J|A|C|S
ARTICLES
Published on Web 09/26/2006

Residual Dipolar Couplings in Short Peptides Reveal Systematic Conformational Preferences of Individual Amino Acids

Sonja Alexandra Dames,[†] Regula Aregger,[†] Navratna Vajpai,[†] Pau Bernado,[‡] Martin Blackledge,[‡] and Stephan Grzesiek*,[†]

Contribution from the Biozentrum, University of Basel, Switzerland, and Institut de Biologie Structurale Jean-Pierre Ebel, Grenoble, France

Received May 23, 2006; E-mail: stephan.grzesiek@unibas.ch

Abstract: Residual dipolar couplings (RDCs) observed by NMR in solution under weak alignment conditions can monitor average net orientations and order parameters of individual bonds. By their simple geometrical dependence, RDCs bear particular promise for the quantitative characterization of conformations in partially folded or unfolded proteins. We have systematically investigated the influence of amino acid substitutions X on the conformation of unfolded model peptides EGAAXAASS as monitored by their $^1H^N-^{15}N$ and $^1H^\alpha-^{13}C^\alpha$ RDCs detected at natural abundance of ^{15}N and ^{13}C in strained polyacrylamide gels. In total, 14 single amino acid substitutions were investigated. The RDCs show a specific dependence on the substitution X that correlates to steric or hydrophobic interactions with adjacent amino acids. In particular, the RDCs for the glycine and proline substitutions indicate less or more order, respectively, than the other amino acids. The RDCs for aromatic substitutions tryptophane and tyrosine give evidence of a kink in the peptide backbone. This effect is also observable for orientation by Pf1 phages and corroborated by variations in $^{13}C^\alpha$ secondary shifts and $^3J_{HNH\alpha}$ scalar couplings in isotropic samples. RDCs for a substitution with the β-turn sequence KNGE differ from single amino acid substitutions. Terminal effects and next neighbor effects could be demonstrated by further specific substitutions. The results were compared to statistical models of unfolded peptide conformations derived from PDB coil subsets, which reproduce overall trends for $^1H^N-^{15}N$ RDCs for most substitutions, but deviate more strongly for $^1H^\alpha-^{13}C^\alpha$ RDCs. The outlined approach opens the possibility to obtain a systematic experimental characterization of the influence of individual amino acid/amino acid interactions on orientational preferences in polypeptides.

Introduction

Weak alignment of molecules dissolved in anisotropic liquid phases[1] has become a powerful tool to directly monitor average net orientations and order parameters of individual bonds by residual dipolar couplings (RDCs). RDCs are proportional to the ensemble average $\langle 3\cos^2(\Theta) - 1\rangle/2$, where Θ is the angle between the internuclear vector and the magnetic field in the laboratory frame. Similar to other applications in physical chemistry,[2] $\langle 3\cos^2(\Theta) - 1\rangle/2$ can be interpreted as a local order parameter S of the internuclear vector relative to an external director. S adopts a value of 1, if there is perfect alignment of the bond along the magnetic field, $-1/2$ if there is perfect alignment perpendicular to the magnetic field, and 0 if for example all orientations are equally probable or if there is perfect alignment along the magic angle $\Theta = 54.7°$. By this geometrical dependence and because usually many different RDCs can be determined, RDCs bear particular promise for a detailed, quantitative characterization of conformations in partially folded or unfolded proteins.

Thus RDCs have revealed residual structure in urea-denatured forms of staphylococcal nuclease[3] and natively unfolded alpha-synuclein,[4] α-helix propensities in the unfolded S-peptide,[5] the acyl-coenzyme A binding protein[6] and myoglobin,[7] and β-turn propensities above the melting transition of the T4 fibritin foldon β-hairpin.[8] For shorter peptides, modest RDCs have been observed and interpreted as a local stiffness of the backbone.[9] RDCs of unfolded polypeptides have been described theoretically by polymer random flight models[10−12] and more recently

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

(1) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114.
(2) Doruker, P.; Mattice, W. L. *J. Phys. Chem. B* **1999**, *103*, 178−183.
(3) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487−489.
(4) Bertoncini, C. W.; Jung, Y. S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1430−1435.
(5) Alexandrescu, A. T.; Kammerer, R. A. *Protein Sci.* **2003**, *12*, 2132−2140.
(6) Fieber, W.; Kristjansdottir, S.; Poulsen, F. M. *J. Mol. Biol.* **2004**, *339*, 1191−1199.
(7) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2004**, *340*, 1131−1142.
(8) Meier, S.; Guthe, S.; Kiefhaber, T.; Grzesiek, S. *J. Mol. Biol.* **2004**, *344*, 1051−1069.
(9) Ohnishi, S.; Shortle, D. *Proteins* **2003**, *50*, 546−551.
(10) Louhivuori, M.; Paakkonen, K.; Fredriksson, K.; Permi, P.; Lounila, J.; Annila, A. *J. Am. Chem. Soc.* **2003**, *125*, 15647−15650.

