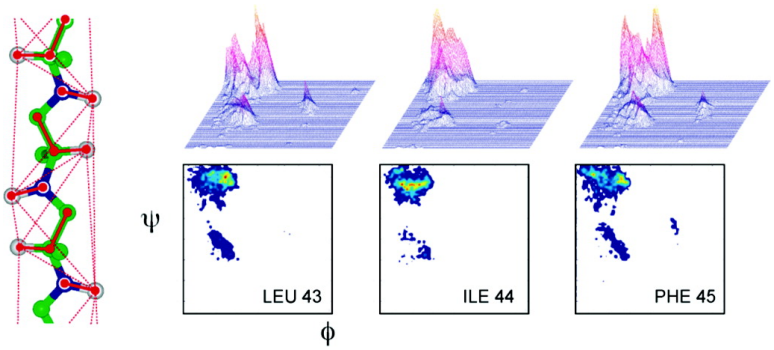
Article

# Mapping the Conformational Landscape of Urea-Denatured Ubiquitin Using Residual Dipolar Couplings

Sebastian Meier, Stephan Grzesiek, and Martin Blackledge

*J. Am. Chem. Soc.*, **2007**, 129 (31), 9799-9807• DOI: 10.1021/ja0724339 • Publication Date (Web): 18 July 2007






## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 7 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 07/18/2007

# Mapping the Conformational Landscape of Urea-Denatured Ubiquitin Using Residual Dipolar Couplings

Sebastian Meier,[†,§] Stephan Grzesiek,*[,†] and Martin Blackledge*[,‡]

*Contribution from the Biozentrum, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland, Institut de Biologie Structurale Jean-Pierre Ebel, CEA, CNRS, UJF UMR 5075, 41 Rue Jules Horowitz, Grenoble 38027, France, and Institute of Molecular Biology and Physiology, August Krogh Building, University of Copenhagen, Universitetsparken 13, DK-2100 Copenhagen, Denmark*

Received April 14, 2007; E-mail: martin.blackledge@ibs.fr; stephan.grzesiek@unibas.ch

***Abstract:*** Characterization of the unfolded state is a fundamental prerequisite for understanding protein stability and folding. We have investigated local conformational sampling in urea-denatured ubiquitin at pH 2.5 by measuring an extensive set of residual dipolar couplings (RDCs) under conditions of partial molecular alignment. Seven experimental RDCs per peptide unit, including complementary fixed-geometry and interproton ($^1H^N-^1H^N$ and $^1H^N-^1H^\alpha$) couplings, were used to investigate the structural properties of the peptide chain. Amino-acid-specific potentials that simultaneously reproduce all RDCs in the molecule are found to sample more extended conformations than the standard statistical coil description. Analysis of $^3J_{HNH\alpha}$ scalar couplings measured under the same conditions suggests that neither polyproline II nor extended $\beta$-regions dominate this additional sampling of extended conformations. Using this approach we propose a model of the conformational landscape throughout the peptide chain of urea-denatured ubiquitin, providing an essential description for understanding the unfolded state.

## Introduction

Characterization of the conformational properties of unfolded proteins provides the basis for understanding both the mechanisms of protein folding and the molecular function of natively unfolded proteins.[1,2] While the vast amount of structural information defining stable protein structures to high atomic resolution has provided a solid basis for relating structure to function in soluble proteins, the unfolded state occupies a conformational landscape that remains relatively poorly understood. This basic information is however central for accurate interpretation of experimental data on protein stability and folding,[3] on protein misfolding and development of associated disease,[4] and on the induced folding upon binding that is fundamental to the molecular function of natively unfolded proteins.[5,6] In this study we are interested in describing the conformational behavior of urea-denatured ubiquitin. Despite the large volume of experimental data measured in chemically denatured proteins, the effect of high concentrations of urea on the local structural and dynamic properties of the peptide chain, and their relation to the mechanism of protein unfolding, remain largely unknown. An accurate representation of conformational sampling in urea-denatured proteins will help to establish fundamental rules for understanding the molecular basis of protein stability and folding.

NMR is particularly well-suited to the study of unfolded proteins in solution,[7] providing site-specific characterization of structural properties along the entire peptide chain.[8−12] Local conformational properties in different partially folded and denatured forms of ubiquitin have been investigated using NMR secondary chemical shifts and scalar couplings[13,14] and spin relaxation.[15,16] Strong correlations were shown to exist in ubiquitin between both *J*-couplings and cross-correlated relaxation rates, sensitive to local backbone dihedral angles, and values calculated by averaging over a statistical coil model derived from a structural database.[13] Recently residual dipolar couplings (RDCs), measured under conditions of partial molecular alignment,[17] and reporting on ensemble- and time-averaged conformational sampling up to the millisecond time

† University of Basel.
‡ Institut de Biologie Structurale Jean-Pierre Ebel, CEA CNRS UJF.
§ University of Copenhagen.

(1) Dill, K. A.; Shortle, D. *Annu. Rev. Biochem*. **1991**, *60*, 795−825.
(2) Dyson, H. J.; Wright, P. E. *Nat. Rev.* **2005**, *9*, 197−208.
(3) Daggett, V.; Fersht, A. R. *Nat. Rev.* **2003**, *4*, 497−502.
(4) Dobson, C. M. *Nature* **2003**, *426*, 884−890.
(5) Fink, A. L. *Curr. Opin. Struct. Biol.* **2005**, *15*, 35−41.
(6) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527−533.
(7) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607−3622.
(8) Neri, D.; Billeter, M.; Wider, G.; Wuthrich, K. *Science* **1992**, *257*, 1559−1563.
(9) Alexandrescu, A. T.; Evans, P. A.; Pitkeathly, M.; Baum, J.; Dobson, C. M. *Biochemistry* **1993**, *32*, 1707−1718.
(10) Vendruscolo, M.; Paci, E.; Karplus, M.; Dobson, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14817−14821.
(11) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494−506.
(12) Choy, W.-Y.; Mulder, F. A. A.; Crowhurst, K. A.; Muhandiram, D. R.; Millett, I. S.; Doniach, S.; Forman-Kay, J. D.; Kay, L. E. *J. Mol. Biol.* **2002**, *316*, 101−112.
(13) Peti, W.; Henning, M.; Smith, L. J.; Schwalbe, H. *J. Am. Chem. Soc.* **2000**, *122*, 12017−12018.
(14) Wirmer, J.; Schwalbe, H. *J. Biomol. NMR* **2002**, *23*, 47−55.
(15) Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *Biochemistry* **1997**, *36*, 13043−13053.
(16) Wirmer, J.; Peti, W.; Schwalbe, H. *J. Biomol. NMR* **2006**, *35*, 175−186.
(17) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114.

range[18,19] have been shown to be sensitive to levels of structure that are difficult to detect using other techniques, thereby offering an increasingly promising tool to quantitatively assess the degree of order present in unfolded proteins.[20−26] In this study we use an extensive set of RDCs measured in ubiquitin, denatured in 8 M urea, and aligned in polyacrylamide gel to develop a description of the conformational landscape of the peptide chain under these conditions.

