JACS ARTICLES

Published on Web 02/15/2010

Side-Chain χ_1 Conformations in Urea-Denatured Ubiquitin and Protein G from ^3J Coupling Constants and Residual Dipolar Couplings

Navratna Vajpai,[†] Martin Gentner,[†] Jie-rong Huang,[†] Martin Blackledge,[‡] and Stephan Grzesiek*[,†]

Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland, and Institut de Biologie Structurale Jean-Pierre Ebel, CEA, CNRS, UJF UMR 5075, 41 Rue Jules Horowitz, Grenoble 38027, France

Received December 7, 2009; E-mail: stephan.grzesiek@unibas.ch

Abstract: Current NMR information on side-chain conformations of unfolded protein states is sparse due to the poor dispersion particularly of side-chain proton resonances. We present here optimized schemes for the detection of $^3J_{H\alpha H\beta}$, $^3J_{NH\beta}$, and $^3J_{C'H\beta}$ scalar and $^1D_{C\beta H\beta}$ residual dipolar couplings (RDCs) in unfolded proteins. For urea-denatured ubiquitin and protein G, up to six ^3J-couplings to $^1H^\beta$ are detected, which define the χ_1 angle at very high precision. Interpretation of the ^3J couplings by a model of mixed staggered χ_1 rotamers yields excellent agreement and also provides stereoassignments for $^1H^\beta$ methylene protons. For all observed amino acids with the exception of leucine, the chemical shift of $^1H^{\beta 3}$ protons was found downfield from $^1H^{\beta 2}$. For most residues, the precision of individual χ_1 rotamer populations is better than 2%. The experimental χ_1 rotamer populations are in the vicinity of averages obtained from coil regions in folded protein structures. However, individual variations from these averages of up to 40% are highly significant and indicate sequence- and residue-specific interactions. Particularly strong deviations from the coil average are found for serine and threonine residues, an effect that may be explained by a weakening of side-chain to backbone hydrogen bonds in the urea-denatured state. The measured $^1D_{C\beta H\beta}$ RDCs correlate well with predicted RDCs that were calculated from a sterically aligned coil model ensemble and the ^3J-derived χ_1 rotamer populations. This agreement supports the coil model as a good first approximation of the unfolded state. Deviations between measured and predicted values at certain sequence locations indicate that the description of the local backbone conformations can be improved by incorporation of the RDC information. The ease of detection of a large number of highly precise side-chain RDCs opens the possibility for a more rigorous characterization of both side-chain and backbone conformations in unfolded proteins.

Introduction

A detailed, quantitative description of the unfolded state of proteins is crucial for understanding protein folding,[1] protein misfolding and aggregation in amyloidogenic diseases such as Alzheimer's and Parkinson's,[2] and function of intrinsically disordered proteins.[3,4] Such a description is both experimentally and theoretically highly challenging, because only a limited number of measurable parameters are available to describe the vast space of possible unstructured conformations.

During recent years, the conformations of the backbone of unfolded proteins have been described in some detail using paramagnetic relaxation enhancements PREs[5,6] and RDCs.[7,8] These two parameters report on well-defined ensemble averages of the long- and short-range backbone geometry and are thus more amenable to a rigorous quantitative interpretation than chemical shifts, NOE, or relaxation data. Both PREs and RDCs have revealed long-range contacts and residual structure in many denatured proteins showing that such states contain structural bias that may drive them toward a folded structure. The correct prediction of such structural propensities of denatured states from the amino acid sequence may be an important step toward solving the protein folding problem.

RDCs offer particular advantages for the characterization of unfolded states, because they do not require additional chemical labeling and can be detected with ease for a large number of internuclear vectors; for example, a recent study showed that up to seven RDCs per peptide unit could be determined for urea-unfolded ubiquitin.[9] For a number of unfolded proteins, trends of backbone RDCs along the polypeptide sequence could be reproduced in structural ensembles created according to the

[†] University of Basel.
[‡] Institut de Biologie Structurale Jean-Pierre Ebel.

(1) Shortle, D. *FASEB J.* **1996**, *10*, 27–34.
(2) Dobson, C. M. *Nature* **2003**, *426*, 884–890.
(3) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756–764.
(4) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31–38.
(5) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170–184.
(6) Mittag, T.; Forman-Kay, J. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.
(7) Shortle, D.; Ackerman, M. *Science* **2001**, *293*, 487–489.
(8) Meier, S.; Blackledge, M.; Grzesiek, S. *J. Chem. Phys.* **2008**, *128*, 052204.
(9) Meier, S.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 9799–9807.

10.1021/ja910331t

amino-acid-specific phi/psi angle propensities in non-alpha, non-beta conformations of PDB structures (PDB coil libraries).[8,10,11] This indicates that the so-called coil model[12,13] is a good, first approximation of the unfolded state ensemble. In turn, deviations from the coil model point to residual order within the unfolded state. Such deviations have revealed highly populated turn conformations in the natively unfolded Tau protein[14] and have shown that urea binding drives the backbone to more extended conformations for ubiquitin.[9] Additional long-range RDCs between amide protons have given evidence for a remaining, significant (10−20%) population of the first β-hairpin (residues 1−18) in (8 M) urea-denatured ubiquitin.[15]

In contrast to the backbone of unstructured polypeptides, experimental information on side chains is rather sparse. Such investigations are severely hampered by the poor dispersion of side-chain signals resulting from the conformational averaging. A small number of $^3J_{H\alpha H\beta}$ couplings have been determined in shorter unfolded peptides without ^{13}C labeling,[16,17] which correlated with predictions from χ_1 coil distributions. However, no stereoassignments of methylene $H^{\beta 2}$ and $H^{\beta 3}$ resonances were obtained. A more advanced study[18] determined heteronuclear $^3J_{NC\gamma}$ and $^3J_{C'C\gamma}$ couplings in urea-denatured lysozyme. Assuming staggered χ_1 rotamers, estimates for their populations were derived for approximately 50 amino acids, which also showed correlations to coil model predictions for most amino acids with the exception of aromatics. Precision in this analysis was limited by the lack of precise Karplus coefficients for $^3J_{NC\gamma}$ and $^3J_{C'C\gamma}$ and the fact that the three populations $p_{-60°,60°,180°}$ (two independent parameters because $p_{-60°} + p_{60°} + p_{180°} = 1$) were determined from only two experimental values.

