

Repeating Aspartic Acid Residues Prefer Turn-like Conformations in the Unfolded State: Implications for Early Protein Folding

Bridget Milorey, Harald Schwalbe, Nichole O'Neill, and Reinhard Schweitzer-Stenner*

Cite This: *J. Phys. Chem. B* 2021, 125, 11392–11407

Read Online

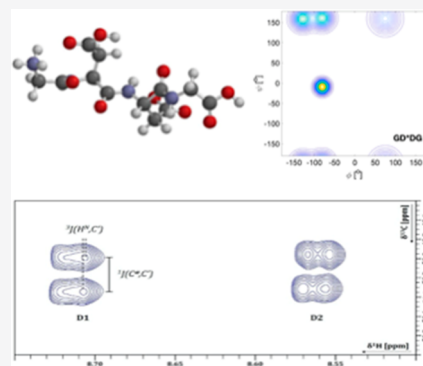
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Protein folding can be described as a motion of the polypeptide chain in a potential energy funnel, where the conformational manifold is narrowed as the chain traverses from a completely unfolded state until it reaches the folded (native) state. The initial folding stages set the tone for this process by substantially narrowing the manifold of accessible conformations. In an ideally unfolded state with no long-range stabilizing forces, local conformations (i.e., residual structures) are likely to drive the folding process. While most amino acid residues tend to predominantly adopt extended structures in unfolded proteins and peptides, aspartic acid exhibits a relatively high intrinsic preference for turn-forming conformations. Regions in an unfolded polypeptide or protein that are rich in aspartic acid residues may therefore be crucial sites for protein folding steps. By combining NMR and vibrational spectroscopies, we observed that the conformational sampling of multiple sequentially neighbored aspartic acid residues in the model peptides GDDG and GDDDG even show an on average higher propensity for turn-forming structures than the intrinsic reference system D in GDG, which suggests that nearest neighbor interactions between adjacent aspartic acid residues stabilize local turn-forming structures. In the presence of the unlike neighbor phenylalanine, nearest neighbor interactions are of a totally different nature in that they decrease the turn-forming propensities and mutually increase the sampling of polyproline II (pPII) conformations. We hypothesize the structural role of aspartic residues in intrinsically disordered proteins in general, and particularly in small linear motifs, that are very much determined by their respective neighbors.



INTRODUCTION

Turn motifs are found abundantly in folded proteins and exert key structural requirements for the three-dimensional architecture of a protein. Reversed β -turns are of particular importance for the formation of β -hairpins. Since some dihedral angles of turn forming residues lie either within (type I and III, $i+1$, type III, $i+2$) or close to the right-handed helical region of the Ramachandran plot (type I and II', $i+2$), one can expect that they might be relevant for the initial phase of folding into helical structures.¹ Less prominent α -turns (aspartate/aspartic acid and asparagine) and ST-turns (serine and threonine) occur close to the N-termini of α -helices and around metal-protein complexes.²

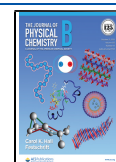
The question arises whether a high propensity for turn-supporting conformations is an intrinsic property of amino acid residues which would thus be able to enrich and to form turns in the early nucleation phase of the folding process. Such a view would be at variance with the conventional understanding of unfolded proteins and peptides which assumes that each residue samples the full sterically allowed region of the Ramachandran plot and that with the exception of glycine and proline none of the natural residues exhibit a peculiar structural propensity.³ Such a "random coil" supporting distribution would contain turn-forming conformations but the respective propensities particularly for type I/I' and II/II' turn supporting structures

would be very low and practically identical for all residues with the exception of proline and glycine. However, a series of NMR studies on short peptide segments by Dyson and Wright showed more than 30 years ago that local interactions cause the formation of turns in unfolded proteins, suggesting that it might constitute the early steps of protein folding processes.^{4–6} They used antipeptide antibodies to locate peptide sequences that could act as these folding initiation sites. For instance, the polypeptide sequence YPYDVPDYA from influenza virus hemagglutinin was thus found to populate β -turns quite substantially.⁷ Supplementary NMR studies using shorter peptide sequences (four to six amino acid residues in length) showed that local interactions cause the formation of turns in peptides that contain amino acid residue sequences of turn structures found in proteins. This observation suggests that the formation of turns (or turn-forming structures on the residual level) is indeed important during the early steps of protein

Received: July 20, 2021

Revised: September 14, 2021

Published: October 7, 2021



folding. Moreover, NMR and circular dichroism experiments revealed nascent helix-like turn conformations as initiation sites for the formation of helices. These results are in line with a variety of protein folding studies showing that turns are frequently formed in a very early phase of the folding process.^{4–6} Two similarities among amino acid residues commonly found in turn motifs of proteins are that they mostly reside on the protein surface and that they are hydrophilic.⁸ In a study searching a database of proteins containing different folds, residues with short, polar side chains were found more often at all positions in β -turns. In addition to proline and glycine, aspartic acid/aspartate (D) is among the most prevalent β -turn-forming residues.¹

The work of Dyson, Wright, and colleagues provided very early evidence that amino acid residues may differ with regard to their conformational propensities in the unfolded state. In the meantime this notion has been confirmed by a plethora of studies on very short model peptides.^{9–14} In this context, our research groups investigated a set of GxG model peptides, where x is a guest residue.^{13,15} In line with highly contested findings of other groups, we discovered that among the investigated residues, alanine stands out with its high propensity for polyproline II (pPII).^{13,16–20} The others have a slightly more balanced distribution among extended structures. Results of most recent work on GGG suggest that the propensity for pPII is to a significant extent engrained in the peptide/protein backbone if it is fully exposed to water.²¹ Amino acid side chains amplify or reduce this peculiar propensity. While residues with aliphatic and aromatic side chains predominantly sample the upper left quadrant of the Ramachandran plot (pPII and β -strand), residues with short, polar side chains and hydrogen bonding capacity (D, S, T, N) populate turn-forming structures significantly more.^{15,22} Protonated GDG stands out regarding the central residue's remarkably high propensity for turn-forming structures, especially for the somewhat less prominent asx-turn-like conformation shown in Figure 1 which resides in

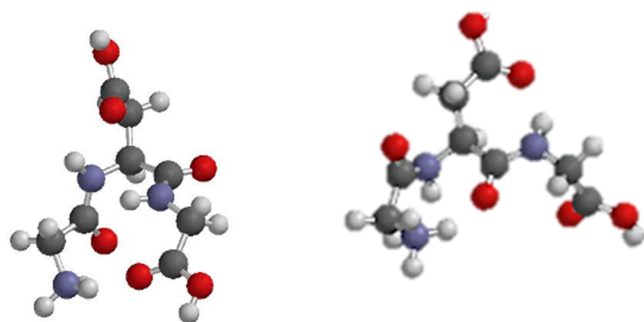


Figure 1. Type I/II' ($i+2$) β turn-forming conformation (left) and asx-turn conformation (right) of protonated GDG.

the upper right quadrant of the Ramachandran plot.^{15,22} In addition, a less pronounced though still significant population of conformations resembling type I/II' ($i+2$) β -turn conformations (cf. Figure 1) is clearly visible in the Ramachandran plot of GDG. It is apparently stabilized by a hydrogen bond between the C-terminal carboxylic acid and the N-terminal NH_3^+ group. In longer peptides this hydrogen bond might be substituted by CO–NH hydrogen bonding.

The studies on GDG and other tripeptides revealed interesting information about conformational preferences and the underlying thermodynamics. However, they cannot be

directly transferred to unfolded peptides and proteins. While propensities in the latter are affected by nonlocal interactions, local nearest neighbor interactions (NNIs) occur even in short peptides. Here and in earlier papers the term “nearest neighbor” refers to adjacent amino acid residues, not adjacent peptide groups as in the theory of Flory.²³ They are ignored in classical random coil models which rely on the isolated pair hypothesis (IPH) borrowed from polymer physics. Whereas rather comprehensive data sets of intrinsic conformational propensities of amino acid residues have been published over the last 15 years (*vide supra*), our understanding of NNIs on a residue level and their significance for protein folding is still underdeveloped. This issue was for quite some time solely addressed by utilizing either computational or biostatistical tools (i.e., analysis of conformational distributions in coil libraries).^{19,24–29} Experimental data became available only more recently. Toal et al. combined vibrational and NMR spectroscopy to obtain Ramachandran plots of nonterminal residues in cationic GxyG peptides.³⁰ They mostly focused on pairs of different aliphatic residues or of aliphatic and polar residues. The latter are of particular interest because NNIs were found to significantly reduce the intrinsic capabilities of protonated aspartic acid (D, henceforth referred to as aspartic acid) to sample turn-forming conformations stabilized by side chain–backbone hydrogen bonding.^{14,22} More recently, we started to augment the set of Ramachandran plots of Toal et al. by investigating interactions between like residues. For the cationic arginine (R) containing peptides GRRG and GRRRG, we found that NNIs stabilize more extended conformations of R. This result explains at least in part the noncoil behavior of polyarginine segments in protamine sequences.³¹

