Supporting Information

Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition

R. Bryn Fenwick[†], Santi Esteban-Martín[†], Barbara Richter[#], Donghan Lee[‡], Korvin F. A. Walter[‡], Dragomir Milanovic[‡], Stefan Becker[‡], Nils A. Lakomek[‡], Christian Griesinger[‡] and Xavier Salvatella[†,§]

[†]Joint BSC-IRB research programme in Computational Biology. Institute for Research in Biomedicine – IRB Barcelona, Parc Científic de Barcelona, Baldiri Reixac 10, 08028 Barcelona, Spain [#]Department of Chemistry, University of Cambridge, Lensfield Rd, Cambridge CB2 1EW, UK [‡]Max Planck Institut fur Biophysikalische Chemie, Am Fassberg 11, 37077 Göttingen, Germany [§]Institució Catalana de Recerca i Estudis Avançats – ICREA, Barcelona, Spain.

Details of the simulation protocol used for the determination of ERNST

The ensemble was generated using an in-house modified version of the molecular simulation package CHARMM (version c35). The X-ray crystal structure of ubiquitin (1ubq[1]) was protonated and solvated in a 4 Å explicit water shell containing 577 TIP3P water molecules. The shell was maintained during the simulation using the CHARMM MMFP module[2]. An atom-based truncation scheme with a list cutoff of 13 Å, a non-bond cutoff of 12 Å and a Lennard-Jones smoothing function initiated at 8 Å were used.

The ensemble simulations were carried out with a time step of 1fs and consisted of 12 simulated annealing cycles with T_{min} = 300 K and T_{max} = 500 K. The first of such cycles was started, using different random seeds for each ensemble members from N copies of the X-ray structure of the ubiquitin (1ubq). The final N configurations were saved for analysis and used as starting configurations for the following cycle. The heating and cooling stages had durations of 45 and 90 ps, respectively.

Details of the NOEs and of the implementation of the NOE restraint are available in Ref. 3. Details of the implementation of the RDC restraint are available in Refs 4 and 5. The 36 sets of NH RDCs used to restrain the simulations are listed in Ref 6. Given that the globular domain of ubiquitin does not experience shape changes in the sub-ms timescale[7] a single tensor was used to describe the alignment of the protein in each of the 36 media. The RDCs corresponding to residues in the C-terminal disordered tail of ubiquitin were not used as restraints. The ensemble, therefore, does not accurately represent the structural properties of the tail.

The optimal value of N was determined by assessing the ability of the ensembles determined at increasing values of N to predict the results of independent NMR experiments. The results, show in Fig S1 for $^{3h}J_{NC'}$, indicated that N = 64 maximized agreement with experiment; this value was therefore used for the structure calculations used in this work, that lead to the determination of ERNST.

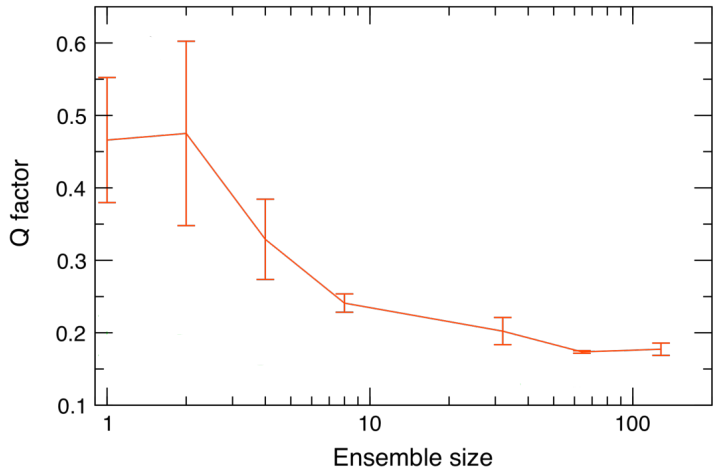

Fig S1: Ability of ensembles determined at increasing values of N to predict $^{3h}J_{NC'}$ data.