In the present study, we have improved the description of χ_1 conformations in unfolded proteins by optimized heteronuclear experiments involving β-protons, which are able to resolve most methylene $H^{\beta 2}$ and $H^{\beta 3}$ pairs. Stereoassignments and χ_1 angle information could be obtained for the predominant part of residues in urea-denatured ubiquitin and protein G from an extensive set of up to six three-bond scalar couplings ($^3J_{NH\beta 2,3}$, $^3J_{C'H\beta 2,3}$, and $^3J_{H\alpha H\beta 2,3}$). A combined analysis of all ^3J couplings according to the staggered conformer model yields individual populations with a maximal error of 2%. This analysis is corroborated by independent $^1D_{C\beta H\beta 2,3}$ RDC data detected in strained polyacrylamide gels.[19,20] These side-chain RDCs agree well with theoretical RDCs calculated from the ^3J-derived χ_1 conformer distribution and a coil model ensemble of backbone conformations generated by the program *Flexible*-Meccano.[11] The obtained χ_1 conformer populations cluster around coil model averages, but individual variations in particular for serines and threonines of up to 40% are significant and indicate sequence- and residue-specific preferences.

Materials and Methods

Sample Preparation and NMR Spectroscopy. $^{15}N/^{13}C$-labeled human ubiquitin and protein G (GB1 sequence ^{1}MQYKLILNGK ^{11}TLKGETTTEA ^{21}VDAATAEKVF ^{31}KQYANDNGVD ^{41}GEW-TYDDATK ^{51}TFTVTE) were prepared according to standard protocols.[21] Ubiquitin NMR samples contained 1.0 (0.6) mM $^{15}N/^{13}C$-labeled protein in 10 mM glycine, 8 M urea, pH 2.5, 95/5% H_2O/D_2O for measurement under isotropic (anisotropic) conditions. Protein G samples contained 0.6 mM $^{15}N/^{13}C$-labeled protein in 10 mM glycine, 7.4 M urea, pH 2.0, 95/5% H_2O/D_2O. Residual alignment of urea-denatured proteins was achieved by introducing the protein solutions into 7% (w/v) polyacrylamide gels and horizontal compression (aspect ratio 2.9:1) in NEW-ERA sample tubes[22] yielding maximal $|^1D_{NH}|$ RDCs of about 13 Hz for both proteins.

All NMR experiments were carried out at 298 K on a Bruker Avance DRX 800 spectrometer equipped with a TCI cryoprobe. Spectra were processed with NMRPipe[23] and evaluated with NMRView[24] and PIPP.[25]

Assignments. Assignments of urea-denatured ubiquitin (BMRB entry 4375)[26] and protein G[27] were transferred to our sample preparations and extended by a combination of CBCA(CO)NH,[28] HNCO,[29] HBHA(CO)NH,[28] and HNHB[30] experiments. To obtain higher resolution, the constant time ^{15}N acquisition period was increased in these experiments to about 40 ms. Note that for CBCA(CO)NH, HNCO, and HBHA(CO)NH, this still achieves a transfer of about 95% via the $^1J_{NC'}$ (∼15 Hz) coupling. Almost complete assignments of all $^1H^N$, ^{15}N, $^{13}C'$, $^{13}C^\alpha$, $^{13}C^\beta$, $^1H^\alpha$, and $^1H^\beta$ resonances were obtained from this procedure. Missing assignments mainly comprise amino acids preceding proline or were due to signal degeneracy of some geminal protons. The obtained chemical shifts are close to the published data with the exception of residues in the vicinity of the mutated T2Q site in protein G. They also extend the previous data by the stereoassignments of β-methylene protons and the $^{13}C'$ chemical shifts (protein G). The assignments are deposited in the BMRB data bank under accession numbers 16626 (ubiquitin) and 16627 (protein G).

Determination of Scalar and Residual Dipolar Coupling Constants. ^3J scalar couplings carrying information on the χ_1 angle of the denatured proteins were obtained from modified versions of quantitative $^3J_{NH\beta}$-HNHB,[30] $^3J_{C'H\beta}$-HN(CO)HB,[31] and $^3J_{H\alpha H\beta}$-HAHB-(CACO)NH[32] experiments. $^1D_{C\beta H\beta}$ RDCs for χ_1 angle information were determined from an HBHA(CO)NH[28] experiment, where an IPAP detection scheme[33] was introduced into the mixed constant

(10) Jha, A.; Colubri, A.; Freed, K.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099–13104.
(11) Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002–17007.
(12) Serrano, L. *J. Mol. Biol.* **1995**, *254*, 322–333.
(13) Smith, L.; Bolin, K.; Schwalbe, H.; MacArthur, M.; Thornton, J.; Dobson, C. *J. Mol. Biol.* **1996**, *255*, 494–506.
(14) Mukrasch, M.; Markwick, P.; Biernat, J.; Bergen, M.; Bernado, P.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 5235–5243.
(15) Meier, S.; Strohmeier, M.; Blackledge, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2007**, *129*, 754–755.
(16) West, N. J.; Smith, L. J. *J. Mol. Biol.* **1998**, *280*, 867–877.
(17) Mathieson, S. I.; Penkett, C. J.; Smith, L. J. *Pacific Symp. Biocomputing* **1999**, 542–553.
(18) Hennig, M.; Bermel, W.; Spencer, A.; Dobson, C. M.; Smith, L. J.; Schwalbe, H. *J. Mol. Biol.* **1999**, *288*, 705–723.
(19) Sass, H.; Musco, G.; Stahl, S.; Wingfield, P.; Grzesiek, S. *J. Biomol. NMR* **2000**, *18*, 303–309.
(20) Tycko, R.; Blanco, F.; Ishii, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9340–9341.
(21) Sass, J.; Cordier, F.; Hoffmann, A.; Cousin, A.; Omichinski, J.; Lowen, H.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 2047–2055.
(22) Chou, J.; Gaemers, S.; Howder, B.; Louis, J.; Bax, A. *J. Biomol. NMR* **2001**, *21*, 377–382.
(23) Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.
(24) Johnson, B.; Blevins, R. *J. Biomol. NMR* **1994**, *4*, 603–614.
(25) Garrett, D.; Powers, R.; Gronenborn, A.; Clore, G. *J. Magn. Reson.* **1991**, *95*, 214–220.
(26) Peti, W.; Smith, L.; Redfield, C.; Schwalbe, H. *J. Biomol. NMR* **2001**, *19*, 153–165.
(27) Frank, M. K.; Clore, G. M.; Gronenborn, A. M. *Protein Sci.* **1995**, *4*, 2605–2615.
(28) Grzesiek, S.; Bax, A. *J. Biomol. NMR* **1993**, *3*, 185–204.
(29) Grzesiek, S.; Bax, A. *J. Magn. Reson.* **1992**, *96*, 432–440.
(30) Archer, S.; Ikura, M.; Torchia, D.; Bax, A. *J. Magn. Reson.* **1991**, *95*, 636–641.
(31) Grzesiek, S.; Ikura, M.; Clore, G.; Gronenborn, A.; Bax, A. *J. Magn. Reson.* **1992**, *96*, 215–221.
(32) Lohr, F.; Schmidt, J.; Ruterjans, H. *J. Am. Chem. Soc.* **1999**, *121*, 11821–11826.
(33) Ottiger, M.; Delaglio, F.; Bax, A. *J. Magn. Reson.* **1998**, *131*, 373–378.

