR, 352 residues), collagenase-3 (1EUB, 171 residues), nuclease inhibitor A (1J57, 143 residues), plastocyanin (1AG6, 99 residues), purine nucleoside phosphorylase (1A9O, 289 residues), stromelysin catalytic domain (1B8Y, 167 residues), goose lysozyme (153L, 185 residues), and ferredoxin reductase (1A8P, 257 residues).

and ν (0.588) is an exponential scaling factor (*29*). The R_g values of high-resolution Protein Data Bank (PDB) structures, with a sequence length similar to those of our constructs, were calculated using CRYSOL (*30*).

Model Building and Experimental Data Fitting. The scattering patterns from the tau constructs were analyzed using the recently developed ensemble optimization method (EOM), which takes the flexibility into account by allowing for the coexistence of multiple conformations in solution (*31*). EOM selects appropriate ensembles of configurations from large pools of random models of the protein. Representative backbone models were created for constructs ht40, ht23, and K23 using Flexible Meccano, which was shown to give an adequate conformational sampling for natively unfolded proteins (*32*). The side chains were added using SCCOMP (*33*). The models for the smaller constructs were selected as appropriate portions of the larger ones. The models of ht40 were used to select the conformations of K32, K16, and K18; ht23 served as a model for K27, K17, K19, K25, K44 and K10, and K23 for K25. With this approach, it was possible, apart from reducing the computational time, to fit simultaneously multiple experimental data to the same model sets.

The theoretical scattering intensity of the models was calculated using CRYSOL (*30*). GAJOE from the EOM package (*31*) employed a genetic algorithm to select from the pool of structures the ensembles of curves (and subsequently three-dimensional models) such that the averages over the ensembles fitted the experimental data. In this context, a chromosome is the ensemble of structures and the structural models and SAXS patterns are the genes. Cycles of mutations and crossings among the different "chromosomes" help in exploring a variety of combinations. The goodness of fit, for each individual experimental curve, is characterized by the discrepancy

$$\chi^2 = \frac{1}{N-1}\sum_j \left[\frac{I_{exp}(s_j) - cI_{calc}(s_j)}{\sigma(s_j)}\right]^2$$

where N is the number of experimental points, c is a scaling factor, $I_{exp}(s_j)$ and $I_{calc}(s_j)$ are the experimentally determined and ensemble-averaged calculated intensities, respectively, and $\sigma(s_j)$ is the experimental error at momentum transfer s_j. This goodness of fit guides the genetic algorithm by allowing only the ensembles with the best scores to proliferate to the next generations of the genetic algorithm.

With the assumption that the domains of the protein retain their overall structure even as parts of larger or smaller constructs, simultaneous fitting of different models was also performed. The constructs were grouped in three categories (Figure 1): (i) constructs with all four repeats (ht40, K32, K16, and K18), (ii) constructs with three repeats (ht23, K27, K17, K19, K44, and K10 with the addition of K25, which corresponds to the N-terminus of ht23), and (iii) repeatless tau (K23 with the ht23 N-terminus K25). By running EOM multiple times, we were able to compare the R_g distributions of the selected structures versus the original pool. The matrices of average $C_\alpha - C_\alpha$ distances between each pair of amino acids were also calculated for the selected models and compared with the original pool.

RESULTS

Overall Structure of Wild-Type Constructs. Figure 2 displays the processed scattering curves of all the constructs measured (without those of point mutants). All the profiles are featureless, and the Kratky plots [$s^2I(s)$ vs s] are without apparent peaks (Supporting Information Figure 3), which is typical for unfolded proteins. A globular folded protein of ~400 residues (to recall, ht23 and ht40 have 352 and 441 residues, respectively) would have an R_g ranging typically from 2 to 2.5 nm. In contrast, we observe values of 5.3 and 6.5 nm for ht23 and ht40, respectively, and 3.5 and 3.8 nm for K19 and K18, respectively (lysozyme with 129 residues has an R_g of 1.5 nm). The experimental R_g values in Table 1 were calculated by the indirect transformation program GNOM as described in Materials and Methods, as the indirect