The CHARMM input files necessary to reproduce the results are available in http://lmb.irbbarcelona.org or directly from the authors.

Details of the equations used for the back-calculation of the NMR parameters

The 640 structures of ERNST were aligned relative to one another prior to validation using backbone RDCs,[6] scalar couplings[8] and cross-correlated relaxation rates.[9]

RDCs ($D_{HNC'}$ and $D_{NC'}$)

A single tensor was used to describe the alignment of ubiquitin in each alignment medium. This was fit to the experimental NH RDCs of residues in elements of secondary structure using single value decomposition. The level of agreement between ERNST and a given set of RDCs was computed as quality factor Q.[8] The values of Q listed in Table 1 do not include the RDCs of residues in the tail as their alignment cannot be described by a single tensor.[7]

Scalar couplings ($^3J\phi$)

To quantify the level of agreement of ERNST with scalar couplings related to the backbone torsion angle ϕ the following RMSD was computed (Eq. 1). In it i iterates through the list of scalar couplings available for this protein ($^3J_{HN\text{-}HA}$, $^3J_{HN\text{-}C'}$, $^3J_{HN\text{-}CB}$, $^3J_{C'\text{-}HA}$, $^3J_{C'\text{-}C'}$, $^3J_{C'\text{-}CB}$) and j iterates through the residues for which each type of coupling is available. A_i, B_i and C_i we taken from Ref 8.

$$RMSD\left(^3J_\phi\right) = \left(\sum_{i=1}^{N_{sets}}\sum_{j=1}^{N_{res}}\left(\left\langle A_i cos^2\phi_j + B_i cos\phi_j + C_i\right\rangle - J_j^{exp}\right)^2\right)^{1/2} \quad (1)$$

Trans-hydrogen bond scalar couplings ($^{3h}J_{NC'}$)

These couplings were back-calculated using the equation derived by Grzesiek and Barfield[10] using *ab initio* methods.

Cross-correlated relaxation rates ($R_{NH,NH}$ and $R_{NH,C\alpha H\alpha}$)

The rates of static structures were back-calculated using Eq. 2.[11] In it μ_0 represents the magnetic susceptibility of vacuum, $\hbar$ Planck's constant, γ_X the gyromagnetic ratio of nucleus X, r_{XH} the distance between nucleus X and H, τ_c the correlation time of ubiquitin and $\theta_{XH,YH}$ the angle between bond vectors XH and YH. The contribution of sub-ms uncorrelated dynamics to $R_{NH,NH}$ was introduced as a correction factor S^{RDC} that was computed from ERNST as in Ref 12. It is well-established that NH order parameters computed from ensembles obtained using RDCs need to be corrected because part of the motion of the bond vector is absorbed by the tensor optimization routine.[5] This correction was introduced in the calculation of $R_{XH,YH}$ involving NH bond vectors as S^{libr}, with a value of 0.95.[5] In the calculations of RMSD($R_{XH,YH}$) τ_c was optimized to maximize agreement with experiment; the values of τ_c obtained in this way were in all cases in agreement with independent experiments.

$$R_{XH,YH} = \left(\frac{\mu_0\hbar}{4\pi}\right)\left(\frac{\gamma_X\gamma_H}{r_{XH}^3}\right)\left(\frac{\gamma_Y\gamma_H}{r_{YH}^3}\right)P_2\left(cos\theta_{XH,YH}\right)\frac{2\tau_C}{5}S_{XH}^{RDC}S_{YH}^{RDC}S_{XH}^{libr}S_{YH}^{libr} \quad (2)$$

The rates of ensembles were instead back-calculated using Eq. 7.[11] Angular brackets indicate ensemble-averaging.

