Supporting information for

Structural impact of proline-directed pseudo-phosphorylation at AT8, AT100 and PHF1 epitopes on 441-residue tau

Stefan Bibow[+], Valéry Ozenne°, Jacek Biernat[#], Martin Blackledge°, Eckhard Mandelkow[#] and Markus Zweckstetter[+,§,*]

[+] Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Goettingen, Germany; [§] DFG Center for the Molecular Physiology of the Brain, 37073 Goettingen, Germany; [#] Max Planck Unit for Structural Molecular Biology, c/oDESY, Notkestrasse 85, 22607 Hamburg, & DZNE, German Center for Neurodegenerative Diseases, c/o CAESAR, Ludwig-Erhard-Allee 2, 53175 Bonn, Germany; ° Institut de Biologie Structurale Jean-Pierre Ebel, CEA-CNRS-UJF UMR 5075, 41 Rue Jules Horowitz, Grenoble 38027, France

* mzwecks@gwdg.de

Materials and Methods

Sample preparation

Protein expression in *E. coli* and purification of 441-residue tau have been described elsewhere.[1,2] Mutants of tau (A15C, A72C, A125C, A178C, A239C, V256C, A322C, S352C, A384C, S416C and S199E_S202E_T205E_T212E_S214E_S396E_S404E (E-mutant) and E-mutant_C15, E-mutant_C72, E-mutant_C125, E-mutant_C178, E-mutant_C239, E-mutant_C256, E-mutant_C322, E-mutant_C352, E-mutant_C384, E-mutant_C416) were generated by using the QuikChange site-directed mutagenesis kit (Stratagene). The introduced mutations were confirmed by DNA sequencing. $^{13}C/^{15}N$-labeled samples were prepared from *E. coli* cells grown in M9-based minimal medium containing $^{15}NH_4Cl$ and ^{13}C glucose.[1]

Spin labelling of tau

To label 441-residue tau cysteine-containing mutants with the nitroxide spin label MTSL (Toronto Research Chemicals), DTT was removed before labelling from the buffer by using size exclusion chromatography (PD-10 columns, GE Healthcare), and the proteins were equilibrated in PBS buffer (pH 7.4). Free sulfhydryl groups were reacted with a 5-fold molar excess of the MTSL solubilised in ethyl acetate, at 21 °C for 2.5 h. Unreacted spin label was removed by using PD-10 columns equilibrated in 50 mM Na phosphate buffer (pH 6.8), and spin-labelled proteins were concentrated by using Amicon Ultra-15 (molecular weight cutoff, 3,000) (Millipore).

NMR spectroscopy

NMR spectra were acquired at 278K on a Bruker Avance 900 spectrometer equipped with a cryogenic probe. All samples were in pH 6.8, 50 mM phosphate buffer. Aggregation did not occur under these low temperature conditions. NMR data were processed and analyzed using NMRPipe [3] and Sparky 3 (T. D. Goddard and D. G. Kneller, http://www.cgl.ucsf.edu/home/sparky). Backbone resonance assignment of 300 μM $^{13}C/^{15}N$-labeled S199E_S202E_T205E_T212E_S214E_S396E_S404E (E-

mutant) tau was obtained from 3D (HA)CANNH (80 [F1] x 40 [F2] x 1024 [F3] complex data points, ns:4 and the experimental time was 35h) and 3D HNN (60 [F1] x 60 [F2] x 1024 [F3] complex data points, ns:8 , experimental time: 73h) experiments in combination with the previously obtained assignment of the wild-type protein.[1,4,5] Secondary shift values were calculated as the differences between measured C^{α} chemical shifts and the empirical random coil value for the appropriate amino acid type.[6] Random coil values were corrected for residues followed by a proline.[7] In addition, random coil values for histidines, glutamates, and aspartates were taken from Wishart and Sykes,[7] as the chemical shifts of these residues are particularly sensitive to pH. The C^{α} chemical shifts were referenced to DSS.

$^{3}J_{HNHA}$ scalar couplings were measured using the intensity modulated HSQC (32 scans, relaxation delay 1.2 ms, 2τ= time for evolution of $^{3}J_{HNH\alpha}$: 18 ms). Coupling values were calculated from the intensity ratios using the relation $I_{coupled}/I_{decoupled}= \cos(\pi\ ^{3}J_{HNH\alpha}\ 2\tau)$.[8] Secondary $^{3}J_{HNH\alpha}$ scalar couplings were calculated as the difference between experimental $^{3}J_{HNH\alpha}$ scalar couplings and random coil values.[9]

PRE measurements were performed at a protein concentration of 15 μM. We previously showed that at this low concentration, intermolecular interactions do not contribute to the PRE profiles.[1] PRE effects were extracted from the peak intensity ratios between two 2D ^{15}N-^{1}H HSQC NMR spectra acquired in the presence of the nitroxide radical and after addition of 4 mM DTT (heated to 42 °C for 10 min before measurement) to the same sample. Addition of DTT will cleave the MTSL tag from the cysteine residue, such that the spin label is no longer attached to the protein and the protein is in the diamagnetic state. We previously showed that oxidation of the MTSL tag with ascorbic acid gives very similar results in case of tau.[1] For the PRE profiles the peak intensity for every residue of the HSQC with the tag attached was divided by the peak intensity from the HSQC with the cleaved tag (PRE= I_{para}/I_{dia}).

NMR diffusion experiments of *wt* and E-mutant tau were recorded on a Bruker 600 Avance spectrometer equipped with a cryogenic probe using the PGSTE-WATERGATE pulse sequence.[10] Gradient strengths were incremented linearly in 16 steps from 25% to 95% of the maximum gradient strength with 32 scans for each increment. Experiments were performed in triplicates. Protein concentrations were 100 μM. To estimate the hydrodynamic radius, the diffusion of the protein lysozyme ($D_{reference}$) was measured on the same spectrometer, with the same protein concentration (100 μM), the same buffer and at the same temperature. Hydrodynamic radius values of *wt* and E-mutant tau were then calculated using the hydrodynamic radius of lysozyme ($R_{reference}$ = 18.9 Å) [11] as an external reference according to $R_{protein} = (D_{reference}/D_{protein})\ R_{reference}$.

PRE-based ensemble generation

The algorithm Flexible-meccano was used to create explicit ensembles of molecules that sample the conformational space available to the disordered protein Tau.[12] Unbiased conformational ensembles of 30,000 structures were calculated, and effective relaxation rates for each conformer calculated in the presence of the different spin probes. MTSL sidechain flexibility was incorporated by allowing the sidechain to sample all sterically allowed rotamers, and averaging the relaxation rates of each backbone conformation as previously described.[1] Transverse relaxation rates for each conformer were calculated and transformed into intensity ratios. The effects of different levels of side chain dynamics are shown in Figure S4 for cysteine mutants

at position 15 and 75. No significant differences in the profile of the long-range contact maps were found for different levels of MTSL sidechain dynamics.
The correlation time for the electron-nuclear interaction was set to 5 ns, in broad agreement with previous studies of unfolded proteins using PREs.[13] Exchange between individual backbone conformers was assumed to be fast on the chemical shift and relaxation rate timescale, so that average relaxation rates were calculated over the selected ensemble. Sub-ensembles of 200 structures were selected on the basis of agreement with respect to the experimentally measured intensity ratios using the genetic algorithm ASTEROIDS.[14] Extensive cross validation was used to ensure the ability of the approach to predict PRE data and to determine the optimal number of structures within the ensemble. Entire data sets, corresponding to one cysteine mutant, were removed from the analysis and these values were back-calculated from sub-ensembles selected using the remaining 10 data sets (Figures S5-S8).