10.1021/ja063606h CCC: $33.50

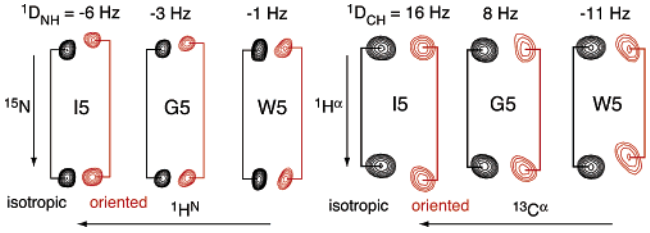

Figure 1. Detection of amino acid specific order in peptides EGAAXAASS from RDCs. $^1D_{NH}$ and $^1D_{CAHA}$ RDCs are obtained from the difference in doublet splittings of nondecoupled natural abundance $^1H-^{15}N$ (left) and $^1H-^{13}C$ HSQCs (right) of peptides EGAAXAASS (X = I, G, W) under isotropic (black) and anisotropic (red) conditions in mechanically strained polyacrylamide gels. ^1H (^{13}C) decoupling was omitted during the ^{15}N ($^1H^\alpha$) evolution period.

by statistical models derived from coil subsets of the Protein Data Bank.[13,14] At present, a systematic experimental characterization of the influence of individual amino acids on the RDC-derived local and global order of polypeptides is lacking.

Here, we have systematically investigated the effect of various single and multiple amino acid substitutions on the conformation of short peptides as monitored by $^1H^N-^{15}N$ and $^1H^\alpha-^{13}C^\alpha$ RDCs detected at natural abundance of ^{15}N and ^{13}C in strained polyacrylamide gels.[15,16] The RDCs show specific dependencies on the amino acid substitutions that for the investigated cases correlate to steric or hydrophobic interactions with adjacent amino acids. The determination of RDC-derived order parameters in conjunction with a systematic variation of the amino sequence opens the possibility for a rigorous experimental characterization of individual amino acid/amino acid interactions in polypeptides.

Results

The peptides used in this study were all derived from the sequence EGAAXAASS where X is the amino acid under investigation. This sequence was based on the rationale of providing neutral next neighbor alanine residues for X and making the peptides water-soluble by hydrophilic residues at their N- and C-terminal ends. In total, 14 amino acids X were investigated; they comprise G, V, L, I, P as aliphatic; T, N, Q as polar; K, D, E as charged; and Y, W, H as aromatic residues.

$^1D_{NH}$ and $^1D_{CAHA}$ RDCs of amide $^1H-^{15}N$ and $^1H^\alpha-^{13}C^\alpha$ internuclear vectors were determined as the difference of the respective doublet splittings in nondecoupled $^1H-^{15}N$ and $^1H-^{13}C$ HSQCs of anisotropic and isotropic samples. Figure 1 shows typical examples for the peptides X = I, G, W. Despite a very similar overall alignment (see below), the RDCs of the three peptides vary significantly in the vicinity of amino acid X, e.g., $^1D_{CAHA}$ equals 16 Hz for I5 but −11 Hz for W5.

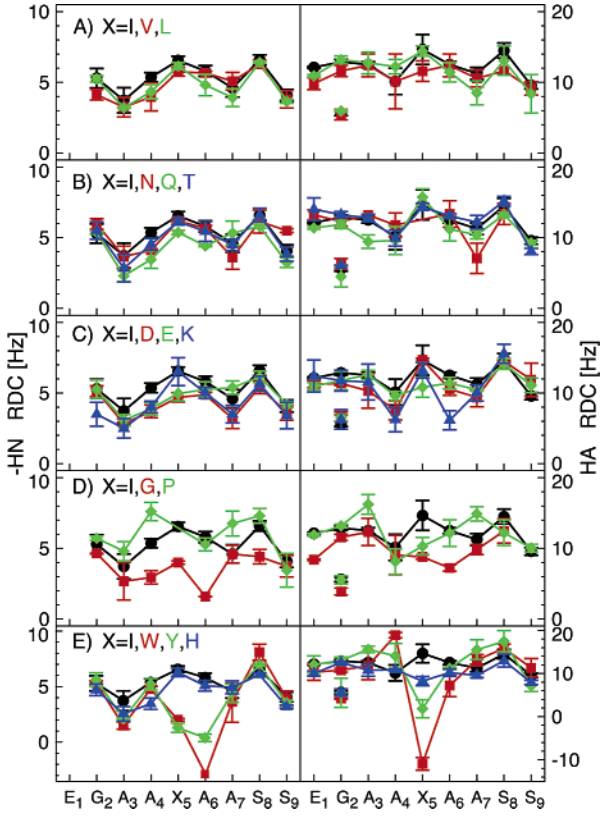

Figure 2. Sequential $^1D_{NH}$ (left) and $^1D_{CAHA}$ (right) RDCs in oriented peptides EGAAXAASS. Experimental RDCs are given for aliphatic (A), hydrophilic (B), charged (C), G and P (D), and aromatic (E) substitutions of the residue X. The specific amino acid substitutions are marked. For comparison, the behavior of the X = I substitution is shown in all panels. For G2, the separately observable $^1D_{CAHA}$ RDCs of both $^1H^\alpha$ protons are shown. For the X = G substitution (D) only one $^1H^\alpha$ resonance is observable. The corresponding average $^1D_{CAHA}$ is shown (see text). Error bars indicate statistical errors from repeated experiments.