$$R_{XH,YH} = \left\langle\left(\frac{\mu_0\hbar}{4\pi}\right)\left(\frac{\gamma_X\gamma_H}{r_{XH}^3}\right)\left(\frac{\gamma_Y\gamma_H}{r_{YH}^3}\right)P_2\left(cos\theta_{XH,YH}\right)\frac{2\tau_C}{5}\right\rangle S_{XH}^{libr}S_{YH}^{libr} \quad (3)$$

Experimentally determined values of $R_{NH,C\alpha H\alpha}$ data

$R_{NH,C\alpha H\alpha}$ cross-correlated relaxation rates were measured using experiments previously published.[13] Each sub-spectrum of the 3D HNCA type-$R_{NH,C\alpha H\alpha}$ measurement of 3 mM ^{13}C, ^{15}N-labeled ubiquitin at pH 6.8 and 308 K was recorded with 85(t_1) * 33(t_2) * 512(t_3) complex points, t_{1max}=23.5 ms, t_{2max}=19.3 ms, t_{3max}= 61.1 ms, and inter-scan delay of 1.0 s on a Bruker Avance 600 MHz spectrometer equipped with a QCI-probe head. The time domain data were multiplied with square cosine functions in the all dimensions and zero-filled to 2048 * 256 * 2048 complex points. Two sub-spectra were added and subtracted to yield zero quantum (ZQ) and double quantum (DQ) spectra, respectively. All processing was done using the program NMRPipe. Intensities of multiplets in ZQ and DQ spectra were extracted using the program CARA (Rochus Keller, http://www.nmr.ch).

Res	$R_{NH,C\alpha H\alpha}$	Res	$R_{NH,C\alpha H\alpha}$	Res	$R_{NH,C\alpha H\alpha}$	Res	$R_{NH,C\alpha H\alpha}$
3	-9.7 ± 1.6	**20**	-7.5 ± 0.6	**39**	-4.8 ± 0.7	**57**	-1.6 ± 0.9
4	-15.7 ± 1.5	**21**	-4.9 ± 0.8	**40**	-12.0 ± 0.8	**58**	-2.9 ± 0.8
5	-14.5 ± 2.2	**22**	-10.1 ± 1.5	**41**	-11.4 ± 1.1	**59**	-12.1 ± 0.7
6	-13.7 ± 1.7	**23**	-1.8 ± 1.8	**42**	-13.3 ± 2.6	**60**	-2.0 ± 1.1
7	-11.7 ± 1.2	**25**	-6.0 ± 0.9	**43**	-14.9 ± 2.1	**61**	-9.4 ± 1.2
8	-2.8 ± 2.5	**27**	-2.9 ± 0.8	**44**	-15.0 ± 2.3	**62**	-12.4 ± 1.2
9	-9.9 ± 5.5	**28**	-3.9 ± 0.9	**45**	-12.6 ± 1.4	**63**	2.1 ± 1.4
11	-6.7 ± 1.1	**29**	-5.5 ± 0.8	**46**	-3.7 ± 7.0	**65**	-5.6 ± 0.7
12	-13.1 ± 3.4	**30**	-4.6 ± 1.0	**48**	-13.5 ± 1.0	**66**	-9.3 ± 2.0
13	-17.4 ± 7.0	**31**	-1.3 ± 1.1	**50**	-7.9 ± 1.7	**67**	-14.9 ± 2.3
14	-14.4 ± 1.6	**32**	-2.1 ± 0.7	**51**	-13.6 ± 2.4	**68**	-15.0 ± 2.0
15	-14.4 ± 1.3	**33**	-6.8 ± 0.7	**54**	-14.2 ± 1.1	**70**	-14.4 ± 2.1
17	-13.1 ± 1.2	**34**	-13.1 ± 1.0	**55**	-12.3 ± 1.7	**74**	-1.2 ± 2.1
18	-14.8 ± 1.4	**36**	-10.4 ± 0.9	**56**	-0.2 ± 1.3		

Table S1: Experimental values of $R_{NH,C\alpha H\alpha}$ measured for ubiquitin in Hz.