The algorithm selects an ensemble of structures using the following fitness function compared to the experimental data.

$$\chi^2_{asteroids} = \sum_k \left(\Delta I^k_{calc} - \Delta I^k_{\exp}\right)^2 \quad (1)$$

Where:

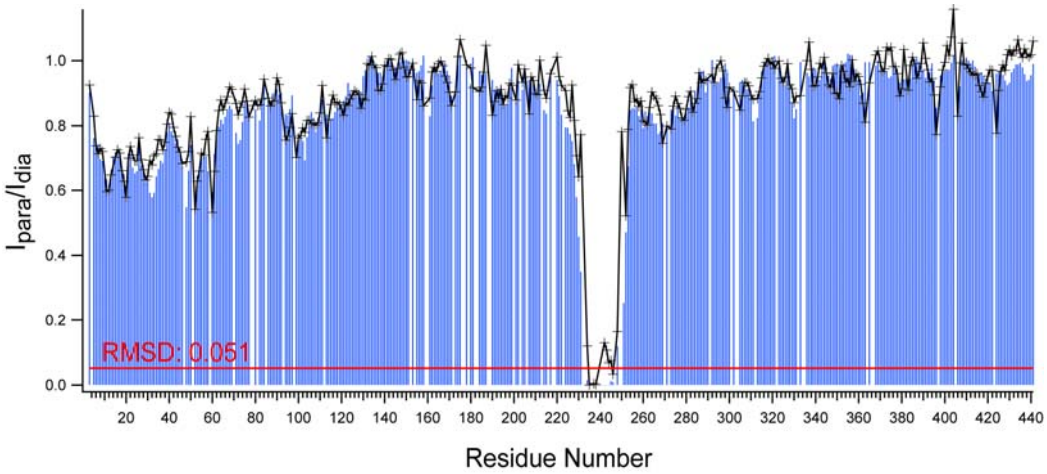
$$\Delta I^k_{calc} = \frac{I^k_{ox}}{I^k_{red}} = \frac{\Gamma^k_{2,red}\exp\left(-\Gamma^k_{2,para}t_m\right)}{\Gamma^k_{2,red}+\Gamma^k_{2,para}} \quad (2)$$

A mixing time (t_m) of 10ms was used. $\Gamma_{2,\mathrm{red}}$ is the intrinsic transverse relaxation rate of the observed proton spin and was estimated on the basis of an intrinsic linewidth measurement of 4Hz.[1] The final ensemble is obtained from generations of ensembles that undergo evolution and selection using this fitness function. Each generation comprises 100 different ensembles of 200 structures. The identical procedure was applied for both samples, and contact maps showing a logarithmic comparison between the selected ensemble and the original unbiased ensemble are used to compare the data. Average distances between sites are represented in terms of the following metric:

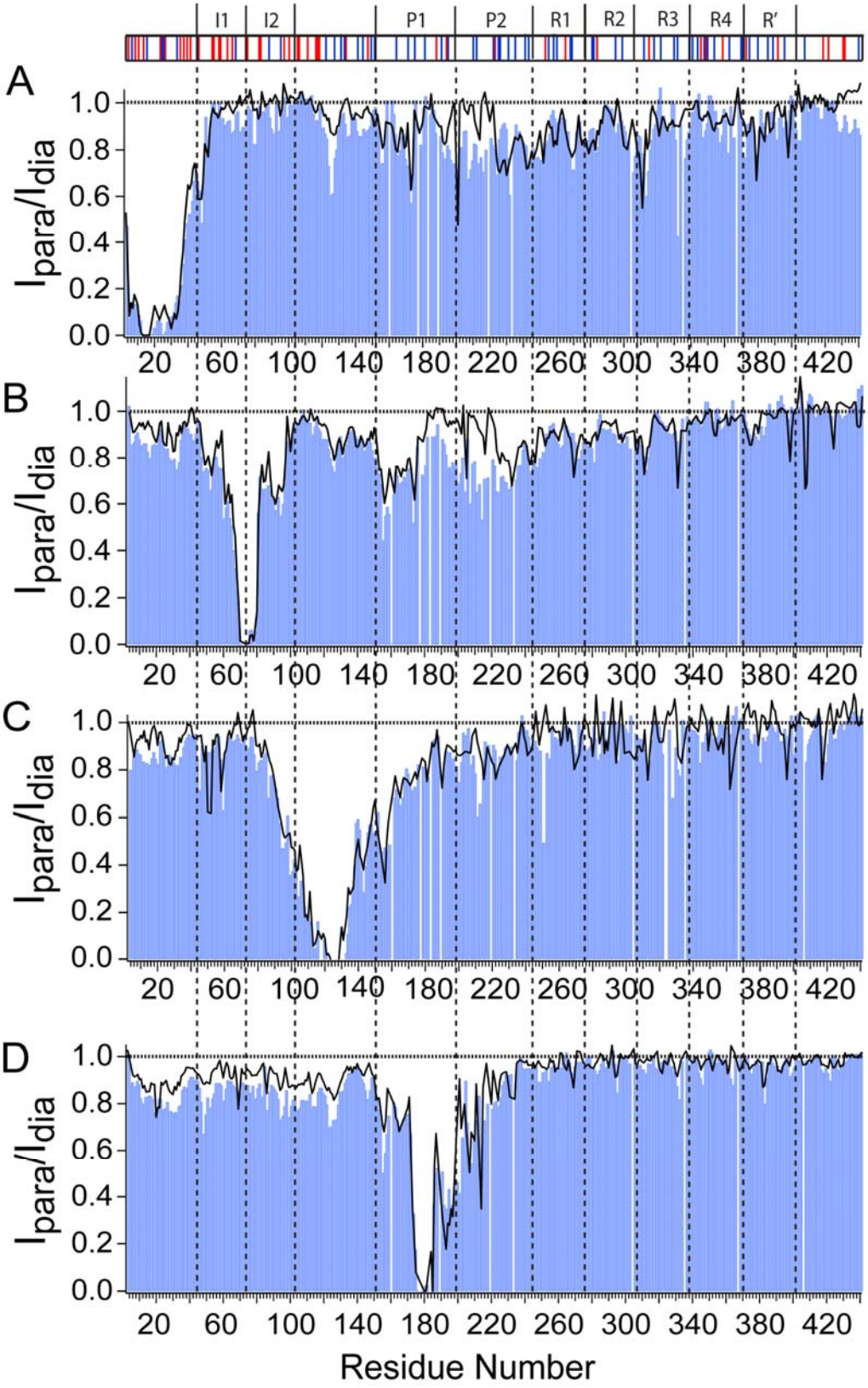
$$\Delta_{ij} = \log\left(\left\langle d_{ij}\right\rangle / \left\langle d^0_{ij}\right\rangle\right) \quad (3)$$

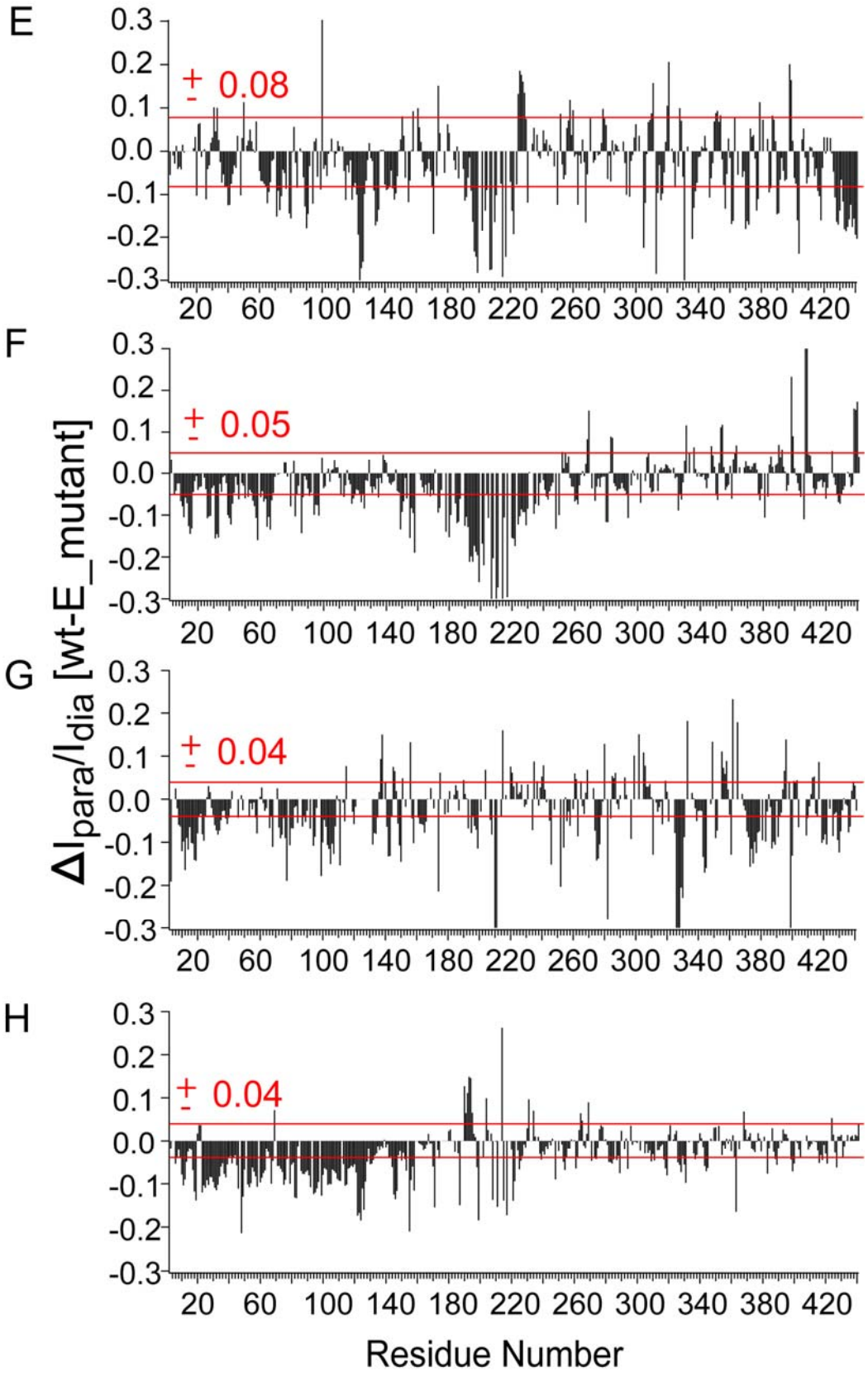
Where d_{ij} is the distance in any given structure of the ASTEROIDS ensemble between sites i and j, and d^0_{ij} is the distance in any given structure of the reference ensemble (with no ASTEROIDS selection) between sites i and j.