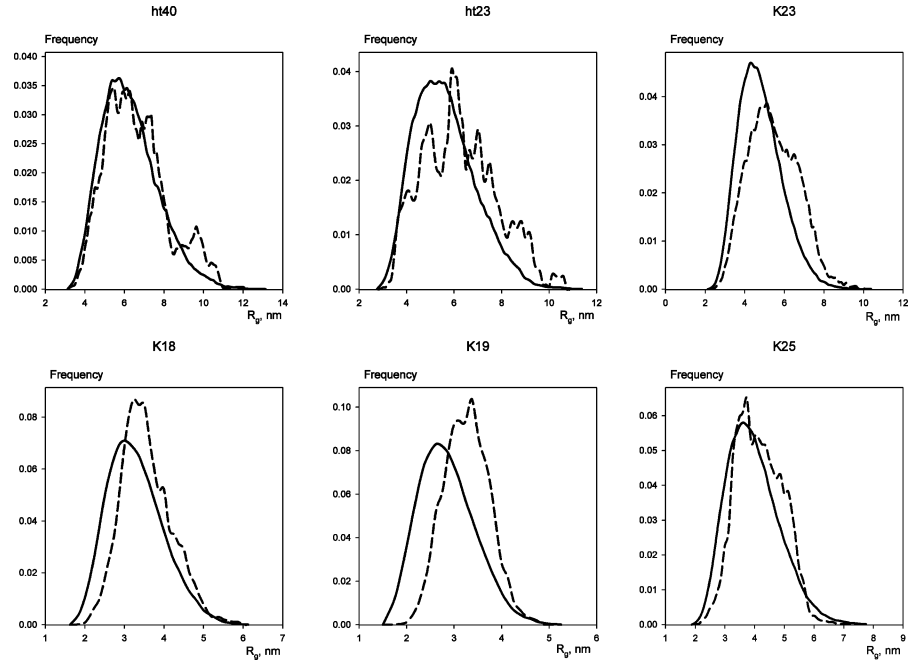

FIGURE 3: Radius of gyration distributions of the pools (—) vs the selected structures (---) using EOM (the histograms were smoothed using a sliding average). The integral of the area defined by the histograms equals 1.

transform approach (*34*) provides better accuracy of R_g determination compared to conventional Guinier analysis. The R_g values of all the constructs are compared to the expected values of random coils with the same numbers of residues (*29*) in Table 1. It should be noted that the experimentally determined R_g for unfolded proteins does not correspond to a specific configuration but rather results from the average over different coexisting conformations in solution. Table 1 indicates that the larger constructs (ht23 and ht40) have R_g values close to or smaller than the predicted random coil values, which are shown by ref *29* to agree well with the experimental data from chemically unfolded proteins (note that as tau constructs are not denatured using chemicals the random coil estimations cannot be considered ideal predictions, but they do provide a useful guidance). In contrast, smaller constructs containing the repeat domain have values either slightly or (K18 and K19) considerably larger than the random coils of the same length. We can see a clear trend to increasingly larger ratios when the protein is truncated from the full length to just the repeat domain (for either the three- or four-repeat constructs). Note that all the experimental R_g values exceed significantly those of globular proteins of equivalent lengths (examples of the latter R_g values calculated from the appropriate PDB models are given in Table 1 for comparison).

Combination of Information from Different Constructs. Already, the model-independent data given here suggest that the repeat domain must adopt extended conformations. As illustrated in Figure 2, it was possible to obtain good fits to the experimental data (χ values range from 1 to 1.5) with multiple-curve modeling using EOM. From the ensembles of models, selected after several independent runs of the program, the R_g distributions from the selected models were compared to the distributions in the original pools (largest and smallest constructs as well as repeatless tau are shown in Figure 3; the rest of the wild-type constructs are shown in Supporting Information Figure 4). For full length constructs (ht40 and ht23), little difference from the pool distribution was observed (typical models of ht40 selected from the random pools are displayed in Figure 4 to demonstrate that the protein is indeed rather unstructured). Remarkably though, for the smaller constructs, especially for those containing the repeats, the selected R_g distribution was clearly shifted to larger values indicating conformations of these fragments, which are more extended than random coils. Note that ht23 is still shifted to somewhat higher values, but this shift is much less pronounced than those for the shorter constructs, K19, K27, and K17. We are therefore dealing with an unfolded protein which appears more extended than a random coil. This means that, even though the protein is still highly flexible, the propensity to adopt extended conformations is much higher than that observed