time $^1H^{\alpha/\beta}$ evolution/$^1H^{\alpha/\beta} \rightarrow {}^{13}C^{\alpha/\beta}$ transfer period. $^1D_{C\beta H\beta}$ RDCs were calculated as the difference in couplings observed under anisotropic conditions in strained polyacrylamide gels and a second experiment under isotropic conditions. Details of these experiments are given in Supporting Information Figures S1−S4. Similar to the assignment experiments, acquisition times in the indirect ^{15}N and $^1H^{\alpha/\beta}$ dimensions were set to 40 and 30 ms, respectively, to obtain sufficient resolution. Each experiment for scalar and dipolar couplings was carried out twice, and the reported coupling constants and error estimates refer to mean and standard deviations from such repeated experiments. The *J*-coupling constants were not corrected for effects of scalar relaxation of the second kind,[34] because these are expected to be small due to the fast effective correlation times in unfolded proteins.

Analysis of ^3J Coupling Constants and RDCs. Analysis of the ^3J data according to a model of staggered χ_1 rotamers (see main text) was carried out using in-house written Matlab (MathWorks, Inc.) routines and its QUADPROG function for constrained linear minimization. For the analysis of RDC data, theoretical RDCs from steric alignment were calculated for the three staggered rotamers by an in-house written C program[35] from an ensemble of 50 000 unfolded protein structures generated by the *Flexible*-Meccano program.[11] Experimental RDCs were then compared and fitted to these predicted values by the same staggered rotamer model (see main text).

Coil Library Rotamers. The experimentally derived χ_1 rotamer populations were also compared to average populations from a protein coil library. This library was downloaded as version 20080310_pc20_res1.6_R0.25 generated on 3/10/2008 from the Rose lab server[36] and contained 16 856 protein fragments from nonhomologous proteins of X-ray structures with resolution better than 1.6 Å in non-alpha, non-beta conformations.

Results and Discussion

χ_1 Torsion Angle Information from Scalar Couplings and RDCs. The NMR analysis of unfolded proteins in solution is made difficult by the low spectral dispersion resulting from conformational averaging. In particular, this applies to side-chain resonances. Thus, so far, the potential of side-chain protons has not fully been used as a source of conformational information. Here, we have probed the χ_1 torsion angles in urea-denatured ubiquitin and protein G by ^3J scalar and residual dipolar couplings involving H^β protons. Scalar couplings were detected by modified versions of the quantitative-$^3J_{NH\beta}$ HNHB,[30] $^3J_{C'H\beta}$ HN(CO)HB,[31] and $^3J_{H\alpha H\beta}$ HAHB(CACO)NH[32] experiments (Supporting Information). These experiments overcome the problems of low dispersion in both backbone and side-chain resonances by making use of the long transverse relaxation times in the unfolded state to achieve maximal frequency resolution. Acquisition times of 30−35 ms for $^1H^\alpha$ and $^1H^\beta$ resonances and of 40 ms for ^{15}N in the indirect dimensions proved to be sufficient to resolve most of the overlap in the side-chain experiments. Additional information from similarly optimized 3D CBCA(CO)NH, HBHA(CO)NH experiments was used to establish sequential assignments. Figure 1 shows the good resolution of the side-chain $^1H^\beta$ resonances in the quantitative-*J* spectra for residues L7 and K10 of urea-denatured protein G. The intensity ratios of the resonances in these spectra yield information on the χ_1 torsion angles and the stereospecific assignments for β-methylene protons. The high population of $-60°$ χ_1 conformations is obvious from the higher intensities

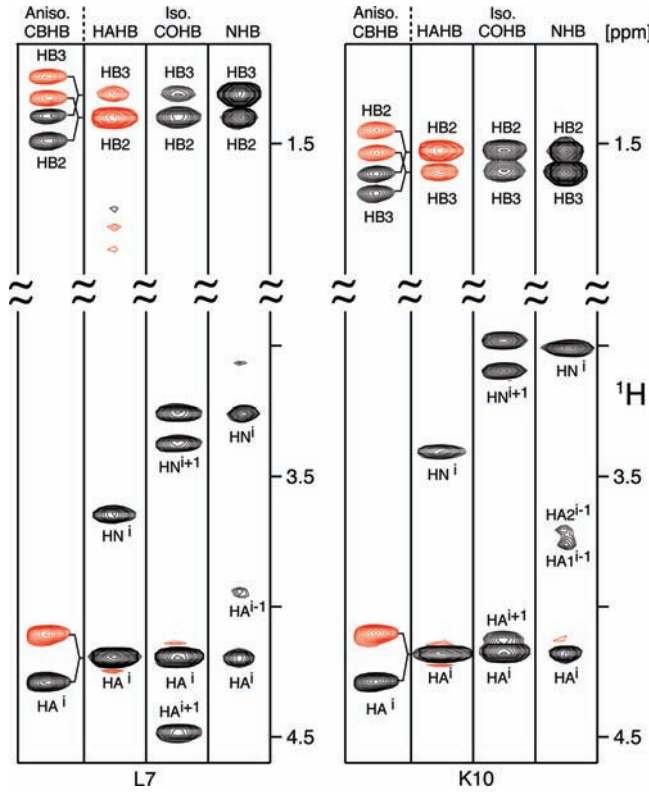

Figure 1. Strip plots of 3D spectra used for determination of $^3J_{H\alpha H\beta}$, $^3J_{NH\beta}$, $^3J_{C'H\beta}$ coupling constants and $^1D_{C\beta H\beta}$ RDCs in urea-denatured protein G. Data are shown for two typical residues L7 (left) and K10 (right) from the quantitative $^3J_{H\alpha H\beta}$-HAHB(CACO)NH (HAHB), $^3J_{C'H\beta}$-HN(CO)HB (COHB), and $^3J_{NH\beta}$-HNHB (NHB) experiments recorded under isotropic conditions as well as from the $^1D_{C\beta H\beta}$-IPAP-HBHA(CO)NH (CBHB) experiment recorded under anisotropic conditions. Resonances are labeled with assignment information. Amide proton signals are folded in the indirect proton dimension to reduce experimental time. These signals are split by the $^1J_{NH}$ coupling for the HN(CO)HB experiment. For the quantitative-*J* HAHB, COHB, and NHB experiments, resonances shown in red are negative signals. For the IPAP-HBHA(CO)NH experiment, red is used to distinguish the $^1H^{\alpha/\beta}$ upfield from the downfield (black) components.

of the $^1H^{\beta2}$ resonances relative to the $^1H^{\beta3}$ resonances in the quantitative-$^3J_{H\alpha H\beta}$ HAHB(CACO)NH experiment, which corresponds to $^3J_{H\alpha H\beta2} > {}^3J_{H\alpha H\beta3}$, from the inverse situation in the quantitative-$^3J_{NH\beta}$ HNHB experiment, and from more equal intensities in the quantitative-$^3J_{C'H\beta}$ HN(CO)HB experiment.