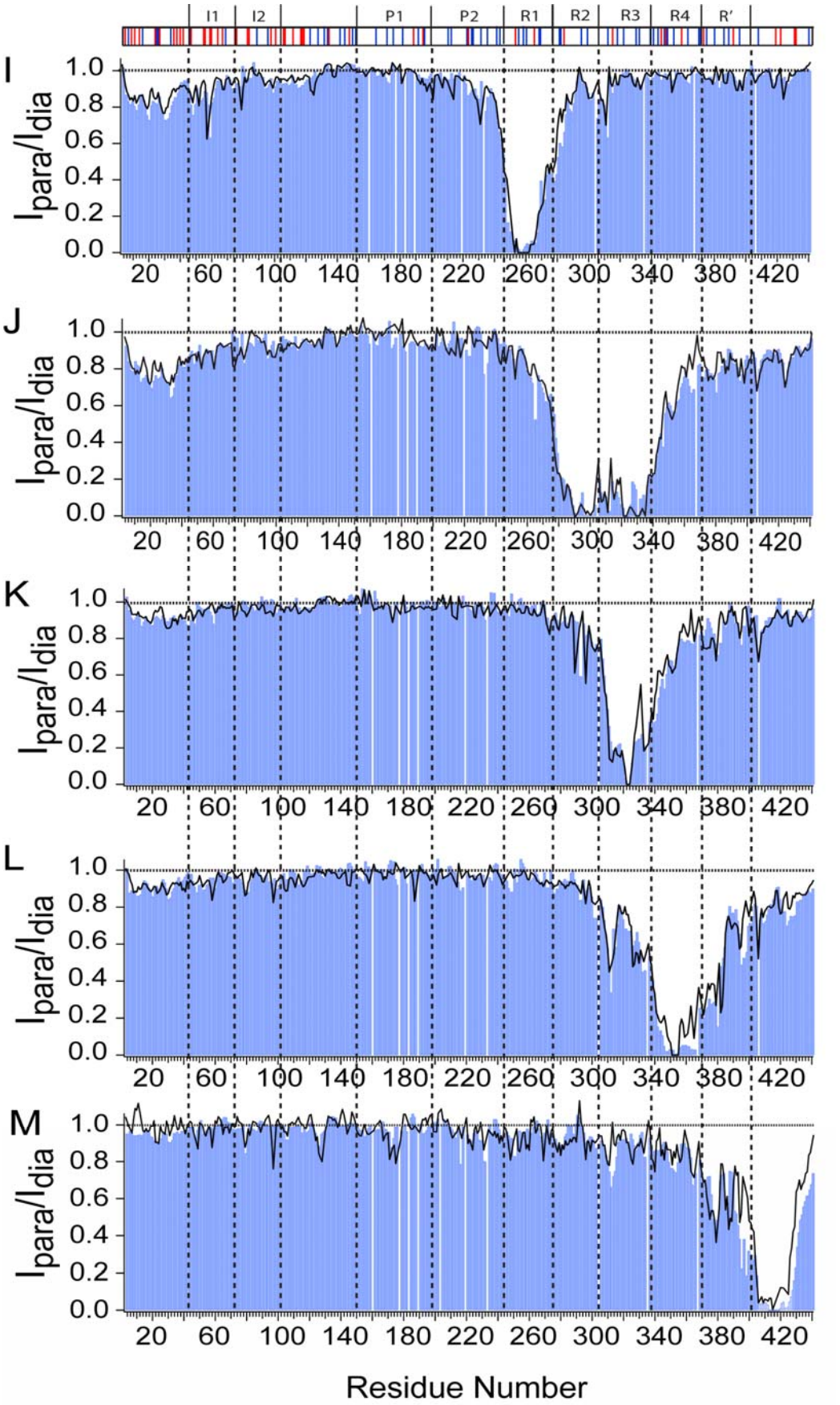
Supporting Figures

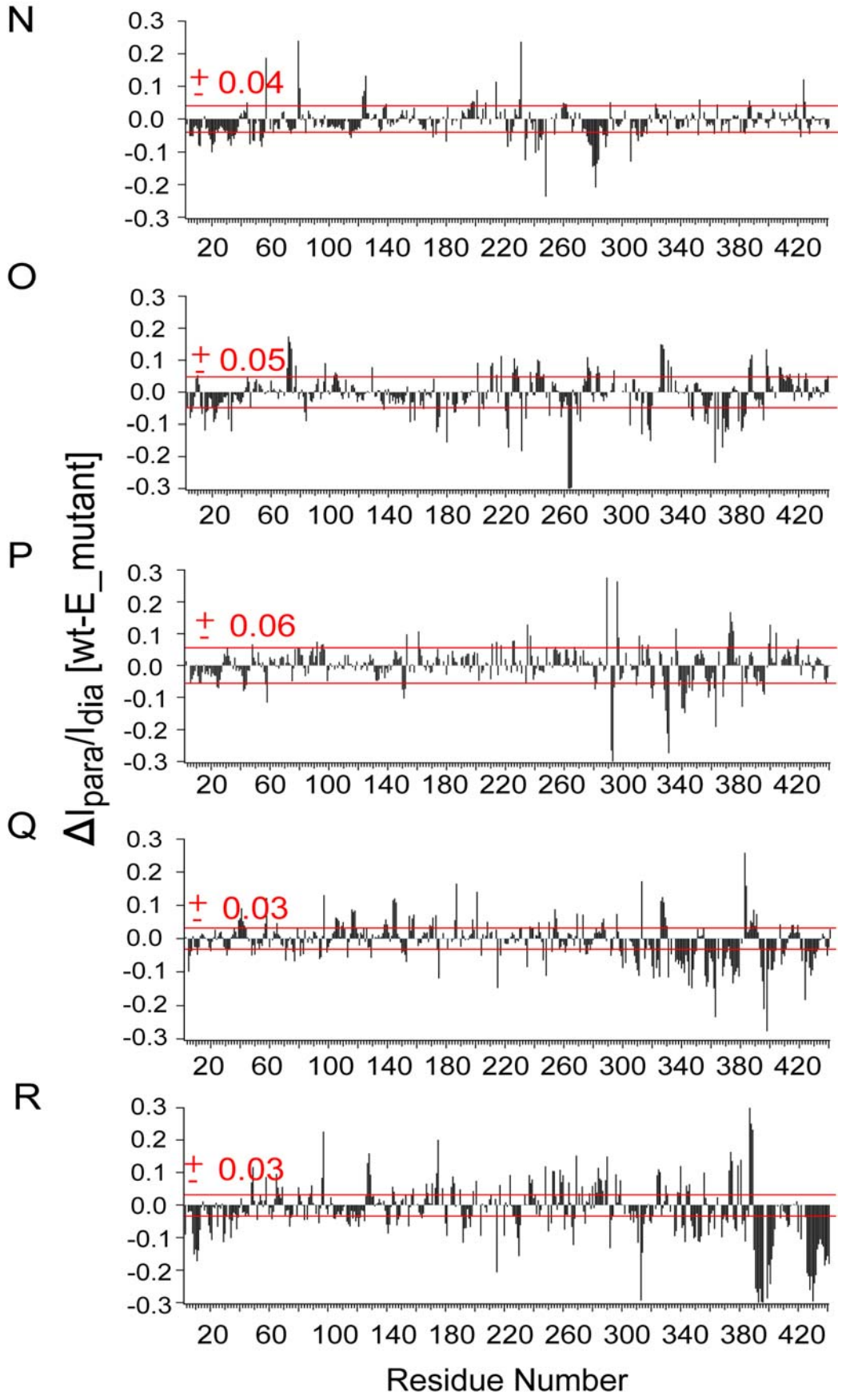


Supporting Figure S1. Comparison of the intensity ratios of two samples with MTSL@C239 either quenched with ascorbic acid (black line) or treated with DTT (blue bars). The root mean square deviation between the intensity ratios was 0.051. The data demonstrate that two separate measurements with two different treatments of the MTSL introduce only minor differences, supporting the high accuracy of the data.