Correlation between the experimental and back-calculated values of $^{3h}J_{NC'}$

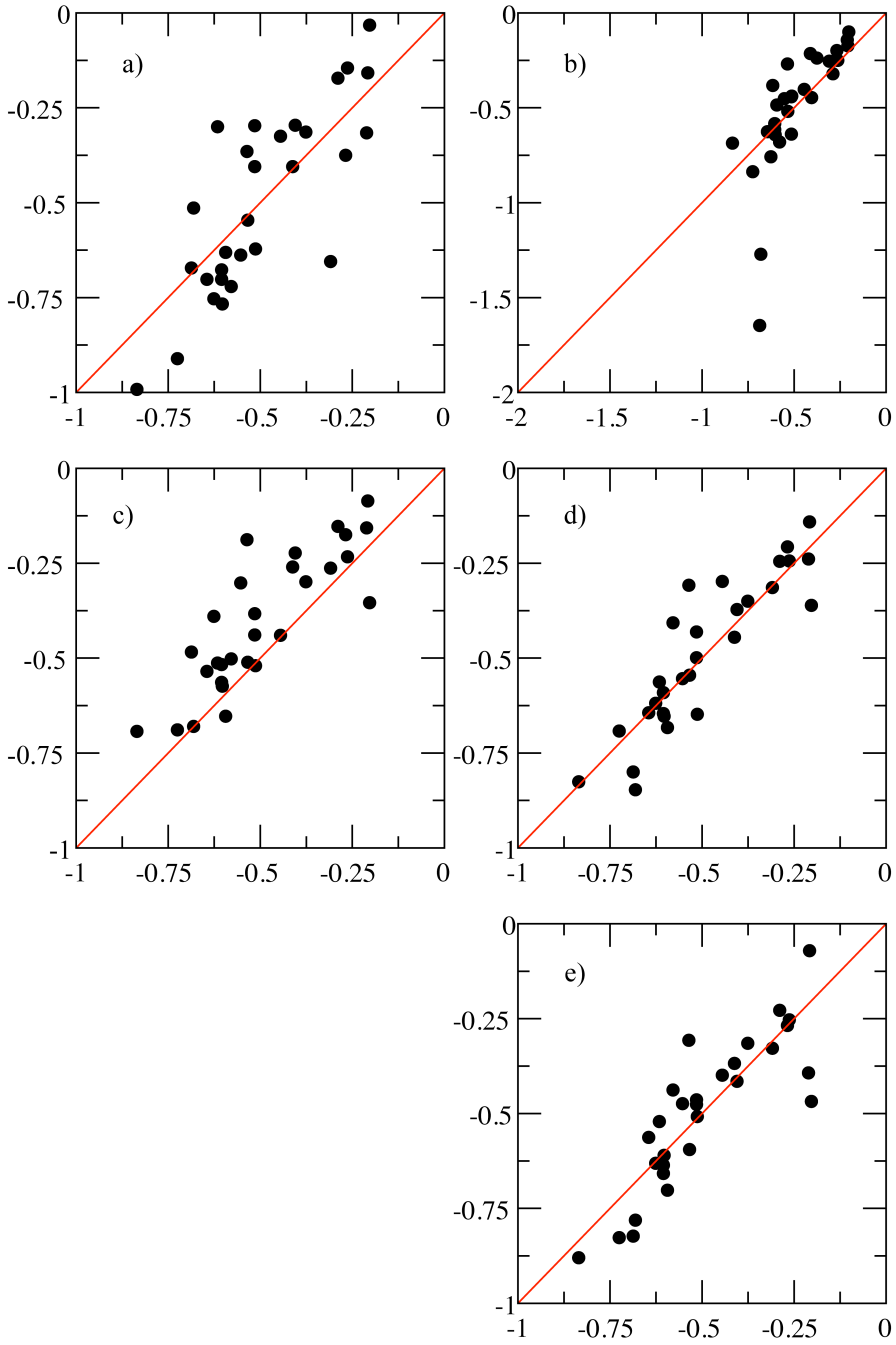

Fig S2: Correlation between the experimental (x axis) and back-calculated (y axis) values of $^{3h}J_{NC'}$ for a) the X-ray structure of ubiquitin - 1ubq b) the NMR structure – 1d3z c) the EROS ensemble – 2k39 d) ERNST e) the MD ensemble obtained without restraints.