In this paper, we focus on how the conformational propensity of a protonated aspartic acid residue is modified by like and unlike neighbors. As stated above, the Ramachandran distribution of aspartic acid in GDG is peculiar in that it exhibits an above average sampling of the turn-forming conformations shown in Figure 1.^{13,22} From the truncated coil library of Sosnick and colleagues (no helices and sheets considered), we obtained the Ramachandran plot for aspartic acid/aspartate with glycine neighbors (reminiscent of a GDG peptide) and a plot obtained by adding the distributions for all up- and downstream neighbors of the considered data set (Figure S1).³² The former is clearly dominated by conformations found in type I and II' β -turns. The same can be said for the map obtained by integrating over all neighbors, but a minority polyproline II population is now notable. A similar dominance of in type I and II' β -turns can be inferred from coil library-based Ramachandran plots reported by Ting et al.²⁹ Toal et al. showed that the turn propensity of the aspartic acid residue is significantly reduced if G is replaced by aliphatic neighbors such as valine (V), leucine (L), or protonated lysine (K).³⁰ K is also counted here as aliphatic because only its end group is polar. Here, we evaluate the influence of nearest neighbors on the conformational ensembles of aspartic residues in the model peptides GDDG, GDDDG, GDFG, and GFDG. The choice of self-neighbors is motivated by the general goal to augment the work of Toal et al. by an investigation of interactions between like residues. Further support for this choice is provided by the observation of the presence of, e.g., DD or even DDD segments in disordered motifs.^{33–35} Moreover, we were wondering whether putting multiple aspartic acid residues in a row might maintain or even increase turn-forming propensities, in contrast to what Toal et al. reported about the influence of unlike mostly

aliphatic nearest neighbors.³⁶ Choosing phenylalanine (F) as neighbors sets a counterpoint owing to its aromatic character. Thus, far, we have not explored the influence of aromatic residue on any amino acid residue. Generally, coil library results suggest that F should have an even stronger influence on the D distribution than alanine, valine, and leucine. Biological relevance is added to the investigation of DF by the occurrence of this motif in intrinsically disordered segments.³⁵

Ramachandran plots of each residue in the investigated GDDG, GDDDG, and GDFG/GFDG peptides were obtained by a combined use of NMR *J*-coupling constants and respective amide I' profiles in the IR, Raman, and vibrational circular dichroism (VCD) spectra. All experiments in our study were performed at acidic pH, in line with earlier experiments on short peptides. This is mostly being done for technical reasons: at acidic conditions the NMR signal of amide groups is measurable with a good signal-to-noise over a broad temperature range while the protonation states of the terminal groups ensure a sufficient dispersion of the amide I infrared absorption peak. Several lines of evidence have been provided by our group for the notion that the terminal charges do not affect the conformational backbone distributions of the aliphatic central residues of tripeptides.^{22,37,38} We also found that the influence is negligible for aspartic acid residues.²² However, our results show that in the ionized state of the aspartate residue in GDG (denoted as GDⁱG) the β -turn-like conformation is eliminated and the asx-turn population is reduced.²² We assign this to the combined influence of the terminal and side chain charges. This effect would be absent for D-containing sequences in longer peptides and proteins. We therefore consider the cationic peptides used in this study as a better model system than the corresponding state populated at neutral pH. In order to further validate our reasoning, we explore the conformational manifold of the blocked aspartate dipeptide for which NH₃⁺ and COOH are substituted by methyl groups.

MATERIALS AND METHODS

Materials. Glycyl-aspartic acid-aspartyl-glycine (GDDG), glycyl-aspartyl-aspartyl-aspartyl-glycine (GDDDG), glycyl-aspartyl-phenylalanyl-glycine (GDFG), and glycyl-phenylalanyl-aspartyl-glycine (GFDG) were all custom synthesized by Genscript. The blocked aspartic acid dipeptide (Ac-D-NHMe) was purchased from Bachem with >98% purity. For vibrational spectroscopy experiments (IR, Raman, VCD) the investigated peptides were dissolved in D₂O and the pD was adjusted with DCl to values between 1.9 and 2.4 (recorded pH*-value range: 1.5–2.0). The peptide concentration was adjusted to 150 mM and 100 mM for the tetrapeptides and pentapeptide, respectively. For UV CD experiments, we prepared 10 mM solutions of GDDG and GDDDG in H₂O. Acidic pH-values between 1.8 and 2.0 were obtained by the addition of HCl. For ¹H NMR experiments, peptides were dissolved in an aqueous solution of 90% H₂O/10% D₂O at a concentration of 100 mM, and the pH was adjusted to a value between 1.5 and 2.0. Peptides with isotope-labeled amino acids were synthesized using materials purchased from Cambridge Isotopes (Cambridge, Massachusetts, USA). Solid-phase peptide synthesis was carried out with an Applied Biosystems 433A peptide synthesizer. After synthesis, the peptides were cleaved from the resin with 90% trifluoroacetic acid (TFA) and precipitated with ice-cold ether. Once the solid peptide was dry, a sample was analyzed with electrospray ionization mass spectrometry (ESI-MS) to confirm the peptide product by molecular weight. The peptide product

was purified with reverse-phase HPLC before finally undergoing an acid exchange to replace the TFA with HCl. The purified peptides were dissolved at a concentration of 5–10 mM in a solvent of 90% H₂O/10% D₂O, and the pH was adjusted to between 1.5 to 2.0. For all NMR measurements, the used D₂O solution contained 0.1% 4,4-dimethyl-4-silapentanesulfonic acid which was used as internal standard.

Vibrational Spectroscopies Measurements. The Raman spectra were obtained with the 514.5 nm radiation of a Spectra-Physics (Mt. View, CA, USA) argon laser (200 mW). The laser beam was directed into a Renishaw confocal microscope and focused onto a thin glass coverslip with a 20 $\times$ objective. The scattered light was filtered with a 514.5 notch filter. Spectra were recorded between 1400 and 1800 cm⁻¹. A background spectrum of the D₂O/DCl mixture was recorded on the same day as and later subtracted from each peptide spectrum. The spectra were recorded as an average of five measurements using the WiRE 3.3 Renishaw software.

FTIR and VCD spectra were recorded with a Chiral IR/VCD spectrometer from BioTools using a 48 μ m cell and peptide concentrations of 150 mM for tetrapeptides and 100 mM for pentapeptides. The concentration of the aspartic acid dipeptide was 100 mM. All VCD spectra exhibited a highly nonlinear baseline even after the subtraction of the background which indicates some internal birefringence. This was corrected for by fitting the baseline to a cubic polynomial function which was then subtracted from the spectrum in the amide I' region. Note that the prime sign indicates that the experimental amide I profiles were obtained with peptide dissolved in D₂O. In what follows, we use the term amide I' if we refer to experimental data, while we use amide I for the results of our calculation and for general characteristics of the mode.

NMR Spectroscopy. The ¹H measurements were recorded with a 600 MHz Bruker AV600 spectrometer. The heteronuclear NMR measurements were performed using either a 600 or 800 MHz Avance Bruker NMR spectrometer at the Biomolecular Magnetic Resonance Center (BMRZ) of the Johann Wolfgang Goethe University (Frankfurt, Germany). First, a series of measurements for backbone assignment were performed. These measurements allow for chemical shift assignments of all backbone atoms. Then, the E.COSY and *J*-modulated HSQC measurements described previously¹³ were acquired to determine the following set of coupling constants: ³*J*(H^N,H ^{α}), ³*J*(H^N,C'), ³*J*(H ^{α} ,C'), ³*J*(H^N,C ^{β}), and ¹*J*(N,C ^{α}). All ³*J* coupling constants depend on the dihedral angle φ , though in a vastly different way. ¹*J*(N,C ^{α}) is an indicator of the dihedral angle ψ .

UV Circular Dichroism Spectroscopy Measurements. The UVCD spectra were measured at different temperatures on a JASCO J-1200 spectropolarimeter in a 100 μ m International Crystals Laboratories (ICL) cell, from 180 to 300 nm with a 100 nm/min scan speed, 1 s response time, 0.05 data pitch, and 1 nm bandwidth.

Data Analysis. The detailed procedure for our data analysis was described recently for a previous study³¹ but will be summarized here for the sake of readability. The analysis starts with the *J*-coupling constants measured for each nonglycine residue. The experimental coupling constants, which represent a population average, were used to optimize conformational distributions that contained a set of subdistribution associated with beta-strand (β -strand), polyproline II (pPII), asx-turn (asx), and inverse-gamma-turn (*i* γ) conformations. While pPII and β -strand are sampled by nearly all amino acid residues, asx-

turns appear in the Ramachandran plots of protonated D and of C.¹⁵ For the fifth conformation we considered the entire space covered by right-handed helical conformations and by structures found at the $i+2$ residues of β -turn type I/II' conformations (β I/ β II _{$i+2$}). While low populations of the former have been obtained for many GxG and GxyG peptides,^{13,15,39} sampling of the latter has been observed for GDG. The populations for each conformation were determined by modeling the respective subdistributions by two-dimensional Gaussian functions positioned at locations on a Ramachandran plot that are representative of the considered conformations.⁴⁰ The statistical weights of these conformational subdistributions were used as free parameters subject to normalization in a nonlinear least-squares fit using the lsqarefit module of Matlab 2019b. The statistical weight (population) values obtained by the fitting procedure are a measure of residue propensities for the respective secondary structure.