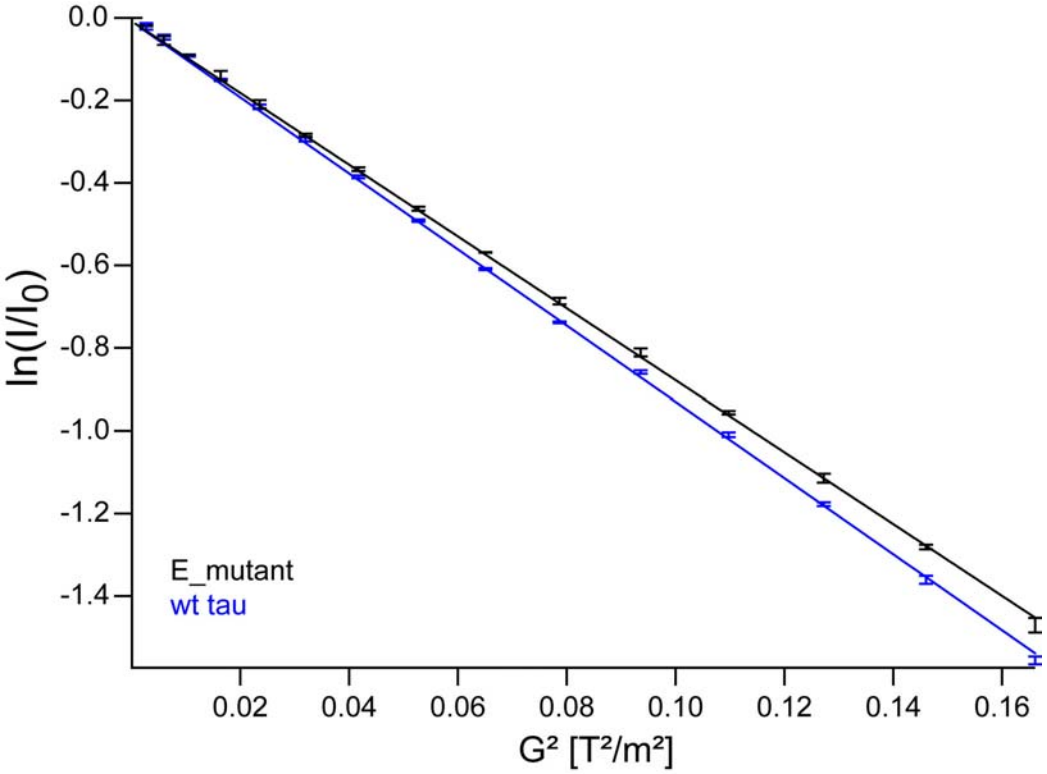

I1
I2
P1
P2
R1
R2
R3
R4
R′
A
B
C
D
I_{para}/I_{dia}
1.0
0.8
0.6
0.4
0.2
0.0
20
60
100
140
180
220
260
300
340
380
420
Residue Number


E
F
G
H
± 0.08
± 0.05
± 0.04
± 0.04
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
20 60 100 140 180 220 260 300 340 380 420
ΔIpara/Idia [wt-E_mutant]
Residue Number

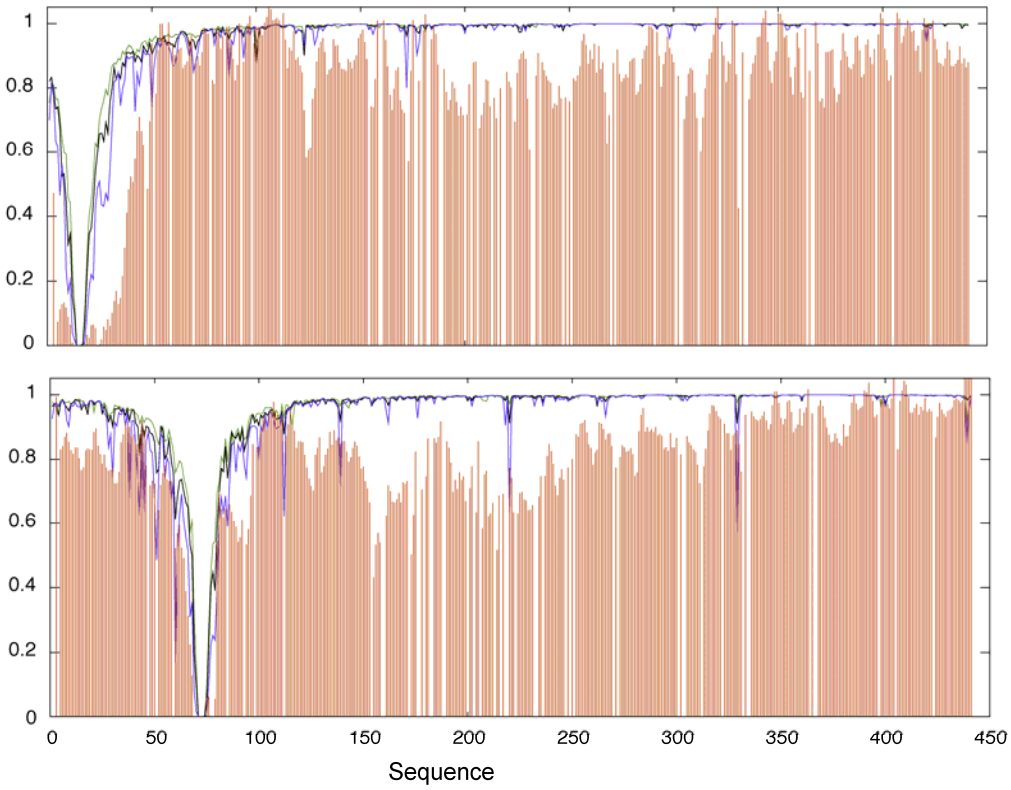

I1
I2
P1
P2
R1
R2
R3
R4
R'
I
J
K
L
M
I_{para}/I_{dia}
1.0
0.8
0.6
0.4
0.2
0.0
20
60
100
140
180
220
260
300
340
380
420
Residue Number


N
O
P
Q
R
± 0.04
± 0.05
± 0.06
± 0.03
± 0.03
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
20 60 100 140 180 220 260 300 340 380 420
ΔIpara/Idia [wt-E_mutant]
Residue Number

