J|A|C|S ARTICLES

Published on Web 12/16/2009

Intrinsic Propensities of Amino Acid Residues in GxG Peptides Inferred from Amide I′ Band Profiles and NMR Scalar Coupling Constants

Andrew Hagarman,[†] Thomas J. Measey,[†] Daniel Mathieu,[‡] Harald Schwalbe,[‡] and Reinhard Schweitzer-Stenner*,[†]

Department of Chemistry, Drexel University, 3141 Chestnut Street, Philadelphia, Pennsylvania 19104, and Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance (BMRZ), Johann Wolfgang Goethe University, Max-von-Laue-Strasse 7, 60438 Frankfurt, Germany

Received July 13, 2009; E-mail: rschweitzer-stenner@drexel.edu

Abstract: A reliable intrinsic propensity scale of amino acid residues is indispensable for an assessment of how local conformational distributions in the unfolded state can affect the folding of peptides and proteins. Short host−guest peptides, such as GxG tripeptides, are suitable tools for probing such propensities. To explore the conformational distributions sampled by the central amino acid residue in these motifs, we combined vibrational (IR, Raman, and VCD) with NMR spectroscopy. The data were analyzed in terms of a superposition of two-dimensional Gaussian distribution functions in the Ramachandran space pertaining to subensembles of polyproline II, β-strand, right- and left-handed helical, and γ-turn-like conformations. The intrinsic propensities of eight amino acid residues (x = A, V, F, L, S, E, K, and M) in GxG peptides were determined as mole fractions of these subensembles. Our results show that alanine adopts primarily (~80%) a PPII-like conformation, while valine and phenylalanine were found to sample PPII and β-strand-like conformations equally. The centers of the respective β-strand distributions generally do not coincide with canonical values of dihedral angles of residues in parallel or antiparallel β-strands. In fact, the distributions for most residues found in the β-region significantly overlap the PPII-region. A comparison with earlier reported results for trivaline reveals that the terminal valines increase the β-strand propensity of the central valine residue even further. Of the remaining investigated amino acids, methionine preferred PPII the most (0.64), and E, S, L, and K exhibit moderate (0.56−0.45) PPII propensities. Residues V, F, S, E, and L sample, to a significant extent, a region between the canonical PPII and (antiparallel) β-strand conformations. This region coincides with the sampling reported for L and V using theoretical predictions (Tran et al. *Biochemistry* **2005**, *44*, 11369). The distributions of all investigated residues differ from coil library and computationally predicted distributions in that they do not exhibit a substantial sampling of helical conformations. We conclude that this sampling of helical conformations arises from the context dependence, for example, neighboring residues, in proteins and longer peptides, some of which is long-range.

Introduction

All attempts to understand and reproduce protein folding events are generally based on Anfinsen's theory that the folded structures of proteins and peptides are determined by their primary amino acid sequence.[1] One possible strategy to particularly rationalize the relationship between secondary structure and amino acid composition utilizes the concept of conformational propensity, which can be described as a quantitative measure of the occurrence of a considered amino acid in helical, sheet, and turn segments of folded proteins.[2] It is clear, however, that the thus obtained propensity values reflect both intrinsic properties of the respective amino acids as well as the influence of their environment, which involves local and nonlocal interactions.[3−5] However, these propensities do not reflect the earliest step of secondary structure formation, which generally has to pass through the bottleneck of a nucleation process in its initial stage. For coil⇔helix transitions, the latter is accounted for by the nucleation parameters σ and v in the canonical Zimm−Bragg[5] and Lifson−Roig[6] theories, respectively. In the absence of any nonlocal interactions, which could be present in, for example, collapsed conformations,[3] nucleation parameters should reflect the conformational distributions of amino acids in unfolded proteins and peptides.

The classical textbook understanding of the unfolded state suggests that all natural amino acid residues, with the exception of proline and glycine, nearly uniformly sample the sterically

[†] Drexel University.
[‡] Johann Wolfgang Goethe University.

(1) Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. N. *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 1309–1314.
(2) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13*, 211.
(3) Dill, K. A.; Shortle, D. *Annu. Rev. Biochem.* **1991**, *60*, 795–825.
(4) Kihara, D. *Protein Sci.* **2005**, *14*, 1955–1963.
(5) Minor, D. L., Jr.; Kim, P. S. *Nature* **1996**, *380*, 730–734.
(6) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963.

 10.1021/ja9058052

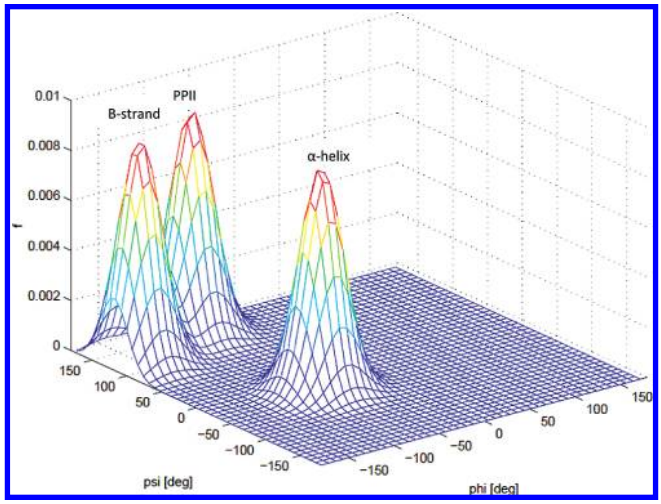

Figure 1. Three-dimensional Ramachandran plot of Gaussian distributions centered on representative conformations of PPII, β, and α.

allowed region of the Ramachandran plot.[7,8] This region encompasses a major fraction of the upper left quadrant (extended conformations) and a smaller, although still substantial, fraction of the lower left quadrant (right-handed helical conformations) (Figure 1). This maximal entropy view of the unfolded state has led to the conclusion that, for example, unfolded proteins in water can be considered as flexible polymers in a so-called good or θ solvent, which does not exhibit any residual structure and can therefore be considered as a random coil.[9,10] Scheraga and co-workers preferred the term statistical coil to avoid the misconception that the allowed Ramachandran space is practically isoenergetic.[11] Irrespective of the used terminology, the above model indicates that nucleation parameters would not depend on the specific properties of amino acid residues. However, this notion has become increasingly questionable. On the basis of far-UV circular dichroism spectra, Tiffany and Krimm[12,13] hypothesized more than 40 years ago that polypeptides such as poly-L-lysine and poly-L-glutamic acid exhibit some degree of local order (TK hypothesis), caused by the rather extensive sampling of polyproline II (PPII)-like conformations by the respective amino acid residues. This hypothesis implies that the sampling of the conformational space by some amino acids is more restricted than indicated by the Ramachandran plots of, for example, Brandt and Flory[8] and by many more recently performed MD simulations of dipeptides.[14−16]

In support of the TK hypothesis, polyproline II-like conformations[17−22] clustering in a trough centered around $\phi = -70°$ and $\psi = 150°$ (Figure 1) have been identified as a subensemble, which, at least for some amino acid residues (e.g., alanine and lysine), dominate the Ramachandran plot.[19,23−26] It is clear that a very high propensity for a substantially restricted region of the Ramachandran plot of adjacent amino acid residues can give rise to the local formation of residual structure, the occurrence of which has indeed been detected in unfolded proteins by NMR spectroscopy.[27,28] Various lines of evidence suggest that the nonrandom distributions of amino acid residues can differ from each other,[21,29−32] so that the occurrence of residual structures can depend on the amino acid composition of a peptide or protein. Hence, a reliable description of the unfolded states of peptides and proteins requires the knowledge of intrinsic conformational propensities of all of the natural amino acid residues.

Generally, the analysis of coil libraries is considered to yield reliable information about conformational propensities of amino acid residues in unfolded proteins.[30,33−35] This view is based on the argument that nonlocal interactions can be "averaged out", if a large basis of different structures is considered.[36] However, as previously demonstrated,[33,37] the conformational distributions obtained for individual amino acid residues in coil libraries depend on whether or not one considers regular secondary structures.[29,34] Even in nonregular structures, the propensity of a residue is affected by the properties of its nearest and second-nearest neighbors, so that it only partially reflects an intrinsic property.[23,35,38−42] The determination of the latter

(7) Ramachandran, G. N.; Ramachandran, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95–99.
(8) Brant, D. A.; Flory, P. J. J. *J. Am. Chem. Soc.* **1965**, *87*, 2791–2800.
(9) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley & Sons: New York, 1969.
(10) Tanford, C. *Adv. Protein Chem.* **1968**, *23*, 121–282.
(11) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 150–167.
(12) Tiffany, M. L.; Krimm, S. *Biopolymers* **1968**, *6*, 1767–1770.
(13) Tiffany, M. L.; Krimm, S. *Biopolymers* **1969**, *8*, 347–359.
(14) Duan, Y.; Wu, C.; Chowdury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
(15) Hovmöller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr.* **2002**, *D58*, 768–776.
(16) Zagrovic, B.; Lipfert, J.; Sorin, E. J.; Millett, I. S.; van Gunsteren, W. F.; Doniach, S.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11698–11703.
(17) Rucker, A. L.; Creamer, T. P. *Protein Sci.* **2002**, *11*, 980–985.
(18) Rucker, A. L.; Pager, C. T.; Campbell, M. N.; Qualls, J. E.; Creamer, T. P. *Proteins: Struct., Funct., Genet.* **2003**, *53*, 68–75.
(19) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190–9195.
(20) Chen, K.; Liu, Z.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15352–15357.
(21) Eker, F.; Griebenow, K.; Cao, X.; Nafie, L.; Schweitzer-Stenner, R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10054–10059.
(22) Eker, F.; Griebenow, K.; Cao, X.; Nafie, L.; Schweitzer-Stenner, R. *Biochemistry* **2004**, *43*, 613–621.
(23) Shi, Z.; Shen, K.; Liu, Z.; Kallenbach, N. R. *Chem. Rev.* **2006**, *106*, 1877–1897.
(24) Stapley, B. J.; Creamer, T. P. *Protein Sci.* **1999**, *8*, 587.
(25) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J. Am. Chem. Soc.* **2007**, *129*, 1179–1189.
(26) Schweitzer-Stenner, R. *J. Phys. Chem. B* **2009**, *113*, 2922–2932.
(27) Wright, P. E.; Dyson, H. J.; Lerner, R. A. *Biochemistry* **1988**, *27*, 7167–7175.
(28) Shortle, D. *Adv. Protein Chem.* **2002**, *62*, 1–23.
(29) Avbelj, F.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5742–5747.
(30) Jha, A. K.; Kolubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099–13104.
(31) Shi, Z.; Chen, K.; Liu, Z.; Ng, A.; Bracken, W. C.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17964–17968.
(32) Tran, H. T.; Wang, X.; Pappu, R. V. *Biochemistry* **2005**, *44*, 11369.
(33) Serrano, L. *J. Mol. Biol.* **1995**, *254*, 322–333.
(34) Jha, A. K.; Colubri, A.; Zaman, M. H.; Koide, S.; Sosnick, T. R.; Freed, K. F. *Biochemistry* **2005**, *44*, 9691–9702.
(35) Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. *Nat. Struct. Biol.* **1995**, *2*, 596–603.
(36) Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. *J. Phys. Chem.* **1996**, *100*, 2661–2666.
(37) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494–506.
(38) Pappu, R. V.; Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12565–12570.
(39) Zaman, M. H.; Shen, M.-Y.; Berry, R. S.; Freed, K. F.; Sosnick, T. R. *J. Mol. Biol.* **2003**, *331*, 693–711.
(40) Avbelj, F.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 19067–10972.
(41) Makowska, J.; Rodziewicz-Motowidlo, S.; Baginska, K.; Vila, J. A.; Liwo, A.; Chmurzynski, L.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1744–1749.

