J|A|C|S
ARTICLES
Published on Web 03/27/2007

Highly Populated Turn Conformations in Natively Unfolded Tau Protein Identified from Residual Dipolar Couplings and Molecular Simulation

Marco D. Mukrasch,[†] Phineus Markwick,[‡] Jacek Biernat,[§] Martin von Bergen,[§] Pau Bernadó,[‡,||] Christian Griesinger,[†] Eckhard Mandelkow,[§] Markus Zweckstetter,*[,†,⊥] and Martin Blackledge*[,‡]

Contribution from the Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany, Institut de Biologie Structurale Jean-Pierre Ebel, CEA, CNRS, UJF UMR 5075, 41 Rue Jules Horowitz, Grenoble 38027, France, Max-Planck-Arbeitsgruppen für strukturelle Molekularbiologie Hamburg, Notkestrasse 85, c/o DESY, 22067 Hamburg, Germany, and DFG Research Center for the Molecular Physiology of the Brain (CMPB), Göttingen, Germany

Received December 16, 2006; E-mail: martin.blackledge@ibs.fr; mzwecks@gwdg.de

Abstract: Tau, a natively unstructured protein that regulates the organization of neuronal microtubules, is also found in high concentrations in neurofibrillary tangles of Alzheimer's disease and other neurodegenerative disorders. The conformational transition between these vastly different healthy and pathological forms remains poorly understood. We have measured residual dipolar couplings (RDCs), *J*-couplings, and nuclear Overhauser enhancement (NOE) in construct K18 of tau, containing all four repeat domains R1-R4. NH^N RDCs were compared with prediction on the basis of a statistical model describing the intrinsic conformational sampling of unfolded proteins in solution. While local variation and relative amplitude of RDCs agrees with propensity-based prediction for most of the protein, homologous sequences in each repeat domain (DLKN, DLSN, DLSK, and DKFD in repeats R1−R4) show strong disagreement characterized by inversion of the sign of the central couplings. Accelerated molecular dynamic simulations (AMD) in explicit solvent revealed strong tendencies to form turns, identified as type I β-turns for repeats R1−R3. Incorporation of the backbone dihedral sampling resulting from AMD into the statistical coil model closely reproduces experimental RDC values. These localized sequence-dependent conformational tendencies interrupt the propensity to sample more extended conformations in adjacent strands and are remarkably resistant to local environmental factors, as demonstrated by the persistence of the RDC signature even under harsh denaturing conditions (8 M urea). The role that this specific conformational behavior may play in the transition to the pathological form is discussed.

Introduction

Natively unfolded proteins play key roles in the development of neurodegenerative diseases,[1] often characterized by the transformation from soluble, physiological forms to aggregated, pathological forms of the same protein.[2] Amyloid plaques or neurofibrillary tangles (NFTs) found in brains of patients with Alzheimer's disease are fibrous assemblies made up of Aβ peptide[3] or microtubule associated protein (MAP) tau in the form of paired helical filaments (PHFs).[4] In its physiological form tau is an intrinsically unfolded protein believed to regulate the organization of neuronal microtubules.[5] The molecular mechanism associating the conformational behavior of tau in solution with the process of pathological aggregation remains poorly understood, principally due to the experimental difficulties encountered when studying unfolded polypeptide chains using classical tools of structural biology.

A number of isoforms of protein tau are found in Alzheimer PHFs, varying in length from 352 to 441 amino acids.[6] Three or four imperfect repeat domains are present in the C-terminal half of all isoforms. The repeat domains form the core of PHFs and have been proposed to be largely responsible for PHF assembly.[7] Although tau is essentially hydrophilic, two short, weakly hydrophobic hexapeptides occurring at the beginning of the repeat domains 2 and 3 are known to play an essential role in the formation of PHFs and were subsequently proposed

[†] Max Planck Institute for Biophysical Chemistry.
[‡] Institut de Biologie Structurale Jean-Pierre Ebel.
[§] Max-Planck-Arbeitsgruppen für strukturelle Molekularbiologie.
[||] Current address: Institut de Recerca Biomèdica, Josep Samitier 1-5, 08028-Barcelona, Spain.
[⊥] DFG Research Center for the Molecular Physiology of the Brain (CMPB).

(1) Dobson, C. M. *Phil. Trans. R. Soc., Ser. B* **2001**, *356*, 133−145.
(2) Kelly, J. W. *Curr. Opin. Struct. Biol.* **1998**, *8*, 101−106.
(3) Koo, E. H.; Lansbury, P. T., Jr.; Kelly, W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9989−9990.
(4) Mandelkow, E.-M.; Mandelkow, E. *Trends Cell Biol.* **1998**, *8*, 425−427.
(5) Gustke, N.; Trinczek, B.; Biernat, J.; Mandelkow, E.-M.; Mandelkow, E. *Biochemistry* **1994**, *33*, 9511−9522.
(6) Trinczek, B.; Biernat, J.; Baumann, K.; Mandelkow, E.-M.; Mandelkow, E. *Mol. Biol. Cell* **1995**, *6*, 1887−1902.
(7) Wille, H.; Drewes, G.; Biernat, J.; Mandelkow, E.-M.; Mandelkow, E. *J. Cell. Biol.* **1995**, *118*, 573−584.

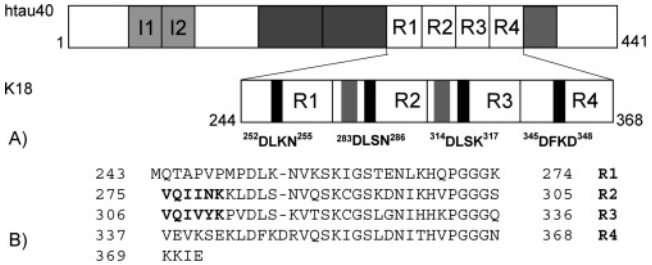

Figure 1. Tau protein and K18. (A) Position of the K18 construct with respect to the htau40 isoform of tau. Tau contains four pseudorepeat domains (R1−R4) involved in binding microtubules and forming the core of PHFs. The position of the amino acids showing specific conformational behavior on the basis of RDCs is shown. The position of hexapeptides at the start of repeats R2 and R3 are shown as gray boxes. (B) Primary sequence of the K18 construct, showing the position of repeat domains 1−4, containing 30−31 residues each. The position of hexapeptides at the start of repeats R2 and R3 are shown in bold type.

as seeds of aggregation (Figure 1).[8] Recent studies of the repeat domains contained in constructs K18 and K19 used CD and FTIR spectroscopy[9] and subsequently NMR chemical shifts[10,11] to identify β-sheet propensities in these peptides. These regions were also shown to mediate the interaction with microtubules and polyanions.[10]

Nuclear magnetic resonance (NMR) is uniquely suited to the study of unfolded proteins in solution,[12] giving access to site-specific averages over all sampled conformations of the ensemble.[13−15] In particular residual dipolar couplings (RDCs)[16] report on time and ensemble-averaged conformations[17−19] and can be used to understand both the structure and dynamics of disordered proteins.[20,21] RDCs are sensitive to the presence of structure in flexible chains at levels that are difficult to detect using other techniques and offer an increasingly promising tool to determine quantitative levels of order in unfolded proteins.[22−25] Following from previous NMR studies of tau protein in solution,[10,11,26,27] in this study we use RDCs to characterize conformational behavior in the K18 construct of the protein.