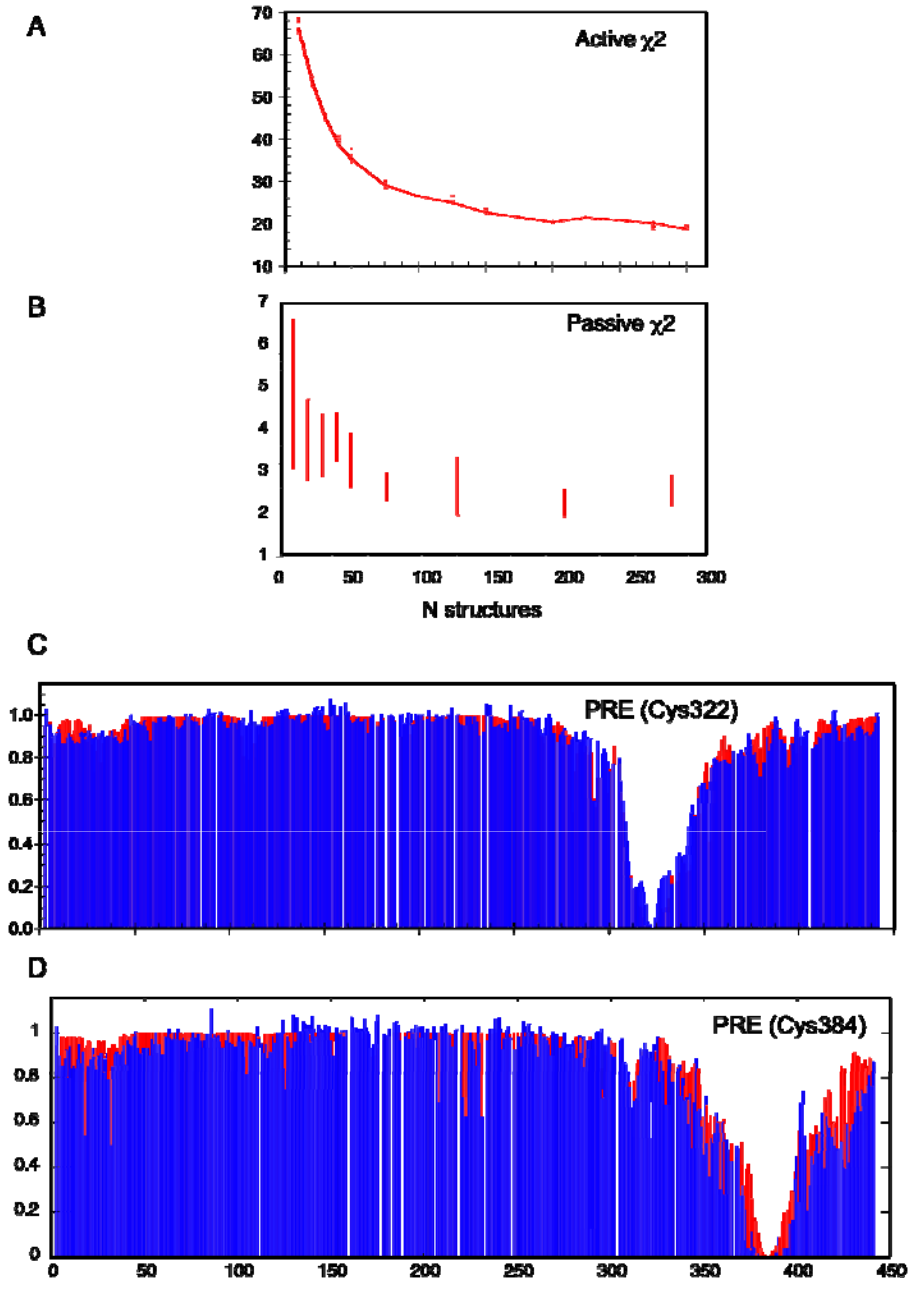
Supporting Figure S2. PRE broadening profiles (A-D and I-M) of amide protons of *wt* (blue) and E-mutant 441-residue tau (black line) with MTSL attached to position 15 (A), 72 (B), 125 (C), 178 (D), 256 (I) 291 and 322 (J) 322 (K), 352 (L) and 416 (M) with the corresponding differences (*wt* – E-mutant) for MTSL at position 15 (E), 72 (F), 125 (G), 178 (H), 256 (N), 291 and 322 (O), 322 (P), 352 (Q) and 416 (R). The error margins in the difference plots represent the maximal uncertainty of the measurement (maximal value of the sum of the errors in the intensity ratios was taken). Decreases in peak intensity ratios that occur far from the site of spin-labelling (>10 residues) are indicative of long-range contacts between the spin-label and distant areas of sequence. The domain organization on top (inserts I1, I2; proline-rich regions P1, P2; repeat domain R1-R4, pseudorepeat R') highlights the location of negative (red) and positive (blue) charges.



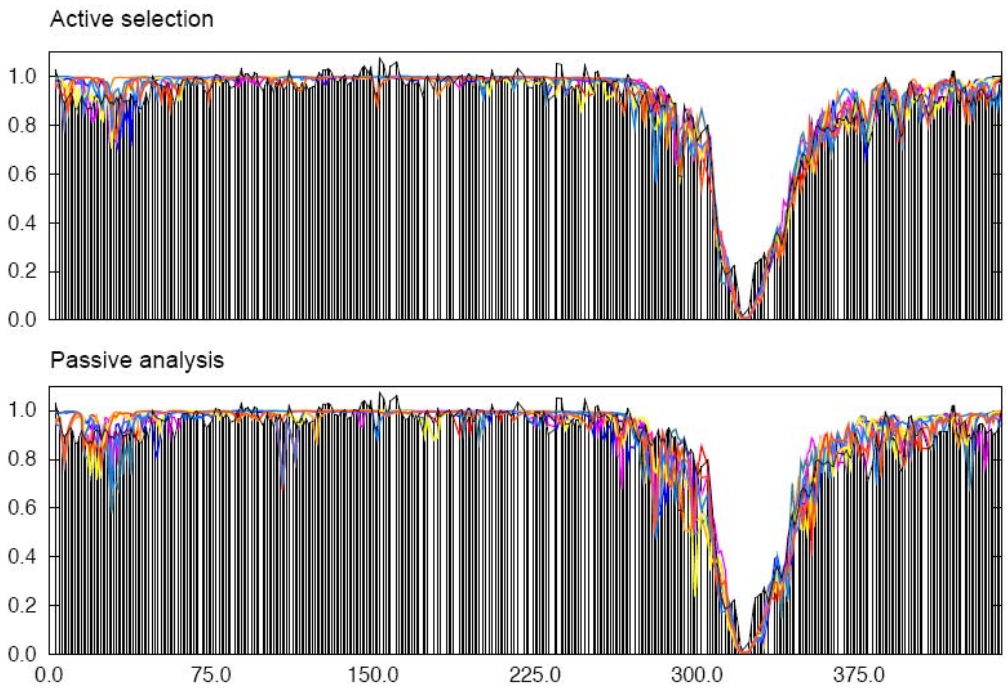
Supporting Figure S3. NMR diffusion experiments on *wt* tau and E-Mutant tau. The magnetization decay is faster for *wt* tau when compared to E-mutant tau pointing to a smaller hydrodynamic radius of *wt* tau.



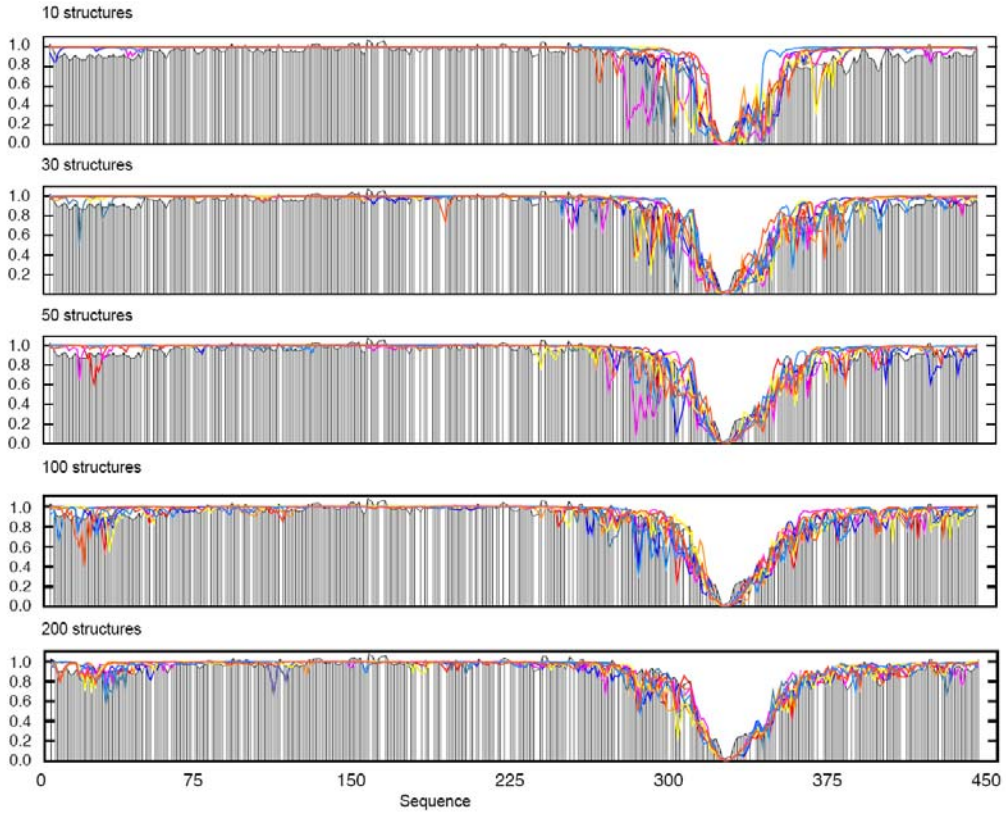
Supporting Figure S4. Comparison of the effects of different levels of side chain dynamics for cysteine mutants at position 15 and 75. Green: all sterically possible rotamers were selected for each conformer in the ensemble, black: only those rotamers within a radius of 15 Å from the average position of the MTSL, blue: only those conformers within a radius of 5 Å from the average position of the MTSL. All of these profiles were averaged over 50000 flexible meccano conformers, calculated without selection against the experimental data. Red represents the experimentally measured ratios. No significant differences in the profile of the long-range contact maps were found for different levels of MTSL side chain dynamics.