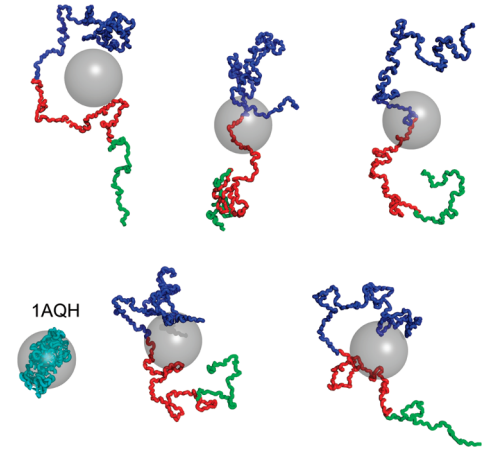

FIGURE 4: Representative ht40 structures. The N-terminus (residues 1−243) is colored blue, the repeat domain (residues 244−368) red, and the C-terminus (residues 369−441) green. The model of a folded protein with a similar number of residues in the crystal structure (α-amylase, 1AQH, R_g = 2.4 nm) is displayed as a cyan C_α chain for comparison. Spheres (semitransparent gray) with an R_g of 2.4 nm (sphere radius of 3.1 nm) are positioned at the mass center of each model. This figure was prepared using PyMol (*39*).

Table 2: Average R_g Values of Selected Models of the Repeat Domain and Equally Sized Flanking Regions[a] in the Full Length Proteins

	ht40			ht23		
amino acid range	1−129	115−243	244−372 (rd)	1−156	146−243	244−372 (rd)
R_g	3.2 ± 0.7[b]	3.1 ± 0.5	3.5 ± 0.3	2.5 ± 0.4	2.9 ± 0.6	3.5 ± 0.7

[a] Only from the N-terminus, since the C-terminus after the repeat domain is too small. [b] The deviations of the R_g estimates are computed as the variation in the R_g values computed from the portions of appropriate length taken from the EOM-selected models, and these deviations reflect the natural sampling of conformational space in these models.

for a random chain. Given that the structure of the small constructs is "embedded" in the larger ones (the repeat domain accounts for one-fourth to one-third of the full length of the protein), these results indicate that the rest of the full length protein should be less extended than the random coil, e.g., by having contacts between the N- and C-termini (*35*). In ref *35*, it is shown that the repeat domain is in the proximity of the repeat domain and the N-terminus is close to the C-terminus, creating a "paperclip" structure. The K25 construct (N-terminus) behaves nearly like a random chain. The reason K23 (not containing the repeat domain) appears more extended than the other large constructs, ht23 and ht40, could be, at least partly, attributed to the fact that the N- and C-termini in K23 are unphysiologically attached, potentially preventing the aforementioned N−C contacts. The extended nature of the repeat domain is further supported by the R_g values of the selected models for the repeat domain compared to equally sized domains in the full length protein (Table 2).

The presence or absence of the second repeat seems to also play a pivotal role in the structure of tau. The inspection of the ratio of the $C_\alpha-C_\alpha$ distances of the selected models versus the pool (Figure 5) reveals large heterogeneity in the distance distribution, especially in the repeat domain. While for ht23 the area with the largest $C_\alpha-C_\alpha$ distances is in the repeat domain itself, this is not the case for ht40 where the largest separations are found between the repeat domain and the residues preceding it in the sequence. This may seem to be puzzling at first, but apparently the number of repeats changes drastically the spatial arrangement of the protein. Recent findings (*36*) suggested the presence of turns at repeats R1−R3, and different numbers of turns may lead to different topologies of the chain conformation adjacent to the repeat domain.

Comparison of the Point Mutants with the Wild-Type Proteins. ht23 pathologic pseudophosphorylated mutant S214E (numbering based on the largest isoform ht40−441 residues) was measured along with the wild-type protein and also with a five-site pseudophosphorylated mutant "AT8-AT100" (E mutations at positions 199, 202, 205, 212, and 214). No significant differences were observed between the scattering patterns (and between R_g values) of the three constructs in solution, indicating that the pseudophosphorylation at some of the noticeable epitopes elevated in Alzheimer's disease does not significantly alter the overall shape. A similar behavior was observed for the pseudophosphorylated construct K32 AT8-AT100 (Table 1 shows the corresponding R_g values).