### Conformational Sampling in Unfolded Proteins

The physical interpretation of RDCs in terms of levels of order in unstructured proteins remains the subject of some discussion. RDCs have a simple dependence on the internuclear dipolar vector:

$$D_{ij} = -\frac{\gamma_i\gamma_j\mu_0 h}{8\pi^3}\left\langle \frac{P_2(\cos\theta(t))}{r_{ij}^{\,3}} \right\rangle \quad (1)$$

where $\theta$ is the instantaneous angle of the internuclear vector with respect to the magnetic field, and the average is taken over all sampled orientations up to the millisecond range in all members of the ensemble. The frequently observed pattern of negative $^1D_{NH}$ RDCs measured along the unfolded chain in longitudinally stretched gels, with larger couplings at the midpoint of the chain tapering off to negligible values at the chain termini, can be understood in terms of average bond orientations that are expected to be more often perpendicular than parallel to the static field. Gaussian chain models have been used to present related observations in more formal terms.[21] Fine structure that departs from this general distribution has been interpreted in terms of local turn conformations resulting in positive $^1D_{NH}$, due to inversion of the sign of the angular average[22−24] (the bond vector has a higher probability to align with the field). Such interpretations have the advantage of being intuitively accessible, describing the general features of expected distributions of $^1D_{NH}$ in unfolded proteins. In order to further understand the significant site-by-site dispersion observed for RDCs in unfolded proteins, it is useful to consider two recently presented approaches that explicitly account for the heteropolymeric nature of the peptide chain.[27,28] Both methods use random sampling of distinct amino-acid-specific $\phi/\psi$ propensities, derived from a structural database of nonsecondary structural elements in folded proteins, to construct a large number of conformers of the protein of interest. One of these approaches, termed flexible-meccano (FM), explicitly calculates an ensemble of conformers that can then be used to predict experimental properties for the disordered chain in solution. In particular, RDCs were calculated using a shape-based alignment prediction[29] applied to each copy of the ensemble using the expression

$$D_{ij} = -\frac{\gamma_i\gamma_j\mu_0 h}{16\pi r_{ij}^{\,3}}\left(A_a(3\cos^2[\Theta-1]) + \frac{3}{2}A_r\sin^2\Theta\cos 2\Phi\right) \quad (2)$$

where $A_a$ and $A_r$ are the axial and rhombic components of the alignment tensor of the conformer and $\{\Theta,\Phi\}$ the polar coordinates of the internuclear vector in the tensor frame. $D_{ij}$ is then averaged over all copies. This approach was initially applied to a two-domain protein, one of which was entirely unstructured, providing a quantitative analysis of the degree of order present in the unstructured domain, and closely reproducing both $^1D_{NH}$ and $^2D_{C'NH}$ couplings throughout the protein. Agreement was also found between experimental and simulated $^1D_{NH}$ couplings measured in a number of natively and urea-unfolded proteins.[27,30] The reproduction of experimental RDCs from both natively and chemically unfolded proteins establishes FM and related approaches as basic tools for studying unfolded proteins in solution, providing baseline values that result directly from the conformational properties of the primary sequence that can be considered as "random coil". Departure from these values can be interpreted in terms of specific local or long-range conformational behavior.

Despite these advances, very few studies have extended the investigation of RDCs in unfolded proteins beyond the analysis of $^1D_{NH}$ couplings. In this study, we seek to extend our understanding of the structural behavior of unfolded proteins by measuring a more extensive set of RDCs (up to seven couplings per peptide unit) in urea-denatured ubiquitin. These data are combined with FM analysis to provide a detailed description of the conformational phase space sampled by the chain under these conditions.

As discussed above, RDCs depend on the angular probability function of the dipolar orientation (eq 1), that can be numerically simulated by averaging eq 2 for each conformer. This provides access to the effective angular average with respect to the magnetic field that we will denote $\langle g(\Theta,\Phi)\rangle$ such that:

$$D_{av} = -\frac{\gamma_i\gamma_j\mu_0 h}{16\pi r_{ij}^{\,3}}\langle g(\Theta,\Phi)\rangle = -\frac{\gamma_i\gamma_j\mu_0 h}{8\pi r_{ij}^{\,3}}\langle P_2(\cos\theta)\rangle \quad (3)$$

As we demonstrate below, despite the apparent complexity of this function, examination of the behavior of $\langle g(\Theta,\Phi)\rangle$ can provide useful insight.

### Results and Discussion

**One-Bond (Fixed Distance) RDCs.** Three sets of one-bond RDCs ($^1D_{NH}$, $^1D_{C\alpha H\alpha}$, and $^1D_{C\alpha C'}$) were measured in urea-denatured ubiquitin aligned in stretched polyacrylamide gel. The data are shown in Figure 1 in comparison to predicted couplings determined using the statistical coil-based approach, flexible-meccano (FM).[27] The ability of the coil model to reproduce experimental data is also summarized in Table S1 in the Supporting Information. In the absence of a quantitative estimate of the absolute degree of alignment in the sample, RDC

(18) Prestegard, J. H.; al-Hashimi, H. M.; Tolman, J. R. *Q. Rev. Biophys.* **2000**, *33*, 371−424.
(19) Blackledge, M. *Progr. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 23−61.
(20) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487−489.
(21) Louhivuori, 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) Dames, S.; Aregger, R.; Vajpai, N.; Bernado, P.; Blackledge, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2006**, *128*, 13508−13514.
(26) Mukrasch, M. D.; Markwick, P. R. L.; Biernat, J.; von Bergen, M.; Bernado, P.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 5235−5243.
(27) 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.
(28) Jha, A. K.; Colubri, A.; Freed, K.; Sosnick, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099−13105.,
(29) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791−3792.
(30) Bernado, P.; Bertoncini, C.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2005**, *127*, 17968−17969.

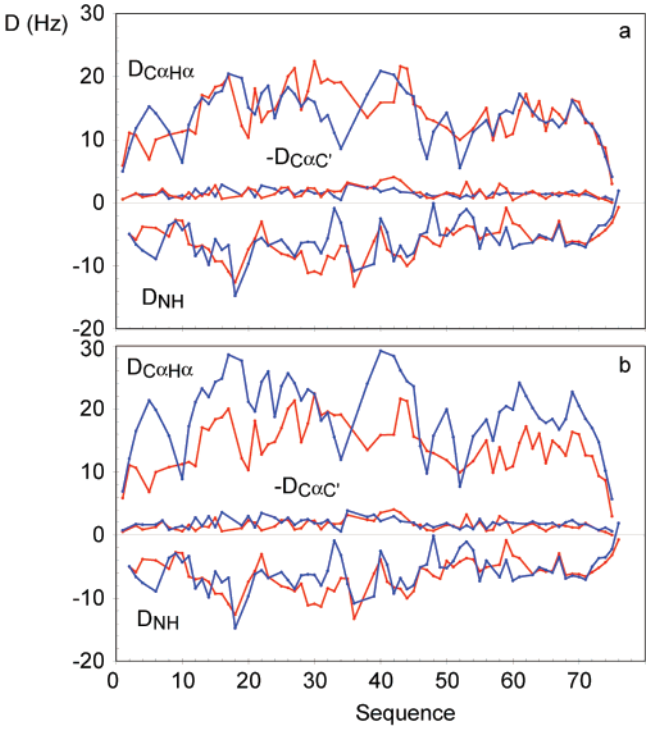



***Figure 1.*** Fixed geometry RDCs measured in 8 M urea unfolded ubiquitin at pH 2.5 aligned in stretched polyacrylamide gel. (a) Experimental $^1D_{NH}$, $^1D_{C\alpha H\alpha}$, and $^1D_{C\alpha C'}$ RDCs (red) compared to simulation using the FM approach (blue) with standard database sampling. The three datasets were scaled independently to best match the experimental data (scaling values in Table S1, Supporting Information). (b) RDCs shown in (a) using the same scaling factor for all couplings. The best matching scaling factor was optimal for $^1D_{NH}$.

