JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

COMMUNICATION

pubs.acs.org/JACS

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 Milovanovic,[§] 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 Road, Cambridge CB2 1EW, U.K.

[§]Max Planck Institut fur Biophysikalische Chemie, Am Fassberg 11, 37077 Göttingen, Germany

[∥]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Supporting Information

ABSTRACT: Long-range correlated motions in proteins are candidate mechanisms for processes that require information transfer across protein structures, such as allostery and signal transduction. However, the observation of backbone correlations between distant residues has remained elusive, and only local correlations have been revealed using residual dipolar couplings measured by NMR spectroscopy. In this work, we experimentally identified and characterized collective motions spanning four β-strands separated by up to 15 Å in ubiquitin. The observed correlations link molecular recognition sites and result from concerted conformational changes that are in part mediated by the hydrogen-bonding network.

The detection of motional correlations in proteins has been hampered by difficulties in experimentally determining the fluctuations of atomic coordinates.[1] However, it has been shown that NMR cross-correlated relaxation rates (CCRs) can provide a direct route to the characterization of correlated dynamics in biomolecules.[2−5] We chose to study correlated backbone motions in the protein ubiquitin by analyzing a conformational ensemble that is in good agreement with CCRs measured for this protein as well as with trans-hydrogen-bond scalar couplings ($^{3h}J_{NC'}$). Ubiquitin has a well-characterized function in targeting proteins for degradation. During this process, this protein recognizes a structurally diverse set of partners with a mechanism based on conformational selection.[6]

To study the presence of long-range correlated motions in the backbone of ubiquitin, we determined an ensemble of conformations that represents the structural heterogeneity of this protein on the submillisecond time scale.[7] The ensemble was obtained using ensemble molecular dynamics (MD) simulations[8−10] restrained by previously reported 2663 nuclear Overhauser effects (NOEs)[11] and 1971 NH residual dipolar couplings (RDCs).[12] The structure calculations were carried out with the CHARMM27 force field[13] in explicit solvent[14] using an in-house-modified version of the CHARMM molecular simulations software.[15] To ensure that the structure of ubiquitin was well-reproduced, the NOEs were restrained over pairs of structures.[16] The RDCs, which report on both the orientation and structural fluctuations of NH bond vectors,[17] were restrained by a quadratic potential[8,9] that penalizes deviations between the experimentally measured and ensemble-averaged RDCs. The five independent tensor elements that define the alignment of ubiquitin in each of the 36 alignment media[12] were optimized simultaneously with the atomic coordinates. The ensemble refinement protocol was started from the X-ray structure of ubiquitin (PDB code 1UBQ[18]) and consisted of simulated annealing (SA) cycles [for details, see the Supporting Information (SI)].

We first identified the minimal number of conformations that fitted the NMR data.[8] For this purpose, we carried out ensemble simulations starting from the X-ray structure of ubiquitin (1UBQ) using the SA protocol described in the SI and ensemble sizes ranging from 1 to 128. We then cross-validated the resulting ensembles against NMR parameters not used as restraints and obtained an optimal size of 64 (Figure S1 in the SI). Finally, to obtain a sufficiently large number of conformations to analyze correlated motions with statistical confidence (see below), we performed 12 consecutive SA cycles with an ensemble size of 64. After discarding the first two SA cycles to minimize the bias introduced by the first steps of the simulation, we obtained a 640-member ensemble that we named *e*nsemble *r*efinement for *n*ative proteins using a *s*ingle alignment *t*ensor (ERNST).

As shown in Table 1, ERNST is in good agreement with previously reported backbone scalar and H_NC' and NC' dipolar couplings[11] that were not used as restraints. The level of agreement is comparable to those for static structures (PDB code 1D3Z) and dynamic ensembles (PDB code 2K39) that, unlike ERNST, were refined using some of these RDCs. ERNST is also in very good agreement with both the NMR and X-ray structures of ubiquitin and is of lower structural heterogeneity than related ensembles[7] as a result of the nature of the protocol that was used to generate the ensemble.