Sequential RDCs. Sequential $^1D_{NH}$ and $^1D_{CAHA}$ RDCs for the 14 investigated peptides EGAAXAASS are shown in Figure 2. It is evident that the aliphatic side chains of the amino acids I, V, and L at position X5 (Figure 2A) all lead to very similar RDC profiles and thus indicate similar average orientations of the $^1H^\alpha-^{13}C^\alpha$ and $^1H^N-^{15}N$ internuclear vectors. The absolute values of $^1D_{NH}$ show a bell-shaped, almost 2-fold increase in the center of the peptide (A4, X5, A6). This increase is equivalent to an increase in the average $\langle 3\cos^2(\Theta) - 1\rangle/2$ for the respective N−H bond vectors, which may be caused by a stiffening and/or a kink of the backbone due to the larger side chain at position X5. Toward both peptide termini the $^1D_{NH}$ RDCs decrease. However, the penultimate residues G2 and S8 show increased $^1D_{NH}$ RDCs, which are analyzed in more detail below. The increase for these residues deviates from the expected profile for a random chain polymer,[11] where terminal fraying causes a continuous bell-shaped decrease of RDCs from the peptide center toward the termini.

In comparison to the $^1D_{NH}$ RDCs, the variation of $^1D_{CAHA}$ RDCs along the peptide chain is less pronounced. A weak (about 20%) increase in $^1D_{CAHA}$ is also observed at position X5 for the case of X = I and L, but it is undetectable for X = V. It is remarkable that in all investigated cases the two G2 $^1H^\alpha$ nuclei showed distinguishable resonances and strongly differing $^1D_{CAHA}$ values. This indicates that the averages of the two $^1H^\alpha-$

(11) Louhivuori, M.; Fredriksson, K.; Paakkonen, K.; Permi, P.; Annila, A. *J. Biomol. NMR* **2004**, *29*, 517−524.
(12) Fredriksson, K.; Louhivuori, M.; Permi, P.; Annila, A. *J. Am. Chem. Soc.* **2004**, *126*, 12646−12650.
(13) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099−13104.
(14) Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002−17007. Epub 12005 Nov 17011.
(15) Tycko, R.; Blanco, F. J.; Ishii, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9340−9341.
(16) Sass, H. J.; Musco, G.; Stahl, S. J.; Wingfield, P. T.; Grzesiek, S. *J. Biomol. NMR* **2000**, *18*, 303−309.

$^{13}C^{\alpha}$ directions are not identical on the time scale of the inverse of the RDC value (hundreds of milliseconds) and hence that backbone flexibility at the position of G2 is restricted. Unfortunately, an experimental stereoassignment of the $^{1}H^{\alpha 2}$ and $^{1}H^{\alpha 3}$ nuclei ($H^{\alpha 2}$ corresponds to H^{α} in nonglycine amino acids) was not possible, since both showed very similar ROEs and $^{3}J_{HNH\alpha}$ coupling constants.

The substitution of residue X by polar residues N, Q, and T (Figure 2B) or by the charged residues D, E, and K (Figure 2C) does not alter the observed profile of sequential $^{1}D_{NH}$ and $^{1}D_{CAHA}$ RDCs in a significant way. For all of these substitutions, there is a very similar increase of $^{1}D_{NH}$ RDCs in the center of the peptide as well as at G2 and S8. Likewise, less pronounced variations are observed for the $^{1}D_{CAHA}$ RDCs.

A significantly different profile is observed for the substitutions X = G, P (Figure 2D). When the side chain at position X5 is replaced by the hydrogen of a glycine, the (absolute) increase of $^{1}D_{NH}$ RDC values for residues 5, 6, and 8 is abolished and the $^{1}D_{CAHA}$ RDC profile becomes even flatter. Due to averaging or symmetry, the two $^{1}H^{\alpha}$ resonances of G5 are completely indistinguishable even at 800 MHz ^{1}H frequency, and Figure 2D shows only their average $^{1}D_{CAHA}$. In strong contrast to glycine, the proline substitution induces markedly elevated (absolute) $^{1}D_{NH}$ RDCs for the preceding residue A4 and for residue A7. In addition, there is a significant increase of $^{1}D_{CAHA}$ RDCs for A3 and A7. The observed behavior is consistent with the expectation that substitution by the least sterically hindered glycine should lead to a loss of local order, whereas the proline ring should induce additional orientational order by a stiffening of the polypeptide chain.[11] A comparable increase of local RDCs around prolines has also been observed in the case of the thermally unfolded β-hairpin of foldon.[8]

The most significant variations in RDC profile were observed for the aromatic substitutions X = W, Y (Figure 2E). As compared to the adjacent amino acids, X5 and A6 $^{1}D_{NH}$ as well as X5 $^{1}D_{CAHA}$ RDCs are strongly reduced and even change sign to slightly [X5 = Y: $^{1}D_{NH}$(A6)] or even pronounced [X5 = W: $^{1}D_{CAHA}$(W5), $^{1}D_{NH}$(A6)] negative values. The abrupt changes of RDCs along the polypeptide chain indicate a strong kinking or bulging of the peptide backbone around residues X5 and A6.[11] No such pronounced behavior was observed for the substitution of X5 by histidine (Figure 2E), which displays a profile similar to that of isoleucine albeit with a reduced $^{1}D_{CAHA}$ RDC. Unfortunately, it was not possible to test the substitution by phenylalanine, since this peptide was not sufficiently water-soluble for natural abundance ^{13}C and ^{15}N detection.