distributions are normally scaled to facilitate comparison to experimental values. In this case we find that, while individual RDC types are well reproduced in terms of the distribution of RDCs along the sequence, they are not optimally scaled by the same factor in each case. When a common scaling factor is used such that the $^1D_{NH}$ couplings are best reproduced, $^1D_{C\alpha H\alpha}$ are overestimated by more than 40% (Figure 1b). Similarly $^1D_{C\alpha C'}$ couplings are overestimated by approximately 35%. One possible explanation for this could be differential dynamic fluctuations of the different bond vectors, possibly due to urea binding to the amide groups and restricting their motion. However, if this discrepancy were due to mobility alone, effective angular order parameters for $C^\alpha-H^\alpha$ bonds would be 40% higher than for $N-H^N$ bonds, requiring very specific dynamic modes throughout the protein. The interspin distances chosen to predict each coupling can of course also contribute to the differential scaling effect. The distances used here for $C^\alpha-H^\alpha$ and $N-H^N$ bonds are taken from the dynamically averaged values determined by Bax and co-workers (1.117 Å and 1.041 Å, respectively).[31] The effect of assuming even a rather short distance of 1.01 Å results in an effective differential shift in $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ scaling of 9%, indicating that an inaccurate estimate of interbond distances cannot realistically account for the observed effects.

An alternative explanation is identified from simulation: As the peptide plane associating two sequential amino acid backbone elements is essentially flat and the $C^\alpha$ junctions tetrahedral, $\langle g(\Theta,\Phi)\rangle$ is influenced by two factors: the sampling

(31) Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 12334−12341.

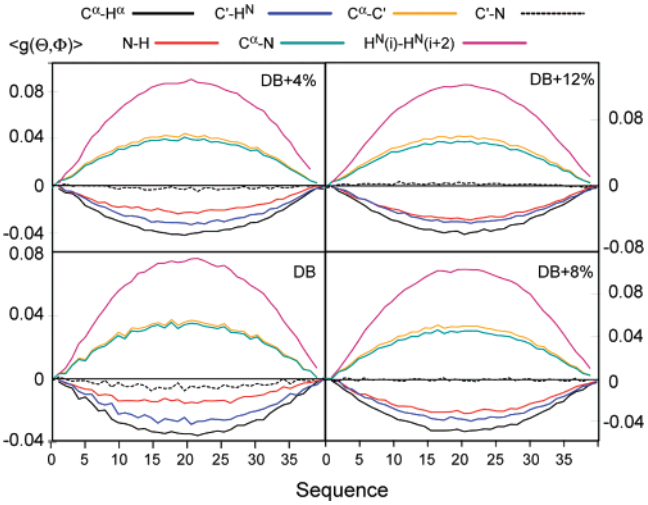



***Figure 2.*** Dependence of the angular contribution $\langle g(\Theta,\Phi)\rangle$ on conformational sampling. $\langle g(\Theta,\Phi)\rangle$ was calculated using the shape-based approximation to the alignment tensor[54] for a poly-Ala polymer, averaged over 50 000 conformations using standard FM database sampling of Ala (bottom left-hand panel). The three other panels show the effects of introducing 4−12% additional conformational sampling of more extended regions of Ramachandran space $\phi,\psi$ = (−105° ± 40°, 135° ± 15°), encompassing both $\beta$-sheet and PPII populations).

of the peptide chain (defined in terms of $\phi/\psi$ distributions) and the orientation of the bond vector within each peptide unit. The angular contribution $\langle g(\Theta,\Phi)\rangle$ to the effective average coupling simulated using FM is shown in Figure 2 for a number of different couplings in a poly-Ala molecule at different degrees of extendedness. $\langle g(\Theta,\Phi)\rangle$ is averaged over the FM ensemble using the standard database conformational sampling for Ala residues. The results demonstrate that, while the vectors implicated in the tetrahedral junction ($C^\alpha-H^\alpha$, $C^\alpha-N$, $C^\alpha-C'$) have similar absolute values, with either negative or positive signs, the remaining couplings present in the peptide plane ($N-H$,$C'-H^N$,$C'-N$) have lower angular averages. It is useful to recall that $\langle g(\Theta,\Phi)\rangle$ is identical to the angular average $2\langle P_2(\cos\theta)\rangle$ in eq 3 and therefore reports on the predicted average vector orientation with respect to the magnetic field. In some cases, for example the $^1D_{CN}$ coupling, the average orientation is sufficiently close to the "magic angle" that predicted values are close to zero. When the chain is forced to sample more extended conformations (Figure 2) such that the peptide chain is likely to align more strongly in the direction of the magnetic field, the relative averages change such that the $^1D_{NH}$ angular average approaches the $^1D_{C\alpha H\alpha}$ average, and the $^1D_{C'N}$ coupling actually goes through zero and changes sign. It is therefore evident that the differential scaling of different coupling types required to reproduce experimental data could be related to the degree of extendedness of the chain.

We note that the largest positive $\langle g(\Theta,\Phi)\rangle$ values for one-bond couplings ($^1D_{C\alpha N}$, $^1D_{C\alpha C'}$) are of similar absolute value as the smallest negative $\langle g(\Theta,\Phi)\rangle$ values ($^1D_{C\alpha H\alpha}$). Since $P_2(\cos\theta)$ has a lower bound of −0.5, the average orientations relative to the field must lie in a cone where $|P_2(\cos\theta)| < 0.5$ (48° < $\theta$ < 90°). Thus, the range of $P_2(\cos\theta)$ between 0.5 and 1 (0° < $\theta$ < 48°) is not effectively sampled by $^1D_{NH}$, $^1D_{C\alpha H\alpha}$, $^1D_{C\alpha C'}$, $^1D_{C\alpha N}$, $^1D_{C'N}$, or $^2D_{C'NH}$, implying also that interactions that are more coaxial with the magnetic field would sample angular order with greater sensitivity (up to a factor of 2). This is verified, at least by simulation, by calculating the expected angular average

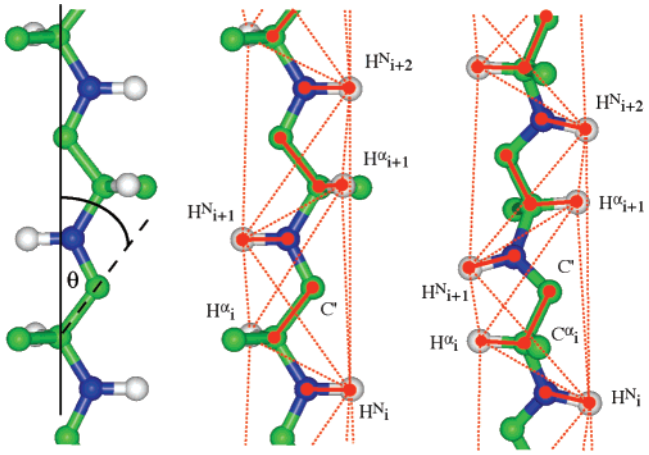



***Figure 3.*** Strands of extended and $\beta$-sheet conformations showing the measured couplings used in this study to define the conformational space native to urea-denatured ubiquitin.