We recently developed a statistical coil model, termed flexible−meccano (FM), for the prediction of experimental properties of disordered proteins in solution.[28] FM describes the intrinsic conformational sampling of unfolded proteins by constructing an ensemble from randomly sampled residue-specific ϕ/ψ propensities present in a folded protein database. Close agreement was found between experimental and simulated RDCs for a number of unfolded proteins, establishing FM as a basic tool for studying unfolded proteins in solution.[29]

Although statistical sampling from structural databases is well suited to the study of highly disordered systems such as natively unfolded proteins, local conformational preferences that diverge from predicted behavior cannot easily be characterized from such an analysis. Examples of this kind of behavior are found here in four homologous regions of the construct K18 of tau protein, manifest as local inversion of the sign of measured RDCs. To study the molecular basis of these specific structural propensities, we have complemented the FM approach with state-of-the-art molecular simulation techniques. While classical molecular dynamics (MD) calculations are routinely used to simulate fast stochastic motions in the sub-microsecond range in folded proteins, we have used a biased-potential simulation approach, known as accelerated MD,[30] to statistically sample the much larger conformational subspace available to unfolded proteins. The backbone dihedral angle distributions extracted from simulation are then incorporated into the FM molecular construction algorithm. This combination of molecular simulation and statistical conformational sampling provides rare insight into the structure and dynamics of tau protein and holds great promise for the study of unfolded proteins in general.

Results and Discussion

Tau aggregation from the healthy, unfolded form to pathological filament-forming PHFs is driven by a transition from unfolded to β sheet structure.[31] Recent studies have identified two hexapeptides, located around the second and third repeat sequences in the core region of tau (^{275}VQIINK^{280} and ^{306}VQIVYK^{311} in repeats R2 and R3, respectively), that are sufficient for PHF formation, represent a minimal tau−tau interaction motif,[9] and have been shown to mediate the interaction of tau with microtubules (Figure 1).[10] Understanding the native behavior of the core domain of the intrinsically unfolded form of tau in solution is therefore critical to identifying the basis of this conformational transition. Here we have measured NH^N RDCs in strained gels, interproton nuclear Overhauser enhancement (NOE), and $^3J_{NH-\alpha H}$ couplings from K18, a construct comprising all four repeat domains and therefore containing the key regions responsible for both normal and pathological activity (Figure 2). Note that in this figure we present RDCs in an inverted form, to facilitate comparison with the other experimental measurements.

The expected distribution of RDCs in unfolded proteins[21] is globally followed in K18, with negative couplings throughout the chain. The experimental data nevertheless present notable

(8) von Bergen, M.; Barghorn, S.; Li, L.; Marx, A.; Biernat, J.; Mandelkow, E.-M.; Mandelkow, E. *J. Biol. Chem.* **2001**, *276*, 48165−48174.
(9) von Bergen, M.; Freidhoff, P.; Biernat, J.; Heberle, J.; Mandelkow, E.-M.; Mandelkow, E. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *97*, 5129−5134.
(10) Mukrasch, M.; Biernat, J.; von Bergen, M.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M. *J. Biol. Chem.* **2005**, *280*, 24978−24986.
(11) Eliezer, D.; Barré, P.; Kobaslija, M.; Chan, D.; Li, X.; Heend, L. *Biochemistry* **2005**, *44*, 1026−1036.
(12) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607−3622.
(13) Wirmer, J.; Schlörb, C.; Schwalbe, H. In *Protein Folding Handbook, Part I, Vol. I*; Buchner, J., Kiefhaber, T., Eds.; Wiley-VCH: New York, 2005.
(14) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 476−477.
(15) Choy, W. Y.; Forman-Kay, J. D. *J. Mol. Biol.* **2001**, *308*, 1011−32.
(16) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114.
(17) Prestegard, J. H.; al-Hashimi, H. M.; Tolman, J. R. *Q. Rev. Biophys.* **2000**, *33*, 371−424.
(18) Blackledge, M. *Prog. NMR Spectrosc.* **2005**, *46*, 23−61.
(19) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Bruschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098−6107.
(20) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487−489.
(21) Louhivouri, M.; Pääkkönen, K.; Fredriksson, K.; Permi, P.; Lounila, J.; Annila, A. *J. Am. Chem. Soc.* **2003**, *125*, 15647−15650.
(22) Fieber, W.; Kristjansdottir, S.; Poulsen, F. M. *J. Mol. Biol.* **2004**, *339*, 1191−1199.
(23) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J. A.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2004**, *34*, 1131−1142.
(24) Meier, S.; Güthe, S.; Kiefhaber, T.; Grzesiek, S. *J. Mol. Biol.* **2004**, *344*, 1051.
(25) Bertoncini, C. W.; Jung, Y. S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1430−1435.
(26) Landrieu, I.; Lacosse, L.; Leroy, A.; Wieruszeski, J.-M.; Trivelli, X.; Sillen, A.; Sibille, N.; Schwalbe, H.; Saxena, K.; Langer, T.; Lippens, G. *J. Am. Chem. Soc.* **2005**, *128*, 3575−3583.
(27) Sibille, N.; Sillen, A.; Leroy, A.; Wieruszeski, J.-M.; Mulloy, B.; Landrieu, I.; Lippens, G. *Biochemistry* **2006**, *45*, 12560−12572.
(28) Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002−17007.
(29) Bernado, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2005**, *127*, 17968−17969.
(30) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2005**, *127*, 13778−13779.
(31) von Bergen, M.; Barghorn, S.; Niernat, J.; Mandelkow, E.-M.; Mandelkow, E. *Biochem. Biophys. Acta* **2005**, *1739*, 158−166.