$^{13}C^{\alpha}$ Secondary Shifts and $^{3}J_{HNH\alpha}$ Scalar Couplings. To obtain additional insights into the intrinsic conformational preferences of the investigated peptides, their $^{13}C^{\alpha}$ secondary shifts were also analyzed (Figure 3, left side). In general, the $^{13}C^{\alpha}$ secondary shifts for all peptides are very similar and close to random coil values. In all cases, N- and C-terminal $^{13}C^{\alpha}$ shifts are altered by −0.6 ppm and +2 ppm, respectively, due to the terminal $-NH_3^+$ and $-COO^-$ substitutions. As expected, the presence of the proline ring for the substitution X5 = P shifts the $^{13}C^{\alpha}$ resonance of the preceding A4 by −2 ppm (Figure 3D). It seems remarkable that, in most cases (with the exception of G, N, H), the X5 residues have moderately negative (−0.5 ppm) $^{13}C^{\alpha}$ secondary shifts, whereas their A4 and A6 neighbors have $^{13}C^{\alpha}$ secondary shifts that are zero or moderately positive. This

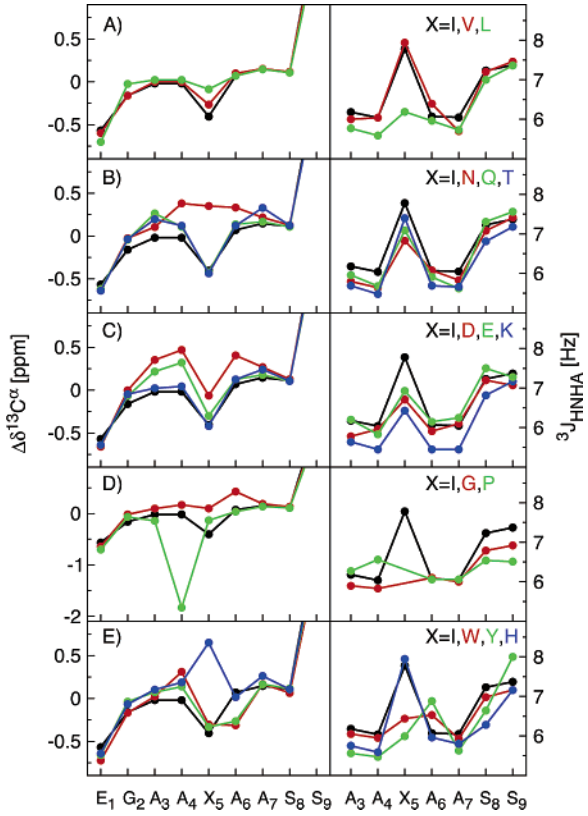

Figure 3. $^{13}C^{\alpha}$ secondary shifts ($\Delta\delta^{13}C^{\alpha}$) and backbone $^{3}J_{HNH\alpha}$ scalar couplings in peptides EGAAXAASS. $\Delta\delta^{13}C^{\alpha}$ values (left) are calculated as the difference of the observed shift and the random coil shift.[17] $^{3}J_{HNH\alpha}$ values (right) were determined from resolved splittings of the $^{1}H^{N}$ resonances in $^{1}H^{N}-^{15}N$ HSQCs at natural abundance of ^{15}N. Data are shown for aliphatic (A), hydrophilic (B), charged (C), G and P (D), and aromatic (E) substitutions of the residue X. For comparison, data of the X = I substitution are shown in all panels.

may indicate that the X5 residues adopt a slightly more extended conformation than their neighbors.[17] A further remarkable feature is that the $^{13}C^{\alpha}$ secondary shifts of A6 are reduced to about −0.3 ppm for the aromatic substitutions X5 = W, Y (Figure 3E). No such shift is observed for any of the other peptides. Thus, these aromatic substitutions entail a specific interaction on the C^{α} position of the next residue, which can be either a conformational preference or a ring current shift. Similarly, the observed variations of the X5 $^{13}C^{\alpha}$ secondary shifts for X5 = H or N must be related to specific interactions of the respective side chains.

In addition to $^{13}C^{\alpha}$ shifts also three-bond $^{1}H^{N}-^{1}H^{\alpha}$ *J*-couplings ($^{3}J_{HNH\alpha}$) were analyzed as an independent indication for phi angle preferences (Figure 3, right side). Consistent with earlier studies and a random coil sampling of conformational space,[18] the observed $^{3}J_{HNH\alpha}$ values are in the range from 6 to 8 Hz for residue X5, whereas the adjacent alanine residues A3, A4, A6, and A7 have smaller values (<∼6 Hz) for all substitutions of X5 with the exception of tyrosine and tryptophane. According to the Karplus relation for $^{3}J_{HNH\alpha}$, this indicates that these alanines have less extended phi angles than X5. In contrast, alanine residues A6 for the aromatic substitu-

(17) Spera, S.; Bax, A. *J. Am. Chem. Soc.* **1991**, *113*, 5490−5492.
(18) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977−8991.