We started the above-described data analysis by revisiting the earlier reported conformational distributions of the D residue in cationic GDG and of the F residue in cationic GFG.^{15,22,39} We refitted the corresponding J -coupling constants and amide I' profiles by using the Karplus parameters of Ding and Gronenborn⁴¹ rather than the earlier employed Wimper-Schwalbe parameters for $^1J(\text{N,C})$.^{22,39,42} To this end, we allowed the mole fractions of the subdistributions to vary while leaving the positions and widths of the earlier distribution unchanged. For GDG, this analysis led to a slight change of the statistical weights of the considered subdistributions which are described in the **Results and Discussion**. The first fits to the J -coupling constants of the investigated tetra- and pentapeptides were performed with the φ and ψ positions and half-widths of the corresponding subdistributions of GDG and GFG as fixed parameters. Subsequently, we performed these fits for various manually changed positions of the Gaussian maxima and halfwidths until the best reproduction of coupling constants was reached. The qualities of the fits were judged based on the respective reduced χ^2 -values.⁴³

Once these initial Ramachandran plots were created, we used these outputs to simulate amide I profiles for IR, VCD, and Raman. As described in detail in earlier papers, the respective band profiles reflect the excitonic coupling between excited vibrational states of local amide I' oscillators and thus the geometry of the residues between the interacting peptide modes. Through-bond nearest neighbor interactions are calculated using the algorithm reported by Schweitzer-Stenner.⁴⁰ Non-nearest neighbor interactions were described with the transition dipole mechanism of Krimm and co-workers.⁴⁴ Contrary to the NMR coupling constants, the amide I profiles are a superposition of profiles assignable to different peptide, not individual residue conformations. The amide I profiles were therefore calculated based on the entire conformation of the peptide, not for individual residues. Keeping the computational time to a minimum required the use of a truncated model, which ignores regions of the Ramachandran plot that are not sampled by the peptides' residues. It is described in detail by Milorey et al.³¹

Generally, the first simulations based on the output of the J -coupling analysis did not fully account for the experimental band profiles. Particularly for the VCD signal, discrepancies were always obtained. Therefore, we started an iterative procedure during which the subdistribution positions were adjusted to so that the simulated spectra reproduced the measured amide I' profiles and the corresponding J -coupling constants. The entire data analysis process (starting with J -coupling constants)

involved about 1–4 iterations until the best reproduction of data was obtained. It is noteworthy in this context that our fitting to the J -coupling constants revealed some uncertainties regarding the ψ coordinate of the subdistribution in the right-handed helical/type I/II' β -turn region. We obtained fits of similar quality with ψ -coordinates varying between -30° and 30° for this subdistribution. However, this uncertainty was minimized when we used the output of the NMR analysis to reproduce the experimentally obtained Raman, IR, and VCD profiles of amide I' in the above iterative procedure. The Raman and IR profile have a specific shape if the conformational ensemble is dominated by extended structures such as pPII and β -strand.⁴⁵ Moreover, the choice of the ψ coordinate of the pPII and of the helical/turn-forming distribution conformations heavily impacts the magnitude of the VCD signal. Since the aspartic acid residues in GDDG and GDDDG were found to populate extended structures to a lesser extent than residues such as arginine or alanine,^{13,46} simulations of their amide I' profiles were found to depend on optimizing the distribution positions for the helix/turn-forming conformations.

Obtaining statistical errors of statistical weights and subdistribution positions is not straightforward. This notion particularly applies to the former since normalization makes the respective parameters linear dependent. In addition, there are correlation effects between subdistribution positions and statistical weight. In order to at least estimate the uncertainties of both we varied first the statistical weights in a systematic way by concomitantly subtracting and adding the same amount to two statistical weights of the ensemble. The χ^2 function was subsequently calculated for this parameter set. For the positions of the considered subdistributions we varied the φ -value of the maximum in an interval of $\pm 10^\circ$ and calculated the respective χ^2 -values.

The upper and lower limit of the error interval corresponded to increases of the χ^2 function by 50%. The statistical error was the estimated as the mean of the upper and lower deviation from the value obtained by the fitting process. The respective ψ -values obtained from the fit represent a compromise between the minimization requirements for J -coupling and the VCD of amide I'. They therefore do not represent a minimum of the individual χ^2 function. We know from earlier simulations that small changes of the ψ -coordinates for pPII and β -strand lead to significant deviations from either experimental $^1J(\text{NC}')$ value (by an increase of ψ) or from the VCD signal (by a decrease of ψ). A statistical uncertainty of $\pm 2^\circ$ generally reflects this fact. ψ -values of the considered turn conformations are subject to the same competition but with less sensitivity. Statistical error of ca. $\pm 5^\circ$ reflect this slightly larger flexibility. Half-width changes of $\pm 5^\circ$ generally lead to the deterioration of fits if the corresponding structure is significantly populated.

Comparison of Ramachandran Plots. The similarities/dissimilarities between Ramachandran plots were assessed using the Hellinger distance.^{29,47} It has values between 0 and 1, where 0 indicates identical and 1 indicates orthogonal or exactly opposite distributions. Due to the natural constraints on the explorable regions of the Ramachandran plot caused by steric hindrance around the peptide backbone, the Hellinger distance will never approach 1 for two different amino acid residues. We adopt the scaling of Schweitzer-Stenner and Toal,⁴⁷ which ranks the similarity of the Ramachandran plots as follows:

Very similar	$H \leq 0.1$
Moderately similar	$0.1 < H \leq 0.25$
Moderately dissimilar	$0.25 < H \leq 0.4$
Very dissimilar	$0.4 < H$

It should be noted at this point that the Hellinger distance is more sensitive to changes in the positions of subdistributions than it is to changes of their statistical weight. In other words, two Ramachandran plots with basins (i.e., subdistributions) at the same positions but significantly different distributions of statistical weight might still be judged moderately similar, while changes of the basin positions can make them moderately dissimilar or even dissimilar. This issue will be discussed in concrete terms below in connection with our data.

In addition to our use of the Hellinger distance we compared the Ramachandran distributions of GDG, GFG and the investigated peptides by calculating their conformational entropy. For individual residues it is calculated as

$$S = -R \sum_{\varphi=-\pi}^{\pi} \sum_{\psi=-\pi}^{\pi} P(\varphi, \psi) \ln P(\varphi, \psi) \quad (1)$$

where R is the gas constant. In the absence of any nearest neighbor interactions the total entropy of a peptide would be the sum of individual entropies of the respective residues. If, as expected, nearest neighbor interactions affect the conformational distributions of the investigated peptides, the total conformational entropy of a peptide would reflect correlation effects which can be accounted for by conditional probabilities.⁴⁸ Here we take them into account by comparing the entropy of residue pairs with the individual entropy of residues in GxG and the investigated tetra and pentapeptides. The former was calculated as follows:

$$S_{ij} = -R \sum (P_{ij} \cdot \ln P_{ij}) \quad (2)$$

The sum runs over the components of a row vector that is being calculated as the dot product of the probability matrix P_{ij} of the pair of the residue i and j with a matrix containing the corresponding logarithmic values. The elements of the probability matrix reflect the combined probability of the i^{th} residue having adopted a backbone conformation (φ_i, ψ_i) and the j^{th} residue having adopted a conformation with (φ_j, ψ_j) . This combined probability is written as the Kronecker product of individual residue probabilities:

$$P_{ij}(\varphi_i, \psi_i, \varphi_j, \psi_j) = P_i(\varphi_i, \psi_i) \otimes P_j(\varphi_j, \psi_j) \quad (3)$$

As shown by Baxa et al.,⁴⁸ the absolute value of the entropy depends on the mesh size of the square into which one divides the Ramachandran plot for the purpose of the calculations. For our calculations the mesh size is 2° for φ and ψ . Here, we eliminate this ambiguity by calculating difference between the combined entropy of pairs and the sum of their individual entropies. This provides us with a measure of the entropy change caused by nearest neighbor interactions.

RESULTS AND DISCUSSION

This paper is organized as follows. First, we present the results of the conformational analysis of GDDG and GDDDG, based on our analysis of J -coupling constants and amide I' profiles in IR, Raman, and VCD spectra. In a second step, we report and analyze the temperature dependence of the chemical shift of amide protons which we utilize as a probe of intrapeptide hydrogen bonding. Third, UVCD data of GDDG and GDDDG

are presented to further corroborate rather peculiar result of this analysis. Fourth, we discuss the obtained temperature coefficients of the chemical shift of amide protons. Fifth, we use IR, VCD, and NMR data of GDDG and GDDDG to determine how an aromatic neighbor affects the conformational propensity of D. Sixth, we compare the results of this study with the one of Toal et al.³⁶ and Duitch et al.⁴⁹ who reported changes of the conformational propensities of aspartic acid and aspartate residues induced by NNIs with aliphatic and polar side chains. Finally, we discuss whether findings obtained with fully protonated aspartate side chains are representative of the properties of the corresponding peptides with ionized side chains.

NNIs between Like Residues I: Conformational Analysis of GDDG and GDDDG. We measured the J -coupling constants and amide I' profiles of cationic GDDG and GDDDG as described in the [Materials and Methods](#) and in our previous paper.³¹ A representative two-dimensional NMR spectrum of GDDG is shown in [Figure S1](#). The experimental J -coupling constants are listed in [Table 1](#). The amide I' band profiles of the respective IR, VCD, and Raman spectra are shown in [Figure 2](#). In all these spectra amide I', overlaps with a very intense band at the high wavenumber side which reflects the combined contribution of all CO-stretching modes of the aspartic acid side chains and the C-terminal carboxylate group. The noncoincidence between the respective amide I' peak positions in the IR and Raman spectrum with the former at a lower wavenumber than the latter is indicative of significant conformational sampling of the upper left quadrant of the Ramachandran plot.⁴⁵ The deviation of the VCD data from the zero line below 1600 cm^{-1} reflects a drift of the baseline which could not be totally eliminated with our baseline correction procedure. It does not indicate the presence of a VCD signal.