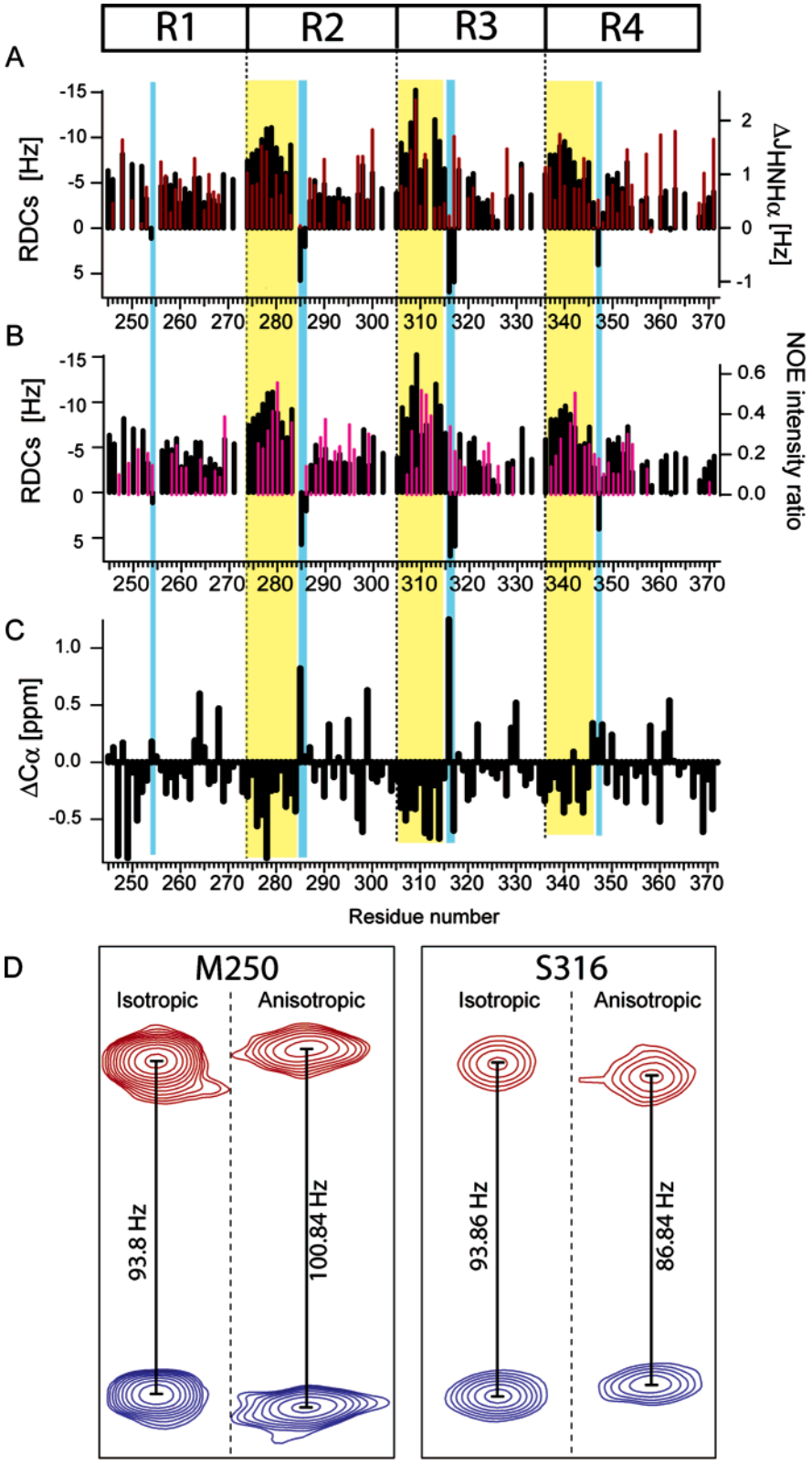

Figure 2. Summary of experimental data measured from the K18 construct of tau. Three regions exhibiting large negative RDC values (indicating extended structure) are highlighted in yellow. Inverted RDC values (turn propensity) are indicated in blue. (A) RDCs measured in polyacrylamide gel (black) compared to secondary $^3J(H^NH^\alpha)$-coupling values (red). (B) RDCs measured in polyacrylamide gel (black) compared to $H^N_i-H^\alpha_{i-1}$ NOE cross-peak intensities (magenta) relative to diagonal cross-peak intensity. (C) C^α secondary chemical shift differences from random coil values,[10] indicating occurrence of secondary structure elements, are depicted as comparison. (D) Isotropic and anisotropic (aligned in polyacrylamide gel) doublets from K18.

characteristics, with larger amplitude RDCs occurring in the four regions at the beginning of the repeats, 248−252, 275−283, 306−314, and 337−342. The two central strands correspond to the weakly hydrophobic repeat regions previously estimated to exhibit β-sheet propensities on the basis of CD, FTIR,[9] and NMR secondary chemical shifts.[10] Large negative

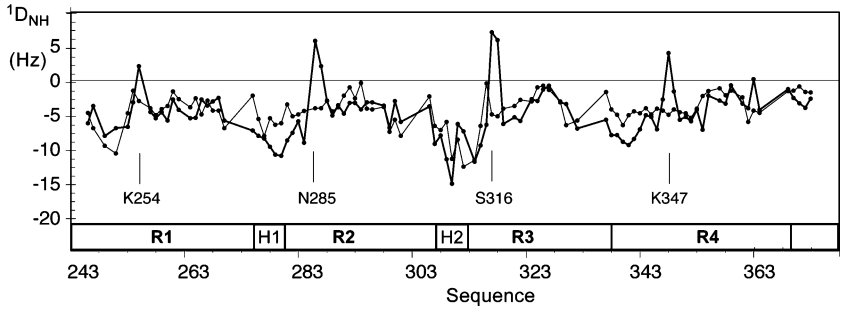

Figure 3. Residual dipolar couplings measured in K18 compared to FM simulation. Experimental NH^N RDCs from tau K18 construct aligned in polyacrylamide gel (thick line), compared to predicted NH^N RDCs using the statistical coil model FM (thin line). The 50 000 conformations were simulated and RDCs predicted using PALES. Average RDCs are uniformly scaled to reproduce experimental values. Only values for which experimental data are available are shown. The position of the domains R1−R4 and H1, H2 are indicated. Experimental error is estimated at 0.2 Hz.