Two pathologic point mutants were measured for the four repeat construct (K18): P301L and ΔK280. While the former mutant behaves in solution practically the same way as the wild type, the ΔK280 mutant gives a very large R_g [7.86 nm (Table 1)] significantly (more than 2 times) exceeding the R_g value expected for a random coil of this length. Apparently, being a "pro-aggregation mutant", ΔK280 is capable of spontaneously aggregating in solution without the addition of polyanions (*16*). Interestingly, this "phenotype" can be rescued by introducing two additional point mutations. If the isoleucines at positions 277 and 308 are replaced with prolines, the triple mutant behaves in solution the same way as the wild type (Supporting Information Figure 2).

DISCUSSION

Comparison with the Results of Other Structural Methods. While the results shown and the conclusions drawn here are important on their own, it is useful to compare and combine them with the findings of other methods suitable for the analysis of unfolded systems. As recently shown by NMR (*19*), the repeat domain of tau has a propensity to form partial β-structure even in solution. Given that a β-strand is a rather extended rigid structure with residues arranged almost on a straight line, this fact may explain the observed extended conformation of the repeat domain, compared to a random coil. Our SAXS results unambiguously demonstrate that this propensity is mainly located at the repeat domain and not the other domains of the protein. More recent findings by NMR (*36*) reveal type I β-turns in the repeat domain. Although this may seem contradictory at first, the existence of these turns does not necessarily imply that the repeat domain's structure does not have an overall extended shape. However, the existence of these turns would probably indicate that the residues in the repeat domain not involved in these turns are even more stretched. This rigidity of the domain could be a perfect prerequisite for the packing when the protein aggregates to PHFs as soon as the charge of the residues is neutralized (e.g., using heparin in vitro). These extended structures can be easily transformed to β-sheets found in the PHFs, while the turns of the free protein can be used as hinges for tighter packing (*37*) (essentially leading to a shorter length of the repeat domain in PHFs than when the protein is free). Extended rodlike structures reported for free tau by rotary shadowing EM (*38*) could be easily formed by the packing of the repeats, which represents also a quite useful structural arrangement for binding to microtubules.

The global hairpin (paperclip) structure in solution supported by fluorescence resonance energy transfer (FRET) and electron paramagnetic resonance (EPR) data (*35*) seems also to be explained by our models. Although the extended nature of the repeat domain might have resulted in a more extended shape of the full length tau, this is not the case, and according to the SAXS data, the overall size of tau is compatible with that of a random coil. It is therefore conceivable that, at least transiently, the protein could fold bringing together the N- and C-termini as described in ref *35*, thus reducing the overall size of the molecule as seen by SAXS. The more extended shape of K23 (a construct lacking the repeat domain) could