In addition to ^3J couplings, also RDCs yield information on the χ_1 conformations. The detection of $^1D_{C\beta H\beta}$ RDCs induced in strained polyacrylamide gels proved particularly easy from an IPAP-HBHA(CO)NH experiment (Supporting Information). Figure 1 shows as an example the data of this experiment on the oriented urea-denatured protein G. The quality of the spectra is excellent and allowed the unambiguous determination of 91 (61) $^1D_{C\beta H\beta}$ RDCs in unfolded ubiquitin (protein G).

In total for ubiquitin (protein G), 353 (246) $^3J_{NH\beta}$, $^3J_{C'H\beta}$, $^3J_{H\alpha H\beta}$, and $^1D_{C\beta H\beta}$ couplings (Supporting Information Table S1) could be derived from the quantitative analysis of the spectra. These data cover 82% of all side chains with variable χ_1 angle, that is, 55 out of 68 (ubiquitin) and 40 out of 46 (protein G) non-(Gly, Ala) residues.

Analysis of ^3J Coupling Constants by Staggered χ_1 Rotamer Populations. Our analysis of the ^3J couplings in terms of side-chain conformations assumes as a first approximation that the conformations are a population mixture of three staggered χ_1 ($-60°$, $+60°$, $180°$) rotamers. Previously deter-

(34) Bax, A.; Vuister, G. W.; Grzesiek, S.; Delaglio, F.; Wang, A. C.; Tschudin, R.; Zhu, G. *Methods Enzymol.* **1994**, *239*, 79–105.
(35) Huang, J.-r.; Grzesiek, S. *J. Am. Chem. Soc.* **2010**, *132*, 694−705.
(36) Fitzkee, N.; Fleming, P.; Rose, G. *Proteins* **2005**, *58*, 852–854.

mined amino-acid-specific Karplus coefficients[37] were used to predict theoretical $^3J_{H\alpha H\beta i}$, $^3J_{NH\beta i}$, and $^3J_{C'H\beta i}$ coupling constants for these rotamers according to the Karplus relation:

$$^3J_{ij}^{calc}(\chi_1) = C_{ij}^0 + C_{ij}^1 \cos(\vartheta_{ij}(\chi_1)) + C_{ij}^2 \cos(2\vartheta_{ij}(\chi_1)) \quad (1)$$

where $\vartheta_{ij}(\chi_1)$ is the intervening dihedral angle between the nuclei i and j in a $^3J_{ij}$ coupling for a specific side-chain torsion angle χ_1. For all side chains, only single sets of resonances were observed. Thus, the side chains are in fast exchange on the time scale of the chemical shift, that is, faster than milliseconds. Accordingly, the observed coupling constants should be population averages over the individual rotamers:[38−40]

$$\langle ^3J_{ij}^{calc} \rangle = \sum_{\chi_1=-60°,60°,180°} p_{\chi_1} {}^3J_{ij}^{calc}(\chi_1) \quad (2)$$

where $p_{-60°,60°,180°}$ are the individual populations.

To derive these populations from the experimental $^3J^{exp}$ couplings, their deviation from the calculated average $\langle ^3J^{calc} \rangle$ was minimized with respect to $p_{-60°}$, $p_{60°}$, and $p_{180°}$ by a constrained linear least-squares fit of the target function

$$\chi^2 = \frac{1}{N} \sum_{ij} \left(\frac{^3J_{ij}^{exp} - \langle ^3J_{ij}^{calc} \rangle}{\sigma_{ij}} \right)^2 \quad (3)$$

under the conditions $p_{-60°}+p_{60°}+p_{180°} = 1$ and $0 \le p_{-60°,60°,180°} \le 1$. In eq 3, σ presents the statistical, experimental error of the coupling constant obtained from a repetition of the experiment, the summation runs over all individual nuclei i and j, for which a ^3J coupling could be determined for an individual side chain, and N indicates the total number of measured ^3J values.

The stereospecific assignments of geminal $H^{\beta 2}$ and $H^{\beta 3}$ protons in urea-denatured ubiquitin and protein G were not known prior to the current analysis. This information was also derived from the ^3J couplings by carrying out the fit procedure for both possible stereo assignments and using the assignment that corresponded to the lower χ^2 value. Typically these χ^2 values were about 10−20 times smaller than the values for the swapped assignment, such that discrimination was achieved easily. The stereospecific assignments obtained by this method comprise 82% of all β-methylene protons, corresponding to 41 and 25 residues in ubiquitin and protein G, respectively.

Figure 2 shows the experimental $^3J_{NH\beta}$, $^3J_{C'H\beta}$, and $^3J_{H\alpha H\beta}$ coupling constants and their values according to the fit of eq 3 for a number of residues in ubquitin and protein G (the complete data for both proteins are shown in Supporting Information Figures S5 and S6). For most residues, the agreement between experimental and predicted data is excellent with average RMSDs between measured and predicted ^3J-values of less than 0.3 Hz. This indicates not only a high precision of both experimental ^3J data and Karplus coefficients, but also validates the staggered rotamer model as a reasonable approximation for the side-chain conformations. In total, 53 (39) χ_1 rotamer populations could be derived for those amino acids in ubiquitin (protein G) for which at least two ^3J coupling constants had been measured (Supporting Information Table S2). The constrained linear fit also provides error estimates for the $p_{-60°}$, $p_{60°}$, and $p_{180°}$ populations derived by propagation from the statistical experimental error. These errors range between 0.01 and 0.02, indicating a very high precision of the population estimates.