RDCs are associated with locally more extended conformations, due to the form of the angular probability function of the dipolar orientation governing RDC averaging, $\int P_2(\cos\theta)\,d\Omega$. In more extended parts of the chain the NH interaction direction will tend to be perpendicular to the external field that is also the direction of the longer axes of the anisotropic cavities in strained gels. This function becomes more negative when more β-extended structures are populated. Larger RDCs measured in K18 are therefore in qualitative agreement with β-sheet propensity in these regions. A similar distribution of RDCs is found when the protein is aligned in bacteriophage (Supporting Information), with small differences probably due to additional electrostatic alignment in this medium.[32] In the following we will discuss the somewhat simpler interpretation of gel-aligned RDCs. Importantly the current results also indicate that all four repeat domains contain homologous sites (L253−K254, L284−S285, L315−S316, and L346−K347) exhibiting RDCs with inverted sign. The occurrence of such sign inversion in unfolded chains has been interpreted as an increase in local propensity to form turns or α-helical conformations.[22−24,33]

Comparison of experimental ^3J couplings (dependent on the distribution of ϕ angles) with random coil prediction confirms the presence of specific conformational behavior in these same regions. While positive $\Delta^3J_{NH-H\alpha}$ ($J_{exp}-J_{random\ coil}$) values dominate throughout, indicating a tendency toward extended structure, low values, again indicative of turns, are found for K254, N285, and S316. Notably, despite following the general pattern of the measured RDCs elsewhere, sequential $H^{\alpha}_{i-1}-H^{N}_{i}$ nOes measured along the chain do not show evidence of differential behavior in these regions (vide infra).

Comparison with Prediction Using Flexible−Meccano. Two similar approaches have been developed to describe local conformational sampling in unfolded chains.[28,34] One of these approaches (FM) is used here to provide more detailed insight into the local conformational behavior of K18. FM can be used to predict RDCs in unfolded chains using statistical sampling of amino-acid-specific conformations extracted from a structural database. RDCs simulated using this approach are compared to the experimentally measured values in Figure 3. Simulated values are scaled by a common factor to match the experimental points. Both local variation and relative amplitude of RDCs along the sequence are broadly reproduced by prediction. Inverted couplings measured in the vicinity of homologous positions of the four regions L253−K254, L284−S285, L315−S316, and L346−K347 are however not predicted by the model. We conclude that conformational sampling in these regions is more complex and is due to interactions that are not described by residue-specific propensities.

In an initial approach we modified the FM algorithm to search Ramachandran space at key amino acid positions. RDCs from the fragment 270−300 were extracted from FM simulations of the entire protein, in which the central amino acid, S285, sampled different conformational wells. Figure 4 illustrates the dependence of RDCs on ϕ/ψ sampling, with inverted signs of NH RDCs in β-extended and α-helical regions. Although this analysis verifies that turn conformations can give rise to the experimentally measured RDCs in these four regions, it is also evident that unique descriptions of conformational sampling cannot be extracted exclusively on the basis of NH RDCs. To gain further insight into the specific behavior of these regions and better understand the physical basis of the local conformational preferences, we have therefore used MD simulation techniques. Due to the extensive conformational freedom available to unfolded proteins, it is not clear whether experimental parameters such as RDCs can be reproduced from standard MD. Here we have used accelerated MD (AMD) simulations in explicit water to predict the behavior of short peptides centered on the positions of interest.

Accelerated Molecular Dynamics Simulations. AMD uses biased potential sampling to enhance the probability of transition between low-energy conformational states and therefore offers an efficient method for accessing both longer time scale motions and broader statistical sampling.[30]

Backbone dihedral sampling for regions in the vicinity of the proposed turns is shown in Figure 5 and compared to the sampling present in the database-dependent 50 000-strong FM ensemble. Before analyzing regions where specific local structures are suspected, we verified that conformational sampling from AMD generally reproduces the propensity-based populations in regions where the database conformations do reproduce experimental RDCs. Examples are shown in Figure 5 where ϕ/ψ populations of the amino acids preceding and succeeding the regions of interest strongly resemble dihedral angle sampling extracted from the FM analysis. Histograms of the ψ-angle distribution for these amino acids are shown in Figure 6. Although exceptions exist, in particular P251 where AMD

(32) Skora, L.; Cho, M. K.; Kim, H. Y.; Becker, S.; Fernandez, C. O.; Blackledge, M.; Zweckstetter, M. *Angew. Chem., Int. Ed.* **2006**, *45*, 7012−7015.
(33) Louhivouri, M.; Fredriksson, K.; Pääkkönen, K.; Permi, P.; Annila, A. *J. Biomol. NMR* **2004**, *29*, 517−524.
(34) Jha, A. K.; Colubri, A.; Freed, K.; Sosnick, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099−13104.