Equations used for the calculation of the circular correlation coefficients

The correlations from the ERNST ensemble were analyzed by calculating the circular correlation coefficient for combinations of ϕ and ψ angles for each pair of residues across all ensemble members. The circular correlation coefficient is defined as

$$\rho_{xy}^{circular} = \frac{\sum_{i=1}^{N} sin\left(x_i - \bar{x}\right) \cdot sin\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{N} sin\left(x_i - \bar{x}\right)^2 \cdot \sum_{i=1}^{N} sin\left(y_i - \bar{y}\right)^2}}$$

Where the value of the circular mean is,

$$\bar{x}^{circular} = a\tan 2\left(\left[\sum_{i=1}^{N} sin\, x_i\right], \left[\sum_{i=1}^{N} cos\, x_i\right]\right)$$

The torsion angles were defined by the heavy atoms in the backbone of the protein. For the correlation matrices we present only the significant correlations calculated for the experimental-wide significance level ($p < 0.05$) using the Bonferroni correction.

The circular variance was calculated as a measure of the circular dispersion, where the value of 1 indicates that there is no preferred direction, while the value of zero indicates that only one direction exists.

$$\sigma^{2circular} = 1 - \frac{\sqrt{\left(\sum_{i=1}^{N} sin\, x_1\right)^2 + \left(\sum_{i=1}^{N} cos\, x_1\right)^2}}{N}$$

Supplemental correlation plots

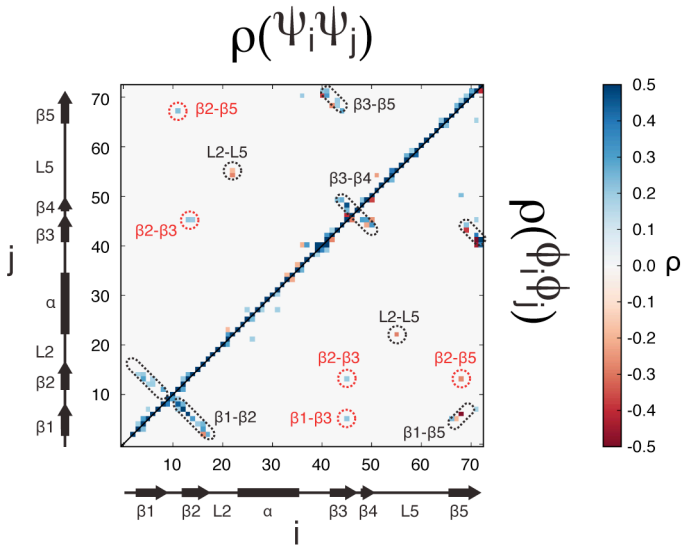

Fig S3: Circular correlation coefficients[14] (ρ) of φ and ψ in ubiquitin. Ellipses indicate hydrogen-bonded pairs of residues and red circles indicate residues with long-range correlations. Correlations between φ of different residues are shown below the diagonal and correlations between ψ are shown above the diagonal.