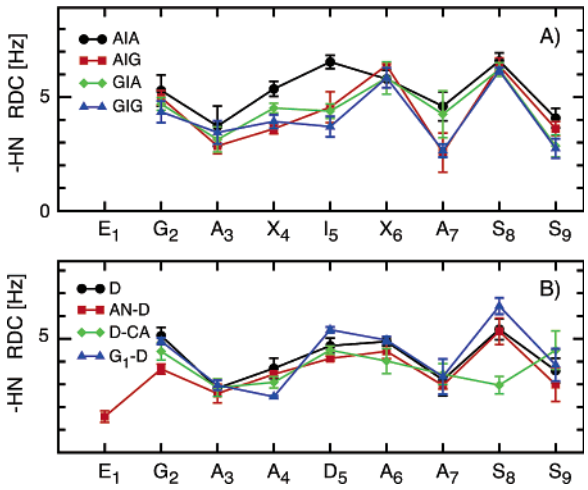

Figure 4. Next neighbor (A) and terminal effects (B) on sequential peptide RDCs. (A) Variation of the central sequence AIA of EGAAIAASS to AIG, GIA, and GIG. (B) Variation of the termini of the peptide EGAADAASS [D] to amidated C-terminus [D-CA], acetylated N-terminus [AN-D], and substitution E1G [G_1-D].

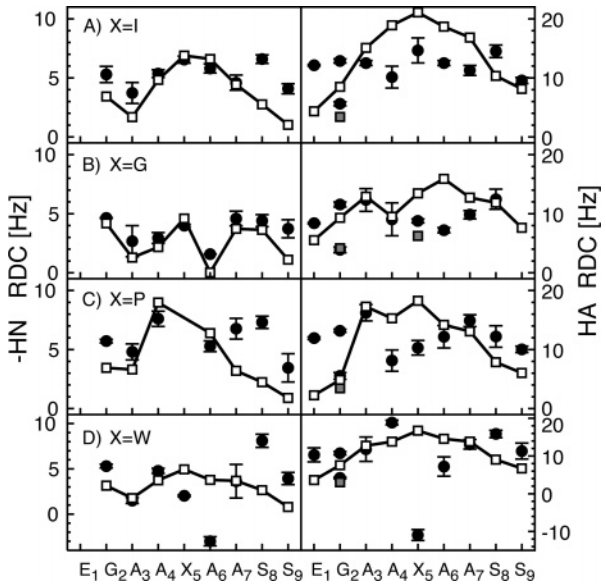

Figure 5. Comparison of experimental RDCs in oriented peptides EGAAXAASS and the predictions from the statistical coil model. Experimental $^1D_{NH}$ (left) and $^1D_{CAHA}$ (right) RDCs are indicated as filled circles, and predicted RDCs, as open squares. (A−D) Single amino acid substitutions X = I, G, P, and W, respectively. For G2 and G5, predictions of $^1D_{CAHA2}$ and $^1D_{CAHA3}$ are shown as open and gray squares, respectively. For each peptide, RDCs of the coil model were scaled by eye such that the $^1D_{NH}$ RDCs best fitted the experimental data. The same scaling factor was used for $^1D_{NH}$ and $^1D_{CAHA}$ RDCs.

tions X5 = Y and W have larger $^3J_{HNH\alpha}$ values (~6.5 Hz) than X5 (Figure 3E). This signifies that the phi angle of A6 is more extended than in the case of the nonaromatic substitutions. More extended conformations cause negative $^{13}C^\alpha$ secondary shifts,[17] consistent with the observed values of about −0.3 ppm. Thus both $^{13}C^\alpha$ shifts and $^3J_{HNH}\alpha$ values for X5 = Y and W show that the backbone conformation differs from the nonaromatic substitutions.

Next Neighbor Effects. The continuous increase of RDCs around the peptide center for most investigated peptides indicates that orientational preferences are transmitted between neighboring residues. To test the effect of the neighbors next to residue X5, we have replaced alanine residues 4 and 6 by glycines. Figure 4A shows a comparison of the sequential $^1D_{NH}$ RDCs of the peptide EGA**AIA**ASS and variations of the form EGA**GIA**ASS, EGA**AIG**ASS, and EGA**GIG**ASS. Whereas the overall alignment is still very similar for the GIA, AIG, and GIG substitutions, it is evident that both a preceding or a following glycine strongly reduce the $^1D_{NH}$ RDC of I5 in the peptide center. In addition, the substitutions A6G (AIG and GIG) reduce the $^1D_{NH}$ RDC of A7. These observations show that the next neighbors have a profound influence on the local backbone propensities and that the larger side chain of alanine induces larger RDCs for the neighboring residues as compared to the smaller glycine. When interpreted as a dynamic effect, this indicates that alanine restricts the neighboring residue more than glycine.

Terminal Effects. To gain insight into the reason for the elevated RDCs at the penultimate residues G2 and S8, we have replaced the C-terminal carboxylate by carboxamide (−CO−NH_2, −CA) and the N-terminal ammonium by acetamide (H_3C−CO−NH−, AN−). Figure 4B shows a comparison of the X5 = D peptide with unmodified termini and its AN− and −CA substitutions. The C-terminal carboxamide abolishes the increased $^1D_{NH}$ RDC for residue S8. Therefore the increased orientation of the S8 N−H vector can be clearly related to interactions between residue S8 and the C-terminal carboxylate group. For the N-terminal AN− substitution only a minor reduction of the G2 $^1D_{NH}$ RDC is observed. A replacement of the large N-terminal residue glutamic acid by glycine (E1G) did not cause any reduction either (Figure 4B). Hence the reason for the increased G2 $^1D_{NH}$ RDC is less obvious. However, together with the strongly differing $^1D_{CAHA}$ RDCs for the two $^1H^\alpha$ protons, it indicates a particular dynamic behavior or conformational preference for G2. This may be caused either by an uncharged N-terminal effect or by an interaction with the following two alanines (see next paragraph).