The refitting of the experimental data of GDG using the $^1J(\text{NC}^\alpha)$ Karplus parameters of Ding and Gronenborn⁴¹ yielded propensities which are slightly different from earlier reported values.²² The pPII population increased from 0.2 to 0.3, while the β -population decreased by the same amount from 0.48 to 0.38. The propensities of the turn-like conformations remained unchanged.

The amide I' band profiles that resulted from this fitting procedure are all shown in [Figure 2](#). The agreement with the experimental profiles is quite satisfactory. It should be noted that the profiles of the CO stretching band above 1700 cm^{-1} has been heuristically modeled by a single Gaussian profile even though contributions from the side chains and the N-terminus might differ regarding their respective wavenumber position. The CO-mode of the C-terminus is generally not VCD active. However, the VCD spectra of GDDG and GDDDG display a negatively biased positive couplet which we just accounted for by two Gaussian profiles of opposite signs. They most likely indicate coupling between the CO-modes of the side chains, which explain the absence of the couplet in the spectrum of GDG.¹⁴

The optimized set of coupling constants and the corresponding reduced χ^2 -values are listed in [Table 1](#). Generally, χ^2 -values below 2.0 are indicative of good fits.⁵⁰ Our fits achieved numbers well below this threshold for D2 of GDDG and all residues of GDDDG. The comparatively large χ^2 -value for D1 of GDDG are due to the deviation between the experimental and computed value for $^3J(\text{H}^{\text{N}}\text{C}')$. A better reproduction of the very low value of this coupling parameter would increase the discrepancy between experimental and computed values for all other 3J coupling parameters. We recently observed for GRRG that the

Table 1. *J*-Coupling Constants of the Indicated Peptides, Determined Experimentally (Exp) from NMR Spectra and Calculated Using Our Gaussian Model (Gauss) (Statistical Errors in Parentheses)

J -coupling [Hz]										
GDDG				GDDDG			GDFG		GFDG	
	D1	D2		D1	D2	D3	D1	F2	F1	D2
$^3J(\text{H}^X, \text{H}^{\alpha})$	7.41 (± 0.03)	7.86 (± 0.05)		7.36 (± 0.03)	7.60 (± 0.05)	7.95 (± 0.03)	7.56 (± 0.03)	7.88 (± 0.07)	6.84 (± 0.06)	7.96 (± 0.06)
Gauss	7.37	7.76		7.35	7.45	7.81	7.65	7.72	6.82	7.73
$^3J(\text{H}^N, \text{C}')$	0.50 (± 0.05)	1.51 (± 0.10)		0.81 (± 0.15)	0.91 (± 0.05)	0.76 (± 0.05)	0.31 (± 0.05)	0.86 (± 0.10)	1.01 (± 0.13)	0.54 (± 0.13)
Gauss	0.77	1.4		1.0	0.80	0.76	0.55	0.97	0.84	0.78
$^3J(\text{H}^{\alpha}, \text{C}')$	2.95 (± 0.07)	2.35 (± 0.34)		2.89 (± 0.07)	1.07 (± 0.27)	2.75 (± 0.13)	3.02 (± 0.13)	1.14 (± 0.13)	2.15 (± 0.07)	2.42 (± 0.27)
Gauss	3.06	2.39		3.01	2.02	2.83	3.07	2.1	2.1	2.36
$^3J(\text{H}^N, \text{C}^{\beta})$	1.49 (± 0.11)	1.10 (± 0.44)		1.32 (± 0.06)	1.39 (± 0.22)	1.54 (± 0.05)	1.37 (± 0.06)	1.1 (± 0.06)	1.93 (± 0.33)	1.60 (± 0.11)
Gauss	1.60	0.95		1.45	1.49	1.41	1.63	1.35	1.82	1.41
$^1J(\text{N}_i, \text{C}_i^{\alpha})$	11.65 (± 0.12)	11.52 (± 0.15)		11.72 (± 0.13)	11.37 (± 0.16)	11.48 (± 0.16)	11.65 (± 0.12)	11.73 (± 0.16)	11.69 (± 0.13)	11.29 (± 0.16)
Gauss	11.67	11.53		11.71	11.35	11.52	11.53	11.74	11.74	11.24
χ^2	3.89	0.91		0.85	0.75	0.94	5.01	2.89	0.2	0.52

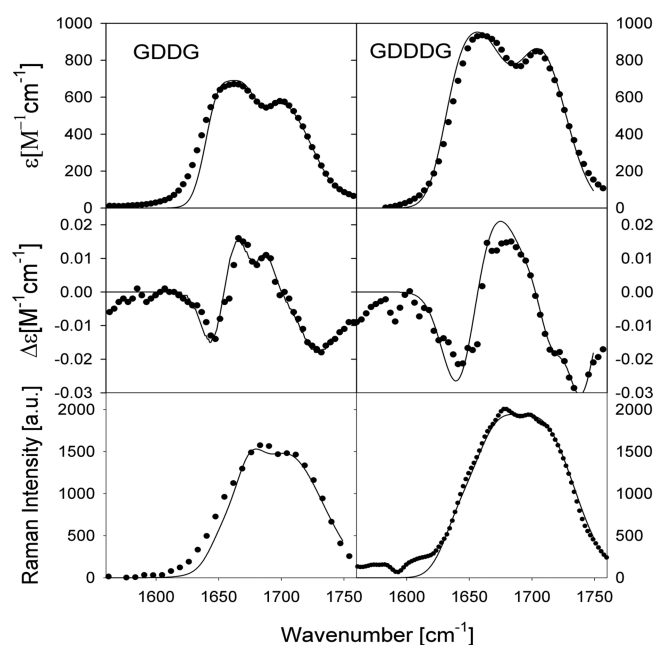


Figure 2. Amide I' profiles (top: IR; middle: VCD and bottom: Raman) for GDDG (left) and GDDDG (right). The experimental data are represented as markers (●) and simulated spectra appear as solid lines.

empirical Karplus curves used for our analysis have difficulties to reproduce experimental values close to their respective minima.³¹

The Ramachandran plots obtained from the fitting procedure are shown in [Figure 3](#).

Table 2 lists the corresponding populations (statistical weights) of D conformations in cationic GDG.¹⁵ A graphic representation can be found in Figure S3. The positions and half-widths of the Gaussian subdistributions are listed in Table 3.

The D residues in GDDG and GDDDG all exhibit relatively high populations for turn-forming structures (β/II_{i+2} , $i\gamma$, and asx-turn). The total population for these turn-like conformations is 0.32 for D in GDG, 0.43 (D1) and 0.27 (D2) in GDDG, and 0.34 (D1), 0.37 (D2), and 0.36 (D3) in GDDDG. All I distributions of the GDDG and GDDDG differ from the one of GDG though to a different extent. Compared with GDG the D1 residue of GDDG shows a significantly higher population of β/II_{i+2} conformations and a lower population for β -strand, while differences between D1 of GDDDG and GDG are not significant. D3 of GDDDG exhibits a reduced asx population and a slightly enriched population of extended conformations (pPII and β -strand). Differences between tripeptide and tetra/pentapeptide Ramachandran plots are most pronounced for the respective D2 residues. Compared with D and D1, D2 of GDDG exhibits a clear dominance of β -strand with a decreased sampling of turn forming conformations and a very weak pPII population. The central residue D2 of GDDDG is peculiar in that it reflects a nearly equal population of pPII, β strand and β/II_{i+2} conformations. There is no population of asx conformations. If the β -turn basin was positioned slight lower on the ψ -axis, the distribution could be interpreted as representing expectations for a random coil supporting ensemble. Our results suggest pronounced NNIs between the aspartic acid residues. The upstream residue depopulates pPII and asx basins while it stabilizes β -strand, whereas the downstream neighbor repopulates pPII and significantly stabilizes β/II_{i+2} conformations.