therefore requires the minimization of nearest neighbor interactions between side chains.

In principle, short peptides are suitable model systems for studying conformational manifolds adopted by individual residues.[43] However, most of the experimental studies aimed at exploring such peptides in structural terms yielded either average conformations[21,22,31,44−51] or ensembles of average structures representing subensembles.[52−54] Only a few of these studies led to the reporting of values reflecting conformational propensities.[31,53,54] Some results (e.g., for alanine) are conflicting.[16,19,41,54,55] Recently, Graf et al. performed a comprehensive study, which explicitly probed the conformational distributions of various oligoalanines and of trivaline by combining NMR spectroscopy with Molecular Dynamics simulations.[25] Their approach exploited the conformational sensitivity of seven *J*-coupling constants to identify the conformational distributions for all residues of these peptides. The results of Graf et al. and a subsequent study of Schweitzer-Stenner,[26] who simultaneously analyzed their *J*-coupling constants and earlier obtained amide I′ band profiles of trialanine and trivaline,[47] clearly revealed that alanine, indeed, has a very high propensity for polyproline II-like conformations in the unfolded state, as proposed by Kallenbach and co-workers.[19] Valine, however, was shown to prefer β-strand-like conformations, in agreement with results from earlier studies of Eker et al.[47,48]

The protocol used for the present study combines NMR and vibrational spectroscopic techniques to determine the conformational distributions of various amino acid guest residues in GxG (x labels the guest residue) peptides. We selected glycine as neighbor to minimize nearest neighbor interactions, so that the obtained propensities can really be considered as intrinsic, which is not strictly the case for the propensity values obtained for homopeptides such as trialanine and trivaline. The determination of intrinsic propensities of amino acid residues in water is necessary to construct a reference system based on which their context dependence in more complex systems can be explored. Moreover, the knowledge of intrinsic propensities will allow computational biochemists to calibrate molecular mechanics force fields so that they can be used for the simulations of unfolded, intrinsically disordered peptides and the formation of secondary structures.[56−58] For the guest residue, x, we selected a representative subset (set 1) of amino acids with aliphatic (A, V, L), aromatic (F), charged (E), and polar (S) residues. We measured six different NMR *J*-coupling constants of the central amino acids, which exhibit different ϕ and ψ-dependencies.[25] Additionally, we measured the amide I′ profiles of the respective polarized Raman, FTIR, and vibrational circular dichroism (VCD) spectra for the subset of peptides. A global analysis of these data was performed by fitting them to distribution models, which can be described as a superposition of two-dimensional Gaussian functions reflecting the peaks of different secondary structure conformations in the Ramachandran plot, that is, PPII, different types of β-strands, right- and left-handed helical, and a variety of turn structures. This model is very sensitive to variations in centers and fractions of distributions due to the simultaneous fit of four ϕ value and two ψ value *J*-coupling constants with different Karplus parameters. The results obtained from the analysis of the residues in set 1 were complemented by analyzing the amide I′ band profiles and the canonical $^3J(H^N,H^\alpha)$ coupling constant of K and M (set 2). In addition to obtaining the propensities of the investigated residues for adopting the considered conformations, we also obtained the distribution functions for the latter, which together reflect the respective Gibbs energy landscape of the residue in an aqueous environment.

Materials and Methods

Materials. L-Glycyl-L-alanyl-L-glycine (GAG), L-glycyl-L-glutamic acid-L-glycine (GEG), L-glycyl-L-lysyl-L-glycine (GKG), L-glycyl-L-phenylalanyl-L-glycine (GFG), L-glycyl-L-methionyl-L-glycine (GMG), and L-glycyl-L-leucyl-L-glycine (GLG) were purchased from Bachem Bioscience Inc. with >98% purity and dissolved with no further purification. L-Glycyl-L-valyl-L-glycine (GVG) and L-glycyl-L-seryl-L-glycine (GSG) were custom synthesized by Genscript Corp. (>98% purity) and purified via dialysis in 100 MWCO dialysis bags (Spectrum Laboratories) in an aqueous HCl medium and subsequent freeze-drying to remove trace amounts of trifluoroacetic acid (TFA).

Vibrational spectra were obtained in acidic (<pD 2.1) D_2O, except for GEG, which was measured at pD = 5.1, with a resulting peptide concentration of 0.2 M. The pD values were obtained via the method of Glasoe and Long[59] using an Accumet micro size standard glass combination electrode with Ag/AgCl and an Accumet pH meter (Fisher Scientific).

For NMR experiments, GAG, GVG, GFG, GEG, GLG, and GSG peptides were ^{13}C carbonyl labeled at residue 1, uniformly ^{13}C and ^{15}N labeled at residue 2, and ^{15}N labeled at residue 3. The C-terminal residue was manually attached to a chlorotrityl resin. The synthesis was carried out on an Applied Biosystems 433A peptide synthesizer using standard Fmoc chemistry. Peptides were purified by reversed-phase HPLC. Products were characterized using electrospray ionization mass spectrometry and analytical HPLC. Resins and activating agents were purchased from Novabiochem. All Fmoc-protected amino acids were purchased from Cambridge Isotope Laboratories. All solvents were of analytical grade and dried over molecular sieves if necessary. All NMR samples were prepared by dissolving the peptides in 90% H_2O/10% D_2O, and the pH was adjusted to 2 with HCl.

Vibrational and SRCD Spectroscopy. The experimental setup for polarized Raman, IR, VCD,[53,54] and UV-CD[53,60] experiments has been previously described in detail. Synchrotron radiation CD

(42) Peti, W.; Smith, L. J.; Redfield, C.; Schwalbe, H. *J. Biomol. NMR* **2001**, *19*, 153–165.
(43) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **1993**, *3*, 60–65.
(44) Woutersen, S.; Hamm, P. *J. Phys. Chem. B* **2000**, *104*, 11316–11320.
(45) Woutersen, S.; Hamm, P. *J. Chem. Phys.* **2001**, *114*, 2727–2737.
(46) Eker, F.; Cao, X.; Nafie, L.; Griebenow, K.; Schweitzer-Stenner, R. *J. Phys. Chem. B* **2003**, *107*, 358–365.
(47) Eker, F.; Cao, X.; Nafie, L.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2002**, *124*, 14330–14341.
(48) Eker, F.; Griebenow, K.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2003**, *125*, 8178–8185.
(49) Chellgren, B. W.; Creamer, T. P. *Biochemistry* **2004**, *43*, 5864–5869.
(50) Grdadolnik, J.; Grdadolnik, S. G.; Avbelj, F. *J. Phys. Chem. B* **2008**, *112*, 2712–2718.
(51) Mikhonin, A. V.; Ahmed, Z.; Ianoul, A.; Asher, S. A. *J. Phys. Chem. B* **2004**, *108*, 19020–19028.
(52) Schweitzer-Stenner, R.; Measey, T.; Hagarman, A.; Eker, F.; Griebenow, K. *Biochemistry* **2006**, *45*, 2810–2819.
(53) Schweitzer-Stenner, R.; Measey, T.; Kakalis, L.; Jordan, F.; Pizzanelli, S.; Forte, C.; Griebenow, K. *Biochemistry* **2007**, *46*, 1587–1596.
(54) Schweitzer-Stenner, R.; Measey, T. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6649–6654.
(55) Makowska, J.; Rodziewicz, S.; Baginska, K.; Makowski, M.; Vila, J. A.; Liwo, A.; Chmurzyński, L.; Scheraga, H. A. *Biophys. J.* **2007**, *92*, 2904–2917.
(56) Gnanakaran, S.; Garcia, A. E. *J. Phys. Chem. B* **2003**, *107*, 12555–12557.
(57) Garcia, A. E. *Polymer* **2004**, *120*, 885–890.
(58) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
(59) Glasoe, P. K.; Long, F. A. *J. Phys. Chem.* **1960**, *64*, 188–190.
(60) Hagarman, A.; Measey, T.; Doddasamayajula, R. S.; Dragomir, I.; Eker, F.; Griebenow, K.; Schweitzer-Stenner, R. *J. Phys. Chem. B* **2006**, *110*, 6979–6986.