Comparison to Random Coil Models. Recently, trends in sequential RDCs of chemically denatured proteins have been reproduced by statistical models derived from backbone conformational frequencies in coil subsets of the Protein Data Bank.[13,14] This implies that local preferences in denatured proteins are closely related to the torsion angle distribution of non-α, non-β structures in folded proteins. Predictions according to such a statistical model[14] are shown in Figure 5 for the cases X5 = I, G, P, W. For I, G, and P, the $^1D_{NH}$ RDCs are reproduced quite well from residue G2 to A7. Apparently, the statistical model also predicts an increase of $^1D_{NH}$ RDCs at residue G2. In the model calculations, the termini were represented by truncated peptide bonds. Therefore the increase at G2 can only be caused by normal sequential interactions. The coil model fails to predict the increase of the $^1D_{NH}$ RDC at residue S8. This is not surprising, since the C-terminal carboxylate group is not represented properly in the calculations. When the van der Waals radius of the C-terminal carbonyl carbon atom was artificially increased to 2.0 Å, an increase of the $^1D_{NH}$ RDC of S8 relative to the values at A7 and S9 was observed (data not shown). This corroborates the experimental result that the stronger $^1D_{NH}$ and $^1D_{CAHA}$ RDCs at residue S8 are caused by interactions with the C-terminal carboxylate group.

For the X5 = I, G, P substitutions, the agreement between the coil model and the experimental values for $^1D_{CAHA}$ is

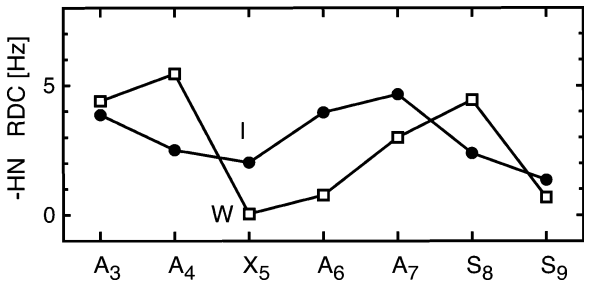

Figure 6. Sequential $^1D_{NH}$ RDCs of the peptide EGAAXAASS observed in Pf1 phages. Data for substitutions X = I are shown as solid circles, and those for X = W, as open rectangles.

considerably worse than that for $^1D_{NH}$ (Figure 5). The reason is currently unclear but may be related to the simplified representation of the side chains by a pseudoatom at the position of the β-carbon. Indeed, variations in the radius of this pseudoatom in the calculations lead to significant variations of the predicted $^1D_{CAHA}$ values (M. Blackledge in preparation).

The statistical model also fails to predict the strong effect of the tryptophane and tyrosine substitutions for both $^1D_{NH}$ and $^1D_{CAHA}$ RDCs around residue X5 (Figure 5D). Apparently, this discrepancy indicates genuine differences between the ensemble of conformations in the gel-oriented peptide solution and the statistical coil ensemble derived from folded proteins.

Orientation by Phages. To investigate whether specific interactions with the orienting medium contribute to the strong variations of the RDCs for the aromatic substitutions, the X5 = I and W peptides were also oriented in a suspension of Pf1 phages[19] (Figure 6). Orientation by the phages is caused by both electrostatic and steric interactions with the highly negatively charged phage surface. Therefore the alignment tensor and the RDCs obtained in phage suspensions usually differ from strained gels[16] or DMPC/DHPC bicelles,[1] which predominantly induce steric alignment. Due to aggregation problems of the phages at the pH of 4.5 used in the gel experiment, the pH had to be raised to 6.0. This rendered the $^1H^N$ resonances of G2 unobservable because of chemical exchange with water protons.

The sequential $^1D_{NH}$ RDCs for the X5 = I peptide in the phage suspension (Figure 6) differ to some extent from the gel orientation in Figure 2A. The profile appears shifted by one residue toward the C-terminus, such that the maximum is located around residues A6 and A7 and the initial decrease occurs from residue A3 to A4 instead of from G2 to A3. The different RDC pattern may indicate either true conformational differences of the peptide in the two different media caused by specific interactions or a changed orientation tensor from the additional electrostatic interactions with the phage. We prefer the latter explanation since (1) the gel yields very similar RDC profiles for almost all peptides (besides G,P,Y,W), which makes specific interactions of the isoleucine peptide with the gel unlikely and (2) the highly negatively charged phages are not expected to interact strongly with the rather hydrophobic isoleucine peptide.

For the X5 = W peptide the $^1D_{NH}$ RDCs in phages (Figure 6) closely resemble the $^1D_{NH}$ RDCs in the strained gels (Figure 2E). Specifically there is also a very sharp decrease of the RDCs after residue A4 to very small values at residues W5 and A6 followed by an increase at residue A7. Since this abrupt change in $^1D_{NH}$ RDCs between A4 and W5 is observed for both alignment media, it is implausible that the evident kink or bulge in the backbone at this position is caused by specific interactions with the medium.