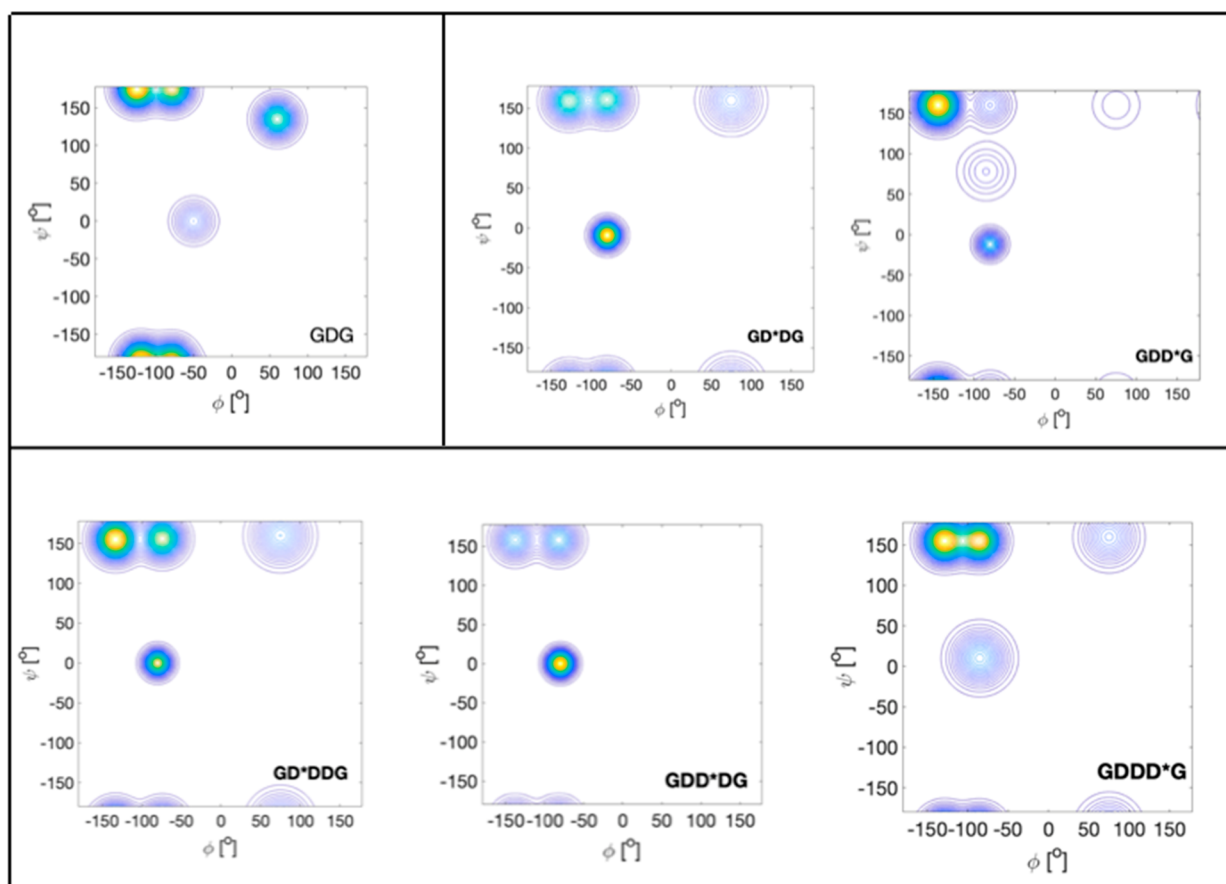


Figure 3. Experimentally determined Ramachandran plots for aspartic acid residues in GDG, GDDG (both in the upper panel), and GDDDG (lower panel). The asterisks indicate the amino acid residue for which the respective plot is shown.

Table 2. Statistical Weight of Gaussian Subdistributions Associated with the Indicated Secondary Structures Obtained from Fits to J -Coupling Constants and Amide I' Band Profiles^a

mole fraction	GDG	GDDG			GDDDG			GFG	GDFG		GFDG	
	D	D1	D2		D1	D2	D3	F	D1	F2	F1	D2
pII	0.30	0.28 ± 0.05	0.12 ± 0.05		0.27 ± 0.05	0.33 ± 0.1	0.31 ± 0.05	0.45	0.55 ± 0.05	0.65 ± 0.1	0.81 ± 0.03	0.36 ± 0.1
β -strand	0.38	0.28 ± 0.05	0.61 ± 0.05		0.38 ± 0.05	0.30 ± 0.1	0.34 ± 0.05	0.45	0.24 ± 0.05	0.35 ± 0.1	0.12 ± 0.03	0.43 ± 0.1
β I/ β I ₂	0.09	0.21 ± 0.03	0.10 ± 0.1		0.13 ± 0.03	0.37 ± 0.05	0.21 ± 0.03	0.05	0	0	0	0.21 ± 0.1
γ	0	0	0.11 ± 0.1		0	0	0	0.05	0	0	0	0.0
asx-turn	0.23	0.23 ± 0.03	0.06 ± 0.05		0.22 ± 0.03	0	0.15 ± 0.03	0	0.21 ± 0.03	0	0.06 ± 0.06	0

^aThe listed errors were estimated by a procedure outlined in the [Materials and Methods](#).

As pointed out in the [Materials and Methods](#), the Hellinger distance provides another tool to compare Ramachandran plots. The provided information is somewhat complementary to the above comparison of mole fractions because the Hellinger distance is more sensitive to changes of the positions of the considered subdistributions. The Hellinger distances relating each D residue in GDDG and GDDDG to the intrinsic reference system D in GDG are listed in [Table 4](#) along with Hellinger distances relating the residues within the same model peptide to each other (for example, relating D1 and D2 in GDDG). The respective values suggest that compared with GDG all residue distributions of the investigated tetra- and pentapeptides are moderately dissimilar with H-values in the upper region of this particular category. For D2, the value is close to the boundary between moderately and very dissimilar. On the contrary, H values of pairs of the same peptide are all in the moderately

similar region. Thus, the Hellinger distances convey indeed a different message from the populations. The reason for this discrepancy becomes obvious from an inspection of the Ramachandran plots. For all residues of GDDG and GDDDG, the ψ -values of pII and the β -strand are at lower values and closer to what one generally sees in a Ramachandran plot. Moreover, the pII and β -strand basins are more clearly separated in the tetra- and pentapeptide than they are in the tripeptide. It is noteworthy in this context that similar observations have been made for other tetrapeptides.^{31,36}

Besides the above-discussed nearest neighbor effects, two aspects of the obtained Ramachandran plots are noteworthy. First, the rather peculiar asx-type structure remains populated in the investigated tetra- and pentapeptide, even though to a different extent for the different residues. The very presence of this conformations leads to very large values for the $^3J(\text{H}^{\alpha}\text{C}')$

Table 3. Positions and Half-Widths of Gaussian Subdistributions Used for the Fitting of *J*-Coupling Constants and Amide I' Profiles of the Indicated Peptides^a

conformation	GDG		GDDG		GDDDG		half-width
	D	D1	D1	D2	D1	D2	
pPII	(-78°, 175°)	(-80°, 161°) ± (2°, 2°)	(-80°, 160°) ± (2°, 2°)	(-74°, 156°) ± (2°, 2°)	(-82°, 158°) ± (2°, 2°)	(-82°, 155°) ± (7°, 2°)	15°
β-strand	(-125°, 175°)	(-128°, 159°) ± (2°, 2°)	(-142°, 160°) ± (2°, 2°)	(-133°, 155°) ± (2°, 2°)	(-138°, 158°) ± (2°, 2°)	(-138°, 155°) ± (7°, 2°)	15°
β/II ₁₊₂	(-50°, 0°)	(-80°, -9°) ± (2°, 5°)	(-80°, -12°) ± (2°, 5°)	(-80°, 0°) ± (10°, 5°)	(-80°, 0°) ± (10°, 5°)	(-80°, 10°) ± (10°, 5°)	10°
inverse γ	(-85°, 78°)	(-85°, 78°)	(-85°, 78°)	(-85°, 78°)	(-85°, 78°)	(-85°, 78°)	20°
asx-turn	(60°, 135°)	(75°, 160°) ± (5°, 5°)	(75°, 160°) ± (5°, 5°)	(75°, 160°) ± (5°, 5°)	(75°, 160°)	(75°, 160°) ± (5°, 5°)	20°
conformation	GFG		GDFG		GFDG		half-width
	F	D1	F2	F1	D2		
pPII	(-74°, 152°)	(-85°, 147°) ± (8°, 2°)	(-83°, 151°) ± (5°, 2°)	(-80°, 151°) ± (2°, 2°)	(-80°, 150°) ± (8°, 2°)		15°
β-strand	(-115°, 120°)	(-115°, 145°) ± (15°, 2°)	(-137°, 151°) ± (5°, 2°)	(-128°, 149°) ± (7°, 2°)	(-127°, 150°) ± (2°, 2°)		15°
β/II ₁₊₂	(-65°, -30°)	(-80°, -9°)	(-80°, -12°)	(-80°, 9°)	(-80°, 12°) ± (2°, 3°)		10°
asx-turn	0	(75°, 160°) ± (10°, 10°)	(75°, 160°) ± (5°, 5°)	(75°, 160°) ± (2°, 2°)	(75°, 160°)		20°
left-handed helix	(-80°, 60°)						10°

^aNote that the same half-widths have been used for φ and ψ . The listed errors were estimated by a procedure outlined in the Materials and Methods.

Table 4. Hellinger Distances for the Indicated Residues of Investigated Tetra- and Pentapeptides^a

GDDG	D1	0.36
	D2	0.39
	D1/D2	0.16
GDDDG	D1	0.35
	D2	0.39
	D3	0.37
	D1/D2	0.13
	D2/D3	0.2
GDFG	D1/D3	0.14
	D1	0.40
	F2	0.30
GFDG	D1/F2	0.14
	F1	0.31
	D2	0.41
	F1/D2	0.14

^aDetails are described in the text.

constant (>2.5 Hz, Table 1), which can only be achieved by a significant population of structures in the right-hand part of the Ramachandran plot. In GDDG, a significant amount of asx population is only present in the plot of D1. In GDDDG, only the terminal residues exhibit asx populations. Our observation strongly suggests an anticooperative effect that prevents the sampling of this conformation in two adjacent residues. This finding is consistent with the absence of this conformation in the Ramachandran plot of the central residue of fully protonated DDD.⁴⁹ Second, it is remarkable that the β I/II₁₊₂ basin is even more populated by all the nonterminal residues of both, GDDG and GDDDG. For D1 of GDDG as well as for D2 and to a lesser extent for D3 of GDDDG the population values are significantly larger than the value observed for GDG. This NNI induced increase of β I/II₁₊₂ seems to be a very D-specific effect, since it is entirely absent in the series of GDyG (y: different non-D guest residues) investigated by Toal et al.³⁶

To further characterize the influence of NNIs in the above peptides, we estimated the difference between the total conformation entropy of the peptide and the sum of the individual entropies of the central residue in GDG. For GDDG we obtained an entropy increase of 1.46 J/mol·K, which produces a Helmholtz free energy contribution of -428 J/mol at room temperature. In GDDDG, however, NNIs reduce the entropy by ca. 2.3 J/mol·K, which corresponds to a Helmholtz energy of 696 J/mol·K.