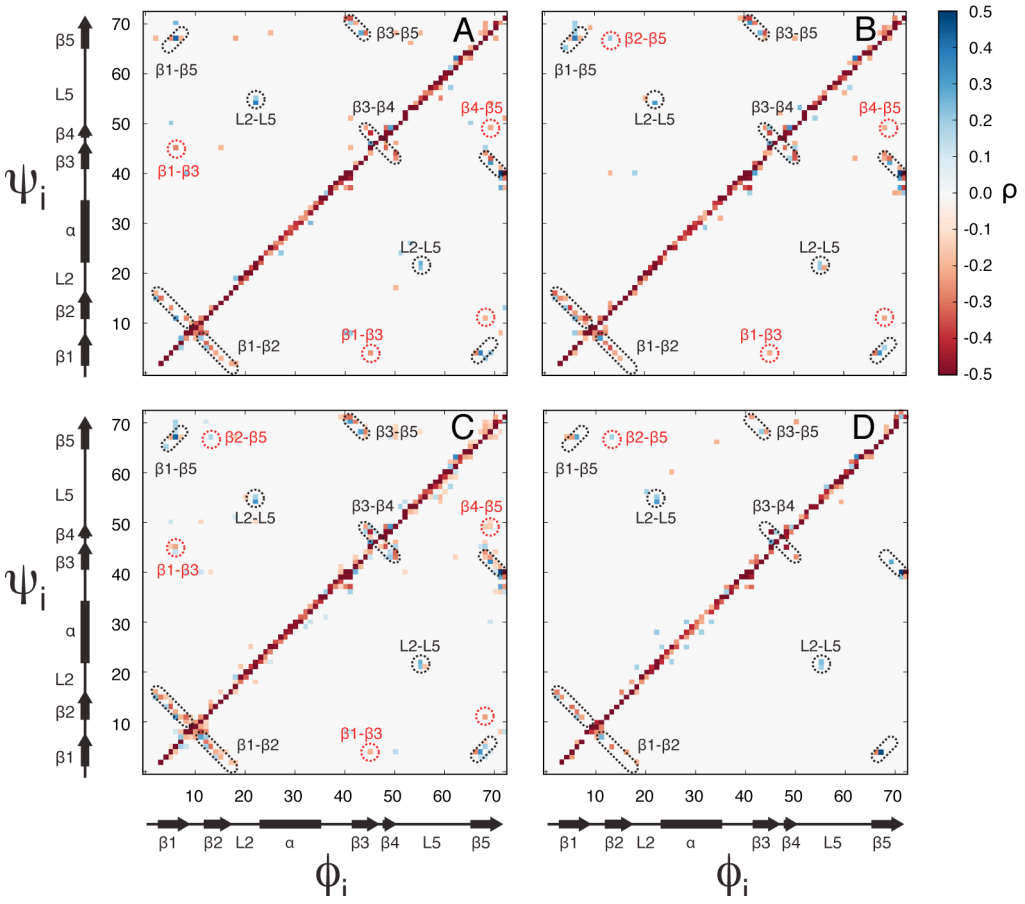

Fig S4: Circular correlation coefficients[14] (ρ) of φ_i and ψ_j in ubiquitin obtained A) Using different random seeds (B) Using ten runs of three cycles started from different random seeds and pooling only the last cycle (C) Combined map for the ensemble described in the main text and ensembles in A and B (n = 1920) (D) Correlations in the ensemble generated without restraints.

Principal component analysis

The structures of ERNST and of an ensemble containing the structures of ubiquitin in complex with its binding partners were fitted to the Cα atoms of the peptide planes 12:13 & 5:6 (β-strands β2 & β1). The covariance matrices were then calculated for the backbone atoms excluding the Cα and Hα atoms. The two extreme projections were calculated along the primary eigenvector and 50 structures interpolated between them. These calculations were made for the other pairs of peptide planes (12:13 & 67:68, 12:13 & 44:45) in the β-strands, β2 & β5 and β2 & β4. The three sets of coordinates, one per pair of peptide planes, were aligned to the structure of ubiquitin and merged to give a single structure containing projections along the primary motion for the three sets of residues. The motions captured in these coordinates are the same for the ERNST and x-ray ensembles.