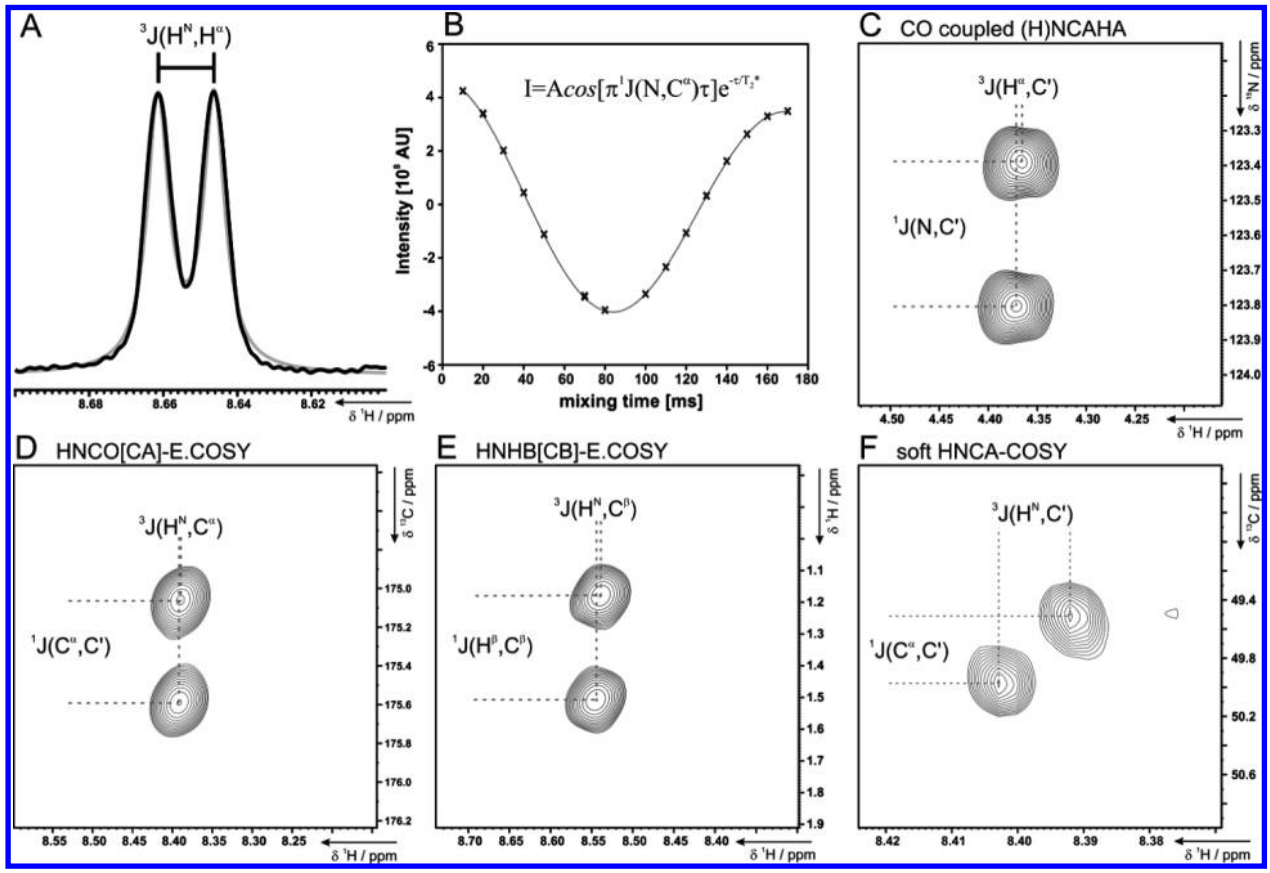

Figure 2. Determination of coupling constants. (A) ^{13}C decoupled 1D for the measurement of $^3J(H^N,H^\alpha)$. The deconvolution result is shown in gray. (B) Intensity fit of *J*-modulated ^{1}H,^{15}N-HSQCs yielding either only $^1J(N_i,C^\alpha_i)$ or both $^1J(N_i,C^\alpha_i)$ and $^2J(N_i,C^\alpha_{i-1})$, depending on the labeling scheme. For the example given, only $^1J(N_i,C^\alpha_i)$ was determined. (C) CO-coupled (H)NCAHA for the measurement of $^3J(H^\alpha,C')$. (D) HNCO[CA]-E.COSY for the measurement of $^3J(H^N,C^\alpha)$. (E) HNHB[HB]-E.COSY for the measurement of $^3J(H^N,C^\beta)$. (F) Soft HNCA-COSY for the measurement of $^3J(H^N,C')$.

(SRCD) experiments were performed at Brookhaven National Laboratories at beamline U-11 using peptide concentrations ranging from 0.01 to 0.1 M, a 0.012 cm path length, 1 nm resolution, and an average of three scans.

NMR Spectroscopy. NMR measurements of GAG, GVG, GFG, GSG, GEG, and GLG have been carried out on a Bruker 400 MHz Avance II spectrometer equipped with a 5 mm HCN triple resonance probe with *z*-gradients. All measurements were performed at 298 K. Spectra were acquired and processed using the program TopSpin Version 2.1. Acquisition and processing parameters are given in the Supporting Information (Table S1).

Proton assignment for all GxG peptides could be obtained from 1D spectra and ^{1}H,^{1}H-TOCSY spectra using a DIPSI-2[61] mixing sequence (Figure S1). The acquisition and processing parameters for all multidimensional NMR are available in the Supporting Information (Table S1). The $^3J(H^N,H^\alpha)$ coupling constant was obtained from a ^{13}C-decoupled 1D spectrum using presaturation for solvent suppression. The exact coupling constant was obtained by Lorentzian deconvolution (see Figure 2A). Depending on the labeling scheme, either only the $^1J(N,C^\alpha)$ or both the $^1J(N,C^\alpha)$ and the $^2J(N,C^\alpha)$ coupling were obtained by measuring a *J*-modulated ^{1}H,^{15}N-HSQC.[62] The intensities for different mixing times were fitted to extract the coupling constant (see Figure 2B). Analysis of the x residue NH resonance yields the $^1J(N,C^\alpha)$ coupling. The same analysis could be performed for the C-terminal glycine NH resonance yielding the $^2J(N,C^\alpha)$ coupling constant. The fitting was performed using the program SigmaPlot and the equation stated in Figure 2B. $^3J(H^\alpha,C')$, $^3J(H^N,C^\alpha)$, $^3J(H^N,C^\beta)$, and $^3J(H^N,C')$ coupling constants were obtained from E.COSY type spectra (see Figure 2C−F).[63−66] Exact peak positions were determined by deconvolution of the respective 1D traces. A full set of the obtained coupling constants is given in Table 1.

The ^{1}H NMR spectra of GKG and GMG were measured with a 500 MHz Varian FT-NMR instrument equipped with a 5 mm HCN triple resonance probe. All spectra were acquired and processed using Varian's VNMR software (v 6.1), with presaturation applied to suppress solvent signals. Peptides were dissolved in 90% H_2O/10% D_2O at concentrations of 0.1 M and acidified with DCl to a pD of 1.5. The D_2O (Sigma-Aldrich) contained 0.05 wt % of 3-(trimethylsilyl) propionic-2,2,3,3-d_4 acid (TSP), which was used as an internal standard. 64 transients were averaged for each sample at 298 K. The $^3J(H^N,H^\alpha)$ coupling constants were determined by deconvolution of the amide proton doublets, using the program MULTIFIT.[67]

Data Analysis. The analysis of the amide I′ band profiles exploits excitonic coupling between the two local amide I′ modes in the tripeptides, which increases the splitting between them and redistributes IR and Raman intensities. The underlying theory as well as the formalism and the empirical parameters used for the simulation of amide I′ band profiles have been described in detail in numerous papers,[45,68−72] to which the interested reader is

(61) Shaka, A. J.; Lee, C. J.; Pines, A. *J. Magn. Reson.* **1988**, *77*, 274.
(62) Wirmer, J.; Schwalbe, H. *J. Biomol. NMR* **2002**, *23*, 47–57.
(63) Löhr, F.; Rüterjans, H. *J. Biomol. NMR* **1995**, *5*, 25–36.
(64) Löhr, F.; Rüterjans, H. *J. Biomol. NMR* **1999**, *13*, 263–274.
(65) Hennig, M.; Bermel, W.; Schwalbe, H.; Griesinger, C. *J. Am. Chem. Soc.* **2000**, *122*, 6268–6277.
(66) Weiseman, R.; Rüterjans, H.; Schwalbe, H.; Schleucher, J.; Bermel, W.; Griesinger, C. *J. Biomol. NMR* **1994**, *4*, 231–240.
(67) Jentzen, W.; Unger, E.; Karvounis, G.; Shelnutt, J. A.; Dreybrodt, W.; Schweitzer-Stenner, R. *J. Phys. Chem.* **1996**, *100*, 14184.

Table 1. Parameters Used for the Simulations of the Vibrational Spectra and Reproduction of the *J*-Coupling Constants for GxG Peptides Where x = A, V, F, S, E, L, K, and M[a]