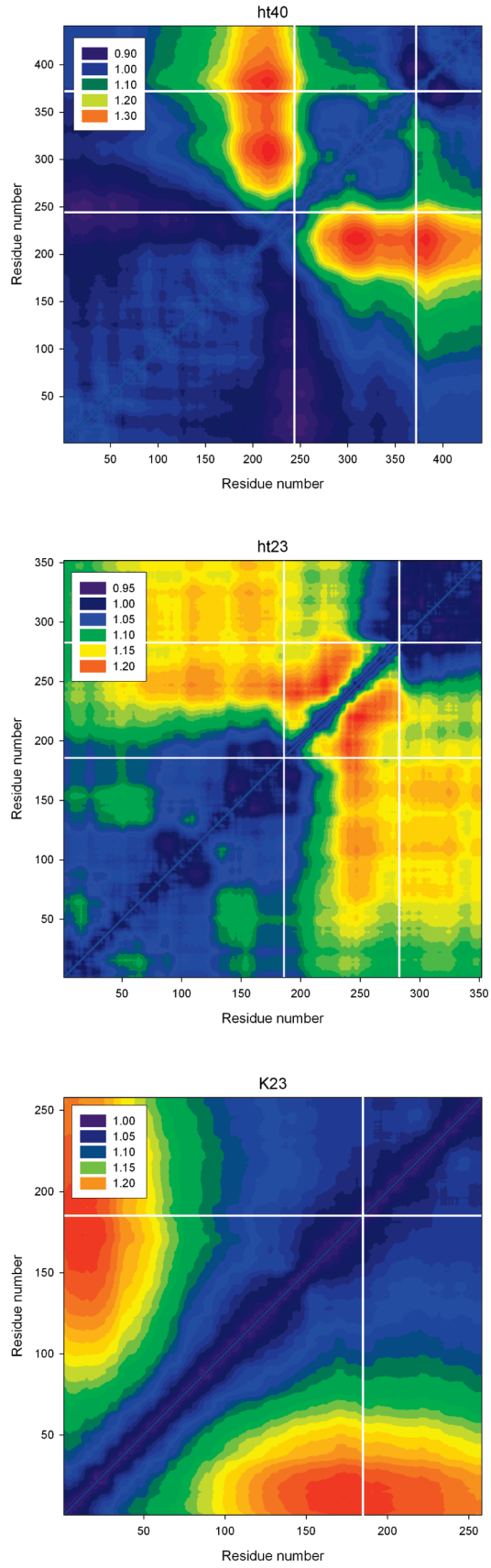

FIGURE 5: C_α–C_α distance plot of ht40, ht23, and K23 using multiple-curve fitting. Each plot point shows the ratio of the average C_α distance of the selected structures to all the structures from the pool. The legend shows the ratio represented by each color. White lines indicate the residues of the repeat domain for ht40 and ht23 and the N- and C-terminal connections for K23. Note that the numbering for ht23 and K23 is not based on the ht40 construct, but it gives the ordinal residue numbers of the two former constructs.

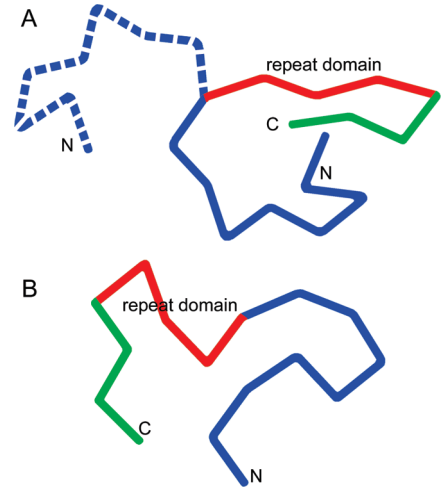

FIGURE 6: Comparison of a schematic model of ht40 with the extended repeat domain and "paperclip" structure (a) with a random one (b). The N-terminal domain is colored blue, the repeat red, and the C-terminus green. The dotted blue segment shows the N-terminus facing away from the rest of the molecule.

be actually caused, at least partly, by the inability to form this paperclip structure (the paperclip structure requires contacts among the N- and C-termini as well as the repeat domain). Figure 6A shows schematically an example ht40 model with extended repeat domain and paperclip structure. If the N-terminus was pointing away from the rest of the protein, then the R_g of the molecule would be significantly larger (granted that the repeats are already quite extended) than that of a random coil with no specific preference for extended repeats (Figure 6B).

Effect of Mutations on the Solution Structure of Tau. All point mutants show little difference from the wild-type constructs. This holds both for the pseudophosphorylation mutants and for the FTDP-17 mutant P301L. The exception to the rule is K18ΔK280, which is however explained by its strong tendency to aggregate caused by the high propensity for β-structure in the repeat domain. This behavior that is very similar to that of the physiological protein indicates that in vivo these mutations have little effect on the shape of the protein in solution. Conversely, any changes in aggregation behavior or in the display of abnormal Alzheimer phospho-epitopes appear to be based on subtle and local changes that do not perturb the overall structure of the protein.

In this study, utilizing point and deletion mutants of tau protein, specific structural propensities of different domains were characterized and the repeat domain was shown to be remarkably extended in comparison with other domains and natively unfolded proteins in general. Our results indicate that the use of SAXS for unfolded proteins, although much less straightforward than for the folded ones, is an invaluable structural analysis tool revealing that the NUPs are not as random as they are often believed to be. Even though the concept of "structure" differs between NUPs and folded proteins, the knowledge of the average structural parameters in solution is important for understanding the functional mechanisms. Multiple SAXS curve analysis from deletion mutants allows one to distinguish between contributions and preferable conformations of individual domains. This possibility allows one to effectively combine the overall structural information from SAXS with the local information

offered by NMR to provide a comprehensive structural picture of NUPs in solution.