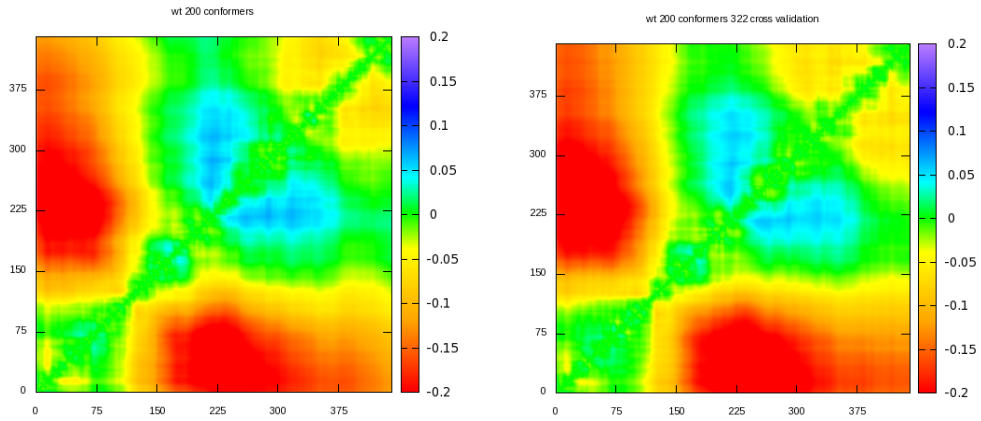
Supporting Figure S5. Cross validation of passive data sets that are not used in the analysis. Calculations were performed for which one of the 11 cysteine mutant data sets was removed from the analysis (passive data set). The ASTEROIDS selection was then performed using the other 10 data sets (active) and the passive data set were back-calculated. (A) $\chi 2$ for the active data sets. (B) $\chi 2$ for the passive data sets. Red lines indicate the range of $\chi 2$. (C) Back-calculated PRE for a calculation where the passive data set was Cys322. (D) Back-calculated PRE for a calculation where the passive data set was Cys384. Experimental data are shown in blue.



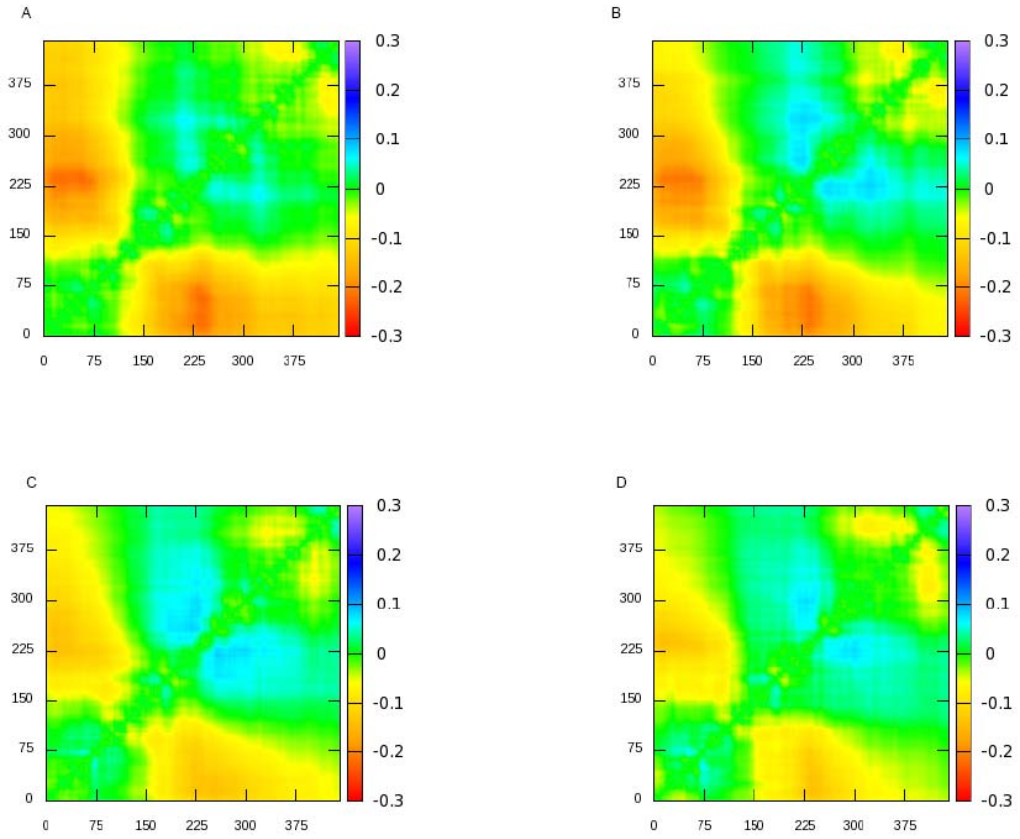
Supporting Figure S6. The difference between the reproduction of the 322 cys PRE mutant data (WT) when included in the fit (active) and when used only for cross-validation (passive). The chi2 over the 8 different runs for the active selection is 1.22±0.09 and for the passive selection 2.4±0.4. The different colours represent different runs of the ASTEROIDS algorithm, with different random seeds.



Supporting Figure S7. Effect of the number of structures on the reproduction of the passive 322 cys mutant PRE data (*wt* tau). The different colored lines show the calculated data from the ensemble for 8 runs of each (different colors represent different runs of the ASTEROIDS algorithm, with different random seeds), and the bars (and black lines) show the experimental data. The figure shows that the ensemble reproduces the 'passive' data better as more structures are used in the ensemble, until, at around 200 structures, the reproduction is similar to the level achieved when these data are used actively (see Figure S6).



Supporting Figure S8. Comparison of contact matrices of *wt* tau calculated with all MTSL mutants (left), and in the absence of the 322 mutant (the one used for the cross validation Figures S5-S7). Color scale as for Figures 2B,C in the main manuscript.



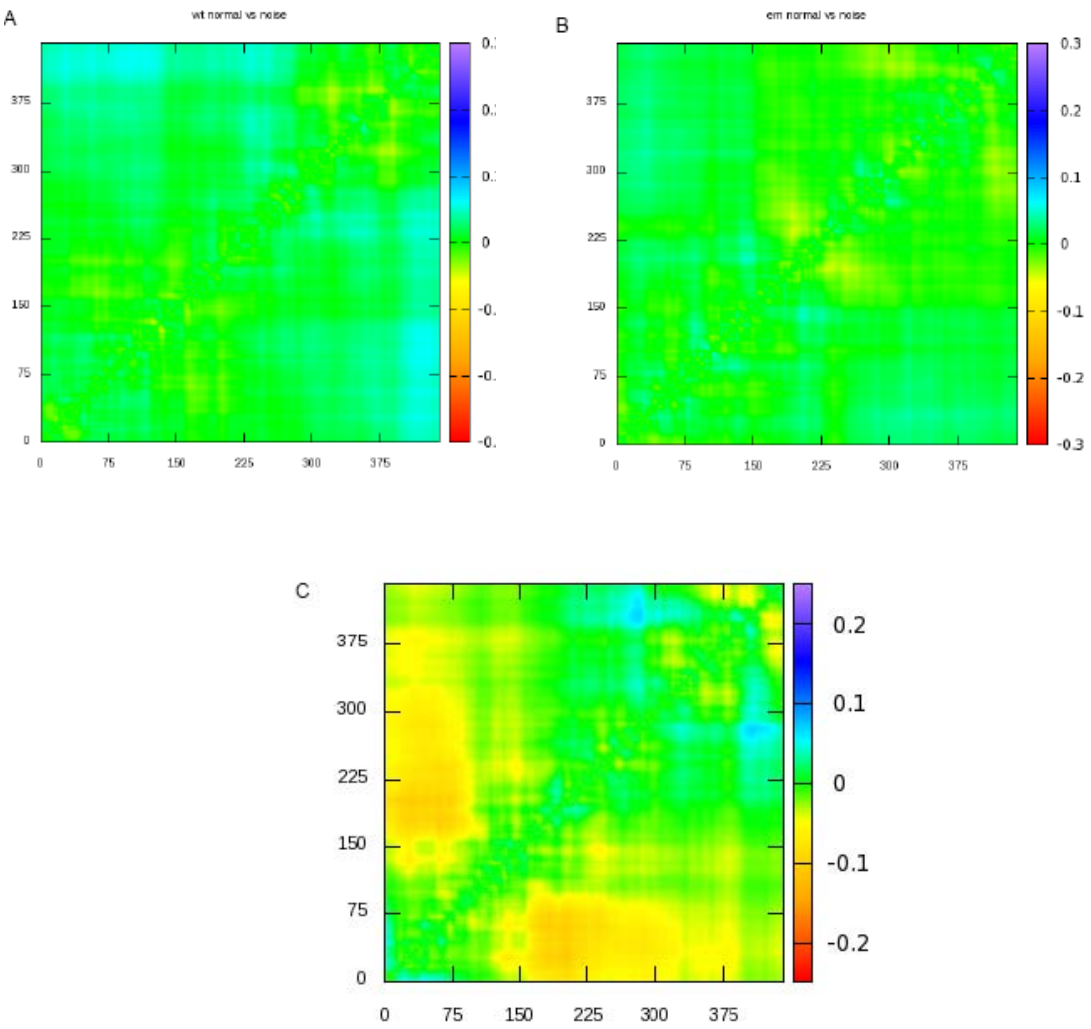
Supporting Figure S9. Effect of Gaussian noise on contact matrices of *wt* and E-mutant htau40.