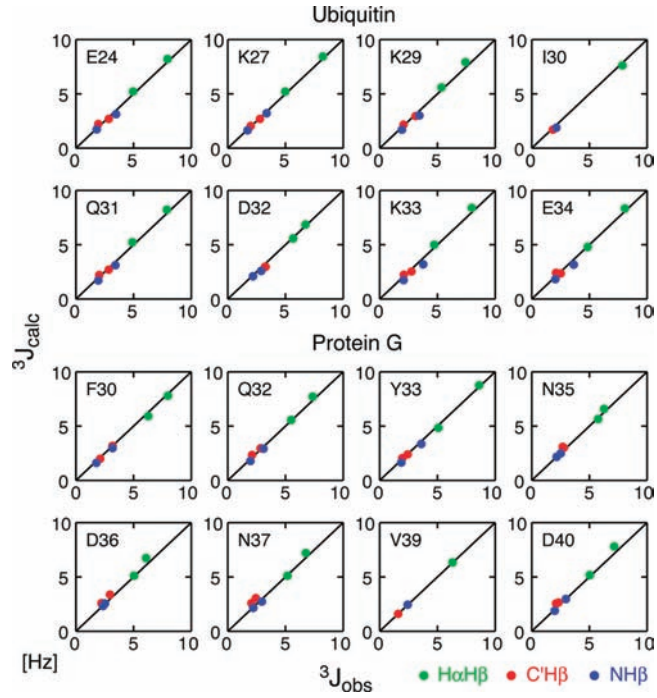

Figure 2. Comparison of experimental (obs) $^3J_{XH\beta}$ (X = C′, N, HA) constants to values derived (calc) according to the χ_1 rotamer population fit of eq 3 for selected residues of urea-denatured ubiquitin (top) and protein G (bottom).

Analysis of RDC Data. To make use of the experimental $^1D_{C\beta H\beta}$ RDCs for the analysis of χ_1 conformations, theoretical RDC estimates for all staggered rotamers were obtained from large simulated ensembles of unfolded ubiquitin or protein G structures. For both proteins, ensembles of 50 000 unfolded structures were generated by the *Flexible*-Meccano program[11] according to the amino-acid-specific phi/psi angle propensities in non-alpha, non-beta conformations of PDB structures (PDB coil library) and omitting structures with sterical clashes. Such ensembles have previously been shown to reproduce the trends of backbone RDCs along the polypeptide sequence.[8,10,11] Because the *Flexible*-Meccano algorithm represents side chains only by a pseudoatom at the C^β position, full coordinates for the C^β and H^β atoms for all staggered rotamers were generated from the N, C^α, and C′ positions using idealized tetrahedral geometry. The alignment tensor for each member k of the ensemble was then calculated on the basis of the assumption of steric exclusion[41,42] using an efficient in-house written algorithm.[35] In brief, this algorithm calculates the maximal extension of the molecule for each direction of the unit sphere. The probability for finding the molecule in a certain orientation is then derived as the volume that can be occupied by the molecule between two infinitely extended, parallel planes relative to the total volume between the planes. The alignment tensor then corresponds to the average over all orientations of

(37) Pérez, C.; Löhr, F.; Rüterjans, H.; Schmidt, J. M. *J. Am. Chem. Soc.* **2001**, *123*, 7081–7093.
(38) Pachler, A. *Spectrochim. Acta* **1963**, *19*, 2085–2092.
(39) Pachler, A. *Spectrochim. Acta* **1964**, *20*, 581–587.
(40) Dzakula, Z.; Westler, W.; Edison, A.; Markley, J. L. *J. Am. Chem. Soc.* **1992**, *114*, 6195–6199.
(41) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791–3792.
(42) van Lune, F.; Manning, L.; Dijkstra, K.; Berendsen, H. J.; Scheek, R. M. *J. Biomol. NMR* **2002**, *23*, 169–179.

second rank spherical harmonics weighted by this probability. Theoretical RDC values were derived for each χ_1 rotamer and in each individual structure as

$$D_{k,ij}^{\mathrm{calc}}(\chi_1) = -\frac{\gamma_i\gamma_j\hbar\mu_0}{4\pi^2}\sqrt{\frac{4\pi}{5}}\sum_{m=-2}^{2} S_{k,m}^{*}\frac{Y_{2m}(\Theta_{k,ij}(\chi_1), \Phi_{k,ij}(\chi_1))}{r_{k,ij}^{3}(\chi_1)} \quad (4)$$

where $D_{k,ij}^{\mathrm{calc}}$ represents the RDC between nuclei i and j for ensemble member k with individual alignment tensor $S_{k,m}$ (written in irreducible form[43]), Y_{2m} are spherical harmonics, $r_{k,ij}$, $\Theta_{k,ij}$, $\Phi_{k,ij}$ are the polar coordinates of the internuclear vector, and $\gamma_{i,j}$ are the nuclear gyromagnetic ratios. These RDC values for the individual structures were then averaged over all N members of the ensemble to obtain an estimate D_{ij}^{calc} for the RDC in the unfolded protein:

$$D_{ij}^{\mathrm{calc}}(\chi_1) = \frac{1}{N}\sum_{k=1}^{N} D_{k,ij}^{\mathrm{calc}}(\chi_1) \quad (5)$$

Because the absolute size of the alignment tensor $S_{k,m}$ is difficult to predict from the experimental conditions, an additional common overall scaling was used such that the mean square difference between measured and predicted average backbone $^1D_{\mathrm{HN}}$ was minimized. Analogously to eq 2, the population average over the individual χ_1 rotamers was then calculated as

$$\langle D_{ij}^{\mathrm{calc}}\rangle = \sum_{\chi_1=-60^\circ,60^\circ,180^\circ} p_{\chi_1} D_{ij}^{\mathrm{calc}}(\chi_1) \quad (6)$$

Figure 3 shows the experimental $^1D_{C\beta H\beta}$ RDCs and their values predicted by eq 6 from the ^3J-derived χ_1 rotamer populations (red ●) for the same residues in ubiquitin and protein G as in Figure 2. For comparison, measured and predicted ^3J-couplings (blue) are also shown. The complete data for both proteins are given in Supporting Information Figures S7 and S8. For many residues, such as E24, K29, I30, Q31, and E34 in ubiquitin and F30, Q32, N35, N37, V39, and D40 in protein G, measured and predicted RDCs agree within about 5 Hz. This agreement is very reasonable when compared to the full variation of about 20−30 Hz of observed RDCs. For other residues like D32 in ubiquitin or D36 in protein G, the deviations from the predictions are clearly larger, but still agree with the trends of the predictions. The correlations between all measured and predicted $^1D_{C\beta H\beta}$ RDCs have a Pearson's correlation coefficient of 0.86 and 0.70 for ubiquitin and protein G, respectively (Figure 4, ●). The correlation can be improved to some extent, when the $^1D_{C\beta H\beta}$ RDCs are also included into the fit of the χ_1 rotamer populations by extending the χ^2 function of eq 3 to the differences between measured and predicted D_{ij} couplings (Figures 3 and 4, ○). In this case, the χ_1 rotamer populations only change by a few percent (not shown), but the correlation coefficient increases to 0.92 (0.76) for ubiquitin (protein G).