We also analyzed the level of agreement of ERNST with NMR parameters sensitive to correlated motions. These included the cross-correlated relaxation rates $R_{NH,NH}$ measured by Bodenhausen and co-workers,[3] the rates $R_{NH,C\alpha H\alpha}$ measured here for ubiquitin using the experiment introduced by Vögeli and Yao[20] (Table S1 in the SI), and trans-hydrogen-bond scalar couplings ($^{3h}J_{NC'}$) previously reported by Cordier and Grzesiek.[21] These

Received: January 17, 2011
Published: June 02, 2011

Table 1. Level of Agreement with NMR Parameters and Structural Analysis of Previously Reported Static (1UBQ[18] and 1D3Z[11]) and Dynamic (2K39[7]) Representations of Ubiquitin, ERNST,[a] and an Ensemble Obtained Using Unbiased MD[b]

	1UBQ	1D3Z	2K39	MD	ERNST
NMR Parameter[c]			Validation		
$Q(D_{HNC'})$	0.29	**0.21**[d]	0.23	0.27	0.23
$Q(D_{NC'})$	0.22	**0.18**[d]	**0.18**	0.19	**0.18**
rmsd($^3J_{\varphi}$)/Hz[e]	0.42	0.38	0.45	0.46	**0.36**
rmsd($^{3h}J_{NC'}$)/Hz[f]	0.15	0.23	0.13	0.10	**0.09**
rmsd($R_{NH,NH}$)/Hz[f]	1.05	0.57	0.50	0.83	**0.46**
rmsd($R_{NH,C\alpha H\alpha}$)/Hz[f]	2.10	1.76	1.66	1.87	**1.61**
Structural Parameter[g]			Structural Analysis		
$\langle rmsd_{ij} \rangle$ (Å)	n.a.	n.a.	1.39	0.88	**0.83**
rmsd from 1UBQ (Å)	0	**0.38**	0.55	0.42	0.40
rmsd from 1D3Z (Å)	0.38	0	0.48	0.39	**0.33**

[a] Coordinates have been deposited in the PDB with code 2KOX. [b] The unbiased MD ensemble was obtained by setting the force constants of all restraints to zero. [c] For each NMR parameter, the structure or ensemble that gave the best agreement is shown in bold. [d] Some of these RDCs were used to refine 1D3Z. [e] Defined in the SI. [f] The equation used for the calculation is given in the SI. [g] The ensemble of minimal heterogeneity and the model most similar to the reference structure are shown in bold.

parameters are useful for validating correlated motions because they average on the same time scale as RDCs and are sensitive to the degree of correlation of the motions of two distinct bond vectors.[2−5,21] In regard to correlated motions in the β-sheet of ubiquitin, $R_{NH,NH}$ and $R_{NH,C\alpha H\alpha}$ values report on the relative motion of NH and CαHα bond vectors in consecutive residues and $^{3h}J_{NC'}$ values report on the relative motion of NH and CO bond vectors of hydrogen-bonded residues in β-strands.

Table 1 shows that cross-correlated relaxation rates involving NHs are sensitive to an accurate representation of the orientation of NH bond vectors.[22] The only structural models of ubiquitin that gave relatively accurate values of $R_{NH,NH}$ and $R_{NH,C\alpha H\alpha}$ are those that were restrained with NH RDCs (1D3Z, 2K39, and ERNST). The $^{3h}J_{NC'}$ data (Figure S2) were quite well reproduced by an unrestrained MD simulation [root-mean-square deviation (rmsd) = 0.10 Hz] and especially by ERNST. This confirms the impact of bond vector dynamics on this NMR parameter.[23] A comparison of the performance of the different structural models available for ubiquitin indicated that ERNST gave the best agreement with the NMR parameters that are sensitive to correlated motions, justifying the use of this ensemble for the investigation of this dynamical property.

To investigate correlated motions on the submillisecond time scale, we analyzed the statistical dependence of the φ and ψ torsion angles of ubiquitin in ERNST (Figure 1A). We observed strong sequential anticorrelations ($\rho \approx -0.5$) of φ_i and ψ_{i-1} for residues flanking amide bonds, especially in β-strands [circular correlation coefficient (ρ) ≈ -0.7]. We also observed sequential correlations of φ and ψ for the same residue in turns, helices, and loops ($\rho \approx -0.4$). The former is due to the planarity of the peptide bond and is associated with the crankshaft motion observed in molecular simulations[24,25] and in an analysis of the dynamics of protein G from RDCs.[26] The latter stems from local steric effects that render φ and ψ of residues in nonextended