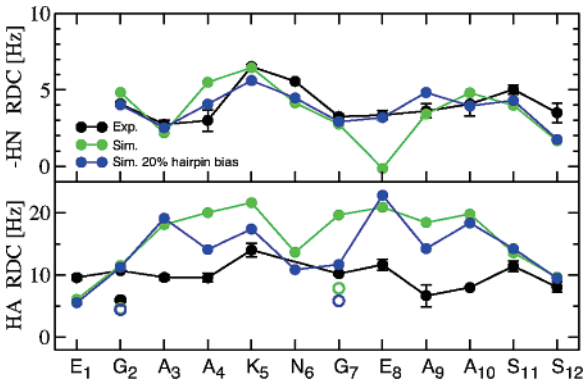

Figure 7. Sequential $^1D_{NH}$ RDCs in the oriented peptide EGAAKNGEAASS containing the foldon KNGE β-turn sequence. Black, experimental data; green, prediction from unbiased coil PDB subset; blue, prediction from a PDB subset biased by 20% toward KNGE β-turns. Predictions for $^1D_{CAHA2}$ and $^1D_{CAHA3}$ of G2 and G5 are shown as closed and open circles, respectively.

Hairpin Substitution. The characterization of orientational preferences in designed peptides by RDCs may ultimately make it possible to follow the peptide folding process depending on sequence and conditions from slight local preferences to more extended stable local structures such as α-helices or β-hairpins. Indeed recently, the gradual loss in local order during thermal denaturation of the foldon β-hairpin could be followed from RDCs.[8] To test whether the foldon β-turn sequence KNGE induces more extended changes in local order within the peptide EGAAXAASS, the sequence KNGE was substituted at position X (Figure 7). The resulting pattern of $^1D_{NH}$ RDCs shows a pronounced maximum at the first two hairpin residues K5 and N6 and a flat profile for most other residues with the exception of the increased values at the penultimate residues G2 and S11 that are also observed for the other peptides. Also similar to the other peptides, the variations in $^1D_{CAHA}$ RDCs are much less pronounced.

The behavior of the $^1D_{NH}$ RDCs is well reproduced by the statistical coil model and even better when the coil model is biased by 20% toward KNGE β-turn configurations (Figure 7). In contrast and again similar to the single amino acid substitutions, the coil model deviates significantly for the $^1D_{CAHA}$ RDCs. The model considerably overestimates $^1D_{CAHA}$ and predicts a more varied pattern with a dip around residue N6. As mentioned, a possible cause for these deviations is the improper modeling of the side chains by the C^β pseudoatoms.

At the present, the limited number of experimental observations (two RDCs per residue) precludes a more quantitative interpretation of the β-turn data. However, the reasonable agreement between $^1D_{NH}$ RDCs and the β-turn-biased coil model is compatible with a significant population of β-turn conformations in the experimental ensemble.

Conclusion

In summary, the reported $^1D_{NH}$ and $^1D_{CAHA}$ RDCs give evidence that single (or multiple) amino acid substitutions alter significantly local structural preferences in short polypeptide

(19) Hansen, M. R.; Mueller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, *5*, 1065−1074.

chains. The observed changes for certain substitutions can be rationalized by specific interactions with neighboring amino acids or terminal groups. For most substitutions besides X5 = W and Y, the observed $^1D_{NH}$ RDCs can be reproduced to a reasonable extent by the statistical coil model. The coil model fails to reproduce $^1D_{CAHA}$ RDCs in a satisfactory way. This may be related to the imperfect modeling of the side chains.

The observed strong variations of RDCs along the polypeptide chain for the tryptophane and tyrosine substitutions must be the result of a kink or bulge of the peptide backbone between residues X5 and A6.[11] Similar strong RDC variations are also observed in phage suspensions used as a second orienting medium. Furthermore, the profiles of $^{13}C^{\alpha}$ secondary shifts and $^3J_{HNH\alpha}$ values obtained under isotropic conditions also deviate from the nonaromatic peptides and point to a more extended conformation of the A6 phi angle. Therefore the observed kink or bulge does not result from specific interactions with the medium. Rather, it must be caused by local interactions of the aromatic side chains with the neighboring alanine amino acids. Indeed, an ROE can be observed between the W5-$^1H^{\epsilon}3$ and the A6-$^1H^{\alpha}$ protons (data not shown). The statistical coil model does not reproduce these strong RDC variations for the aromatic substitutions. Thus there are genuine differences between the conformational ensemble of the unfolded peptides in solution and the current statistical coil model. In contrast, preliminary results from molecular dynamics simulations on fully hydrated peptides are in closer agreement with the observed RDC pattern (N. Vajpai personal observations).

In principle, it should be possible to obtain strong bounds or even entire probability distributions for the torsion angles of an unfolded polypeptide chain from a sufficiently large number of measured RDCs. Currently, this problem is underdetermined since only two parameters per amino acid ($^1D_{NH}$ and $^1D_{C\alpha H\alpha}$) were measured at natural abundance of ^{13}C and ^{15}N. Previously, we have shown that a very larger number ($>\sim$10 per residue) of short-range and long-range RDCs can be detected by the combined use of ^{13}C- and ^{15}N-labeling and perdeuteration.[20,21] The use of other orienting media with different alignment tensors yields additional, independent RDCs of similar number. Such an enormous amount of information has previously yielded new insights on collective slow motions in the folded structure of protein G.[22] When applied to unfolded polypeptides, a similar extensive collection of RDCs should make it possible to overcome the parameter underdetermination and derive distributions for a large number of torsion angles. This would allow a high-resolution thermodynamic description of the structural ensemble of an unfolded polypeptide chain. Current efforts in our laboratories are directed to such a goal.