A - Long-range contacts selected from noise-free PREs back calculated from a 200-strong ASTEROIDS selection of *wt* tau (selected using experimental PREs).

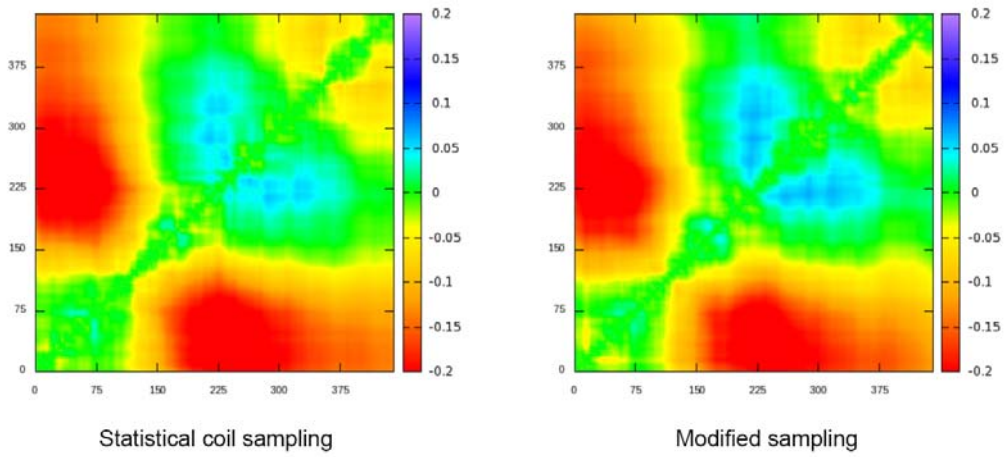
B - Long-range contacts selected from same PREs as in (A) but with noise added from Gaussian noise-based distributions of width 0.016.

C - Long-range contacts selected from noise-free PREs back calculated from a 200-strong ASTEROIDS selection of E-mutant tau (selected using experimental PREs).

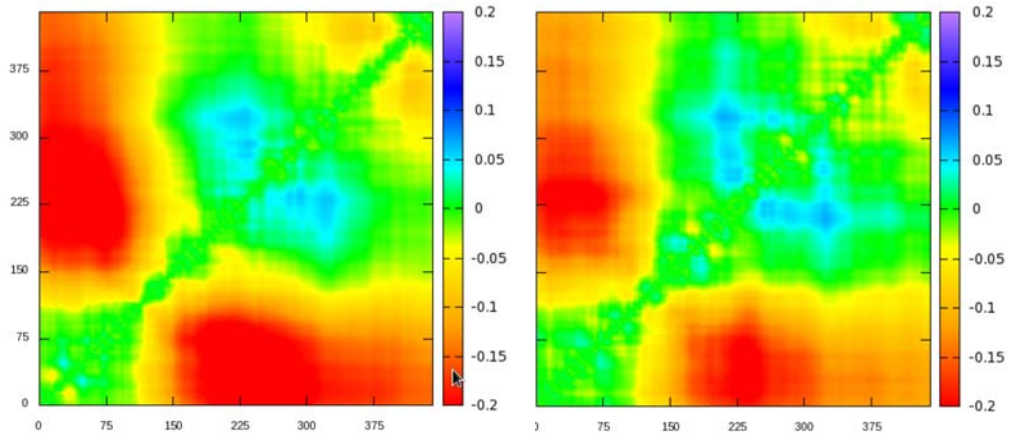
D - Long-range contacts selected from same PREs as in (C) but with noise added from Gaussian noise-based distributions of width 0.016.



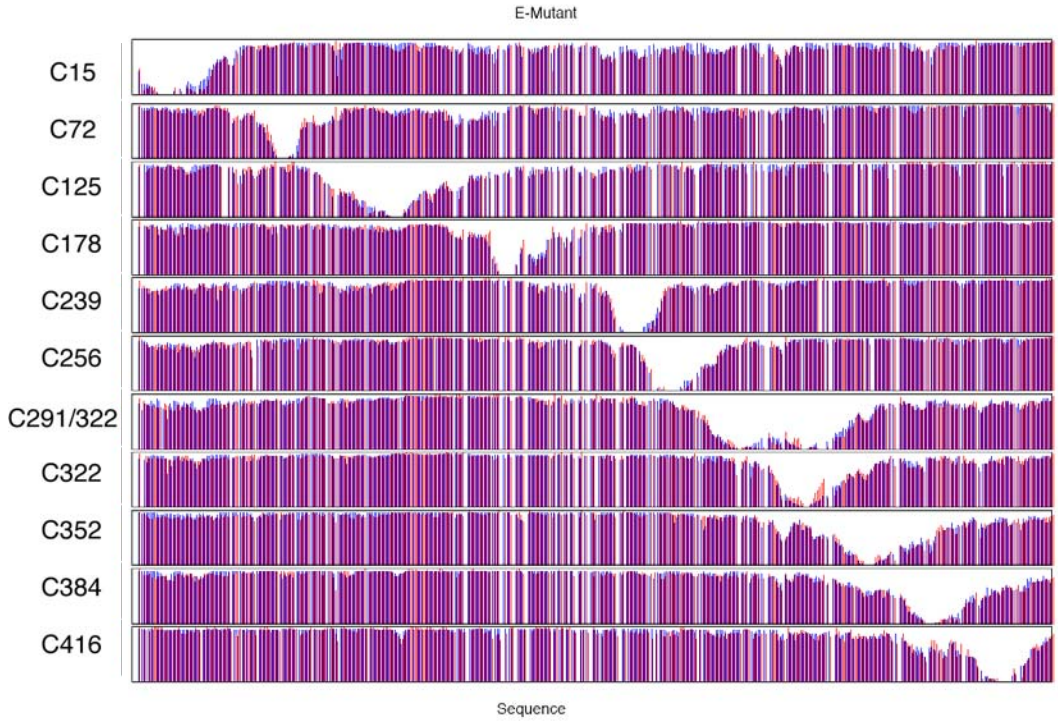
Supporting Figure S10. Effect of Gaussian noise on long-range contacts: difference matrices for Figure S9.
A - Difference matrix between A and B (*wt* tau) of Figure S9.
B - Difference matrix between C and D (E-mutant tau) of Figure S9.
C - Difference matrix between B and D of Figure S9.