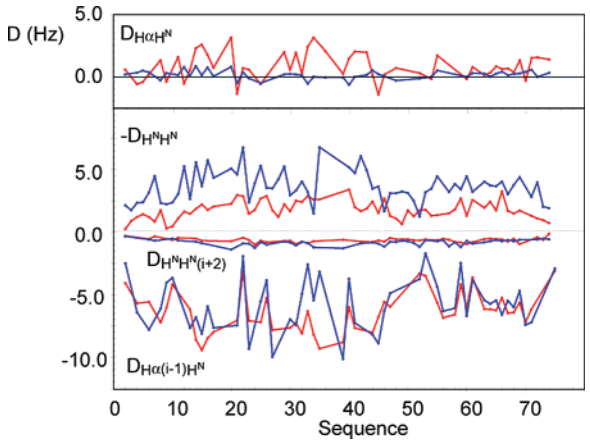



***Figure 4.*** Interproton RDCs measured in 8 M urea-unfolded ubiquitin at pH 2.5 aligned in stretched polyacrylamide gel. Experimental $^1H^N_i-^1H^\alpha_{i-1}$, $^1H^N_i-^1H^N_{i+1}$, and $^1H^N_i-^1H^N_{i+2}$ RDCs (red) compared to simulation using the FM approach (blue) with standard database sampling. The best matching scaling factor ($K = 1.7$) was optimal for $^1H^N_i-^1H^\alpha_{i-1}$.

for couplings between $^1H^N_i-^1H^N_{i+2}$, a vector known to follow the principal direction of $\beta$-sheet and extended chain structures (Figure 3). $\langle g(\Theta,\Phi)\rangle$ values for this vector are more than twice those predicted for bonds attached to the $C^\alpha$ (Figure 2). More generally, these simulations also suggest that measurement of additional probes, in particular interproton RDCs, may provide new and complementary information to one-bond RDCs.

**Long-Range Interproton RDCs.** $^1H^N-^1H^N$ and $^1H^N-^1H^\alpha$ RDCs were therefore measured in urea-denatured ubiquitin and compared to values predicted from the FM ensemble (Figure 4). In addition to sequential and inter-residual RDCs, $^1H^N_i-^1H^N_{i+2}$ couplings were measured throughout the chain. Again, different types of $^1H-^1H$ RDCs require different overall scaling factors to reproduce experimental ranges, suggesting that the simple coil database does not represent the conformational sampling correctly. This is illustrated in Figure 4 (and Table S1) where couplings are scaled to best reproduce sequential $^1H^N_i-^1H^\alpha_{i-1}$ RDCs. We find that standard coil database sampling predicts much larger $^1H^N_i-^1H^N_{i+1}$ RDCs than are actually measured. Sampling of more extended regions of the Ramachandran plot would increase the average distance between sequential amide protons, resulting in smaller couplings, and again may be at the origin of the observed differential scaling.

**A Self-Consistent Description of Conformational Sampling in Urea-Unfolded Ubiquitin.** It appears, therefore, that discrepancies between the ranges of experimental and predicted values for different RDC types indicate that $\phi/\psi$ sampling in the standard coil database does not represent the true conformational sampling appropriate for urea-denatured proteins. In order to test this hypothesis, and to identify a more accurate description of dihedral angle sampling in urea-unfolded ubiquitin, we have repeated the ensemble calculations under conditions where the chain samples more extended conformations.

Explicitly increasing the sampling of polyproline II (PPII) or $\beta$-sheet regions of Ramachandran space for all amino acids shows that the scaling factors ($K$) required to best match simulated and experimental RDCs converge toward similar values for higher levels of extended conformation (Table S1). This is true for both fixed geometry and interproton couplings, although $K$ values cluster around slightly different values (vide infra). The ability to fit data sets independently also improves for all individual coupling types (rms values in Table S1). This information is illustrated in Figure 5, where the reproduction of the couplings is shown with respect to increasing additional population of PPII (additional population of the extended $\beta$ region results in similar effects, shown in figure S1, Supporting Information). In these figures the fixed-geometry couplings are scaled to best match $^1D_{NH}$ couplings, while interproton couplings are scaled to best match $^1H^N_i-^1H^\alpha_{i-1}$ RDCs. The ability to fit all data simultaneously consistently improves with the level of extension, until between 12 and 20% of PPII or $\beta$-sheet conformations are sampled in addition to database levels. The total $\chi^2$ is significantly lower under these conditions compared to standard coil database, as shown in figure S2 in the Supporting Information.

**Comparison with Scalar Couplings Indicates That Neither $\beta$-Sheet nor PPII Dominates Extended Sampling.** In terms of local structure a large volume of experimental data, essentially from optical spectroscopy but also from NMR, has pointed to the importance of the PPII region of the Ramachandran plot in unstructured peptides as well as unfolded proteins[32,33] (see ref 34 for a recent review). Extensive measurements of scalar couplings are available in the literature for urea-unfolded ubiquitin.[13] We have therefore calculated the expected values of measured $^3J_{HNH\alpha}$ couplings for different ensembles that are in similar agreement with measured RDCs. Figure 6 summarizes the results, showing that, while ensembles additionally sampling PPII or $\beta$-sheet do not reproduce the data, predicting either systematically low or high values, a more general increased sampling of ($\phi < 0°$, $50° < \psi < 180°$) reproduces both the general trend, and local detail. On this basis we would propose the latter model as more appropriate, suggesting that the PPII region does not dominate conformational sampling in ubiquitin in the urea-unfolded state.[35]

**A Model for Conformational Sampling in the Urea-Unfolded State.** A model of chain extension is therefore proposed that increases the sampling of the extended regions

(32) Tiffany, M. L.; Krimm, S. *Biopolymers* **1968**, *6*, 1767−1770.
(33) Schellman, J. A.; Schellman, C. G. In *The Proteins*; Neurath, H., Ed.; Academic Press: New York, 1964; Vol. *2*, pp 1−37.
(34) Shi, Z. S.; Chen, K.; Liu, Z.; Kallenbach, N. R. *Chem. Rev.* **2006**, *106*, 1877−1897.
(35) Makowska, J.; Rodziewicz-Motowidło, S.; Baginska, K.; Vila, J. A.; Liwo, A.; Chmurzynski, L.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1744−1749.