Experimental Section

Sample Preparation. HPLC-grade, chemically synthesized peptides EGAAXAASS without isotopic enrichment were obtained from a commercial source and used without further purification. Isotropic samples were prepared as solutions of typically 3−6 mM peptide dissolved in 10 mM sodium acetate-d_5, pH 4.5, 95/5% H_2O/D_2O. Anisotropic conditions were achieved as described[23] by introducing the buffered peptide solutions (final concentration 3 mM) into acrylamide gels (10% w/v) and horizontal compression (aspect ratio 2.9) in NEW-ERA sample tubes. Residual alignment by Pf1 phages[19] was achieved at 20 mg/mL phage concentration, pH 6.0.

NMR Experiments. All NMR experiments were carried out at 25 °C on a Bruker DMX800 MHz spectrometer equipped with a TCI cryoprobe. Peptide resonances were assigned from a combination of ROESY, TOCSY, and natural abundance $^1H^{\alpha}-^{13}C^{\alpha}$ and $^1H^N-^{15}N$ HSQC spectra. $^1D_{NH}$ and $^1D_{CH}$ RDCs were obtained as the difference in $^1H-^{15}N$ and $^1H-^{13}C$ doublet splittings observed under anisotropic and isotropic conditions. The splittings were determined from natural abundance $^1H-^{15}N(^1H$-coupled) and $^1H(^{13}C$-coupled$)-^{13}C$ HSQCs where ^1H or ^{13}C decoupling had been omitted during the respective evolution periods. Total experimental times were 6/17 h (isotropic/anisotropic) for the coupled $^1H-^{15}N$ and 1.5/5 h for the coupled $^1H-^{13}C$ HSQCs. Each experiment was carried out at least twice, and the reported values for the RDCs and the estimates for their statistical errors refer to mean and standard deviations derived from such repeated experiments.

Generation of the Conformational Ensemble. A recently developed algorithm, *Flexible*-Meccano, was used to sample conformational space efficiently. This uses conformational sampling based on amino acid propensity and side chain volume. Consecutive peptide planes and C^{α} tetrahedral junctions are constructed[14] in the inverse direction to the primary sequence, starting from the C-terminal residue to the N-terminus for each peptide. The position of the peptide plane (i) is defined from the C^{α} and C′ atoms of plane ($i + 1$), the selected ϕ/ψ combination, and the tetrahedral angle. Amino acid specific ϕ/ψ combinations are randomly extracted from a database of loop structures, built from 500 high-resolution X-ray structures (resolutions < 1.8 Å and B factors < 30 Å^2)[24] from which all residues in α-helices and β-sheets were removed. Residues preceding prolines were considered as an additional amino acid type due to their specific sampling properties.[25] Amino acid specific volumes were represented by spheres placed at C^{β} (or C^{α} for Gly).[26] In the case of steric clash with another amino acid of the chain, the ϕ/ψ pair is rejected and another set of ϕ/ψ dihedral angles ia selected, until no overlap was found. All ensembles comprise 100 000 conformers, and simulated properties are averaged over all members.

In addition to random sampling of the residue specific ϕ/ψ distribution, additional conformational wells with a specific width and center of the ϕ/ψ distribution can also be sampled. Random conformations are then selected from these distributions at the specified rate (for example 1 in 5 conformers for 20% of the population). In the case of the β-turn, dihedral angles from the NMR-determined structure of foldon were used to define the center of the distribution, and a standard deviation of ±20° was used to define the width. The presence of the structural element is cooperative; that is to say if residue 5 adopts a β-turn conformation, residues 6−8 also adopt this conformation. The remaining amino acids follow the standard residue specific sampling described above. In 20% of the conformers residues 5−8 were constrained to β-turns, and 80% follow the residue specific sampling.

Prediction of RDCs from the Conformational Ensemble

The alignment tensor for each member of the ensemble was calculated based on the assumption of steric alignment using the program PALES.[27] RDCs were calculated with respect to

(20) Meier, S.; Haussinger, D.; Jensen, P.; Rogowski, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2003**, *125*, 44−45.
(21) Jensen, P.; Sass, H. J.; Grzesiek, S. *J. Biomol. NMR* **2004**, *30*, 443−450.
(22) Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Bruschweiler, R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13885−13890. Epub 12005 Sep 13819.
(23) Chou, J. J.; Gaemers, S.; Howder, B.; Louis, J. M.; Bax, A. *J. Biomol. NMR* **2001**, *21*, 377−382.
(24) Lovell, S. C.; Davis, I. W.; Arendall, W. B., III; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins* **2003**, *50*, 437−450.
(25) MacArthur, M. W.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 397−412.
(26) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59−107.
(27) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791−3792.

this tensor using the relationship:

$$D_{ij}(\theta,\phi) = -\frac{\gamma_i\gamma_j\mu_0 h}{16\pi^3 r_{ij,\mathrm{eff}}^3}\left[A_\mathrm{a}(3\cos^2\theta - 1) + \frac{3}{2}A_\mathrm{r}\sin^2\theta\cos 2\phi\right]$$

where A_a and A_r are the axial and rhombic components of the alignment tensor, and θ and ϕ are the polar angles of the vector with respect to the tensor principal axes. Effective RDCs were averaged over all values from the 100 000 conformers.

Acknowledgment. This work was supported by SNF Grants 31-61757 and 31-109712 (S.G.) and a fellowship from the Treubel Fonds (S.A.D.).

JA063606H