	A	V	F	S	E	L	K	M
PPII	0.79 ± 0.03 −74 152	0.40 ± 0.05 −81 153	0.42 ± 0.05 −80 146	0.45 ± 0.03 −79 150	0.54 ± 0.03 −80 143	0.56 ± 0.04 −76 145	0.50 ± 0.02 −66 150	0.64 ± 0.02 −74 150
aβt		0.38 ± 0.03 −99 153	0.40 ± 0.03 −100 146	0.30 ± 0.03 −103 130	0.26 ± 0.03 −97 130	0.24 ± 0.04 −98 145		
Pβ	0.06 ± 0.01 −115 120				0.04 ± 0.01 −140 165		0.41 ± 0.02 −115 145	0.36 ± 0.02 −120 150
α_r	0.05 ± 0.01 −60 −30	0.04 ± 0.02 −60 −30	0.10 ± 0.02 −70 −30	0.10 ± 0.02 −50 0	0.08 ± 0.02 −50 −10	0.04 ± 0.01 −50 −30	0.09 ± 0.02 −65 −30	
α_l		0.07 ± 0.02 60 30		0.15 ± 0.02 70 0		0.10 ± 0.02 55 30		
γ_1	0.05 ± 0.01 −80 60	0.11 ± 0.02 −80 60	0.04 ± 0.02 −85 50		0.04 ± 0.01 −75 60	0.03 ± 0.01 −80 70		
γ_2	0.05 ± 0.01 80 −60		0.04 ± 0.02 85 −50		0.04 ± 0.01 70 60	0.03 ± 0.01 80 −70		
$^3J(H^NH^\alpha)$	6.10 6.11 ± 0.02	7.48 7.46 ± 0.08	7.52 7.45 ± 0.02	7.01 6.99 ± 0.07	7.03 6.99 ± 0.02	6.81 6.78 ± 0.01	6.63 6.60 ± 0.02	7.10 7.08 ± 0.02
$^3J(H^N,C')$	1.19 1.18 ± 0.07	1.04 0.91 ± 0.12	1.00 0.88 ± 0.15	1.13 0.87 ± 0.18	1.13 0.94 ± 0.11	0.98 0.84 ± 0.09		
$^3J(H^\alpha,C')$	1.90 2.02 ± 0.10	2.39 2.33 ± 0.15	2.22 2.20 ± 0.27	2.71 2.77 ± 0.32	2.12 2.07 ± 0.16	2.53 2.45 ± 0.03		
$^3J(H^N,C^\beta)$	2.09 2.32 ± 0.06	1.76 1.59 ± 0.06	1.84 1.79 ± 0.17	1.76 1.71 ± 0.12	1.88 1.59 ± 0.24	1.91 1.75 ± 0.15		
$^1J(N,C^\alpha)$	11.28 11.28 ± 0.07	11.16 11.24 ± 0.02	11.18 11.48 ± 0.04	11.04 11.73 ± 0.05	11.00 11.24 ± 0.03	10.97 10.96 ± 0.10		
$^2J(N,C^\alpha)$	8.35 8.51 ± 0.03	8.17 8.01 ± 0.02	8.23 8.28 ± 0.04	7.89 7.86 ± 0.06	8.24 8.38 ± 0.04	8.14 8.24 ± 0.09		

[a] These parameters include the mole fractions (±relative uncertainty) of the considered subensembles as well as the centers of their (ϕ,ψ) distributions (in degrees), listed in the upper and lower parts of the split cells, respectively, in the top half of the table. Uncertainties stem from redistributing small portions of a given subensemble, while maintaining a good simulation of the vibrational spectra and calculated coupling constants within their statistical error. Simulated (upper sub-cells) and experimental (lower sub-cells) NMR *J*-coupling constants are listed in the bottom half of the table.

referred. While earlier studies analyzed amide I′ profiles in terms of average or representative conformations,[52] we recently linked this analysis to a statistical model, which describes the conformational manifold of the central residue of tripeptides in terms of an ensemble of superimposed two-dimensional Gaussian distribution functions:[26]

$$f_j = \left(\frac{\chi_j}{2\pi\sqrt{|\hat{V}_j|}}\right) e^{-0.5\cdot(\vec{\rho}-\vec{\rho}_j^0)^T\hat{V}_j^{-1}(\vec{\rho}-\vec{\rho}_j^0)} \tag{1a}$$

where

$$\vec{\rho} = \begin{pmatrix}\phi\\ \psi\end{pmatrix} \tag{1b}$$

and

$$\hat{V}_j = \begin{pmatrix}\sigma_{\phi,j} & \sigma_{\phi\psi,j}\\ \sigma_{\phi\psi,j} & \sigma_{\psi,j}\end{pmatrix} \tag{1c}$$

The vector $\vec{\rho}_j^0$ points to the position of the maximum of the *j*th distribution in the (ϕ,ψ) space, and χ_j is its statistical weight, which we use as a quantitative measure of a residue's intrinsic propensity for the conformation *j*. The diagonal elements of the matrix $\hat{V}_j$ are the half-widths at half-maximum of the *j*th distribution along the coordinates ϕ and ψ, and the off-diagonal elements reflect correlations between variations along the two coordinates. If $\hat{V}_j$ is diagonal, the ϕ,ψ projection of the distribution is an ellipse with its main axes parallel to the ϕ and ψ axes.

The expectation value of any observable x (spectral intensities, rotational strengths, *J*-coupling constants) can be written as:

$$\langle x\rangle = \frac{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} x\cdot f(\phi,\psi)\,d\phi\,d\psi}{Z} \tag{2}$$

where *Z* denotes the canonical partition sum.

This approach was successfully applied to trialanine and trivaline,[26] and it has now been used in the present study for simulating the amide I′ profiles as well as the obtained *J*-coupling constants. The centers of the distributions were defined to be located in the following subsections of the Ramachandran plot: (1) PPII ($-60° > \phi_{max,1} \geq -90°$; $180° \geq \psi_{max,1} > 100°$), (2) parallel β-strand (pβ) ($-90° > \phi_{max,2} \geq -130°$; $140° > \psi_{max,2} \geq 100°$), (3) antiparallel β-strand (aβ) ($-130° > \psi_{max,3} \geq -180°$, $180° > \psi_{max,3} \geq 100°$), (4) the transition region between antiparallel β-strand and PPII (aβt) ($-90° > \phi_{max,3} \geq -130°$, $180° > \psi_{max,3} \geq 140°$), (5) right-handed α-helix ($-50° > \phi_{max,5} > -80°$; $-20° > \psi_{max,5} > -40°$), (6) left-handed α-helix ($80° > \phi_{max,6} > 60°$; $40° > \psi_{max,6} > 20°$), and (7,8) type IV β-turn, which we will refer to as γ-turn from here on ($\pm 80° > \phi_{max,7(8)} > \pm 60°$; $\mp 50° > \psi_{max,7(8)} > \mp 60°$). The seventh and eighth conformations have been considered to account for recent experimental and theoretical evidence that amino acids can populate this part of the (ϕ,ψ) space.[73,74] The thus defined conformational regions resemble, to a major extent, those of Tran et al.,[32] the major differences being that our PPII region encompasses what they designated as PII and P_{hyp} and that their $P_{hyp} \rightarrow \beta_p$ region is part of our pβ region. The designations "anti-parallel" and "parallel" β-strand solely indicate that the corresponding region contains the canonical dihedral coordinates found in the corresponding sheet structures.

For each conformation, the corresponding amide I′ band profiles were simulated as the sum of two Gaussian profiles assignable to vibrational transitions into two delocalized excitonic states. To account for inaccuracies in peptide concentrations, the experimental IR and VCD band profiles were scaled according to the dipole strengths obtained by Measey et al.[75] The total intensities of the considered ensemble were then calculated by eq 2. The overlap of amide I′ with bands assignable to the carbonyl stretching mode of the C-terminal and side chain modes (e.g., for F) was accounted for by fitting the former and the latter with empirical Gaussian and Lorentzian band profiles, respectively.

Results

We measured the amide I′ band profiles of IR, isotropic Raman, anisotropic Raman, and VCD spectra for a representative subset of GxG peptides in D_2O at acidic pD. D_2O was used to avoid the overlap with the rather strong IR band of water at 1640 cm^{-1} and the vibrational mixing between amide I and H_2O bending modes.[76] Acidic conditions were selected because GFG is not soluble at concentrations required for the current study and the fact that the $^3J(H^N,H^\alpha)$ constants of some other peptides could not be determined at near neutral pD due to fast H↔D exchange. We chose unblocked, rather than blocked, peptides, because the terminal charges increase the difference between the two amide I′ bands, and thus the spectral resolution. As shown in our earlier papers, the terminal charges do not significantly affect the conformation of the central residue if the latter is aliphatic or aromatic.[21,47] The influence of the terminal charges on charged side chains (i.e., E) will be analyzed below.

Figure 2 represents an example of the NMR experimental spectra, and Figure 3 displays the full set of vibrational spectra. Table 1 lists the complete set of measured *J*-coupling constants for the residues in set 1, that is, x = A, V, F, S, E, and L. Additionally, we measured the $^3J(H^N,H^\alpha)$ constant of x = K and M (Table 1) and present only the VCD spectra for these residues in Figure 4. The remaining spectra can be viewed in the Supporting Information (Figure S2). For the description of the data analysis below, we use the term state for each Gibbs energy minimum of residue x in the (ϕ,ψ) space. This minimum corresponds to a maximum of a two-dimensional Gaussian distribution function, which describes a subensemble of conformations.

In a first step, we employed a two-state model encompassing Gibbs energy minima in the PPII and β-strand region of the Ramachandran space. Following Shi et al.,[31] we obtained the (ϕ,ψ) coordinates of the distribution maxima of the corresponding conformational subensembles from the coil library of Avbelj and Baldwin.[29,77] The centers, widths, and statistical weights of the two superimposed Gaussian distributions were then adjusted to optimize the agreement between the simulated and experimental amide I′ band profiles. In agreement with earlier results,[52−54] the shape and the magnitude of the VCD couplet were found to be the most sensitive of all of the spectral parameters to variations among the investigated amide I′ profiles. In each case, we considered only a single β-strand subensemble associated with one of the three β-strand regions introduced above. Subsequently, the distribution parameters were fine-tuned to minimize the root-mean-square deviation (rmsd) between the experimentally observed and calculated *J*-coupling constants, which is written as:

$$\mathrm{rmsd} = \sqrt{\sum_j \frac{(J_j(\mathrm{exp}) - J_j(\mathrm{sim}))^2}{J_j(\mathrm{exp})^2}} \qquad (3)$$

where J_j(exp) and J_j(sim) are the *j*th experimental and simulated coupling constants, respectively. Generally, this procedure yielded satisfactory reproductions of the experimental data. In a second step, however, motivated by earlier results for trialanine and trivaline,[26] we attempted a further improvement of our simulations by admixing minor fractions of helix and turn-like conformations, which yielded improvements for the fits to some of the data sets. Although small, the admixture of α-helices and γ-turns improved the rmsd of the *J*-coupling constants, yet did not jeopardize the agreement between the experimental and simulated vibrational spectra. The solid lines in Figures 3 and 4 reflect the result for the optimized simulations. The centers of distributions and respective mole fractions used for the simulations are listed in Table 1.