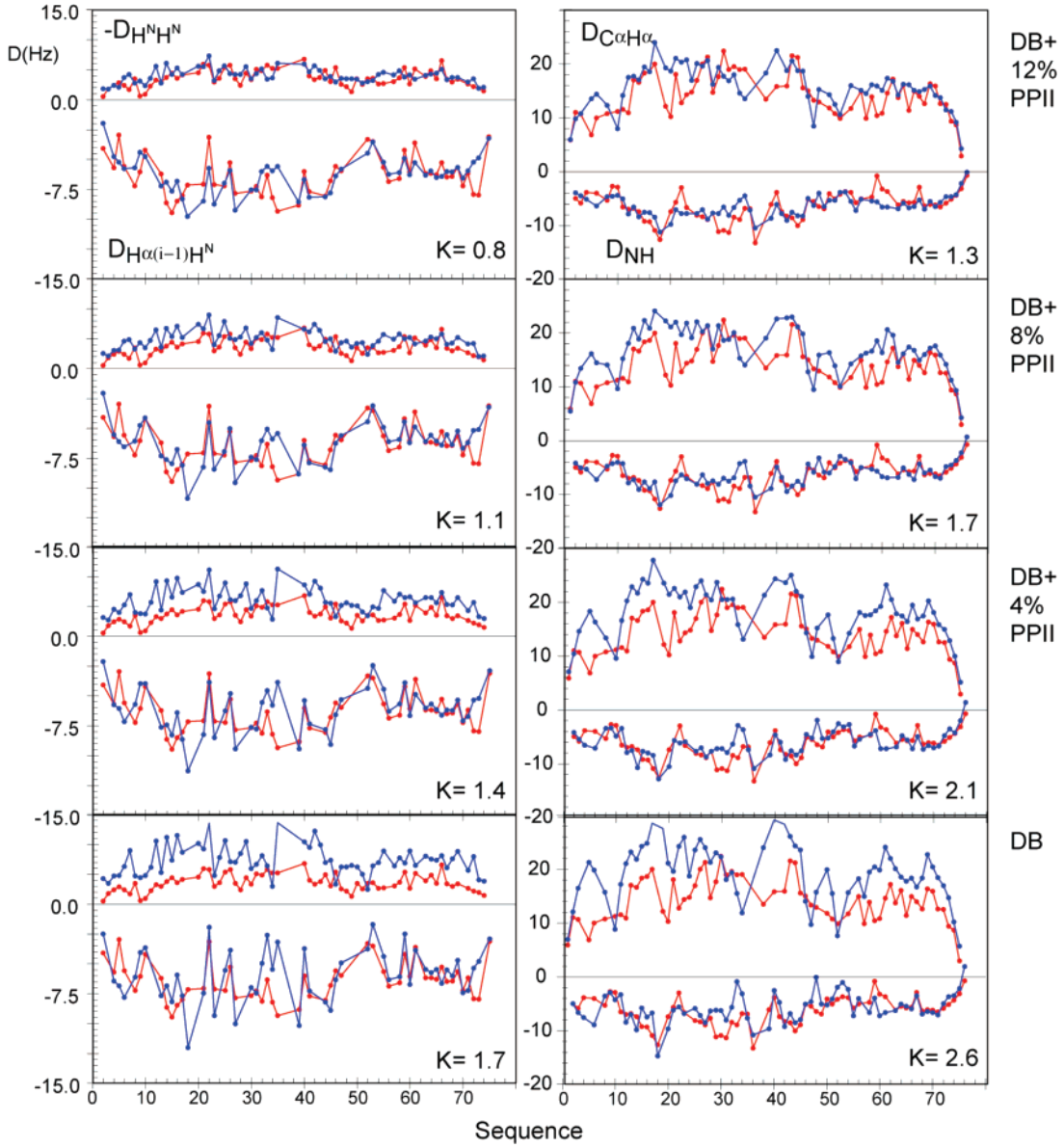



***Figure 5.*** Effect of sampling additional extended conformations on fixed geometry ($^1D_{NH}$ and $^1D_{C\alpha H\alpha}$) and interproton ($^1H^N_i-^1H^\alpha_{i-1}$ and $^1H^N_i-^1H^N_{i+1}$) RDCs. Scaling factors ($K$) are shown. Percentages shown on the right refer to the additional conformations that are specifically assigned to the PPII basin. (Left) Comparison of experimental interproton RDCs with simulated values. Increasing populations of polyproline II conformations $\phi,\psi = (-75 \pm 10°, 150 \pm 10°)$ were used in addition to database propensities. All RDCs were scaled using the $K$ value that best matched the $^1H^N_i-^1H^\alpha_{i-1}$ in each case. (Right) Comparison of experimental $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs with simulated values. Additional populations were used as in (a). All RDCs were scaled to best match the $^1D_{NH}$.

of Ramachandran space ($\phi < 0°$, $50° < \psi < 180°$) for each residue-specific population present in the database, while retaining aspects of the native conformational energy basins for each residue type. The results are again summarized in Table S1 for increasing propensity to sample this region. The results for all measured couplings, under apparently optimal sampling conditions, (around 78% of conformers sampling $\{\phi < 0°, 50° < \psi < 180°\}$ compared to 59% using the standard database distributions) are shown in Figure 7. The ability to reproduce the data simultaneously is shown to be essentially equivalent to similar levels of PPII and $\beta$-sheet for interproton couplings, whereas in the case of fixed geometry RDCs PPII sampling reproduces the data slightly better than the broader sampling (Figure S2). Table S2 in the Supporting Information summarizes the total $\chi^2$ that is significantly lower for this extended sampling than standard coil sampling, for both fixed geometry (25.7 compared to 64.6) and interproton RDCs (24.6 compared to 151.4).

The quality of the individual data reproduction again improves over the entire data set compared to that of the data base sampling shown in Figures 1 and 4. This is true for all couplings, but the improvements in the definition of the fine structure of $^1D_{C\alpha H\alpha}$ and in the overall distribution of intraresidual $^1H^N-^1H^\alpha$ RDCs are particularly clear from these figures. Correlation plots of all measured couplings are shown in Figure 8 for optimal extended sampling compared to the coil database.

On the basis of these results we can propose modified conformational basins for the different amino acids in the primary sequence. Examples are shown for four sequential residues in Figure 9, comparing coil database and optimal extended sampling. These examples demonstrate that, while sampled conformations are more extended, the amino-acid