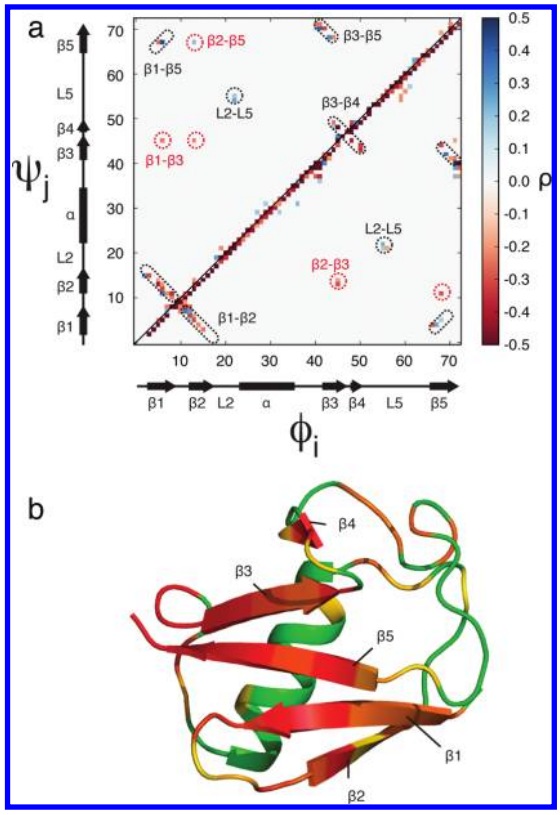

Figure 1. (A) Circular correlation coefficients[19] (ρ) of φ_i and ψ_j in ubiquitin. Black ellipses indicate hydrogen-bonded pairs of residues, and red circles indicate residues with long-range correlations. (B) Solution structure of ubiquitin (PDB code 1D3Z) colored according to the absolute value of the highest nonsequential correlation. Green indicates no correlation and red a correlation with absolute value greater than 0.5.

conformations nonindependent, as shown by Pappu et al.[27] using molecular mechanics calculations. We term the motion that leads to the φ_i/ψ_i correlation the intraresidue motion. The degree to which such steric effects limit motion in helical turns is evident from the circular variances (σ^2)[19] of φ and ψ. These were found to be significantly lower in the loops and α-helix [σ^2(α-helix) = 0.007] than in the β-strands [$\sigma(\beta\text{-strand})^2$ = 0.028], consistent with the relatively high RDC-derived order parameters obtained in other analyses of the dynamics of ubiquitin.[28]

It has been reported that invoking correlations between pairs of hydrogen-bonded residues improved the agreement with experimental $^{3h}J_{NC'}$ data in a study of the dynamics of protein G using the 3D-GAF motional model.[29] Similar correlations have also been observed in an MD simulation of ubiquitin.[30] In agreement with this, we observed correlations between the backbone torsion angles of residues that are distant in sequence but connected through hydrogen bonds (shown in black in Figure 1A). In ERNST, in contrast to in silico observations for ubiquitin,[30] such nonsequential correlations were observed between essentially all pairs of hydrogen-bonded residues in the β-sheet (β2 vs β1, β1 vs β5, β5 vs β3, and β3 vs β4) and also between hydrogen-bonded loops (loops 2 and 5). These correlations are propagated via noncovalent interactions and are generally not as strong ($\rho \approx -0.3$) as sequential correlations.

A large fraction of the nonsequential short-range correlations can be rationalized qualitatively by considering the structure of the protein. When pairs of β-strands are in the antiparallel