The amide I′ profiles shown in Figure 3 for the residues in set 1 look similar with respect to their intensity distributions, but some differences are noteworthy. While GAG, GVG, and GFG exhibit anisotropic Raman profiles with similar intensities for the individual amide I′ bands, the respective profiles of GSG and GEG depict a more intense high-frequency band. The

(68) Schweitzer-Stenner, R. In *Unfolded Proteins. From Denatured States to Intrinsically Disordered*; Creamer, T. A., Ed.; Novalis Press: New York, 2008; pp 101−142.
(69) Schweitzer-Stenner, R.; Measey, T.; Hagarman, A.; Dragomir, I. In *Vibrational Spectroscopy on Peptides and Proteins*; Longhi, S., Uversky, V. N., Eds.; Wiley & Sons: Chichester, 2009; in press.
(70) Choi, J.-H.; Ham, S.; Cho, M. *J. Chem. Phys.* **2002**, *117*, 6821–6832.
(71) Hamm, P.; Lim, M.; DeGrado, W. F.; Hochstrasser, R. *J. Phys. Chem. B* **1999**, *103*, 10049–10053.
(72) Moran, A.; Mukamel, S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 506–510.
(73) Motta, A.; Reches, M.; Pappalardo, L.; Andreotti, G.; Gazit, E. *Biochemistry* **2005**, *144*, 14170–14178.
(74) Gong, H.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3321–3326.
(75) Measey, T.; Hagarman, A.; Eker, F.; Griebenow, K.; Schweitzer-Stenner, R. *J. Phys. Chem. B* **2005**, *109*, 8195–8205.
(76) Sieler, G.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **1997**, *119*, 1720.
(77) Avbelj, F., private communication.
(78) Measey, T.; Schweitzer-Stenner, R. *Chem. Phys. Lett.* **2005**, *408*, 123–127.
(79) Torii, H.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29*, 81–86.

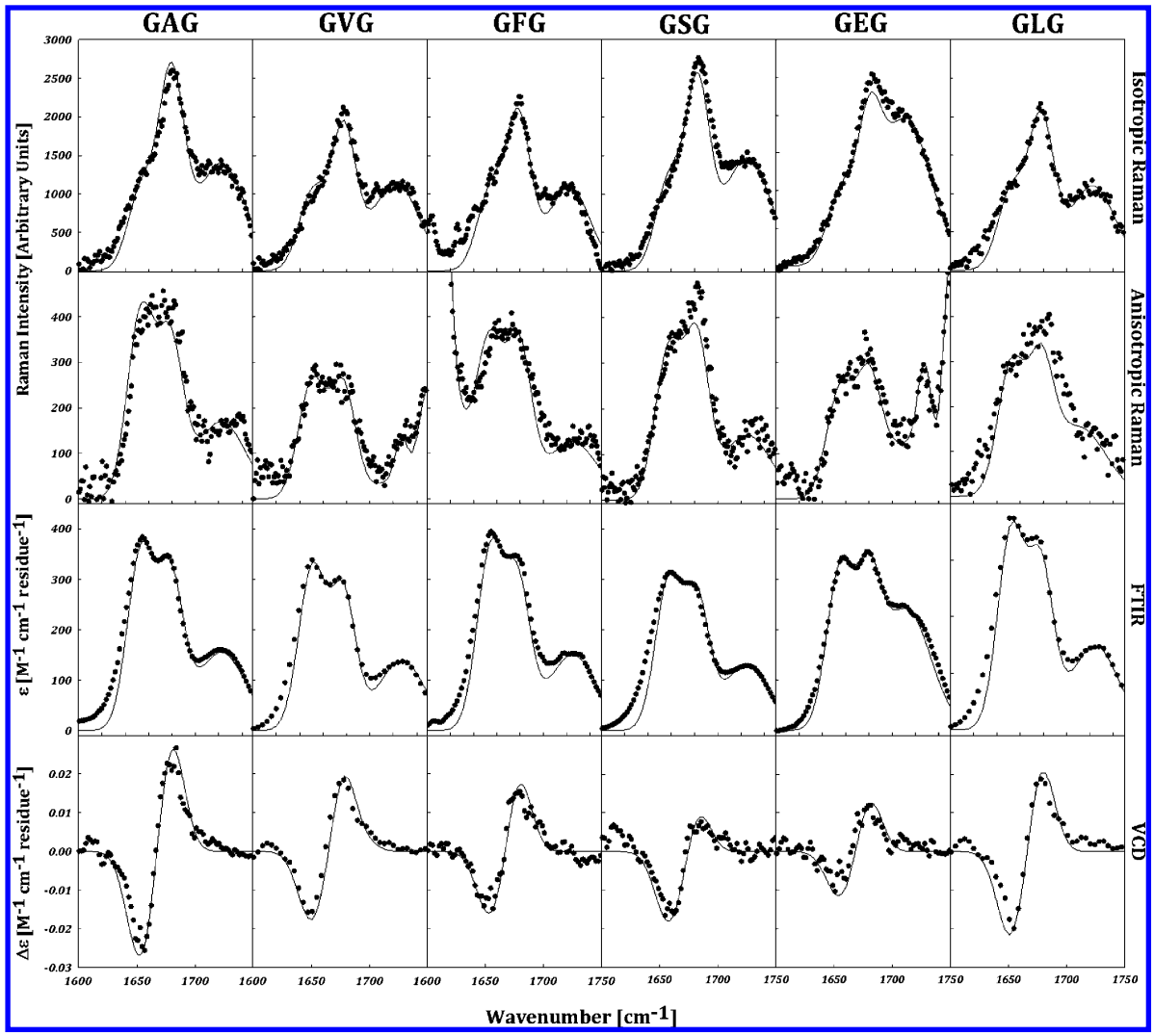

Figure 3. Experimental (dotted line) and simulated (solid line) isotropic (top row) and anisotropic (second row) Raman, IR (third row), and VCD (bottom row) spectra of GAG, GVG, GFG, GSG, GEG, and GLG (as indicated at the top of each column) from 1600−1750 cm^{-1}, encompassing the amide I′ region. The fitting parameters and methods of the simulations are described in the text.

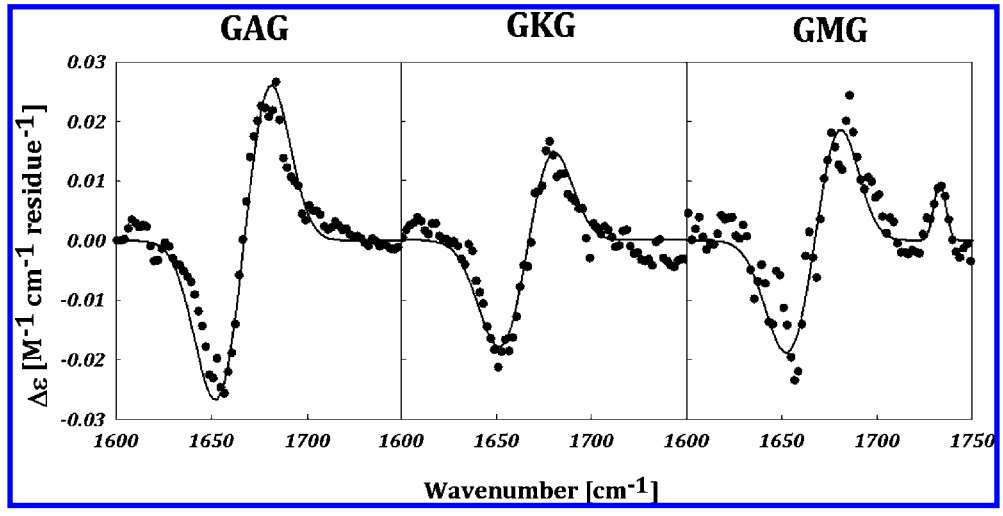

Figure 4. Experimental (dotted line) and simulated (solid line) VCD spectra of GAG, GKG, and GMG from 1600−1750 cm^{-1}, encompassing the amide I′ region. The fitting parameters and methods of the simulations are described in the text.

negative VCD couplets are all indicative of a significant PPII fraction being populated by the respective residues, but the amplitudes of these couplets are clearly different,[78] thus suggesting different conformational distributions. A noncoincidence exists between the isotropic Raman and IR band profiles, with the former displaced toward higher wavenumbers with

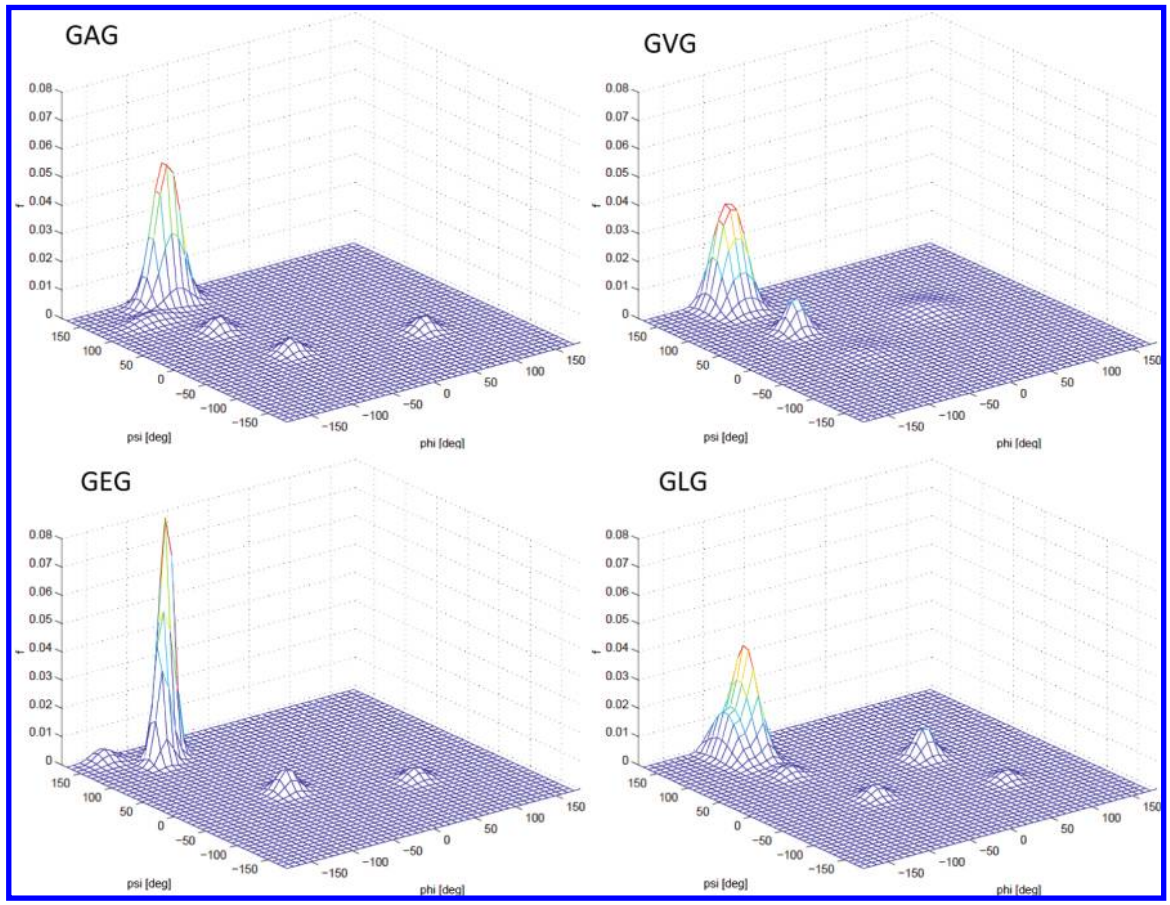

Figure 5. Conformational distributions of the central residue in GAG (top left), GVG (top right), GLG (bottom right), and GEG (bottom left) obtained from the analysis of amide I′ profiles and *J*-coupling constants as described in the text.