Considering the crudeness of the assumptions for the coil model and the steric alignment of the unfolded model ensemble, the agreement between measured and predicted $^1D_{C\beta H\beta}$ RDCs is surprisingly good. This provides an independent confirmation for both the backbone coil model as implemented by the *Flexible*-Meccano algorithm[11] as well as the staggered χ_1

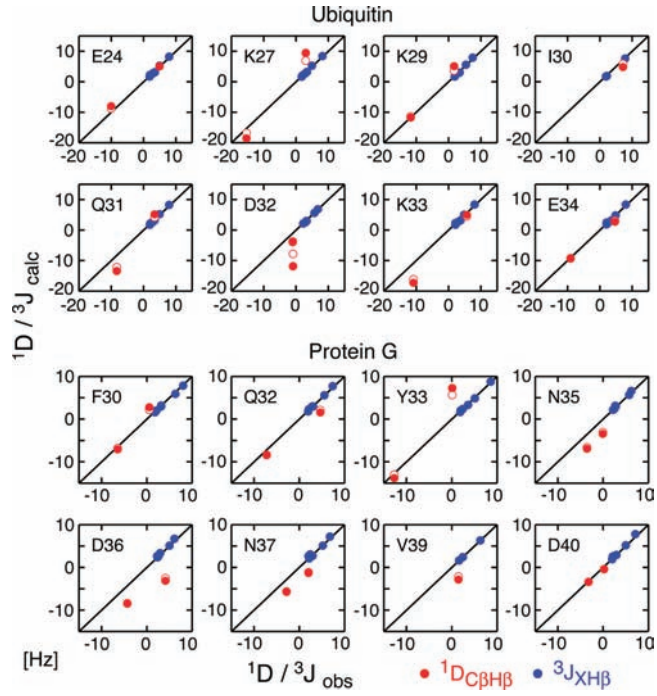

Figure 3. Comparison of experimental (obs) and predicted (calc) $^1D_{C\beta H\beta}$ RDCs (red ●) based on the χ_1 rotamer populations derived from the fit of eq 3 of $^3J_{XH\beta}$ constants and the coil model ensemble. For comparison, experimental and predicted $^3J_{XH\beta}$ couplings are also indicated (blue ●). The same residues are shown as in Figure 2 for urea-denatured ubiquitin (top) and protein G (bottom). Red ○ indicate predictions for $^1D_{C\beta H\beta}$ RDCs, when these RDC data were also included in the rotamer population fit of eq 3 (see text).

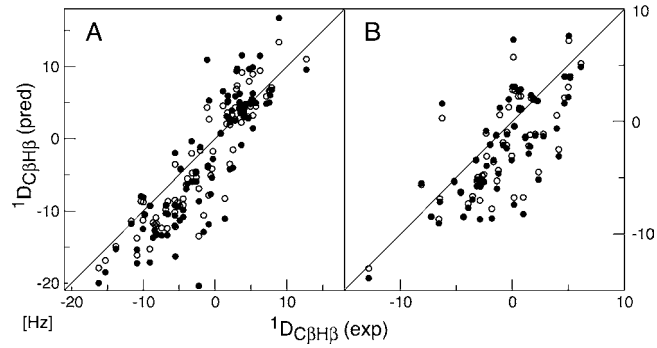

Figure 4. Correlation of all observed experimental (obs) and predicted (calc) $^1D_{C\beta H\beta}$ RDCs (●) based on the χ_1 rotamer populations derived from the fit of eq 3 of $^3J_{XH\beta}$ constants and the coil model ensemble for urea-denatured ubiquitin (A) and protein G (B). The Pearson's correlation coefficient is 0.86 and 0.70 for ubiquitin and protein G, respectively. The "○" indicate predictions for $^1D_{C\beta H\beta}$ RDCs, when these RDC data were also included in the rotamer population fit of eq 3. In this case, the correlation coefficient increases to 0.92 and 0.76 for ubiquitin and protein G, respectively.

rotamer model. However, the deviations, which exceed the experimental errors, also clearly indicate shortcomings of this interpretation. Because the agreement of the staggered rotamer predictions is almost perfect for the ^3J-couplings, it is very likely that the RDC deviations result from inaccuracies of the local backbone geometry predicted by the coil ensemble and from the unknown microscopic details of the alignment interaction, which may not be adequately covered by the simple steric alignment model. In principle, the inaccuracies of the backbone geometry may be reduced by refining the backbone conformations using information from additional backbone RDCs. Such

(43) Moltke, S.; Grzesiek, S. *J. Biomol. NMR* **1999**, *15*, 77–82.

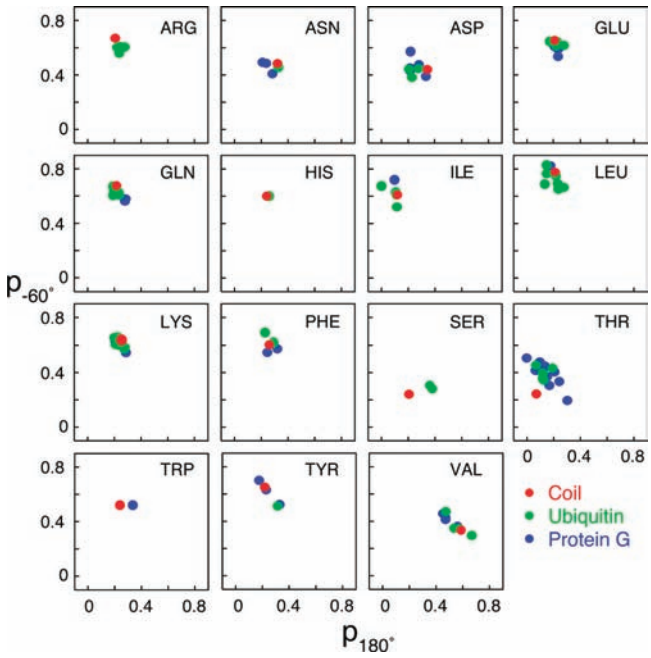

Figure 5. $p_{-60°}$ and $p_{180°}$ χ_1 rotamer populations obtained from the combined fit of $^3J_{NH\beta}$, $^3J_{C'H\beta}$, and $^3J_{H\alpha H\beta}$ couplings of eq 3. Populations are shown separately for ubiquitin (green), protein G (blue), and the average of the PDB coil structures (red).

an approach is currently pursued by constrained molecular dynamics ensemble calculations, which include all of the RDCs as restraints.[35] The ease of detection of a large number of highly precise RDCs on side-chain nuclei may thus make it possible to increase the accuracy of the description of side-chain and backbone conformations in unfolded ensembles beyond the simple coil model.