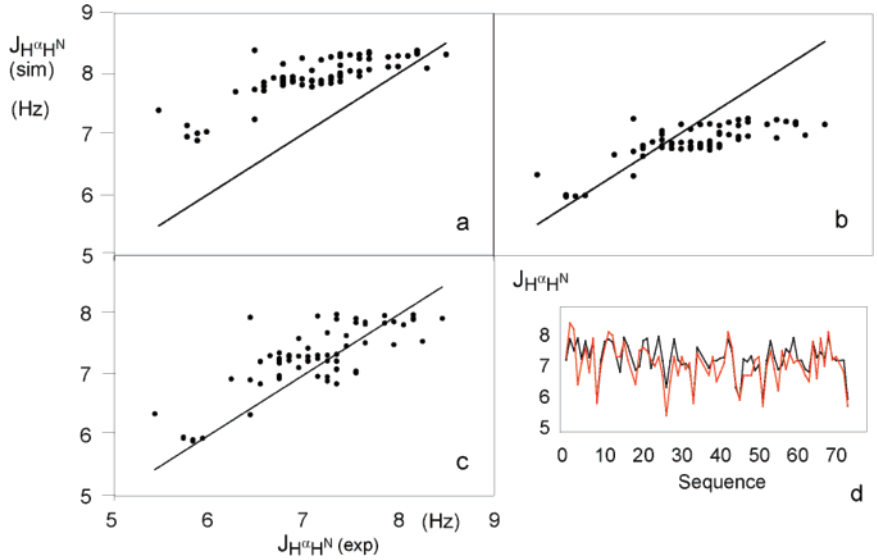



***Figure 6.*** Comparison of expected values of measured $^3J_{HNH\alpha}$ couplings for different ensembles with experimental values.[13] Values were averaged over 10 000 conformations. (a) 12% $\beta$-sheet population in addition to database sampling. (b) 12% PPII population in addition to database sampling. (c) Optimal propensity to sample ($\phi < 0°$, $50° < \psi < 180°$) regions of the database for each amino acid (as used for analysis presented in Figure 7). The standard database shows essentially the same values, as $\phi$ is sampled in a very similar way. (d) Sequence dependence of calculated (black) and experimental (red) couplings shown in Figure 6c.

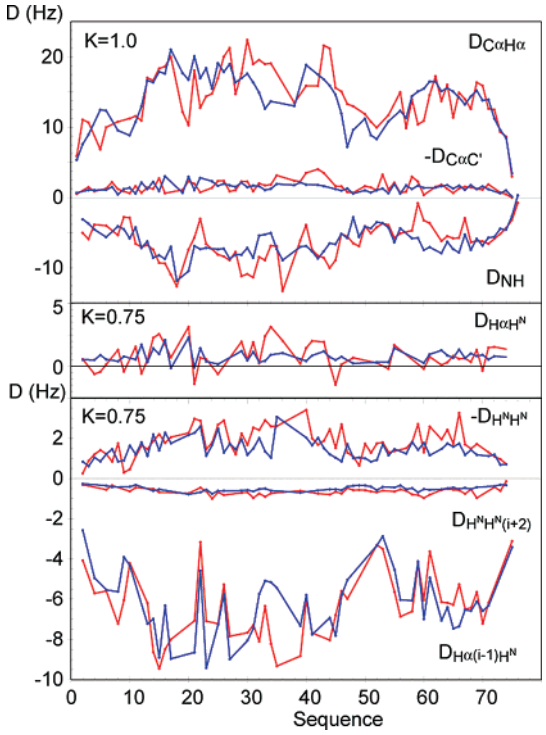



***Figure 7.*** All experimental couplings compared to simulated values calculated using conformational distributions with enhanced propensity to sample ($\phi < 0°$, $50° < \psi < 180°$) regions for each amino-acid-specific population. (a) All fixed geometry RDCs. All values were scaled by the same factor (1.0). (b) All interproton RDCs. All values were scaled by the same factor (0.75).

specificity is retained, and α-helical conformations are still sampled in each case.

This description is consistent with the presence of some population of native conformations for residues 3−15, implicated in the $\beta$-hairpin region of folded ubiquitin, that were previously identified on the basis of scalar couplings, chemical shifts, trans-hydrogen bond scalar couplings, and long-range $^1H^N-^1H^N$ RDCs.[36] To investigate the effect of such native $\phi/\psi$ conformations for these sites, they were explicitly incorporated into 20% of the members of the ensemble, the approximate level estimated from chemical shift data. The results (Supporting Information Figure S3) demonstrate that this level of $\beta$-hairpin does not severely influence the quality of RDC prediction in this region.

**Differential Dynamics Effects for Fixed Geometry and Proton−Proton RDCs.** Although the results are convergent with increased extension, slightly different prefactors are still required for fixed geometry and proton−proton couplings (Figure 7). This effect may be due to specific modes of local backbone dynamics. Indeed peptide plane fluctuations, in particular crankshaft-type motions,[37] would have the potential to quench experimentally measured couplings between protons in sequential peptides. The level to which this occurs depends on the correlation of the motions between adjacent planes, and it is probably not useful to speculate further than to suggest that the effective "differential order parameter" of around 0.75, the ratio between optimal scaling constants for fixed geometry and interproton couplings, is not inconsistent with reasonable levels of backbone dynamics in the unfolded state. This is also consistent with the observation that the $^1D_{C\alpha C'}$ coupling, corresponding to the vector expected to undergo the lowest differential angular fluctuations under this type of motion, shows the highest scaling factor (Table S1).

**Is This Local Extendedness Compatible with Available Biophysical Measurements?** A large volume of experimental data have also been measured to characterize the effects of chemical denaturants on polypeptides. The overall dimensions of guanidinium chloride-unfolded proteins have been extensively studied using small-angle X-ray scattering,[38] allowing a descrip-

(36) Meier, S.; Strohmeier, M.; Blackledge, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2007**, *129*, 745−746.
(37) Fadel, A. R.; Jin, D. Q.; Montelione, G. T.; Levy, R. M. *J. Biomol. NMR* **1995**, *6*, 221−226.
(38) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; Hasan, M. Z.; Pande, V. S.; Ruczinski, I.; Doniach, S.; Plaxco, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12491−12496.