NNIs between Like Residues II: UVCD Spectra of GDDG and GDDDG. We wondered whether this nearest neighbor induced propensities for turn-forming conformations might be visible in the respective UVCD spectra. Generally, spectra of unfolded peptides exhibit a pronounced negative maximum below 200 nm. A weak positive maximum appears around 215 nm, if the corresponding conformational distributions of residues contain significant pPII content.⁵¹ GDG itself still exhibits a typical statistical coil-like CD spectrum with a relatively weak negative maximum at 195 nm, which reflects the low population of pPII-like conformations.⁵¹ Figure 4 shows the UVCD spectra of GDDG and GDDDG as a function of temperature. The spectra of both peptides differ substantially from the one observed for GDG and, e.g., for the tetrapeptide GDVG, which can be considered a representative of the GDyG peptides investigated by Toal et al.³⁶ (Figure S4). The former do not seem to be too different from the spectra generally obtained

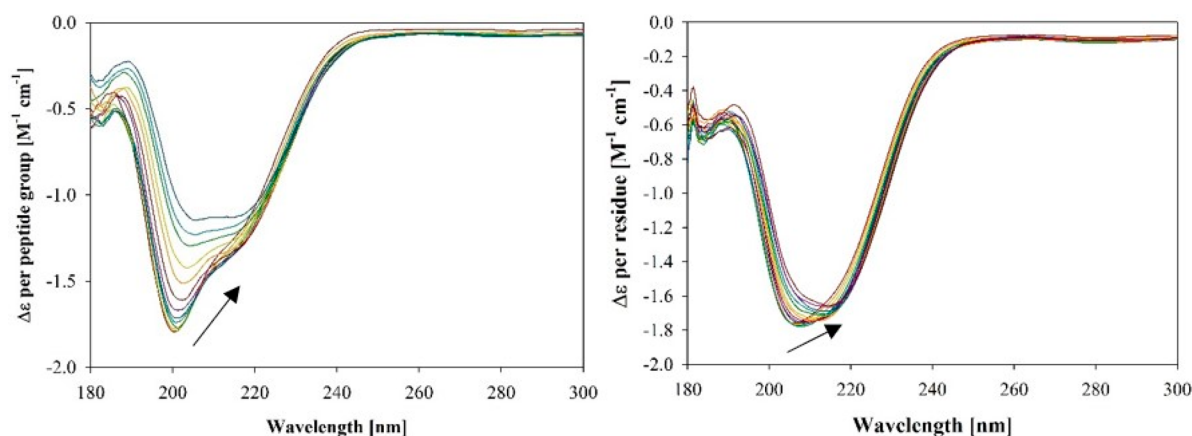


Figure 4. UVCD spectra of GDDG (left) and GDDDG (right) measured as function of temperature between 20 °C (15 °C for GDDDG) and 80 °C in increments of 5 °C. The arrows indicate the change with increasing temperature.

for right-handed helices. In particular the spectra of GDDG are very reminiscent of the type II' β -turn conformation that Hernández et al. observed for the cyclic octapeptide octreotide.⁵² Hence, the CD spectra are fully consistent with the population of these turn conformations inferred from our spectroscopic data. The reason that they dominate the CD spectra of our compounds is the compensation of spectroscopic contributions from pPII and β -strand which already lead to the weak CD signal of GDG.⁵¹ Substantial β -strand sampling as in GDDDG gives rise to a broad negative maximum in the region between 210 and 220 nm which could cause the two negative maxima assignable to the type II' β -turn to merge into a single one. It should also be noted that the CD spectra do not exhibit isodichroic points which distinguishes them from spectra of GxG peptides (including GDG), of GDVG (Figure S4), and of other earlier investigated tri- and tetrapeptides.^{51,53} Their absence suggests that more than two states are involved in the population distribution at higher temperatures. Interestingly, the temperature dependence of GDDG is much more pronounced than the one observed for GDDDG.

NNIs between Like Residues III: Temperature Coefficient of Chemical Shifts. Thus far, we have discussed the results of our structural analysis only in terms of conformational propensities of individual residues and their dependence on nearest neighbors. In view of the peculiar UVCD spectra one might actually wonder whether the investigated peptides are capable of adopting secondary structures on a transient basis. Any formation of a β -turn structure requires that two residues adopt complementary structures at the same time. However, our structure analysis only revealed the population of β I/II_{i+2} conformations. A formation of the latter would require that the D1 residues or alternatively the D2 residue in GDDDG would adopt a conformation with $(\phi, \psi) \sim (-60^\circ, -30^\circ)$ (type I) or $(\phi, \psi) \sim (60^\circ, -120^\circ)$ (type II'). Attempts with the latter did not yield a satisfactory reproduction of our data. The type I option allowed a good reproduction of *J*-coupling constants, but a less than optimal VCD signal. However, the two conformations of the type I residues are actually not that different so it is thinkable if two adjacent residues adopt β I/II_{i+2} simultaneously the resulting structure resembles type I β -turns to some extent. If we ignore the possibility of cooperative effects, the probabilities for such a structure would be 0.02 for GDDG as well as 0.05 (D1D2) and 0.08 (D2D3) for GDDDG. The percentage could be higher if positive cooperativity is operative.

We wondered to what extent combinations of residue structures would be stabilized by hydrogen bonds either between backbone groups or between backbone groups and aspartate side chains. We used the TITAN 1 version of Spartan programs to construct the β - β I/II_{i+2} sequence for GDDG and several turn-forming sequences for GDDDG. To this end, we utilized the coordinates of the respective maxima in the Ramachandran plots in Figure 3. In order to check whether turn-forming conformations derived from our data can facilitate intrapeptide hydrogen bonding, we carried out a semiempirical AM1 optimization with constrained backbone coordinates for all nonterminal residues. For all sequences with asx-turns we also constrained the χ_1 -angle to 180°. Possible hydrogen bonds for these optimized structures are listed in Table S1. While the distances are all in strong bonding regime, the rather large deviations from the ideal 180° geometry should significantly weaken the bonds. Really strong bonds may exist only between NH(3) and C=O(1) in the β - β I/II_{i+2} conformation of GDDG as well as between NH(3) and the oxygen of the OH group of D1 in the in the pPII- β I/II_{i+2}- β I/II_{i+2} conformation of GDDDG. To a slightly lesser extent, the H-bond between NH(4) and O=C(2) in pPII- β I/II_{i+2}- β can also be expected to be comparatively strong. Taken together the numbers in Table S1 suggest that turn-forming conformations in GDDG and particularly in GDDDG are stabilized by a relatively large number of weak hydrogen bonds with significant involvement of peptide-side chain bonds. It should be noted that these low-level calculations solely served the purpose of elucidating the geometry of possible intrapeptide hydrogen bonds. The result does not convey any information about the capability of these interactions to compete with peptide-solvent interactions which are likely to involve strong hydrogen bonds between aspartic acid and water.

Even a transient population of type I/II' turn forming conformations by two consecutive residues would require hydrogen bonding between the amide proton of the C-terminal glycine residue (for GDDG and the (D2D3) option of GDDDG) and of the D3 residue for the (D1D2) option of GDDDG. Such an interaction should cause an increase (decrease of the negative value) of the temperature coefficient of respective chemical shifts. To check for this possibility, we measured the temperature dependence of all NH chemical shifts of GDDG and GDDDG and subjected them to a linear regression using the "linest" function in Excel (cf. Figure S5). The respective temperature coefficients are listed in Table 5.

Table 5. Temperature Coefficients in Units of [ppb/K] for D Residues in GDDG and GDDDG

GDDG			GDDDG			
D1	D2	G (C-term)	D1	D2	D3	G (C-term)
−5.52	−6.8	−5.27	−5.49	−6.89	−5.0	−5.45

Based on a systematic investigation of amide temperature coefficients of turn forming peptides, Dyson and Wright suggested that temperature coefficients higher than -5.5 ppb/K are indicative of some (transient) hydrogen bonding formation.⁵⁴ The numbers in Table 5 reveal, for GDDG, that the amide proton of the C-terminal G falls well into the region for hydrogen bonding. This is in qualitative agreement with the above respective hydrogen bonding analysis (Table S1). For GDDDG, the obtained coefficients indicate that the third D residue (NH(3)) is involved in transient hydrogen bonding. The values for D1 and G are borderline and might still indicate some weak hydrogen bonding. The hydrogen bonding parameters in Table S1 suggest that NH(3) is involved in some strong hydrogen bonding in the pPII- β I/ Π_{i+2} - β I/ Π_{i+2} conformation which has a probability of 0.02. If we assume that this H-bond exists for all β I/ Π_{i+2} - β I/ Π_{i+2} “dimers” irrespective of the conformation of D1, the probability is higher, i.e., 0.077. Hydrogen bonding involving NH(2) is generally weak, which explains its low (very negative) temperature coefficient. NH(4) forms a moderately strong hydrogen bonds with the carbonyl group of the second residue on pPII- β I/ Π_{i+2} - β I/ Π_{i+2} which might be consistent with its borderline temperature coefficient. The borderline coefficient obtained for NH(1) is not fully accounted for by the list in Table S1. However, we should state in this context that depending on its χ_1 -angle aspartic acid can weakly interact with peptide amide groups practically in all conformations. What our data reveal is that β I/ Π_{i+2} conformations add hydrogen bonding interactions between peptide groups which are consistent with the measured temperature coefficients.