ACKNOWLEDGMENT

We thank Dr. Jacek Biernat, Dr. Alexander Marx, and Dr. Eva-Maria Mandelkow (MPG Hamburg) for helpful discussions during this project.

SUPPORTING INFORMATION AVAILABLE

Figures of Guinier and Kratky plots of all wild-type constructs, comparison of scattering patterns at different concentrations, and additional R_g distributions. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Haass, C., and Selkoe, D. J. (2007) Soluble protein oligomers in neurodegeneration: Lessons from the Alzheimer's amyloid β-peptide. *Nat. Rev. Mol. Cell Biol. 8*, 101–112.
2. Margittai, M., and Langen, R. (2004) Template-assisted filament growth by parallel stacking of tau. *Proc. Natl. Acad. Sci. U.S.A. 101*, 10278–10283.
3. Binder, L. I., Guillozet-Bongaarts, A. L., Garcia-Sierra, F., and Berry, R. W. (2005) Tau, tangles, and Alzheimer's disease. *Biochim. Biophys. Acta 1739*, 216–223.
4. Mandelkow, E., von Bergen, M., Biernat, J., and Mandelkow, E. (2007) Structural principles of tau and the paired helical filaments of Alzheimer's disease. *Brain Pathol. 17*, 83–90.
5. Drubin, D. G., and Kirschner, M. W. (1986) Tau protein function in living cells. *J. Cell Biol. 103*, 2739–2746.
6. Butner, K. A., and Kirschner, M. W. (1991) Tau protein binds to microtubules through a flexible array of distributed weak sites. *J. Cell Biol. 115*, 717–730.
7. Mandelkow, E. M., Mandelkow, E., and Milligan, R. A. (1991) Microtubule dynamics and microtubule caps: A time-resolved cryo-electron microscopy study. *J. Cell Biol. 114*, 977–991.
8. Goedert, M., Spillantini, M. G., Potier, M. C., Ulrich, J., and Crowther, R. A. (1989) Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: Differential expression of tau protein mRNAs in human brain. *EMBO J. 8*, 393–399.
9. Goedert, M., Spillantini, M. G., Jakes, R., Rutherford, D., and Crowther, R. A. (1989) Multiple isoforms of human microtubule-associated protein tau: Sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron 3*, 519–526.
10. Eliezer, D., Barré, P., Kobaslija, M., Chan, D., Li, X., and Heend, L. (2005) Residual structure in the repeat domain of tau: Echoes of microtubule binding and paired helical filament formation. *Biochemistry 44*, 1026–1036.
11. Bunker, J. M., Kamath, K., Wilson, L., Jordan, M. A., and Feinstein, S. C. (2006) FTDP-17 mutations compromise the ability of tau to regulate microtubule dynamics in cells. *J. Biol. Chem. 281*, 11856–11863.
12. von Bergen, M., Friedhoff, P., Biernat, J., Heberle, J., Mandelkow, E. M., and Mandelkow, E. (2000) Assembly of tau protein into Alzheimer paired helical filaments depends on a local sequence motif ((306)VQIVYK(311)) forming β structure. *Proc. Natl. Acad. Sci. U.S.A. 97* (10), 5129–5134.
13. von Bergen, M., Barghorn, S., Li, L., Marx, A., Biernat, J., Mandelkow, E. M., and Mandelkow, E. (2001) Mutations of tau protein in frontotemporal dementia promote aggregation of paired helical filaments by enhancing local β-structure. *J. Biol. Chem. 276* (51), 48165–48174.
14. Cleveland, D. W., Hwo, S. Y., and Kirschner, M. W. (1977) Physical and chemical properties of purified tau factor and the role of tau in microtubule assembly. *J. Mol. Biol. 116*, 227–247.
15. Lee, G., Cowan, N., and Kirschner, M. (1988) The primary structure and heterogeneity of tau protein from mouse brain. *Science 239*, 285–288.
16. Barghorn, S., and Mandelkow, E. (2002) Toward a unified scheme for the aggregation of tau into Alzheimer paired helical filaments. *Biochemistry 41* (50), 14885–14896.
17. 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 β-structure. *J. Biol. Chem. 269*, 24290–24297.
18. Santarella, R. A., Skiniotis, G., Goldie, K. N., Tittmann, P., Gross, H., Mandelkow, E., Mandelkow, E., and Hoenger, A. (2004) Surface-decoration of microtubules by human tau. *J. Mol. Biol. 339*, 539–553.