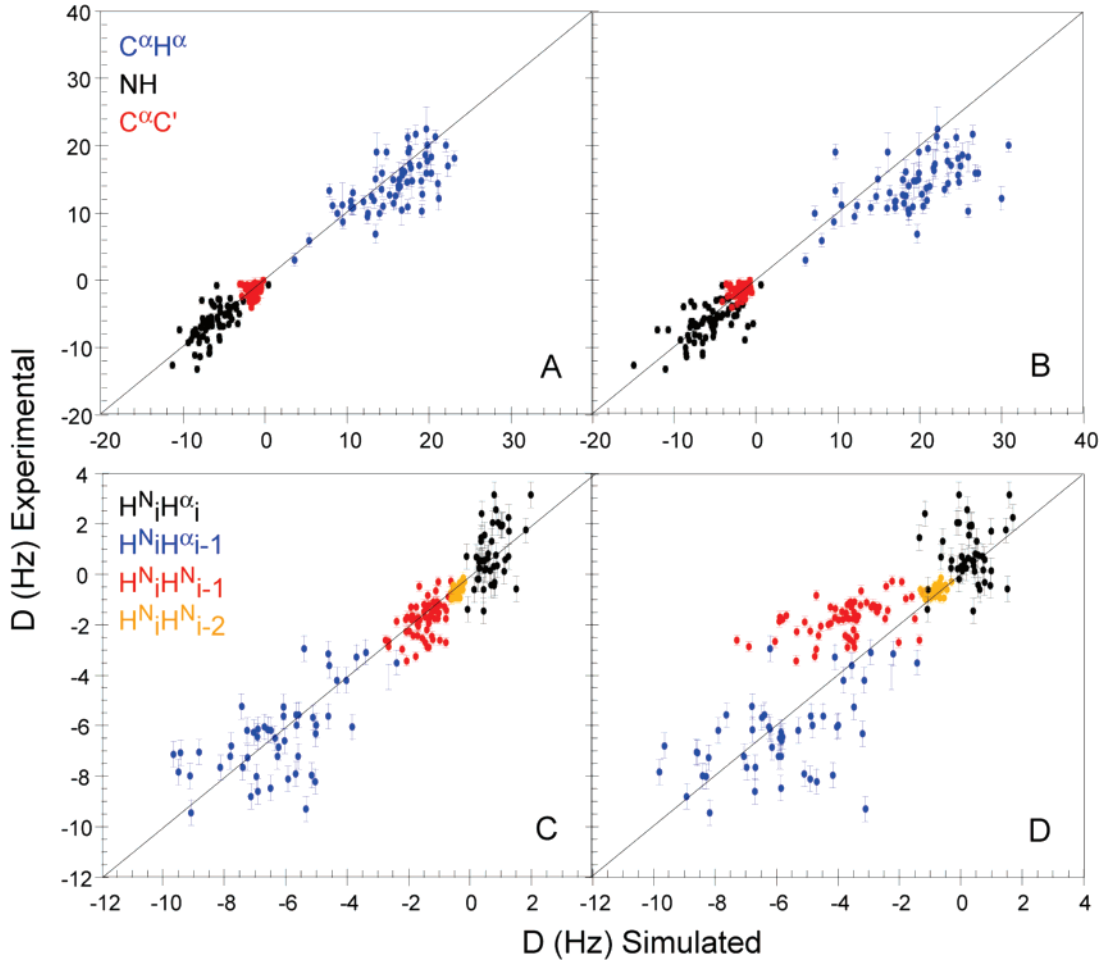



***Figure 8.*** Correlation plot of all measured couplings for sampling under extended conditions (A,C 78% conformers in the range ($\phi < 0°$, $50° < \psi < 180°$)), and the coil database (B,D), with scaling optimized for $^1D_{NH}$ for covalently bound RDCs (A,B) and optimized for $^1H^N_i-^1H^\alpha_{i-1}$ for interproton RDCs (C,D).

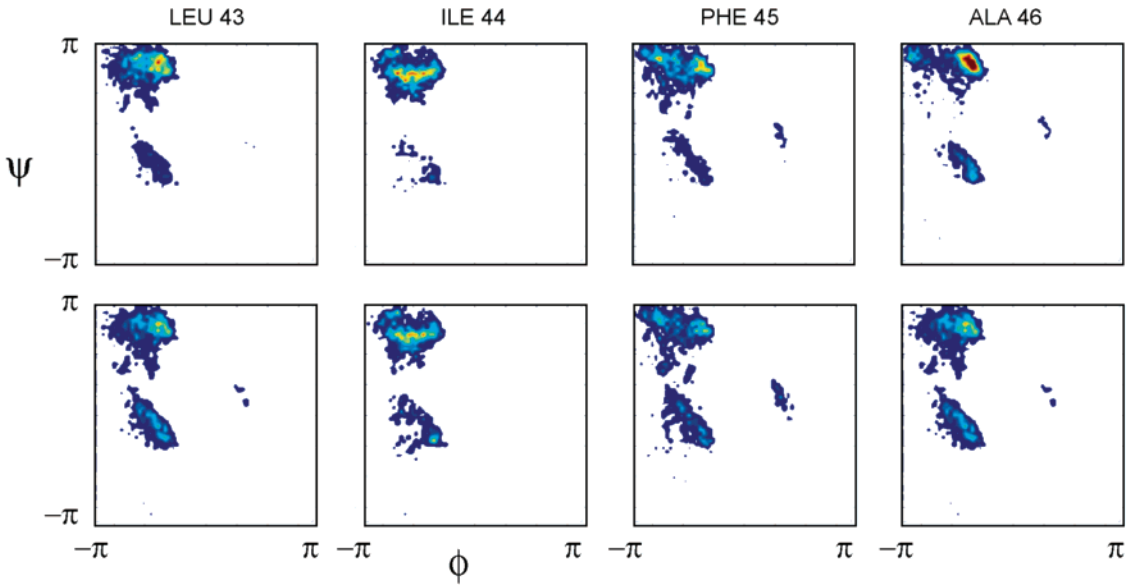



***Figure 9.*** Comparison of $\phi/\psi$ population distribution for four sequential amino acids in the case of database sampling (bottom) and optimal extended sampling used to produce Figure 7 (top). Dark-blue represents lower populations, light-blue/yellow represents intermediate, and red represents higher populations.

tion of the extendedness of unfolded chains on the basis of a simple power law relating the average radius of gyration to the number of amino acids in the chain.[39] FRET measurements on polypeptides and denatured proteins also indicate that denaturants act to extend end-to-end dimensions of peptide chains.[40,41]

(39) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley: New York, 1969.

(40) Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1528−1533.

We have compared the effective radius of gyration ($R_{g,eff}$) from the current simulations with experimental small-angle X-ray scattering data from 6 M guanidinium chloride-unfolded ubiquitin.[42] The effective radius calculated from the FM ensemble that reproduces the experimental NMR data is slightly larger than that measured experimentally (28 Å compared to 26 Å), while the standard database gives slightly smaller effective values (25 Å). We note, however, that simulations that retain the same local sampling but impose a small number of long-range contacts reproduce both $R_{g,eff}$ derived from scattering data and agree equally well with experimental RDCs (data not shown). Similarly, the recently demonstrated presence of a $\beta$-hairpin in urea-unfolded ubiquitin will reduce the effective radius of gyration.

**Extended Local Sampling and the Mechanism of Urea Denaturation of Proteins.** Despite the large volume of experimental data measured on proteins in the presence of urea, the actual mechanism of protein denaturation remains the subject of debate. Two models have been identified. The first invokes the different hydrogen-bonding properties of urea compared to those of water in the vicinity of hydrophobic groups, such that urea acts as a better solvent for groups that are normally buried in water-solvated proteins.[43] An alternative model proposes that urea binds directly to the protein backbone, presumably via hydrogen-bonding interaction with the amide groups, thereby destabilizing hydrogen-bonding networks present in the native fold relative to those in the unfolded state.[44] Although it is not possible to differentiate between proposed models of urea-denaturation of proteins from our results, we note that a systematically more extended conformational sampling behavior could be consistent with a model of urea binding to the peptide backbone, excluding more compact turnlike structures, and occupying important potential hydrogen-bonding sites that stabilize turns.[45]

We can also compare these results with those found for natively unfolded proteins. We have applied the extended conformational sampling appropriate for urea-unfolded ubiquitin to the prediction of $^1D_{NH}$ and $^2D_{C'NH}$ RDCs from protein X, a viral protein with a long, natively unfolded chain whose behavior was previously studied in the absence of urea. The results clearly indicate that couplings simulated under these conditions extensively violate experimental RDCs (Supporting Information Figure S4), whereas standard database propensities reproduce the experimental data very closely.[27] We also note that $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ measured in temperature-denatured ubiquitin (data not shown) require scaling factors ($K = 1.3$ and 1.5) that are closer than those required in the case of urea-unfolded ubiquitin ($K = 1.8$ and 2.6) when using standard database propensities.

In combination these observations strongly support the hypothesis that the structural sampling of proteins in the presence of high concentrations of denaturant is shifted to more extended regions of the conformational potential wells characteristic of each amino acid.