respect to the latter. This noncoincidence is indicative of a dominant sampling of extended conformations associated with the upper left quadrant of the Ramachandran plot.[79]

Our analysis revealed a PPII fraction of 0.79 for alanine, which is only slightly lower than the PPII propensity obtained for the central residue of trialanine (0.84),[26] thus suggesting that the energy of nearest neighbor interactions between alanine residues is small, in agreement with Chen et al.[20] The remaining fraction of the GAG ensemble is almost evenly distributed over Ramachandran plot regions associated with pβ, helical, and γ-turn conformations (Table 1), in agreement with findings for trialanine.[26] The total distribution function is plotted in Figure 5. As one can infer from Table 1, the set of *J*-coupling constants obtained from simulations are sufficiently close to the corresponding experimental values.

The VCD (Figure 3) and the ϕ-dependent $^3J(H^N,H^\alpha)$ coupling constants of GVG and GFG (Table 1) indicate that the PPII fractions of both guest residues are substantially lower than that of alanine. A comparison of the GAG and GVG SRCD spectra in Figure 6 points in the same direction. The PPII conformation yields a couplet in the UV-CD spectra with a characteristic strong negative component at ~198 nm and weaker positive component at ~215 nm. The reduced intensity of the couplet for GVG at both of these wavelengths is indicative of a lower PPII propensity. It was recently shown that the intensity of these components can be directly related to PPII content.[80] The conformational ensembles of both amino acid residues sample the aβt-region (0.38 and 0.40 for V and F, respectively) and PPII region (0.40 and 0.42 for V and F, respectively) evenly. The distributions of valine and phenylalanine exhibit small

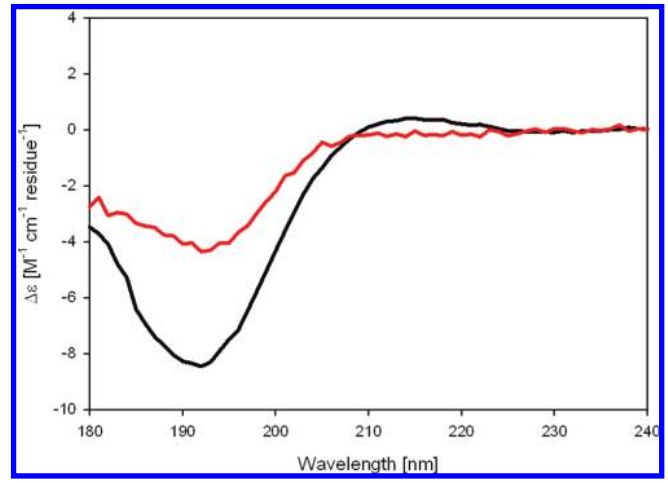

Figure 6. Room temperature synchrotron radiation circular dichroism (SRCD) spectra of GAG (black line) and GVG (red line).

propensities for helical and γ-turn-like conformations (Table 1). The total distribution function of GVG is plotted in Figure 5 for the purpose of illustration. A recent computational study by Xu et al. using a Hamiltonian replica exchange molecular dynamics (HREMD) approach corroborates the notion that valine and phenylalanine have an increased preference for β-like structures relative to alanine.[81]

(80) Woody, R. W. *J. Am. Chem. Soc.* **2009**, *131*, 8234.

Interestingly, the β-strand propensity of V is rather different in GVG and VVV. The latter exhibits a substantially higher β-like propensity (0.68) than the former, a lower value for PPII (0.16), and a higher value for the right-handed helical state (0.16).[26] Moreover, the distribution of the β-strand conformation found by Schweitzer-Stenner is located in the aβ-region ($(\phi,\psi) = (-130°, 135°)$). This discrepancy is somewhat less pronounced, although still significant if one compares GVG with the trivaline (VVV) analysis of Graf et al., who obtained fractions of 0.29, 0.52, and 0.19 for PPII, β-strand, and right-handed helix, respectively.[25] It can, therefore, be concluded that adjacent valine residues in a polypeptide chain shift their mutual propensities toward β-strand, increase the propensity for sampling the right-handed helix region, and decrease the PPII propensity. On the basis of our propensity values, the enhancement of the β-strand propensity of the central valine residue in trivaline would correspond to a Gibbs energy value of 1.2 kJ/mol, at room temperature.

Leucine has a rather long aliphatic side chain, which is less sterically demanding, concerning the backbone hydration, than that of valine. Earlier results obtained for the short peptides ALA and AL indicated a comparable population of the PPII and β-strand region.[21,60] Our analysis of GLG revealed a slightly different picture. Indeed, we found a propensity of 0.56 for PPII, in agreement with Eker et al.[21] However, a substantial fraction (24%) of the ensemble is found in the aβt-region with the center of the corresponding distribution located at $(\phi,\psi) = (-98, 145)$ (cf., Table 1). The corresponding total distribution function therefore suggests that this residue significantly samples the region between the canonical PPII and β-strand conformations, which is in good agreement with recent theoretical predictions.[32]

Serine is particularly interesting to investigate because the side chain is similar to that of alanine except for the replacement of a methyl hydrogen to a hydroxyl group. However, contrary to alanine,[60] this amino acid is a polar residue assumed to adopt mostly a β conformation, as has been reported for another polar residue, threonine.[32] Our results suggest a substantial sampling of the aβt-region (0.30), but the propensity for PPII (0.45) is dominant. Interestingly, the remainder of the ensemble samples distorted right- (0.10) and left-handed (0.15) helical regions.

Charged amino acid residues such as glutamic acid have attracted interest because of their abundance in intrinsically disordered proteins and their alleged propensity for the PPII conformation.[12,13] The values for the fractions obtained from our data set for GEG are, in principle, consistent with this notion, but a PPII fraction of 0.54 indicates that glutamic acid is substantially less inclined to sample the PPII trough than alanine. Sampling of the aβt-region is substantial (0.26) with a small fraction of parallel β-strand. The helical (0.08) and γ-turn-like (0.08) fractions are again small. The total distribution function is shown in Figure 5. The obtained χ-values of GEG reflect the propensity of the protonated residue, which does not resemble its state at physiological pH. Unfortunately, fast H↔D exchange at near neutral pH prevents a measurement of the $^3J(H^N,H^\alpha)$ coupling constant. To check whether there is any influence of the residue charge on the propensity of glutamic acid, we additionally measured and analyzed a full set of amide I′ profiles of GEG at pD = 5.1. We achieved a satisfactory simulation of the amide I′ profiles (Figure S3) by using the exact the fractions, half-widths, and centers of distributions obtained for cationic GEG. This reproducibility indicates that the conformational propensity of glutamic acid does not significantly depend on its protonation state. Generally, this finding underscores the notion that the terminal charges have a limited influence on the conformation of the central residue of tripeptides, most likely because the water molecules in the hydration shell attenuate electrostatic interactions.[82]

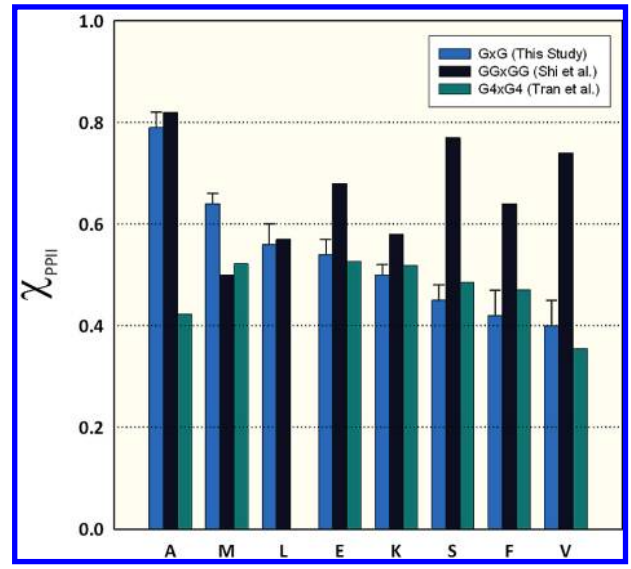

Figure 7. Histograms comparing results for the PPII propensity (χ_{PPII}) of individual amino acids in GxG (this study) (left side: light blue), GGxGG (Shi et al.[31]) (middle: dark blue), and G_4xG_4 (Tran et al.[32]) (right side: green). Amino acids are indicated on the *x*-axis by one-letter codes.