19. Mukrasch, M. D., Biernat, J., von Bergen, M., Griesinger, C., Mandelkow, E., and Zweckstetter, M. (2005) Sites of tau important for aggregation populate β-structure and bind to microtubules and polyanions. *J. Biol. Chem. 280*, 24978–24986.
20. Sibille, N., Sillen, A., Leroy, A., Wieruszeski, J., Mulloy, B., Landrieu, I., and Lippens, G. (2006) Structural impact of heparin binding to full-length Tau as studied by NMR spectroscopy. *Biochemistry 45*, 12560–12572.
21. Schneider, A., Biernat, J., von Bergen, M., Mandelkow, E., and Mandelkow, E. M. (1999) Phosphorylation that detaches tau protein from microtubules (Ser262, Ser214) also protects it against aggregation into Alzheimer paired helical filaments. *Biochemistry 38*, 3549–3558.
22. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) Intrinsically disordered protein. *J. Mol. Graphics Modell. 19*, 26–59.
23. Uversky, V. N. (2002) What does it mean to be natively unfolded? *Eur. J. Biochem. 269*, 2–12.
24. Uversky, V. N. (2002) Natively unfolded proteins: A point where biology waits for physics. *Protein Sci. 11*, 739–756.
25. Roessle, M. W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., Gehrmann, T., Konarev, P., Round, A., Fiedler, S., Hermes, C., and Svergun, D. (2007) Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *J. Appl. Crystallogr. 40*, s190–s194.
26. Konarev, P., Volkov, V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. (2003) PRIMUS: A Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallogr. 36*, 1277–1282.
27. Guinier, A. (1939) La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. *Ann. Phys. 12*, 161–237.
28. Svergun, D. I. (1992) Determination of the regularization parameter in indirect transform methods using perceptual criteria. *J. Appl. Crystallogr. 25*, 495–503.
29. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Hasan, M. Z., Pande, V. S., Ruczinski, I., Doniach, S., and Plaxco, K. W. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A. 101*, 12491–12496.
30. Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995) CRYSOL: A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr. 28*, 768–773.
31. Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. *J. Am. Chem. Soc. 129*, 5656–5664.
32. Bernadó, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W. H., and Blackledge, M. (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. *Proc. Natl. Acad. Sci. U.S.A. 102*, 17002–17007.
33. Eyal, E., Najmanovich, R., McConkey, B. J., Edelman, M., and Sobolev, V. (2004) Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins. *J. Comput. Chem. 25*, 712–724.
34. Glatter, O. (1977) A new method for the evaluation of small-angle scattering data. *J. Appl. Crystallogr. 10*, 415–421.
35. Jeganathan, S., von Bergen, M., Brutlach, H., Steinhoff, H., and Mandelkow, E. (2006) Global hairpin folding of tau in solution. *Biochemistry 45*, 2283–2293.
36. Mukrasch, M. D., Markwick, P., Biernat, J., Bergen, M. V., Bernadó, P., Griesinger, C., Mandelkow, E., Zweckstetter, M., and Blackledge, M. (2007) Highly populated turn conformations in

natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. *J. Am. Chem. Soc. 129*, 5235–5243.

37. Andronesi, O. C., Bergen, M. V., Biernat, J., Seidel, K., Griesinger, C., Mandelkow, E., and Baldus, M. (2008) Characterization of Alzheimer's-like Paired Helical Filaments from the Core Domain of Tau Protein Using Solid-State NMR Spectroscopy. *J. Am. Chem. Soc. 130* (18), 5922–5928.
38. Wille, H., Drewes, G., Biernat, J., Mandelkow, E. M., and Mandelkow, E. (1992) Alzheimer-like paired helical filaments and antiparallel dimers formed from microtubule-associated protein tau in vitro. *J. Cell Biol. 118*, 573–584.
39. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto, CA.

BI800900D