Comparison to Coil χ_1 Populations. Previous analyses of the χ_1 rotamer conformations in unfolded proteins using more limited ^3J data[16−18] have concluded that their populations correlate to the average rotamer populations in the PDB coil conformations. Figure 5 shows the $p_{-60°}$ and $p_{180°}$ χ_1 rotamer populations according to the combined fit of $^3J_{NH\beta}$, $^3J_{C'H\beta}$, and $^3J_{H\alpha H\beta}$ couplings (Figure 2) together with the respective coil averages for all amino acids in ubiquitin and protein G, for which populations could be derived. The populations are indeed in the vicinity of the coil values. However, in some cases, they deviate by more than 30% from the coil values (Figure 6) and also vary around their mean by approximately 10%. These variations are significant considering that the errors of the individual populations are only about 1−2% as estimated from error propagation of the linear least-squares fit. Thus, they must reflect sequence-specific preferences of the side chains along the unfolded polypeptide chain (see below).

As remarked earlier,[18] many amino acids prefer the −60° χ_1 rotamer both in the coil model predictions and in ^3J-derived populations due to the repulsion of substituents at the γ^1 position from the main chain. In our analysis (Figure 5), this is the case for all unbranched amino acids (R, N, D, E, Q, H, L, K, F, W, Y) as well as for the branched isoleucine. These residues have J-derived and coil average $p_{-60°}$ values of 40−80%, where the shortest side-chain asparagine and aspartic acid residues have the lowest and leucine the highest −60° preference. Valines have similar populations for the −60° (30−50%) and +180°

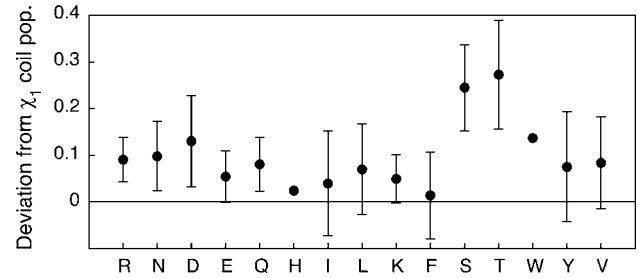

Figure 6. Deviations of the experimentally (^3J-only) derived χ_1 rotamer populations in urea-denatured proteins from the average of the PDB coil structures. The deviations are calculated as $|\vec{p}_{exp} - \vec{p}_{coil}|$ with $\vec{p} = (p_{-60°,60°,180°})$. Average and standard deviations of $|\vec{p}_{exp} - \vec{p}_{coil}|$ are shown for all amino acids for which rotamer populations could be obtained in urea-denatured ubiquitin and protein G.

(40−70%) χ_1 rotamers, which on average places both γ-methyl groups at the furthest distance from the backbone carbonyl.

Exceptions to this behavior are found for serines and threonines. For these residues, the coil model predicts $p_{-60°}$ populations of only 24% (S, T) and much larger $p_{60°}$ populations of 55% (S) and 68% (T). It has been speculated[18] that this is caused by favorable polar interactions between the side-chain hydroxyl and the main-chain amide group in the +60° χ_1 rotamer. Among all amino acid types, the experimentally derived rotamer populations for serines and threonines in urea-denatured ubiquitin and protein G show the strongest deviations from these coil predictions of folded protein structures. Thus, the J-derived $p_{60°}$ populations amount to only about 35% for serine and 40−50% for threonine, whereas $p_{-60°}$ and $p_{+180°}$ are correspondingly higher (Figure 5). The deviations are found for all serine and threonine residues, and no particular correlation to specific locations in the sequence is evident (Supporting Information Figure S9).

It is unlikely that this behavior is an experimental artifact connected to the particular $^{13}C^\beta$ chemical shift of these two amino acids, because the $^{13}C^\beta$ nuclei are not involved in the magnetization pathways of the quantitative $^3J_{NH\beta}$-HNHB, $^3J_{C'H\beta}$-HN(CO)HB, and $^3J_{H\alpha H\beta}$-HAHB(CACO)NH experiments. To further test for systematic errors from the $^3J_{H\alpha H\beta}$-HAHB(CACO)NH experiment, we have also fitted the serine and threonine χ_1 populations by using only the $^3J_{NH\beta}$- and $^3J_{C'H\beta}$-couplings (not shown). The deviations of the resulting populations from the all-^3J-value results are in most cases smaller than 4%, that is, much smaller than the deviations from the coil values. It should be noted that the fit uses specific Karplus coefficients for serines and threonines.[37] Thus, particular effects of the side-chain oxygen on the size of the J-couplings are corrected. Deviations from the χ_1 coil populations for serines and threonines are also evident from the experimental RDCs, since their agreement with predictions from J-derived χ_1 populations is considerably better than with predictions from the coil populations (not shown).

We attribute this genuine difference of serine and threonine χ_1 populations from the coil average to a destabilization of the side-chain hydroxyl/amide interaction in the urea-denatured state, where urea or water could form hydrogen bonds to both groups. Interestingly, also deviations in backbone conformations have been found for these residues in a recent study using RDCs.[44] Such deviations had not been detected in the earlier studies of χ_1 conformations in unfolded proteins,[16−18] possibly due to the more limited precision. In contrast, stronger deviations were observed for aromatic residues in urea-denatured

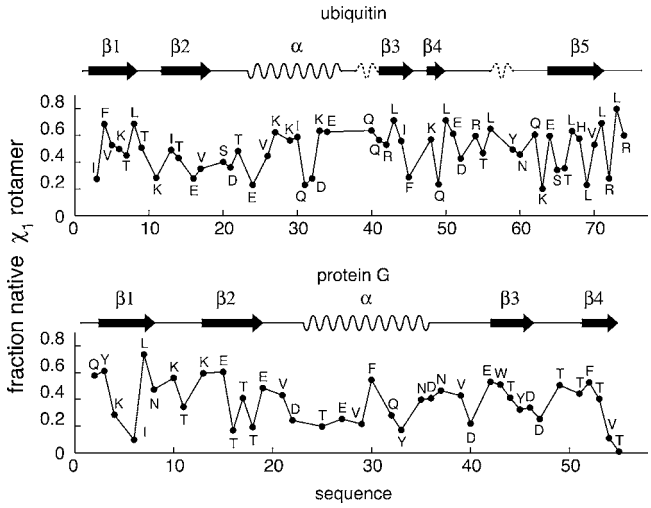

Figure 7. Fraction of native-state χ_1 angles contained in ^3J-derived rotamer populations of urea-denatured ubiquitin (top) and protein G (bottom). Native-state χ_1 angles are approximated as staggered rotamers and taken from the first entry of the folded native NMR structure (PDB code 1d3z)[46] and the 1.1 Å X-ray structure of protein G (PDB code 1igd).[47] The native-state secondary structure of both proteins is shown at the top of the respective panels.