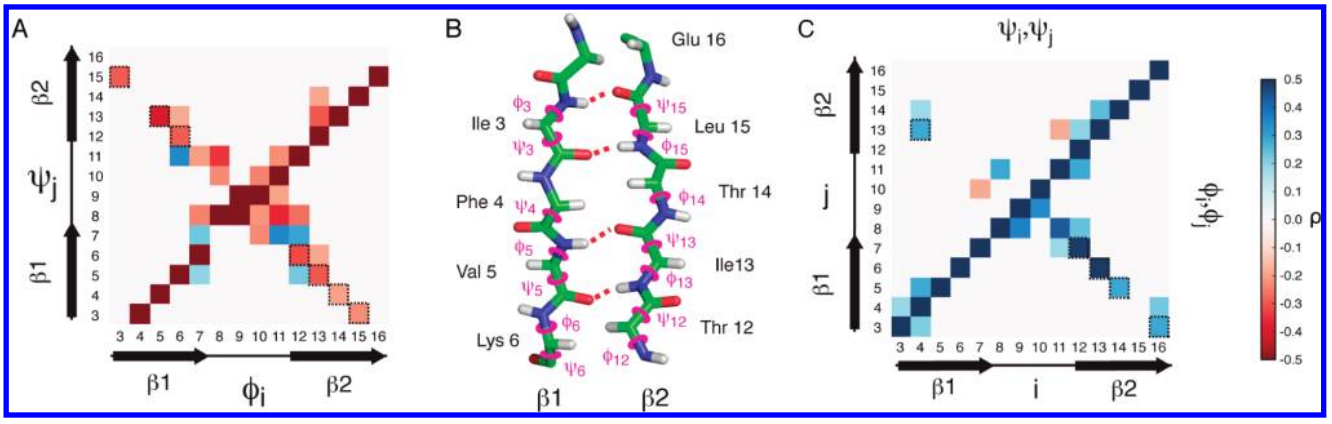

Figure 2. (A) Circular correlation coefficients[19] (ρ) of φ_i and ψ_j in the first two β-strands of ubiquitin. Black dashed squares indicate correlations associated with the β-lever motion either directly or indirectly via crankshaft motions. (B) Structure of β1 and β2 in ubiquitin (PDB code 1D3Z). Torsion angles associated with the β-lever motion are shown in purple. Hydrogen bonds are highlighted by red dashes. (C) Correlations between φ_i and φ_j and ψ_i and ψ_j. Black dashed squares indicate correlations associated with β-lever and crankshaft motions that occur in concert.

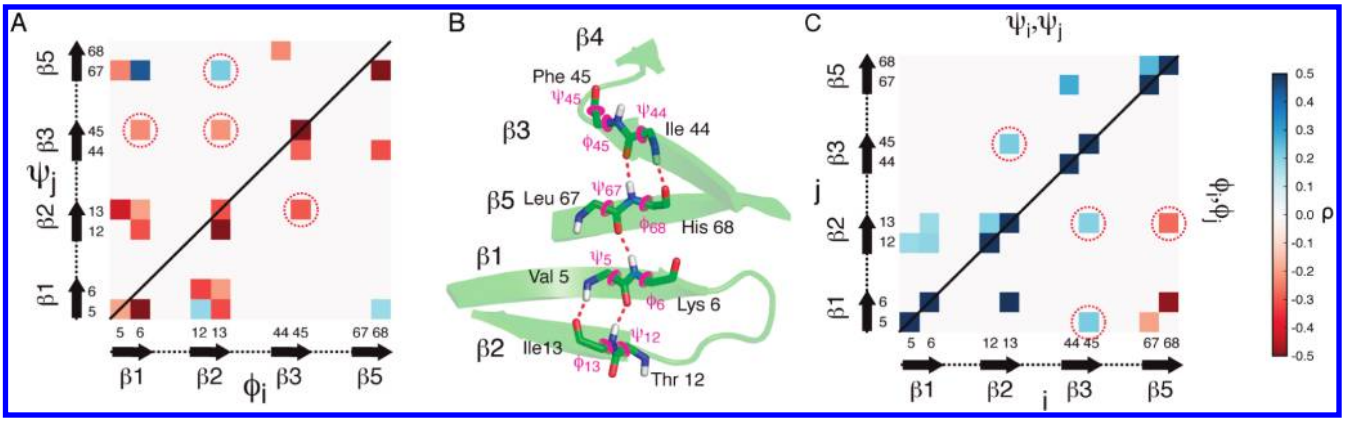

Figure 3. (A) Circular correlation coefficients[19] (ρ) of φ and ψ of residues that are part of the surface patch of ubiquitin involved in binding to UBDs. (B) Representation of the corresponding β-strands (PDB code 1D3Z). (C) Correlation between φ_i and ψ_j of residues that are part of the network. Correlations involving distal residues are indicated by red circles in A and C.

orientation [as is the case for β2 vs β1, β5 vs β3, and β3 vs β4 (Figure 1B)], the φ values of hydrogen-bond donor residues and ψ values of acceptor residues (in-register) are anticorrelated. This motion, here termed β-lever, allows conformational changes to occur while preserving hydrogen bonds. To illustrate its prevalence, we present in Figure 2 strands β1 and β2 in atomic detail as well as an analysis of the correlations observed in this region. In these strands, we also observed correlations between the φ and ψ of residues that are in-register but not hydrogen-bonded, such as Lys6 and Thr12 (Figure 2A), and between the φ and φ (and ψ and ψ) of residues that are out-of-register, such as Lys6 and Ile13 (Figure 2C). These correlations can be attributed to concerted crankshaft and β-lever motions, and their signs can be predicted by consideration of the signs of these two correlations. It should be noted that for parallel β-strands, such as β1 vs β5, the β-lever motion causes instead a positive correlation between the torsion angles of hydrogen-bonded residues (out-of-register) that results from the change in orientation of the two strands (Figure S3).