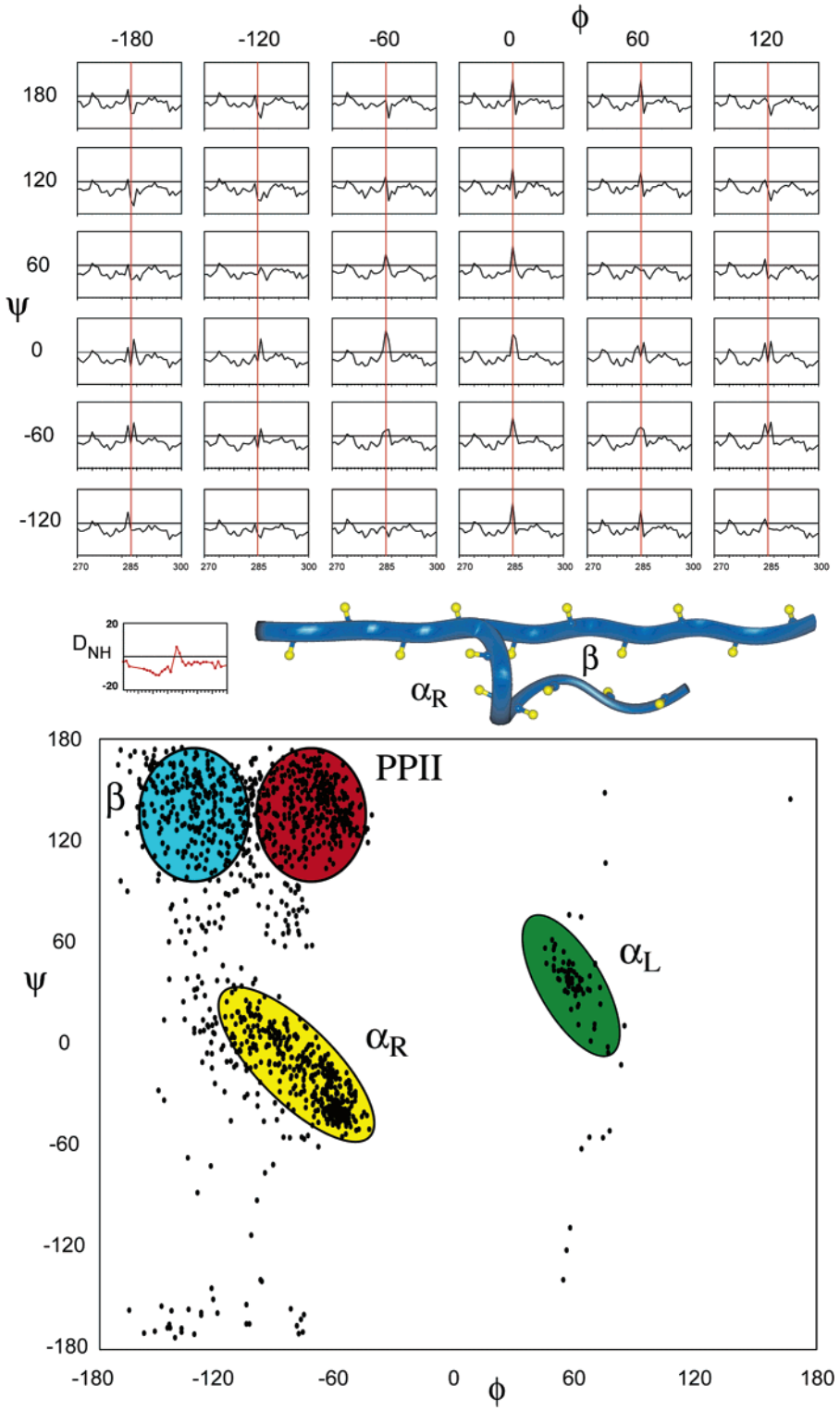

Figure 4. Backbone conformational dependence of NH^N RDCs. Top: Dependence of NH^N RDCs in the range 271−300 on the local conformational sampling of residue S285. The 50 000 conformers were calculated in each case. S285 sampled distributions are centered on the ϕ/ψ values shown, in Gaussian wells of width 10°. All other amino acids obeyed native propensities. The red line indicates the position of S285. The experimental data are shown in the inset (red). Also shown are effective orientations of NH vectors (yellow) in the presence of β-extended and α-helical conformations. Bottom: Ramachandran plot showing conformational distribution drawn from the structural database. The most populated conformational regions discussed in the text are indicated: polyproline II (PPII); β-extended (β); right-handed α helix (α_R); left-handed α helix (α_L).

predicts less extended sampling, for most of the amino acids examined the distributions are similar.

Figures 5 and 6 also reveals that certain amino acids in these four sequences show specific propensities that are significantly different to the native propensities (Figures 5B and 6B). Conformational sampling indicative of type I β turns for residues {D283, L284, S285, N286} and {D314, L315, S316, K317} is observed in repeats R2 and R3, with two consecutive conformations in the α_R region of Ramachandran space, preceded and succeeded by amino acids sampling more extended conforma-

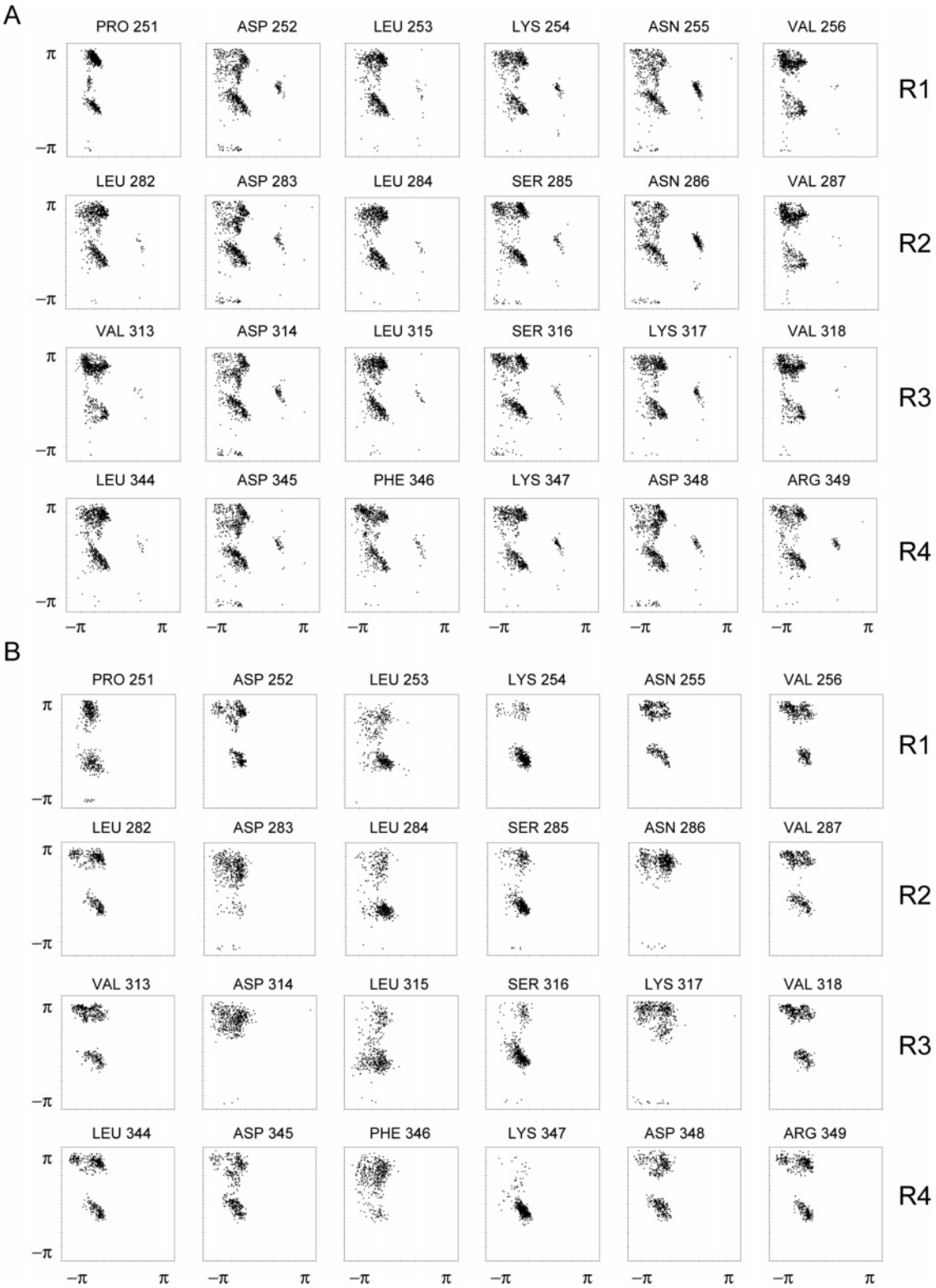

Figure 5. Conformational sampling from AMD and FM simulation. (A) Conformational sampling from the FM simulation for amino acids in the vicinity of the regions exhibiting turn conformations, (251−256), (282−287), (313−318), and (344−349). (B) Sampling from the AMD simulation for the same amino acids as shown in (A).