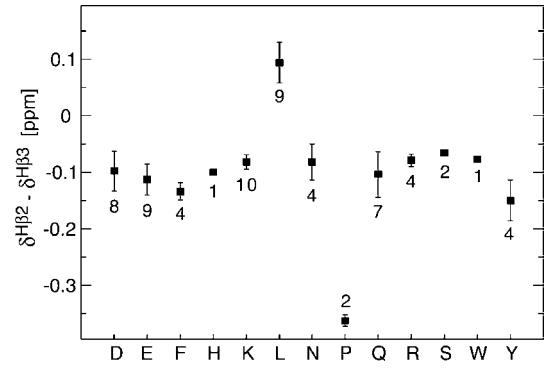

Figure 8. Differences of β-methylene ^{1}H chemical shifts in urea-denatured proteins. Average chemical shift differences of stereospecifically assigned $H^{\beta 2}$ and $H^{\beta 3}$ resonances are shown for all amino acids that could be observed in ubiquitin and protein G. Error bars indicate standard deviations. Data are labeled by the number of observations.

lysozyme.[18] This effect is not clearly visible for the 10 aromatic residues in urea-denatured ubiquitin and protein G (Figure 5), where the deviations from the coil predictions and their individual variations are not stronger than for many other amino acids such as, for example, valines and leucines.

It is expected that the observed variations in the χ_1 populations of the urea-denatured state correspond to sequence-specific structural preferences, such as the recently detected 10−20% population of the native-state, first β-hairpin (residues M1−V17) in ubiquitin.[15] However, the exact extent of the structural preferences is currently difficult to establish, because proper random coil "baseline" χ_1 populations are not known with sufficient precision in solution to be able to interpret population differences on the order of 10% with confidence. Nevertheless, it is clear that a high degree of native-state χ_1 conformations is contained in the observed χ_1 populations of the urea-denatured states. Figure 7 shows this fraction of native-state side-chain conformations in the χ_1 populations when native-state χ_1 angles are approximated by staggered rotamers. High native rotamer populations (>50%) are found for a number of residues in ubiquitin's β-strands β1, β3, β4, and β5, and in its α-helix, as well as in all β-strands of protein G. Average native rotamer populations are 49% for ubiquitin and 38% for protein G. This indicates that the transition to a fully formed structure does not require a particularly high entropic cost.

Stereospecific Methylene H^β Chemical Shifts. The availability of the stereospecific assignments for the $^1H^{\beta 2/3}$ resonances allows one to analyze their chemical shift behavior. Average chemical shift differences, $\delta H^{\beta 2} - \delta H^{\beta 3}$, are shown in Figure 8 for all amino acids with distinct β-methylene resonances that could be observed in both proteins. For all observed amino acids (D, E, F, H, K, L, N, P, Q, R, S, W, Y) with the exception of leucine and proline, $\delta H^{\beta 2}$ is smaller than $\delta H^{\beta 3}$ by about 0.1 ppm. Thus, $H^{\beta 2}$ is usually upfield from $H^{\beta 3}$ (see also Figure 1, K10). Variations of individual amino acids around the mean shift difference are about 0.05 ppm. For prolines, the upfield shift of $H^{\beta 2}$ is stronger (0.36 ppm). In contrast, for leucine, $H^{\beta 2}$ is found downfield of $H^{\beta 3}$ with a mean chemical shift difference of 0.09 ppm. This behavior is evident in Figure 1, where the intensity patterns of the upfield and downfield H^β protons are inverted for L10 relative to K10 in the two quantitative $^3J_{NH\beta}$ and $^3J_{H\alpha H\beta}$ experiments. We speculate that this unusual behavior is related to the extreme bulkiness of the two leucine δ-methyl groups, which restricts the entire side-chain conformations in folded structures to only two strongly populated classes ($\chi_1/\chi_2 = -60°/180°$ or $180°/+60°$) in folded structures[45] and also causes the extremely high (70−80%) p_{-60} and low (1−17%) p_{+60} values in the unfolded proteins (Figure 5).

Conclusion

In summary, we have presented optimized detection schemes for side-chain $^1H^\beta$ resonances in unfolded proteins that yield highly precise structural information about the χ_1 angle from up to six $^3J_{H\alpha H\beta}$, $^3J_{NH\beta}$, and $^3J_{C'H\beta}$ coupling constants and up to two $^1D_{H\beta C\beta}$ RDCs. Interpretation of the detected ^3J couplings in urea-denatured ubiquitin and protein G by a model of staggered χ_1 rotamers[38−40] and previously published Karplus coefficients[37] provides stereoassignments of $^1H^\beta$ methylene protons and yields excellent agreement. This corroborates both the staggered rotamer model and the high precision of the Karplus coefficients. For most residues, the precision of individual rotamer populations is better than 2% as estimated from error propagation. As found in earlier studies,[18] the rotamer populations are in the vicinity of averages obtained from coil regions of folded protein structures. However, individual variations from these averages of up to 40% are highly significant and must originate from sequence- and residue-specific interactions. Particularly strong deviations from the coil average are found for serine and threonine residues, an effect that may be explained by a weakening of side-chain to backbone hydrogen bonds in the urea-denatured state.

The measured $^1D_{H\beta C\beta}$ RDCs correlate well with predicted RDCs based on steric alignment of a coil model ensemble of the unfolded state generated by the program *Flexible*-Meccano,

(44) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 17908–17918.
(45) Ponder, J. W.; Richards, F. M. *J. Mol. Biol.* **1987**, *193*, 775–791.
(46) Cornilescu, G.; Marquardt, J.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836–6837.
(47) Derrick, J.; Wigley, D. *J. Mol. Biol.* **1994**, *243*, 906–918.

where the side-chain conformations had been adjusted according to the *J*-derived χ_1 rotamer populations. This agreement validates the coil model as a good first approximation of the unfolded state. However, deviations between the measured and predicted values also indicate that the local backbone geometries may be improved by incorporation of the additional RDC information. The ease of detection of a large number of highly precise side-chain RDCs should make it possible to obtain such a more accurate description of backbone and side-chain conformations in unfolded states.

Acknowledgment. We thank C. Howald, K. Rathgeb-Szabo, and M. Rogowski for sample preparation, and we also thank H.-J. Sass and S. Kasprzak for helpful discussions. This work was supported by SNF Grant 31-109712.

Supporting Information Available: Details of quantitative $^3J_{NH\beta}$-HNHB, $^3J_{C'H\beta}$-HN(CO)HB, $^3J_{H\alpha H\beta}$-HAHB(CACO)NH, and IPAP-HBHA(CO)NH experiments, figures showing comparisons of experimental and calculated ^3J and RDC values for all observed residues in urea-denatured ubiquitin and protein G, and tables of measured ^3J and RDC values as well as of the fitted χ_1 rotamer populations. This material is available free of charge via the Internet at http://pubs.acs.org.

JA910331T