The short-range crankshaft, intraresidue, and β-lever motions observed in the β-sheet of the ubiquitin backbone are necessary but insufficient for propagation of structural and dynamical information across the long distances required for binding allostery and signal transduction. For these biological processes to operate efficiently, these essential motions must occur in concert to link distal residues dynamically. We observed a number of statistically significant ($p < 0.05$) long-range correlations (shown in red in Figure 1A). These correlations connect strands β2 and β5 (Ile13 and Leu67), strands β1 and β3 (Lys6 and Phe45) and, extraordinarily, strands β2 and β3 (Ile13 and Phe45; Thr14 and Phe45) that span four β-strands of ubiquitin and a distance of ~15 Å (Figure 3 and Figure S5). These properties of ERNST are robust with respect to changes in the simulation protocol (Figure S4).

To determine whether these long-range correlations were a result of the CHARMM27 force field or arose from the information on the structural heterogeneity of ubiquitin contained in the experimental data, we carried out control unrestrained simulations. The resulting in silico ensembles (see Figure S4) contained a fraction of the short-range correlations that we observed in the experimental ensemble and none of the long-range correlations shown in Figure 1. These results indicate that the experimental data act as a correction to the force field, allowing the experimental ensemble to reflect correlated motions that cannot be observed with confidence in silico.[30]

The mechanism by which such correlations operate is likely to involve covalent and noncovalent interactions between atoms in the backbone and side chains of ubiquitin. Despite this

complexity, we were able to trace a direct path of sequential and short-range concerted motions connecting the dihedral angles of the longest-range correlation between residues Ile13 in β2 and Phe45 in β3. This path, which is presented in Figure 3B, involves the β-lever motion of Ile13 and Val5, the crankshaft motion of Val5 and Lys6, the β-lever motion of Lys6 and His68, the β-lever of His68 and Ile44, the crankshaft motion of Ile44 and Phe45, and the intraresidue correlation of Phe45.

The nonsequential correlations, connected as a path, provide channels for the relay of structural and dynamical information across protein structures. These have long been hypothesized to be ubiquitous in biology but remain challenging to detect. To investigate whether the path observed in ERNST can contribute to the function of ubiquitin, we analyzed the relationship between the correlations and the binding modes of ubiquitin binding domains (UBDs). The vast majority of UBDs bind to the surface of ubiquitin around Ile44.[6] Previous studies have shown that the regions of ubiquitin most often found at the binding interface are the loop region between strands β1 and β2, the C-terminal end of strand β5, and the loop region between strands β3 and β4.[7] These regions coincide with both the residues having the largest nonsequential correlations (shown in Figure 2B) and the network of hydrogen bonds that mediate the long-range correlations in Figure 3. We therefore observe that the long-range correlations present in ubiquitin link the residues that are most important for its function.

A further strong indication for the existence of long-range correlated motions in ERNST comes from the observation that a principal component analysis of the residues shown in Figure 3 and those extracted from the ensemble of crystal structures of ubiquitin complexed with recognition proteins revealed similar crankshaft and β-lever motions. The similarity between the β-strand motions in the solution ensemble of ubiquitin and the ubiquitin complexes suggests that these long-range correlations are important for protein−protein recognition.

In conclusion our analysis illustrates how long-range correlations dynamically link distal residues that contribute to the binding interface of a protein and suggests a mechanism by which conformational selection could operate in molecular recognition. The strategy used in this study is general and could be used to obtain atomic-resolution descriptions of signal transduction and binding allostery, as demonstrated by our integrated description of the dynamics, correlations, and biological function of ubiquitin.