It should be noted in this context that the computational work of Porter and Rose showed that the bridge region of the Ramachandran plot which encompasses the considered β I/ Π_{i+2} conformation is more restricted in the unfolded state than generally presented since the absence of intrapeptide hydrogen bonding cannot be compensated by peptide–water interactions due to steric constraints.⁵⁵ We would like to emphasize that the β I/ Π_{i+2} conformations of GDDG and GDDDG lie still in the permissible region of their Ramachandran plot. Only the aspartate residue of GDG adopts a β I/ Π_{i+2} conformation in their forbidden region. However, as indicated in Figure 1, it is stabilized by hydrogen bonding between the residue group so that its population does not conflict with the results of Porter and Rose.

NNIs between Unlike Residues I: Conformational Analysis of GDFG and GFDG. In our earlier study we had provided a detailed analysis on how neighbors with aliphatic (A, V, L) and charged side chains (K) affect the propensity of aspartic acid. We found all these neighbors reduce the turn-forming and increase the pPII fraction of the respective Ramachandran distribution.³⁶ What’s missing is an analysis of how aromatic residues can affect the propensities of D. We opted for phenylalanine as the neighbor since it is a prime candidate for β -strand stabilizing nearest neighbor interactions.^{24,56} With the same approach employed for the structural analysis of GDDG and GDDDG, experimentally determined J -coupling constants

were used to create initial Ramachandran plots for the cationic peptides GDFG and GFDG, which were then optimized by fits to IR and VCD amide I' profiles. We refrained from adding the corresponding Raman spectra to the analyzed data set because of the heavy overlap between the phenylalanine ring mode band and amide I'. Table 1 lists the experimentally obtained and calculated J -coupling values as well as the statistical weights of the considered conformations. The corresponding amide I' IR and VCD profiles are shown in Figure 5. Derivation of the

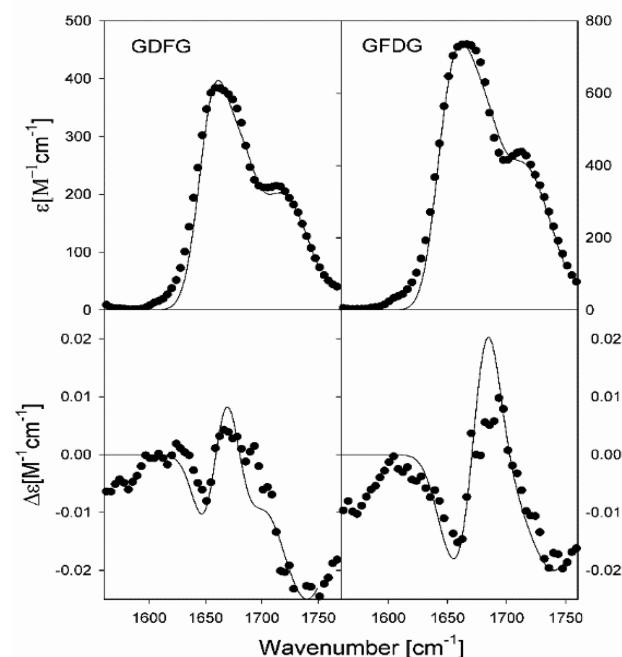


Figure 5. Amide I' IR and VCD band profile of cationic GDFG (left) and GFDG (right). The solid lines represent the simulation of the band profiles with the structural model discussed in the text.

experimental data from the baseline below 1600 cm^{-1} again reflect some yet uncorrected drift of the baseline. For the fitting procedure we utilized the distributions of D and F in GDG and GFG as starting points.^{22,46} They also serve as reference for the discussion of NNIs.

The Ramachandran plots of GDFG and GFDG were compared to GDG and GFG by using the respective statistical weights and Hellinger distances (vide supra). The Ramachandran plots for the nonterminal residues are shown in Figure 6 for both the tri- and tetrapeptides, respectively. Figure S3 visualizes the obtained statistical weights of the considered subdistributions.

For both peptides the J -coupling constants and the amide I' profiles were reasonably well reproduced. The reduced χ^2 -values are somewhat high for GDFG. This can be related again to the very low experimental values of $^3J(\text{H}^{\text{N}}\text{C}')$ and $^3J(\text{H}^{\text{a}}\text{C}')$ obtained for D and F, respectively. These values cannot be fully reproduced with the utilized Karplus curves without totally jeopardizing the results for the other coupling constants and the amide I' profiles. The amide I profiles again contain substantial contributions from the CO stretching mode of side chains C-terminal.

One very notable difference in the Ramachandran plots of the central residues in GDG and GFG is the separation of pPII and β -strand structures adopted by D and the lack of that separation

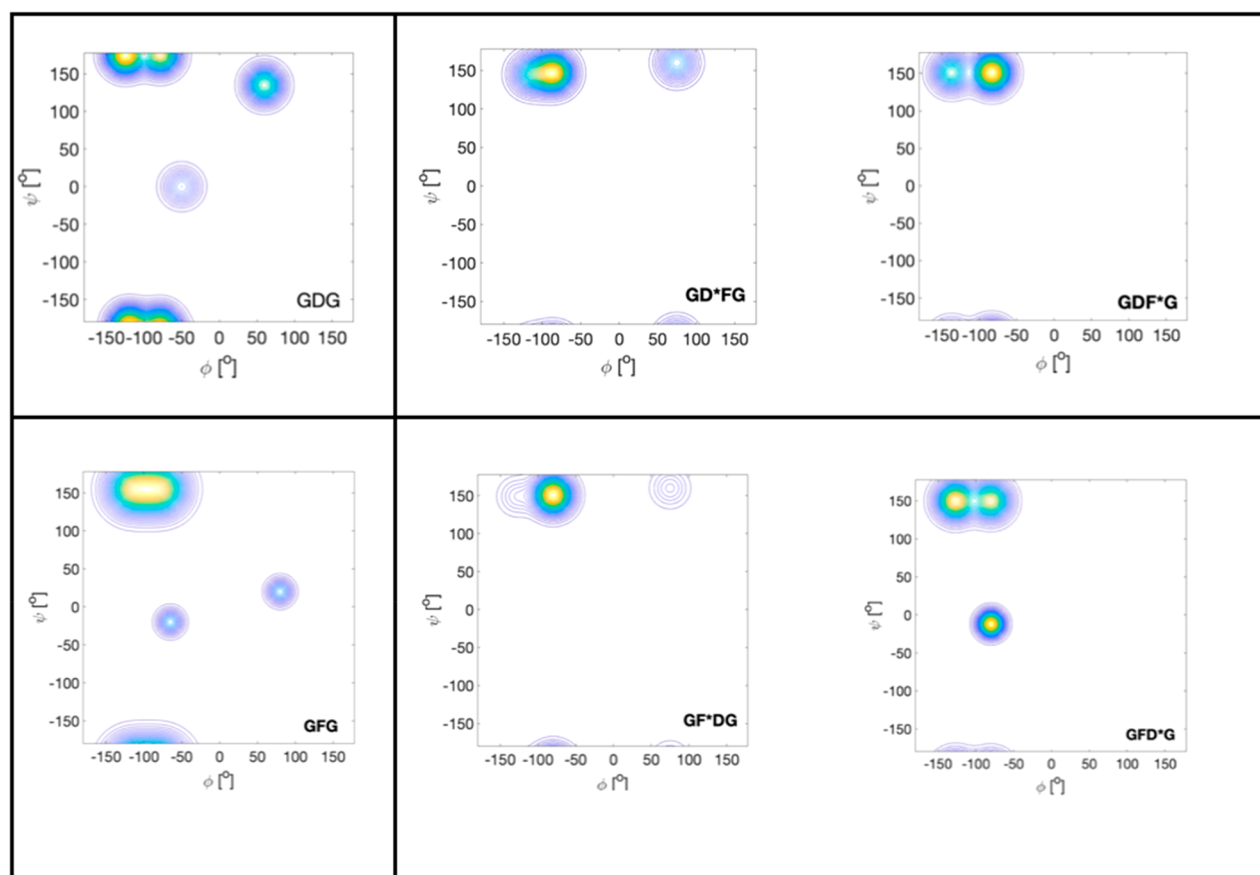


Figure 6. Ramachandran plots of GDG, GDFG (both top), and GFG, GFDG (bottom). The asterisks indicate the residue for which the Ramachandran plot is show.

in F. The inability of F to form α -turns should also be noted. This is not surprising since the phenylalanine side chain does not possess H-bonding capability. The plots in Figure 6 visualize specifics of how NNIs affect the conformational distribution. The statistical weights that are listed in Table 2 and visualized in Figure S7 reveal very surprising results. The corresponding positions and half-widths can be found in Table 3. Apparently, the presence of F at the downstream position makes the distribution of D more normal comparable with the Ramachandran plot of other residues in GxG peptides^{13,39} in that it redistributes population from the β -strand to pPII, which becomes the dominant conformation. This contradicts the expectation that aromatic acid residues shift distributions away from pPII toward the β -strand.⁵⁶ F at the upstream position has a similar effect, but to a lesser extent so that the β -strand remains the dominantly populated conformation. It leaves the β I/II_{i+2} population nearly intact. For F, a D at the downstream position causes a massive shift toward pPII which becomes the dominant conformation. On the contrary, D at the upstream position has a lesser influence on the Ramachandran plot of F, but it still ensures a dominant pPII population. Altogether, the data seem to indicate again a positive correlation between the pPII propensities of the nearest neighbors which we will discuss in more detail below.