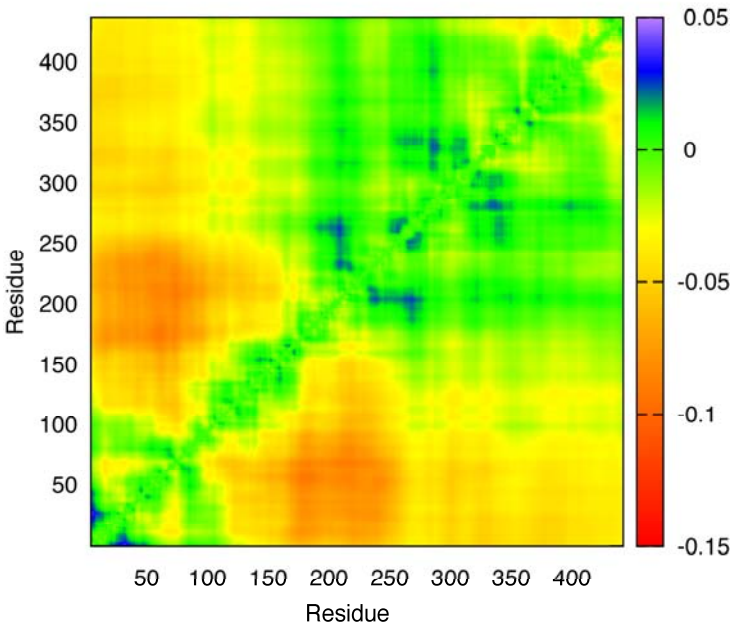
Supporting Figure S11. Effect of modified local sampling on the detection of long-range contacts.
Left: Ensemble containing statistical coil sampling throughout the protein.
Right : Ensemble containing the same sampling as (left), but with the strand of 30 amino acids (395-425) containing the following sampling pattern:
...AXBXAXBXAXBXAXB....
where 'A' corresponds to a level of α-helical sampling that is 20% higher than in the statistical coil (for the relevant amino acid) and 'B' corresponds to a level of β-sheet sampling that is 20% higher than in the statistical coil. 'X' represents statistical coil sampling for the intervening residues. This local sampling was chosen to probe the influence of an arbitrary change in local structure on the ensemble.
The figure demonstrates that with this level of modification of local structure the long-range contacts determined from the experimental PREs are essentially identical.
The matrices represent averages over 8 different runs of ASTEROIDS in both cases, with different random seed initial conditions.



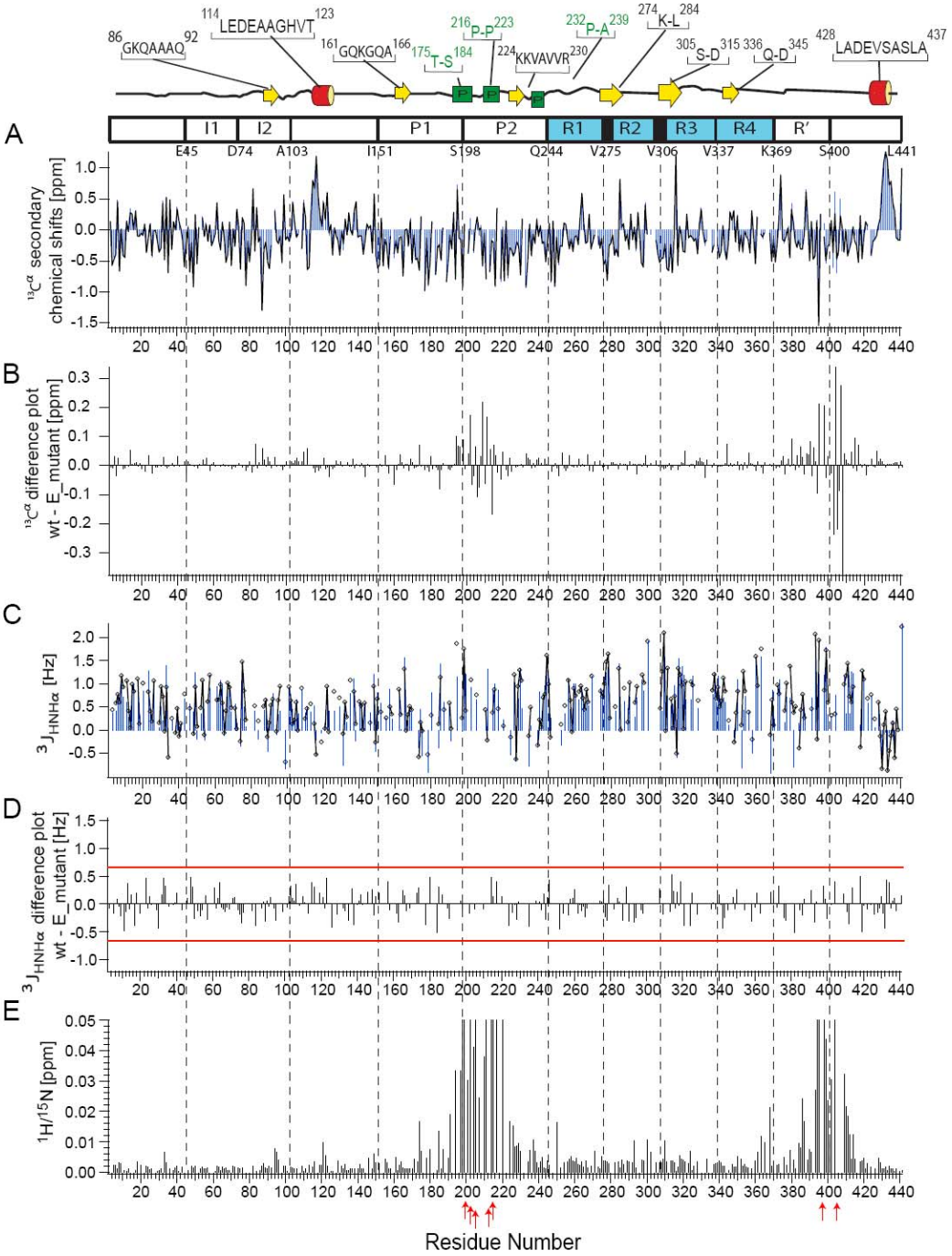
Supporting Figure S12. Effect of modified local sampling on the detection of long-range contacts (2).
Left: Long-range contacts selected from experimental PREs - no local structural preference.
Right: Long-range contacts selected from same PREs with 25% helix between residues 428 and 437.



Supporting Figure S13. Comparison of experimental PREs (red) and values back-calculated from the representative ensemble of E-mutant tau (blue). Positions of the spin-label are indicated on the left.



Supporting Figure S14. Differences in long-range contacts between the representative ensembles of *wt* and E-mutant 441-residue tau (shown in Figure 2b,c).



Supporting Figure S15. Comparison of transient secondary structure in *wt* and E-mutant 441-residue tau. (A) $^{13}C^{\alpha}$ secondary chemical shifts of *wt* (blue) and E-mutant (black) and its difference (B). (C) $^{3}J_{HNH\alpha}$ secondary scalar couplings, i.e. the difference between experimental $^{3}J_{HNH\alpha}$ coupling constants and the amino-acid specific random coil values [9], in *wt* (blue) and E-mutant (black) tau. (D) Differences in $^{3}J_{HNH\alpha}$ couplings constants between *wt* and E-mutant tau. The red line indicates the experimental uncertainty based on the signal-to-noise ratio. (E) Normalized and averaged differences in ^{1}H and ^{15}N chemical shifts between *wt* and E-mutant tau. Red

arrows indicate the seven sites of mutation into glutamic acid. On top, the transient secondary structure in *wt* 441-residue tau is shown schematically.[1]

Supporting References

(1) Mukrasch, M. D.; Bibow, S.; Korukottu, J.; Jeganathan, S.; Biernat, J.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M. *PLoS Biol* **2009**, *7*, e34.
(2) Mukrasch, M. D.; Biernat, J.; von Bergen, M.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M. *J Biol Chem* **2005**, *280*, 24978-24986.
(3) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J Biomol NMR* **1995**, *6*, 277-293.
(4) Panchal, S. C.; Bhavesh, N. S.; Hosur, R. V. *J Biomol NMR* **2001**, *20*, 135-47.
(5) Zweckstetter, M.; Bax, A. *J Am Chem Soc* **2001**, *123*, 9490-9491.
(6) Schwarzinger, S.; Kroon, G. J.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. *J Am Chem Soc* **2001**, *123*, 2970-2978.
(7) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J Biomol NMR* **1995**, *5*, 67-81.
(8) Permi, P.; Kilpelainen, I.; Annila, A.; Heikkinen, S. *J Biomol NMR* **2000**, *16*, 29-37.
(9) 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.
(10) Zheng, G.; Stait-Gardner, T.; Anil Kumar, P. G.; Torres, A. M.; Price, W. S. *J Magn Reson* **2008**, *191*, 159-163.
(11) Pamar, A. S.; Muschol, M. *Biophys J* **2009**, *97*, 590-598.
(12) Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W.; Blackledge, M. *Proc Natl Acad Sci U S A* **2005**, *102*, 17002-17007.
(13) Gillespie, J. R.; Shortle, D. *J Mol Biol* **1997**, *268*, 170-184.
(14) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. *J Am Chem Soc* **2009**, *131*, 17908-17918.