■ ASSOCIATED CONTENT

Supporting Information. Details of the simulation protocol and the equations used to back-calculate the NMR parameters, experimental $R_{NH,C\alpha H\alpha}$ data, additional correlation plots, correlation plots for the $^{3h}J_{NC'}$ data, equations used to compute ρ and σ values, statistical analysis of the correlation of longest range, and complete ref 15. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

cigr@nmr.mpibpc.mpg.de; xavier.salvatella@irbbarcelona.org

■ ACKNOWLEDGMENT

This work was supported through grants from IRB Barcelona, NSF, the Leverhulme Trust, the Wellcome Trust, ICREA, the Max Planck Society, the Fonds der Chemischen Industrie, and the Ministerio de Ciencia e Innovación (CTQ-2009-08850-BQU). The work leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (ERC Grant Agreement 233227). The authors acknowledge the use of computational resources of the Red Española de Supercomputación.

■ REFERENCES

(1) Al-Hashimi, H.; Gosser, Y.; Gorin, A.; Hu, W.; Majumdar, A.; Patel, D. J. *J. Mol. Biol.* **2002**, *315*, 95–102.
(2) Frueh, D. *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *41*, 305–324.
(3) Pelupessy, P.; Ravindranathan, S.; Bodenhausen, G. *J. Biomol. NMR* **2003**, 265–280.
(4) Lundström, P.; Mulder, F. A.; Akke, M. *Proc. Natl. Acad. Sci. U.S. A.* **2005**, *102*, 16984–16989.
(5) Vögeli, B. *J. Chem. Phys.* **2010**, *133*, No. 014501.
(6) Lange, O. F.; Lakomek, N. A.; Farès, C.; Schröder, G. F.; Walter, K. F.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471–1475.
(7) Markwick, P. R. L.; Bouvignies, G.; Salmon, L.; McCammon, J. A.; Nilges, M.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 16968–16975.
(8) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923–2938.
(9) De Simone, A.; Richter, B.; Salvatella, X.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 3810–3811.
(10) Fenwick, R. B.; Esteban-Martín, S.; Salvatella, X. *J. Phys. Chem. Lett.* **2010**, *1*, 3438–3441.
(11) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836–6837.
(12) Lakomek, N. A.; Walter, K. F.; Farès, C.; Lange, O. F.; de Groot, B. L.; Grubmüller, H.; Brüschweiler, R.; Munk, A.; Becker, S.; Meiler, J.; Griesinger, C. *J. Biomol. NMR* **2008**, *41*, 139–155.
(13) Mackerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584–1604.
(14) Beglov, D.; Roux, B. *J. Chem. Phys.* **1999**, *100*, 9050–9063.
(15) Brooks, B. R.; et al. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
(16) Richter, B.; Gsponer, J.; Várnai, P.; Salvatella, X.; Vendruscolo, M. *J. Biomol. NMR* **2007**, *37*, 117–135.
(17) Blackledge, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 23–61.
(18) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531–544.
(19) Jammalamadaka, S. R.; Sengupta, A. *Topics in Circular Statistics*; World Scientific: River Edge, NJ, 2001.
(20) Vögeli, B.; Yao, L. *J. Am. Chem. Soc.* **2009**, *131*, 3668–3678.
(21) Cordier, F.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 1601–1602.
(22) Reif, B.; Hennig, M.; Griesinger, C. *Science* **1997**, *276*, 1230–1233.
(23) Gsponer, J.; Hopearuoho, H.; Cavalli, A.; Dobson, C. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2006**, *128*, 15127–15135.
(24) Go, M.; Go, N. *Biopolymers* **1976**, *15*, 1119–1127.
(25) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–590.
(26) Clore, G. M.; Schwieters, C. D. *Biochemistry* **2004**, *43*, 10678–10691.
(27) Pappu, R. V.; Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U. S.A.* **2000**, *97*, 12565–12570.
(28) Salmon, L.; Bouvignies, G.; Markwick, P.; Lakomek, N.; Showalter, S.; Li, D. W.; Walter, K.; Griesinger, C.; Brüschweiler, R.; Blackledge, M. *Angew. Chem., Int. Ed.* **2009**, *48*, 4154–4157.
(29) Bouvignies, G.; Bernadó, P.; Meier, S.; Cho, K.; Grzesiek, S.; Brüschweiler, R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13885–13890.
(30) Li, D. W.; Meng, D.; Brüschweiler, R. *J. Am. Chem. Soc.* **2009**, *131*, 14610–4611.