Statistical analysis of correlation of longest range

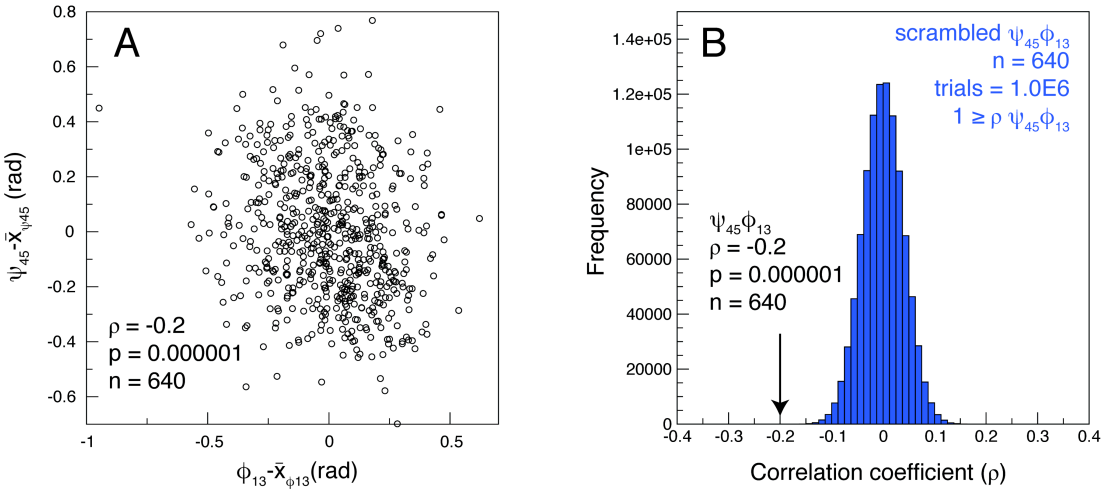

Fig S5: A) Scatter plot of, φ_{13} and ψ_{45}, the torsion angles involved in the correlation of longest range observed in ERNST B) Histogram of the correlation coefficients obtained when the 640 pairs of values of φ_{13} and y_{45} are scrambled randomly 10^6 times. Only once in a million times is a random absolute correlation with $\rho = 0.2$ observed by chance *i.e.* the probability that the correlation coefficient that we observe is untrue is equal to 10^{-6}. Consideration of the multiple comparison correction gives a p-value of 0.01.

References

(1) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531-544.
(2) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, 99050-99063.
(3) Richter, B.; Gsponer, J.; Varnai, P.; Salvatella, X.; Vendruscolo, M. *J. Biomol. NMR* **2007**, *37*, 117-135.
(4) De Simone, A.; Richter, B.; Salvatella, X.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 3810-3811.
(5) Fenwick, R. B.; Esteban-Martin, S.; Salvatella, X. *J Phys Chem Lett* **2010**, *1*, 3438-3441.
(6) Lange, O. F.; Lakomek, N. A.; Fares, C.; Schroder, G. F.; Walter, K. F.; Becker, S.; Meiler, J.; Grübmuller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471-1475.
(7) Salvatella, X.; Richter, B.; Vendruscolo, M. *J. Biomol. NMR* **2008**, *40*, 71-81.
(8) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836-6837.
(9) Pelupessy, P.; Ravindranathan, S.; Bodenhausen, G. *J. Biomol. NMR* **2003**, *25*, 265-280.
(10) Grzesiek, S.; Cordier, F.; Jaravine, V.; Barfield, M. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2004**, *45*, 275-300.
(11) Reif, B.; Hennig, M.; Griesinger, C. *Science* **1997**, *276*, 1230-1233.
(12) Best, R. B.; Vendruscolo, M. *J Am Chem Soc* **2004**, *126*, 8090-8091.
(13) Vögeli, B.; Yao, L. *J. Am. Chem. Soc.* **2009**, *131*, 3668-3678.
(14) Jammalamadaka, S. R.; Sengupta, A. *Topics in circular statistics*; World Scientific: River Edge, N.J, 2001;

Complete citation of reference 15

Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545–1614.