tions. The same pattern is seen for repeat 1, with L253 and K254 populating the helical region, but preceding and following residues more closely resembling their native propensities. In contrast repeat R4 also samples a turn conformation but only at the central residue K347. As the preceding residue tends to more extended conformation, this motif is closest to a type VIII β turn. Since the turns are formed by transition from extended to α_R region, it is not surprising that the $H^{\alpha}{}_{i-1}-H^{N}{}_{i}$ nOes are a poor indicator of this transition.

We can determine whether the conformational sampling observed from AMD is in agreement with the experimental data, by replacing the standard amino-acid-specific propensities used

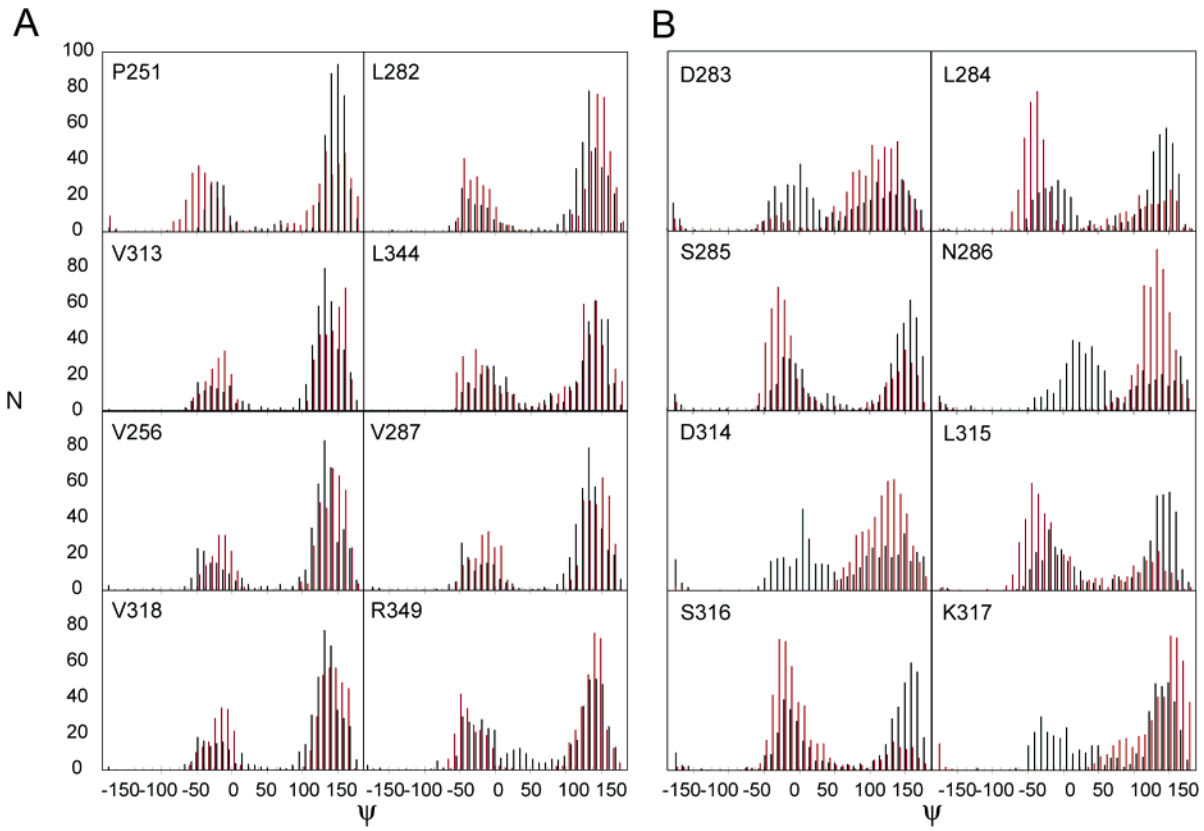

Figure 6. ψ dihedral angle sampling from AMD and amino-acid specific propensities. (A) Histograms showing distribution of 500 ψ angles from FM (black) and AMD simulation (red) of the 8 N and C terminal amino acids of the regions shown in Figure 5A (bin width 9°). (B) Histograms showing distribution of 500 ψ angles from FM (black) and AMD simulation (red) of the 4 central amino acids of regions shown in Figure 5A,B in repeats R2 and R3.

in the FM algorithm with backbone conformations extracted from the trajectory. AMD populations shown in Figure 5B were used to replace the propensity-based sampling (Figure 5A) for the tetrapeptides 252−255, 283−286, 314−317, and 345−348. Database propensities were retained for all remaining residues, and 50 000 conformers were constructed on the basis of this hybrid sampling. The results (Figure 7) show that sampling from AMD is indeed in agreement with the experimental RDCs, at least for repeats R1−R3, indicating that not only the conformations but also their populations are accurately simulated. Importantly, the simulations predict that the configurations of the four regions involve specific conformational sampling in up to four sequential amino acids. Analysis of the AMD simulations failed to identify a unique interaction responsible for the propensity of this generic sequence (DLKN, DLSN, and DLSK in repeats R1−R3, respectively) to induce the observed highly specific conformational properties. Rather a combination of van der Waals obstruction and transient hydrogen bonds were observed as principal factors. Nevertheless this striking result illustrates the power of combining MD simulation and FM-based statistical sampling.