## Conclusions

In this study we have demonstrated that the measurement of multiple RDCs per peptide unit allows detailed mapping of the conformational behavior of unfolded polypeptide chains. The popular statistical coil model, derived from structural databases, is refined to allow a more accurate representation of the conformational sampling of ubiquitin under strong denaturing conditions. A self-consistent model of the conformational landscape of proteins under these conditions is proposed that identifies enhanced access to $\beta$-extended and polyproline II regions of Ramachandran space within the amino-acid-specific conformational wells. We believe that this representation will establish a basis for describing the conformational behavior of unfolded proteins and improve our understanding of the molecular basis of protein stability and folding.

## Methods

**Sample Preparation and NMR Spectroscopy.** Scalar couplings in chemically denatured ubiquitin were obtained from 1 mM samples of $^1$H,$^{15}$N-,$^{13}$C-, and $^2$H,$^5$N-labeled ubiquitin in 10 mM glycine buffer, 8 M urea at pH 2.5 and 25 °C. RDCs were measured on samples of 1.5 mM $^1$H,$^{15}$N-,$^{13}$C- and 2 mM $^2$H,$^{15}$N-labeled ubiquitin aligned in 10% (w/v) polyacrylamide gels[46,47] in 10 mM glycine buffer, 8 M urea, pH 2.5 at 25 °C. $^1D_{NH}$ RDCs were measured with a $^{15}$N IPAP sequence[48] as a 768*($^1$H) × 300*($^{15}$N) data set (where $n$* denotes $n$ complex data points) with acquisition times of 85.2 ($^1$H) and 150.0 ($^{15}$N) ms, respectively. $^1D_{H\alpha C\alpha}$ RDCs were obtained from HN(CO)CA experiments (512*($^1$H) × 145*($^{13}$C) × 26*($^{15}$N) with acquisition times of 55.8, 27.0, and 15.6 ms, respectively), without proton decoupling during a constant time evolution period of 27 ms used to refocus $^1J_{C\alpha C\beta}$ couplings. $^1D_{C\alpha C'}$ RDCs were obtained from HNCO experiments (512*($^1H^N$) × 250*($^{13}$C) × 14*($^{15}$N) points with acquisition times of 55.8, 175, and 8.4 ms, respectively) without decoupling of aliphatic $^{13}$C spins during $^{13}$C′ chemical shift evolution. $^1H^N-^1H^\alpha$ RDCs were measured with a quantitative $J$-correlation HNHA experiment[49] (512*($^1H^N$) × 50*($^1H^\alpha$) × 65*($^{15}$N) points with acquisition times of 42.6, 17.0, and 39.2 ms, respectively), which yields not only the size but also the chemical shift and consequently the identity of the $H^\alpha$ coupling partner. The $^1H^N-^1H^\alpha$ transfer time was 26.7 ms. Long-range $^1H^N-^1H^N$ RDCs were obtained on $^2$H,$^{15}$N ubiquitin with a quantitative COSY measurement similar to the HNHA experiment,[50] using a coherence transfer time of 47 ms and recording 1024*($^1H^N$) × 95*($^1H^N$) × 80*($^{15}$N) points with acquisition times of 110.5, 104.5, and 39.2 ms, respectively. All spectra were recorded on Bruker DRX 600 and 800 MHz instruments equipped with a TXI or TCI probe, respectively. Spectra were processed with extensive zero filling in the "$J$-dimension" using nmrPipe[51] and were evaluated with PIPP.[52]

**Flexible-Meccano Calculations.** 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.[27] Alignment tensors for each conformer were calculated using PALES, an atomic resolution approach to alignment tensor prediction.[29] RDCs were calculated from throughout the molecule using the same program.

(41) Möglich, A.; Joder, K.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12394−12399.
(42) Jacob, J.; Krantz, B.; Dothager, R. S.; Thiyagarajan, P.; Sosnick, T. R. *J. Mol. Biol.* **2004**, *338*, 369−382.
(43) Whitney, P. L.; Tanford, C. *J. Biol. Chem.* **1962**, *237*, PC1735−PC1737.
(44) Schellman, J. A. *Biopolymers* **1978**, *17*, 1305−1322.
(45) Rose, G. D.; Fleming, P. J.; Banavar, J. R.; Maritan, A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16623−16633.
(46) Tycko, R.; Blanco, F. J.; Ishii, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9340−9341.
(47) Sass, H. J.; Musco, G.; Stahl, S. J.; Wingfield, P. T.; Grzesiek, S. *J. Biomol. NMR* **2000**, *18*, 303−309.
(48) Ottiger, M.; Delaglio, F.; Bax, A. *J. Magn. Reson.* **1998**, *131*, 373−378.
(49) Vuister, G. W.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 7772−7777.
(50) Meier, S.; Häussinger, D.; Jensen, P.; Rogowski, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2003**, *125*, 44−45.
(51) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(52) Garrett, D. S.; Powers, R.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson.* **1991**, *95*, 214−220.

FM calculations were also performed using specified $\phi/\psi$ distributions in addition to the standard residue-dependent distributions. Gaussian distributions with a standard deviation of 10° centered on $\beta$-sheet and PPII conformations for $\phi$ and $\psi$ were randomly introduced in a noncooperative manner at the prescribed level. For each combination an entire FM simulation was carried out. The rate of sampling of extended regions ($50° < \psi < 180°$) was also increased for each amino-acid-specific population in the database, by creating databases with 2.0, 2.5, and 3.0 times higher populations for this range. Explicitly the same conformations were repeated in the database to increase the probability of being selected. The $\beta$-hairpin conformation was introduced by sampling narrow ($\sigma = 5°$) Gaussian distributions of $\phi$ and $\psi$ centered on the native conformation for a given percentage of conformers. Structures were introduced such that the entire hairpin was either present or absent.

$^3J$ scalar couplings were calculated using known Karplus-type relationships[53] for each calculated conformer and averaged over the ensemble of values. Effective radii of gyration were calculated as described previously.[27]

**Acknowledgment.** We thank Mark Strohmeier, for help in the initial phases of this project, and Dr. Pau Bernado, for stimulating discussions. This work was supported by the Commisariat à l'Energie Atomique, the CNRS, France and the Université Joseph Fourier, Grenoble, by the French Research Ministry through ANR NT05-4_42781 (M.B.) and by the Swiss National Science Foundation Grant 31-109712 (S.G.).

**Supporting Information Available:** Table S1 showing data reproduction for individual RDC data sets using different conformational sampling conditions; Table S2 showing full statistical analysis of data fitting using extended and standard coil conditions; Figure S1 showing comparison of experimental interproton RDCs with simulated values under conditions of increasing populations of $\beta$-sheet conformations; Figure S2 showing total $\chi^2$ function calculated using common scaling of fixed geometry coupling and H−H couplings as a function of the extendedness of the conformational sampling; Figure S3 showing the effects of including 20% $\beta$-hairpin; Figure S4 showing the effects of increased extended sampling on prediction of RDCs from protein X; tables showing simulated and experimental RDCs. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0724339

(53) Pardi, A.; Billeter, M.; Wüthrich, K. *J. Mol. Biol.* **1984**, *180*, 741−751.
(54) Almond, A.; Axelsen, J. B. *J. Am. Chem. Soc.* **2002**, *124*, 9986−9987.