The above findings and conclusion indicate that a residue like lysine shows a behavior similar to that of glutamic acid, irrespective of the character of the terminal group of the side chain. This hypothesis prompted us to explore the conformational ensemble of GKG, based on the analysis of its amide I′ profiles and $^3J(H^N,H^\alpha)$ coupling constants. We added GMG to this second set to check whether a modification of the hydrophobic chain (S instead of CH_2) makes any substantial difference. The corresponding VCD spectra are shown in Figure 4, with the results of our simulations in Table 1. The remaining spectra are shown in Figure S2. We determined, indeed, that lysine in GKG (Table 1) behaves like glutamic acid, in that it exhibits a PPII fraction of 0.50, and an aβt-fraction of 0.41. Interestingly, methionine has a somewhat higher PPII propensity with a fraction of 0.64, which, together with its aβt-region fraction of 0.36, makes it somewhat comparable with glutamic acid, although the positions of the respective β-strand distributions lie in different subsections of the Ramachandran plot.

The diagram in Figure 7 visualizes the PPII propensities of the amino acid residues investigated in this study and compares it with propensity values reported in the literature, which will be discussed below. The error bars have been estimated by adopting the strategy described by Schweitzer-Stenner.[26]

Discussion

Our investigation of the conformational ensemble sampled by six amino acid residues representing aliphatic (A), branched aliphatic (V), extended aliphatic (L), aromatic (F), polar (S),

(81) Xu, C.; Wang, J.; Liu, H. *J. Chem. Theory Comput.* **2008**, *4*, 1348–1359.

(82) Drozdov, A. N.; Grossfield, A.; Pappu, R. V. *J. Am. Chem. Soc.* **2004**, *126*, 2574–2581.

and charged (E) side chains yields the following picture. The preference of aliphatic side chains for the PPII conformational state is highest for alanine, which has the shortest and least sterically demanding side chain in its class. Sterically, serine is comparable with alanine, but the substitution of a hydrophobic CH by a polar hydroxyl group is sufficient to reduce the PPII content significantly. Aliphatic and sterically demanding residues (V, L, and F) significantly populate the aβt-region. Leucine has a preference for PPII, whereas valine and phenylalanine have a nearly equal preference for both subensembles. Interestingly, the somewhat aliphatic, charged glutamic acid residue behaves like leucine. The fact that the propensity of glutamic acid does not depend on its protonation state suggests that the hydrophilic carboxylate group does not matter significantly in this regard. Interestingly, the sampling of the β-strand region is mostly confined to the pβ and aβt regions. A comparison of GVG and VVV suggests that only nearest neighbor interactions can shift the β-distribution to the aβ-region.

The β-like propensities of V and F, and their possible modulation by nearest neighbors, is important for understanding the role of V_n and F_n segments in proteins as well as in synthetic and natural peptides. It deserves to be mentioned that many lines of experimental evidence suggest that β-propensity of specific residues, that is, V and F, is relevant under certain circumstances where aggregation occurs.[83−85] We have to emphasize, however, that recent investigations by the Hecht group cast some doubt on the notion that propensities of amino acid residues matter for the self-aggregation of peptides.[86,87] Thus, one has again to differentiate between the propensities of amino acid residues before and after the nucleation process. The relationship between both is unclear for β-sheet formations.

The parameters determining the propensities of amino acid residues in unfolded peptides are still a matter of debate. With respect to the stabilization of the PPII conformation, an early proposal was made on the basis of DFT calculations for an alanine dipeptide, which suggested that a bridge comprised of two hydrogen-bonded water molecules connecting the CO and NH groups of neighboring peptide units stabilizes PPII conformations.[88] An alternative explanation for the stability of PPII has been provided by Drozdov et al.[82] These authors performed a very thorough Monte Carlo simulation with an all-atom OPLS force field to explore the conformations sampled by an alanine dipeptide in explicit solvent (i.e., water). Their results suggest that water molecules connecting CO and NH groups of adjacent peptides via hydrogen bonding do not contribute significantly to the solvation free energy of the peptide. Moreover, these simulations revealed that solvation itself does not yield to a preference of PPII over helical conformations. The contribution of solvation to the stabilization of PPII was instead found to be a rather indirect one, in that it neutralizes electrostatic interactions between nonbonded pairs of atoms. As a consequence, steric interactions favoring PPII become predominant. It is unclear whether the results of Drozdov et al. contradict those of Garcia, who performed MD simulations for various polyalanine peptides in explicit water.[57] This work led the author to conclude that PPII is stabilized because it allows an optimal hydration of the considered peptides. Whatever the correct model is, it can be expected that the hydration shell water molecules can be perturbed by the amino acid side chains, so it is understandable that an increasing side chain length reduces the PPII propensity. It is noteworthy in this context that the results of Drozdov et al. indicate that the equilibrium between PPII and β is, in fact, directly affected by solvation. The only residue that does not fit into this category is serine, whose composition is close to alanine but for the replacement of a hydrogen from the methyl group side chain by a hydroxyl group. However, a hydroxyl group can act as an acceptor for hydrogen bonding itself, which, in turn, can disrupt the H_2O network, which stabilizes PPII, for example, in the case of alanine.

Coil libraries are generally considered as suitable tools to explore the propensities of amino acid residues in unfolded peptides and proteins.[29,30,34,40] They are used, for example, to calculate expectation values for *J*-coupling constants, which can then be used with experimental values of residues in unfolded peptides and proteins to identify deviations from a statistical coil behavior.[89,90] However, due to the existence of nearest neighbor and second neighbor interactions, the thus obtained results do not necessarily reflect the intrinsic propensities of amino acid residues, as they were inferred from the present study. A comparison to these values is useful nevertheless. The most noticeable difference between the intrinsic distributions inferred from our data and coil libraries is the fact that the latter exhibit much larger fractions of helical conformations.[29,39] Even the most restricted data set of Jha et al., for which helices, sheets, turns, and preproline residues were omitted, shows helical fractions varying between ca. 0.18 for valine and nearly 0.4 for aspartic acid and threonine (if glycine and proline are neglected). We must therefore conclude that this effective helical propensity is induced by nearest neighbor interactions. With respect to PPII, the values obtained from the restricted coil library are well below the propensities obtained from our study of A, S, K, E, and M, whereas both agree rather well for V, F, and L. Concerning the β-like propensities, the values obtained from the restricted coil library lie well below our values for E and K, while the respective values for V, F, and M are comparable. For A, the value obtained from the coil library (0.25) is much higher than our value (0.06) obtained for the β-strand propensity. This comparison seems to suggest that the propensities of residues with long and bulky side chains are less affected by their respective contexts in a peptide or protein than residues with short (A, S) or charged (E, K) side chains, which is a somewhat surprising observation because bulky residues are considered as more effective in modulating the conformational propensities of their neighbors.[34] It is noteworthy that the centers of our PPII distributions are generally rather close to those obtained from coil libraries. Most of our β-like distributions deviate from the coordinates found in coil libraries. These differences can be viewed in the β distributions of V, F, and L, which were found to be located in the aβt-region $(\phi,\psi) = (-99°, 153°)$, $(-100°, 146°)$, and $(-98°, 145°)$, respectively. These coordinates are rather different from the canonical values obtained in

(83) Bemporad, F.; Taddei, N.; Stefani, M.; Chiti, F. *Protein Sci.* **2006**, *15*, 862–870.
(84) Tjernberg, L.; Hosia, W.; Bark, N.; Thyberg, J.; Johansson, J. *J. Biol. Chem.* **2002**, *277*, 43243–43246.
(85) Tjernberg, L. O.; Naslund, J.; Lindqvist, F.; Johansson, J.; Karlstrom, A. R.; Thyberg, J.; Terenius, L.; Nordstedt, C. *J. Biol. Chem.* **1996**, *271*, 8545–8548.
(86) Kim, W.; Hecht, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15825–15830.
(87) Xiong, H.; Buckwalter, B. L.; Shieh, H.-M.; Hecht, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6349–6353.
(88) Han, W.-G.; Jakanen, K. J.; Elstner, M.; Suhai, S. *J. Phys. Chem. B* **1998**, *102*, 2587–2602.
(89) Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J. Am. Chem. Soc.* **2007**, *129*, 5656–5664.
(90) Bernado, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2005**, *127*, 17968–17969.

parallel and antiparallel β-sheets. For the restricted coil library, the centers of the β distributions for V and F are also not represented by canonical β-sheet values, although the differences are not as pronounced as they are for our values.

Computational studies of conformational propensities have thus far focused mostly on alanine. The work of Tran et al.[32] is an exception in this regard, in that these authors used Monte Carlo simulations to study the propensity of all 20 amino acids and their context dependencies. Here, we focus on the results that they obtained for the residues investigated in the present study in a glycine-rich context. Figure 7 depicts a comparison of PPII propensities discussed here with that reported by Tran et al. It should be noted that in Figure 7, the values in the histogram of Tran et al. include the fractions found in the subensembles of PII and P_{hyp}. As seen in Figure 7, our results correlate well with those of Tran et al., except for alanine and leucine. With respect to PPII, they obtained very similar propensity values between 0.4 and 0.5 (these numbers are the sum of what Tran et al. called P_{II} and P_{hyp}). Leucine again departs from this pattern in that the authors found its distribution dominated by a conformational subensemble located in the transition region between PPII and the antiparallel β-strand coordinate. As indicated above, the respective distribution obtained for GLG attributes 24% of the total ensemble to a region, which heavily overlaps with the $P_{II}/P_{hyp} \Leftrightarrow \beta_{A/P}$ transition region of Tran et al. Interestingly, Tran et al. also obtained a substantial population for V in this region, which also agrees with our results. With the exception of leucine, the simulations of Tran et al. overestimate the helical propensities, mostly at the expense of the β-strand content, which is underestimated.

More recent work of the Pappu group is noteworthy in this context. Vitalis and Pappu employed a new model for implicit solvation together with various force fields to explore the conformational manifold sampled by Ac-x-Nme peptides in water (x again represents different amino acids).[91] They judged the validity of their simulations by comparing them with the respective $^3J(H^N,H^\alpha)$ coupling constants of x. Generally, they found a good agreement between simulation and experiment for OPLS-AA and Amber force fields, but not for alanine for which the $^3J(H^N,H^\alpha)$ was constantly overestimated. This underscores that the peculiarity of alanine is still not fully understood.