These observations agree with previous NMR studies, where chemical shift analysis identified propensity for turns in these regions and slight β-sheet propensity in the strands preceding them.[10] The simulation shown in Figure 7 explicitly incorporated additional β-strand conformations at the levels indicated in the previous publication (16% for the strand 274−283 (R2), 24% for strand 306−313 (R3), and 13% for strand 336−345 (R4)). Calculations run without these additional propensities resulted in slightly decreased couplings in these regions (Supporting Information). We have also run AMD simulations on these three strands and see a slightly higher propensity for extended conformations than found in the FM conformations (data not shown). ϕ-dihedral angle dependent $^3J_{NH-H\alpha}$ couplings also indicate the presence of turns (Figure 2). We have predicted *J*-couplings from throughout the chain and compared these with experimental coupling constants (Figure S4, Supporting Information). The results demonstrate reasonable correlation, both in the turn regions and in the rest of the chain. In addition $H^{\alpha}_{i-2}-H^{N}_{i}$ distances between Q276−I278, D283−S285, S285−V287, K309−V313, and D314−S316 were detected, indicating more compact structures in these regions, in agreement with turn conformations.

Finally to probe the stability of these observed structural propensities, we have repeated RDC measurements on K18 dissolved in 8 M urea. Remarkably, the profile of experimental RDCs is very similar (Supporting Information), with sign inversion at the same positions. The sequence-dependent propensity to form turns is therefore remarkably stable, persisting even in the presence of extreme denaturing conditions.

β Turns in Tau: Inhibitor or Precursor of Filamentous Structure? It is interesting to speculate on the role that these conformational properties may play in the context of the transition between healthy and pathological forms of tau. The second and third turn regions, centered on S285 and S316, immediately follow strands of β-sheet propensity (274−283, 306−313) containing known aggregation−nucleation sites that are sufficient for formation of PHFs. Additional turn regions are expected to exist due to the GGG motif present in the terminal part of the 30 residue repeats, and this motif is preceded

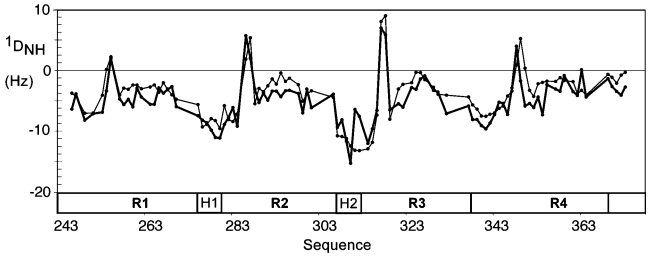

Figure 7. Combined AMD and statistical coil prediction of RDCs. Shown is the prediction of NH^N RDCs from tau K18 construct using a combined statistical coil/MD model (thin line) compared to experimental values (thick). FM was used with specific sampling from the AMD calculation for strands 252−255, 283−286, 314−317, and 345−348. The 50 000 conformations were simulated and RDCs predicted using PALES. Average RDCs are scaled to reproduce the range of experimental values. The position of hexapeptides H1 and H2 is indicated.

by some transient β-structure. Both repeats 2 and 3 therefore contain some propensity for sheet−turn−sheet−turn motifs in solution. β turns may play a role in disrupting the formation of amyloid fibrils,[35] in a similar way to proline residues. Indeed the insertion of a proline into the hexapeptide motif in repeat R3 has been shown to interrupt β-structure.[36] It will therefore be interesting to determine whether the observed turns are present in the aggregated form of tau (some evidence exists for this on the basis of EPR of spin-labeled tau aggregates)[37] or whether they represent an autoinhibitory mechanism to avoid aggregation.

Long-range structural information reporting on the soluble form of tau is available from FRET measurements that have recently been used to demonstrate unusual proximity of sites 291 and 310 and 310 and 322.[38] This is in qualitative agreement with the presence of a turn at position 302 and the highly populated β-turn at position S316 described here, even in the presence of extended sampling either side of the motif. We have also calculated an effective radius of gyration from the ensemble shown here, equal to 35.5 Å, that turns out to agree with experimentally derived values from SAXS studies (personal communication with Dmitri Svergun).

Conclusions

Recent studies have shown the power of statistical coil models to describe the intrinsic conformational sampling of unfolded proteins. The question remained how one could address conformational sampling of disordered states when database approaches did not reproduce experimental data, due to more complex transient conformational sampling. Simulation using standard MD techniques presents a number of shortcomings, in particular restricted sampling that could fail to interrogate the vast phase space available to a disordered system. Here we have used biased potential approaches that allow broader statistical mechanical sampling of the conformational subspace to study the structure and dynamics of four repeat regions of tau whose conformational behavior clearly diverges from prediction on the basis of simple amino-acid-specific propensities.

Four peptides in the repeat domains, ^{252}DLKN^{255}, ^{283}DLSN^{286}, ^{314}DLSK^{317}, and ^{345}DKFD^{348} show high propensities to form turns. AMD strongly indicates that these sequences fold locally and provide otherwise inaccessible detail of the nature of the turn conformations and the participating amino acids. The peptides in repeats 1−3 appear to adopt classical type I β turns and reproduce experimental data very well, while the peptide in repeat 4 resembles a β VIII turn in the simulation but presents poorer reproduction of experimental data. This sequence-dependent structural propensity is remarkably resistant to local environmental factors, as demonstrated by the persistent turn behavior even under harsh denaturing conditions (8 M urea). Importantly these turns all follow regions with β-sheet propensity, including hexapeptides occurring at the beginning of the repeat domains R2 and R3 previously identified as potential seeds of aggregation. Whether the strong propensity to form turns represents an inhibitory mechanism present in the healthy form, to avoid formation of continuous β-fibrils for example, or whether these structured regions are present in the pathological form remains an open and crucially important question.

Methods

Experimental NMR Measurements. K18 protein expression and backbone assignment have been described.[10] ^{15}N−^{1}H residual dipolar couplings were measured in 7% uncharged polyacrylamide gels and in Pf1 phages (5 mg/mL). The sequence of an IPAP-HSQC (16 scans, 1.2 ms as relaxations delay) was used[39] and measured on a Bruker 800 DRX spectrometer. NOESY-HSQC (4 scans, relaxation delay 1.3 ms, mixing time 120 ms) was measured on a Bruker DRX 800 spectrometer. Spectra were recorded at 5 °C and processed and analyzed using NMRPipe[40] and Sparky 3.[41] $^3J(H^N_iH^\alpha_{i-1})$ couplings were measured using quantitative *J* correlation (intensity modulated HSQC) on a Bruker 800 DRX spectrometer (32 scans, relaxation delay 1.2 ms, 2τ = time for evolution of $^3J_{HNH\alpha} = 18$ ms).[42] Couplings were evaluated by using the relation $S_{cross}/S_{diag} = \cos(\pi^3 J_{HNH\alpha} 2\tau)$. Random coil values were subtracted from the couplings to gain direct indication of possible secondary structure elements.[43]

Molecular Dynamics Simulations. Accelerated molecular dynamic (AMD) simulations[30] use biased potential sampling to enhance the probability of transition between low-energy conformational states. The methods have been presented previously and will be briefly described here. Acceleration of dynamic events is provoked by defining a reference or "boost" energy, E_b, lying above the minimum of the potential energy surface. If the potential energy $V(r)$ lies below this boost energy, a bias potential is added to the actual potential, thus accelerating exchange between low-energy conformational states and in the case of unfolded proteins allowing more efficient conformational sampling. The energy modification is given by

$$\Delta V(r) = \frac{(E_b - V(r))^2}{\alpha + (E_b - V(r))} \quad (1)$$

The extent of acceleration is determined by the choice of the boost energy and the acceleration parameter, α. During the course of the simulation the forces are recalculated for the modified potential.