Kallenbach and co-workers reported a comprehensive list of experimentally determined intrinsic amino acid propensities for all 20 naturally occurring amino acids.[31] Their values were obtained by analyzing the corresponding $^3J(H^N,H^\alpha)$ constants of the guest residues in Ac-GGxGG-NH_2 in terms of a mixture of two different (representative) conformations of the PPII and β-strand subensembles. Figure 7 shows a histogram comparing our estimated PPII propensities to those obtained by Shi et al. for the eight amino acids investigated in the current study. Our results agree well with their PPII propensities for alanine, leucine, and lysine, but they are somewhat at variance with their results for glutamic acid and methionine (Table 2). The most significant discrepancies were obtained for F, V, and S. These discrepancies are partially due to differences between the respective $^3J(H^N,H^\alpha)$ coupling constants, in that our values are much higher (by >0.4) for these three residues. The higher coupling constants translate to a larger average negative ϕ angle, corresponding to a larger fraction of β conformations. As a consequence, our PPII propensities are much lower, particularly for V and F, than those reported by Shi et al. (i.e., 0.74 and 0.68, respectively).[31] The differences between the data provided by Kallenbach and our data may also arise from differences in measuring the $^3J(H^N,H^\alpha)$ coupling.

Table 2. Comparisons of $^3J(H^NH^\alpha)$ NMR Coupling Constants and, in Parentheses, PPII Fractions for Individual Amino Acids Indicated for Different Studies

AA	$^3J(H^N,H^\alpha)$ χ(PPII)[a]	$^3J(H^N,H^\alpha)$ χ(PPII)[b]	$^3J(H^N,H^\alpha)$ χ(PPII)[c]	$^3J(H^N,H^\alpha)$[d]	$^3J(H^N,H^\alpha)$[e]	χ(PPII)[f]
A	6.11 (0.79)	5.73 (0.82)	6.02 (0.68)	6.08	6.1	0.63
V	7.46 (0.40)	7.05 (0.74)	7.32 (0.53)	7.55	7.2	0.49
F	7.45 (0.42)	6.97 (0.64)	7.17	7.35	7.3	
S	6.99 (0.45)	6.30 (0.77)	7.05	6.62	7.0	
E	6.99 (0.54)	6.78 (0.68)	7.02	6.50	6.7	
L	6.78 (0.56)	7.15 (0.57)	6.84	6.99	6.8	0.58
K	6.60 (0.50)	7.10 (0.58)	6.85	6.92	7.0	
M	7.08 (0.64)	7.70 (0.50)	7.09	6.97	7.1	0.57

[a] Results found in this study. [b] Results from Shi et al. in a GGxGG context.[31] [c] NMR and vibrational spectroscopic results obtained from Avbelj and co-workers for amino acid dipeptides.[50,99] [d] ^3J-coupling constants from a GGxGG study using 6 M GdmHCl by Plaxco et al.[100] [e] ^3J-coupling constants obtained from distributions of ϕ values from coil libraries.[99] [f] PPII propensities estimated by Creamer and co-workers for the X residue in P_3XP_3 peptides.[92]

It is noteworthy in this context that Avbelj and co-workers used $^3J(H^N,H^\alpha)$ NMR coupling constants, in combination with IR and Raman spectroscopy, to determine conformational preferences of amino acids in a dipeptide context, and reported PPII fractions of 0.68 and 0.53 and β fractions of 0.17 and 0.43 for alanine and valine, respectively.[50] The authors considered three representative conformations and report only minor helical fractions. Qualitatively, their results point into the same direction as ours. Some earlier papers report propensities for residues in a nonglycine context. Creamer and co-workers used a P_3xP_3 host−guest system to derive PPII propensities for A, V, L, and M.[92] The respective values are listed in Table 2. Their PPII propensities, as compared to our results, are lower for alanine and higher for valine. Eker et al. used average conformations of several AxA peptides to illustrate the different propensities of guests residues.[21] The authors then interpreted their results qualitatively in terms of a two-conformer (PPII and β-strand) model. In agreement with our findings, their results suggest rather mixed populations for S, M, and L. As in other earlier studies,[46,47] their obtained PPII propensity of alanine is underestimated. Although some of our previous work estimated PPII content of terminal amino acids, a comparison is nevertheless useful.[60] For the terminal amino acids in alanine-based dipeptides, A, K, L, S, and V, Hagarman et al. reported PPII fractions of 0.63, 0.48, 0.48, 0.36, and 0.35, respectively. These values compare surprisingly well with what we obtained in the current study, except for alanine, in view of the limited method and nature of the considered residues.

(91) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* **2008**, *30*, 673–699.

(92) Kelly, M. A.; Chellgren, B. W.; Rucker, A. L.; Troutman, J. M.; Fried, M. G.; Miller, A.; Creamer, T. P. *Biochemistry* **2001**, *2001*, 14376–14383.

Taken together, new insights emerge from the present study. In contrast to what is indicated by coil libraries[77,93] and MD[94] simulations, all of the investigated amino acid residues have a very limited (<0.15) propensity for helical conformations. It is intriguing to transfer our results into the parameters of the frequently used Zimm−Bragg theory. In this approach, the product, σs, of the nucleation parameter, σ, and the statistical weight of a helical conformation, s, reflects the probability of the formation of a peptide/protein segment comprising three helical residues. Hence, we can estimate, for example, for alanine $\sigma s = (\chi_\alpha/(\chi_{PPII} + \chi_\beta + \chi_\gamma))^3$, which yields 1.5×10^{-4}. This value is considerably lower than those obtained from host−guest experiments ($\sim 8 \times 10^{-4}$). However, one has to consider a rather large uncertainty for any of the mole fractions obtained for the minor species of the conformational ensemble. A χ_α value of 0.085 would already be consistent with the experimental σs-value, which is well in the range of experimental uncertainty. Moreover, computational evidence suggests that a helical conformation does not tolerate a β-strand as its neighbor.[38] It is reasonable to assume that α_L and γ-turns can also be excluded so that $\sigma s = \chi_{\alpha r}{}^3/[(1 - \chi_{\alpha r})\chi^2_{PPII}]$. Thus, a value of 0.077 already reproduces the experimental σs value. Hence, the low helix propensity derived from our data makes perfect sense with respect to known helix⇔coil transition parameters. On the basis of this estimation, one can further conclude that the higher propensities for right-handed helical structures indicated by coil libraries suggest that the initiation of helix formation is generally more likely in larger proteins than in short peptides, which is in accordance with the fact that many helix forming segments of proteins are disordered once they are separated from their protein context.[95]

It should be noted in this context that MD simulations generally overestimate the nucleation parameters of polyalanine peptides,[91] which reflects an overestimation of the population of (right-handed) helical conformers in the unfolded state by many force fields.[16,91] Exceptions from the rule are the modified Amber force field A94/MOD, OPLS/AA/L,[96] and the force field of Cornell et al.[97] modified by Garcia and Sanbonmatsu.[98] It is interesting to note that the modified Amber force field was also found to account for the high PPII propensity of alanine.[56]

The studies of Tran et al.[32] and Shi et al.[31] led to the conclusion that PPII is the default conformation for most amino acid residues. Our results support this notion only for alanine and methionine and to a lesser extent for leucine, glutamic acid, and lysine, but not for serine, valine, and phenylalanine, for which an equal PPII and β-like preference was determined. This leads us to suppose that other branched and/or bulky substituents behave similarly. Alanine is a special case in that its PPII propensity exceeds the corresponding propensities of other residues even more than suggested by coil libraries.[29,33,34] While the distributions of the PPII subensembles are all centered close to the canonical value, the respective β-strand distributions differ in terms of their positions, and some of them depart substantially from the coordinates of parallel and antiparallel β-sheets, in that they populate the region between PPII and canonical β-strand troughs of the Ramachandran plot.

The current study constitutes our attempt to provide a sound basis for exploring the conformational manifolds of unfolded peptides and proteins. The values reported herein reflect a valuable intrinsic propensity scale for a representative subset of amino acids. Furthermore, this work emphasizes the necessity to have an indispensible intrinsic propensity scale for individual amino acids as a basis for comparison of further context dependent studies and the calibration of molecular mechanics force fields for computational biochemistry. The next steps, which are currently underway in our laboratories, comprise the determination of the propensity of additional residues and, most importantly, a detailed investigation of the context dependence of their respective propensities in longer peptides.

Acknowledgment. We thank Dr. Franc Avbelj for allowing us access to his coil library and Dr. Tobin Sosnick for useful discussion and guidance to his coil library website. We would like to thank Jon Trunk from Brookhaven National Laboratories for his assistance with the SRCD measurements. We thank Marie Anders-Maurer and Sarah Mensch (University of Frankfurt) for peptide purification. We would like to thank Jonathan B. Soffer for his assistance in completing the TOC graphic. This work was supported by the NSF (Chem 0804492). The SRCD measurements were supported, in part, by the Drexel University Chemistry Department. H.S. is a member of the DFG cluster of excellence: Macromolecular Complexes. The Center for Biomolecular Magnetic Resonance (BMRZ) is supported by the State of Hesse. D.M. is supported by the Evonik foundation.

Supporting Information Available: NMR assignments; Raman, IR, and VCD spectra; and acquisition and processing parameters. This material is available free of charge via the Internet at http://pubs.acs.org.

JA9058052

(93) Sosnick, T. R., http://godzilla.uchicago.edu/cgi-bin/rama.cgi.
(94) Best, R. B.; Buchete, N. V.; Hummer, G. *Biophys. J.* **2008**, *95*, L07.
(95) Shoemaker, K. R.; Kim, P. S.; Brems, D. N.; Marqusee, S.; York, E. J.; Chaiken, I. M.; Stewart, J. M.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2349–2353.
(96) Gnanakaran, S.; Garcia, A. E. *Proteins* **2005**, *59*, 773–782.
(97) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc. USA* **1995**, *117*, 5179–5197.
(98) Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782–2787.
(99) Avbelj, F.; Gradolnik, S. G.; Grdadolnik, J.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1272–1277.
(100) Plaxco, K. W.; Morton, C. J.; Grimshaw, S. B.; Jones, J. A.; Pitkeathly, M.; Campbell, I. D.; Dobson, C. M. *J. Biomol. NMR* **1997**, *10*, 221–230.