(35) Hosia, W.; Bark, N.; Liepinsh, E.; Tjernberg, A.; Persson, B.; Hallen, D.; Thyberg, J.; Johansson, J.; Tjernberg, L. *Biochemistry* **2004**, *43*, 4655−4661.
(36) von Bergen, M.; Barghorn, S.; Li, L.; Marx, A.; Biernat, J.; Heberle, J.; Mandelkow, E.-M.; Mandelkow, E. *J. Biol. Chem.* **2001**, *276*, 48165−48174.
(37) Margittai, M.; Langen, R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10278−10283.
(38) Jeganathan, S.; von Bergen, M.; Brutlach, H.; Steinhoff, H.-J.; Mandelkow, E. *Biochemistry* **2006**, *45*, 2283−2293.
(39) Kontaxis, G.; Bax, A. *J. Biomol. NMR* **2001**, *20*, 77−82.
(40) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(41) Goddard, T. D.; Kneller, D. G. *SPARKY 3*; University of California: San Francisco, CA.
(42) Permi, P.; Kilpeläinen, I.; Annila, A.; Heikkinen, S. *J. Biomol. NMR* **2000**, *16*, 29−37.
(43) Plaxco, K. W.; Morton, C. J.; Grimshaw, S. B.; Jones, J. A.; Pitkeathly, M.; Campbell, I. D.; Dobson, C. M. *J. Biomol. NMR* **1997**, *10*, 221−230.

AMD simulations were performed for the 15-residue fragments associated with the four regions of interest using an in-house modified version of the sander module in AMBER8.[44] Each system was solvated by approximately 4000 TIP3P water molecules[45] in a periodic box, and the ff99SB force field was employed for the solute.[46] Simulations were carried out at 300 K using a value of 2.0 ps^{-1} for the collision frequency and a time step of 2 fs. All bonds involving hydrogens were constrained using the SHAKE algorithm.[47] A cutoff of 10 Å was used for nonbonded interactions, and long-range interactions were treated with the particle mesh Ewald (PME) method.[48]

AMD simulations were performed for each system for 15×10^6 steps and structures collected at 100-step intervals. Acceleration was applied to the torsion force-field potential only. The boost energy was set 800 kcal·mol^{-1} above the average dihedral angle energy (obtained from an initial 500 ps MD simulation of the system), and the acceleration parameter (α), at 200 kcal·mol^{-1}. AMD simulations were used to calculate free energy surfaces for the backbone dihedral angles. This procedure was repeated starting from multiple different initial peptide structures using the same of level of acceleration to identify the lowest energy regions of conformational space.

All ϕ/ψ pairs were collected for each residue. 2-dimensional ϕ/ψ normalized histograms were constructed as follows: Ramachandran space was divided into 10° by 10° bins. Each ϕ/ψ pair was allocated appropriately. Canonical Boltzmann distributions were obtained by multiplying each ϕ/ψ point by its Boltzmann weighting factor ($e^{\beta\Delta E}$, where $\Delta E = (E_b - V)^2/(E_b - V + \alpha)$. Free energy was calculated using $\Delta G = -RT \ln(N_i/N_0)$, where N_i is the population of a particular bin, N_0 is the most populated bin, and the resulting global free energy minimum by default is always 0 kcal·mol^{-1}. Free energy surfaces were interpolated and a random number generator used to obtain a sample of 500 ϕ/ψ points that accurately represent the Boltzmann weighted ϕ/ψ population.

Generation of the Conformational Ensemble and Prediction of RDCs. Flexible−meccano (FM) uses a Monte Carlo sampling technique based on amino-acid propensity and side chain volume. The details of the algorithm have been presented elsewhere.[28] Alignment tensors for each conformer were calculated using PALES, an atomic resolution approach to alignment tensor prediction.[49] RDCs were calculated from throughout the molecule using the same program.

Incorporation of Specific Ramachandran Sampling. FM calculations were also performed using specific ϕ/ψ distributions for the four amino acids K254, S285, S316, and K347 in place of the standard residue-specific distributions. Gaussian distributions with standard deviation of 10° centered on values ranging from −180° to +180° in steps of 60° were used for both ϕ and ψ. In total 36 entire FM simulations were performed. Calculations were also performed using 2000 ϕ/ψ pairs for residues extracted from the accelerated molecular dynamics simulations of four quindecapeptides with amino acids K254, S285, S316, and K347 in the central position. The effective radius of gyration was calculated from the ensemble using previously described methods.[28]

Acknowledgment. Funded by Deutsche Forschungsgemeinschaft through the DFG-Research Center for Molecular Physiology of the Brain. We thank the Commisariat à l'Energie Atomique, the CNRS, the UJF Grenoble, Agence Nationale de Recherche (Grant NT05-4_42781), the EU through UPMAN, the DFG through Gk782 (C.G. and M.D.M.), the Max Planck Society, and the Fonds of the Chemical Industry for support. M.Z. is supported by a DFG Emmy Noether grant (ZW 71/1-5). P.B. is grateful for a long-term EMBO fellowship, and M.D.M., for a Boehringer Ingelheim fellowship. Donald Hamelberg is thanked for useful discussions, and Nils Lakomek, for help with preparation of the polyacrylamide gel.

Supporting Information Available: Comparison of RDCs measured in tau K18 aligned in polyacrylamide gel and bacteriophage, RDCs simulated without β-sheet propensities, and RDCs measured in tau K18 in 8 M urea, $^3J_{HNHA}$ couplings simulated from the ensemble in comparison to experimental values, and full author list for ref 44. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0690159

(44) Case, D. A.; et al. *AMBER 8*; University of California: San Francisco, CA, 2004.
(45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(46) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(47) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(48) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193−4194.
(49) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791−3792.