The Hellinger distances listed in Table 4 reveal that the Ramachandran plots for D in GDG on one side and of D in GDFG and GFDG are very dissimilar. In this case, this reflects changes of propensities and basin positions. For phenylalanine, the changes are less drastic; the Hellinger distances suggest that

the distributions of F in the tri- and tetrapeptides are moderately dissimilar. Even though propensities are clearly different in the Ramachandran plots of Ds and Fs in the tetrapeptide, the Hellinger distances suggest them to be moderately similar, because the corresponding basin positions are similar.

NNIs in GDFG and GFDG reduce the conformational entropy by 6.8 and 6.05 J/mol·K, which lie way above the values obtained for the investigated homopeptide sequences. The corresponding Helmholtz energies at room temperature are 2.0 and 1.77 kJ/mol.

Unfortunately, the presence of an aromatic residue in GFDG and GDFG rules out the use of the UVCD spectra of these peptides for structure analysis purposes.⁵⁷ We therefore did not measure them for these peptides.

Comparison with Earlier NNI Studies. The occurrence of NNIs and the thus caused violation of the isolated pair hypothesis had been predicted based on bioinformatical and computational results more than 20 years ago.^{19,20,25,28} Our own spectroscopic studies on alanine containing peptide revealed some cooperative effects stabilizing polyproline II conformations in tetraalanine⁵³ and right-handed helical conformations in longer oligoalanine peptides.⁵⁸ Similar findings were reported by Barron and co-workers based on their Raman Optical Activity studies on short oligoalanines.⁵⁹ These experimental results qualitatively confirmed theoretical predictions obtained from MD simulations with a modified Amber 94/MOD force field.^{19,20} In a later study, Toal et al. investigated the influence of downstream residues A, V, K, L, and V on the conformational distribution of D in GDyG tetrapeptides at acidic pH.³⁶ While

the substitution of G by A causes only minor changes, other more sterically demanding side chains cause an increase of pPII at the expense of asx and type I β -turn conformations. Interestingly, the neighbor with the highest intrinsic propensity for β -strand conformations (i.e., valine) produces the largest changes. The influence of valine is even more pronounced regarding the respective Hellinger distance between D in GDG and GDVG, which exceeds 0.5. This indicates that the two Ramachandran distributions are very dissimilar.⁴⁷ On the contrary, the influence of D on V is rather modest and involves only a slight pPII stabilization. As we showed in a more recent paper, the data for GDyG peptides are indicative of an anticooperativity between pPII and β -strand in these peptides.⁶⁰ In other words: the two conformations do not like each other as neighbors.⁶⁰ Thus, NNIs involving D and aliphatic neighbors are clearly distinct from those between adjacent alanine residues which are governed by positive cooperativity.^{19,61} Owing to the considerable involvement of turn-supporting conformations, the NNIs obtained for GDDG and GDDDG are somewhat different and more complex. In both peptides, the D2 residue is mostly affected by its like neighbor(s). In the case of GDDG, β -strand conformations are significantly stabilized at the expense of pPII and turns. Interestingly, the presence of two D neighbors causes a significant stabilization of β I/ β II₊₂ at the expense of asx-turns and to a more limited extent of pPII. This effect is even more pronounced in protonated DDD.⁴⁹

The influence of F neighbors on D residues resembles more the findings for GDyG peptides in that F stabilizes pPII, either at the expense of β -strand (D of GDFG) or turns (D of GFDG). This result shows that F is not always β -strand promoting. The most surprising result obtained for the two F-containing peptides suggests that D as neighbor strongly stabilizes pPII. This is quite different from the influence of D on, e.g., A, V, K, and L where the former mostly affects basin positions and much less conformational propensities.

What are the mechanisms that govern nearest neighbor interactions? The question cannot be fully answered at present. Pappu et al. showed by computational means that residues sampling the right-handed helical region of the Ramachandran plot (type I β -turn included) restrict the conformational space of nearest neighbors (second nearest neighbors in their terminology which follows the one of Flory) by steric effects.²⁸ That could explain the apparent anticorrelation between type I/ β -turn and asx populations in GDDG. However, one wonders why the increased population of type I/ β -turn population in the Ramachandran of D2 in GDDDG affects only the asx population to a significant extent. The theory of Pappu et al. does not account for any cooperativity between β -strand and pPII, which is unlikely caused by any steric effects. Avbelj and co-workers provided compelling computational evidence for the notion that NNIs are caused by changing the solvation of nearest neighbors backbone groups (NH and CO) and side chains.⁵⁶ Their model predicts that increasing solvation causes increased screening of electrostatic interaction between peptide units which normally favors extended β -strand structures. In the presence of a screening solvent, pPII becomes stabilized. The solvation of the backbone itself depends on the solvation free energy of the side chain. A substitution of alanine by, e.g., valine reduces solvation and screening of the respective backbone as well as of the nearest neighbors. This effect is way more pronounced for pPII as it is for the β -strand which explains the obtained anticooperativity between the two conformations.⁶⁰ The role of water is also emphasized regarding the positive cooperativity between the

pPII states of oligo-alanine peptides, which was explained in terms of a water channel around the peptide backbone.^{20,61}

Influence of Side Chain Protonation. As shown in earlier publications, the influence of terminal charges on the conformation of the central residue in GxG peptides is generally rather limited.³⁸ The conformation of the tripeptide AAA has been found to be practically independent of the protonation state of the two termini. The Ramachandran plots of the central residues in GAG and GVG are identical with the ones of the corresponding blocked dipeptides.³⁸ However, the work of Rybka et al. and Dutch et al. seem to indicate that the situation is different for aspartic acid.^{22,49} They found that the Ramachandran plots of the central residue in ionized GDG (GDⁱG) and DDD (DⁱDⁱDⁱ) are significantly different from the corresponding distributions of the respective aspartic acid residues. For GDⁱG, they reported a significant increase of the β -strand and to a lesser extent of pPII population at the expense of asx-turns and β I/ β II₊₂ conformations.²² However, the lack of significant population found for the latter appeared to us as surprising since hydrogen bonding between COO⁻ and NH₃⁺ should still be able to stabilize this particular conformation (Figure 1). We therefore revisited the fits to the amide I' profile and the respective ³J(H^NH ^{α}) value (measured at pH 5.7 to ensure sufficient signal-to-noise for the NMR signal) and found that the distribution reported by Rybka et al. actually overestimated the value of the *J*-coupling constant (8.15 Hz to be compared with the experimental value of 7.47 Hz²²). The only way to reduce the calculated coupling constant without jeopardizing the VCD calculation (which has to produce a vanishing amide I' signal) involved a repopulation of β I/ β II₊₂. A ³J(H^NH ^{α}) value of 7.45 Hz was obtained with the following mole fractions: 0.35 for pPII, 0.45 for β -strand and 0.1 for β I/ β II₊₂ and asx-turns. It should be noted that the fraction of asx-turns was deduced from side chain dependent scalar coupling parameters and used as a fixed value in the fit.²² This result indicates that the deprotonation of the aspartic acid residue causes less drastic changes than reported by Rybka et al. while the revised analysis still indicates gains of β -strand and pPII and a concomitant decrease of the asx-turn population. The newly obtained Ramachandran plot of GDⁱG is shown in Figure S7.

We also revisited the simulations for fully ionized DⁱDⁱDⁱ. It turned out that the very peculiar VCD signal of this compound makes it rather insensitive to any moderate additions of β I/ β II₊₂. The ³J(H^NH ^{α})-value can still be reproduced if we just allow some population transfer from pPII to β I/ β II₊₂. We therefore conclude that the available data for ionized DⁱDⁱDⁱ are insufficient to rule a population of β I/ β II₊₂.

The influence of the terminal groups on the conformational distribution of aspartic acid residues in GDG can be inferred from the structural analysis of the aspartic acid dipeptide (DdP). Rybka et al. showed that its conformational distribution resembles the one of cationic GDG at acidic pH.²² We therefore wondered whether the deprotonation of D alone would cause significant changes of its Ramachandran distribution. To explore the influence of aspartic acid's ionization state on its conformational distribution, we measured and analyzed the IR and VCD profile of ionized DⁱdP in D₂O at pD = 7.1. The measured profiles are shown in Figure 7. The amide I IR profile differs from the one of cationic and GDⁱG in that the amide I wavenumber of the N-terminal peptide group is at much lower wavenumbers in the absence of a charged N-terminus (1630 instead of 1670 cm⁻¹). The VCD signal is weak and barely above the noise level, but it is slightly more pronounced than the earlier obtained

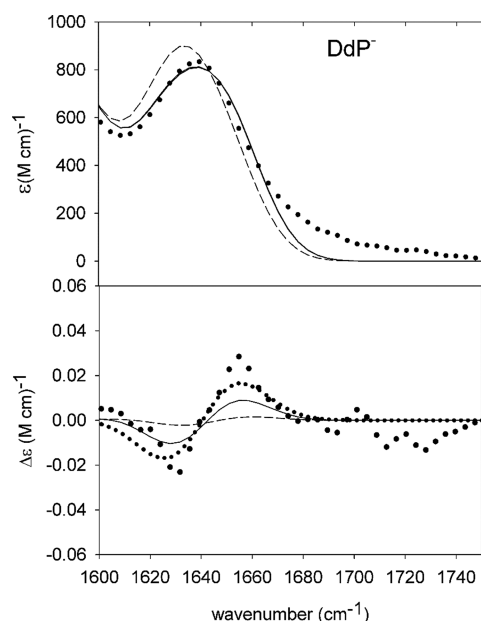


Figure 7. IR and VCD spectrum of the ionized aspartic acid dipeptide in D_2O measured at a pD of 7.1. The solid line represents a simulation of the amide I' band profiles based on the conformational distribution of protonated GDG. The dashed line resulted from a slightly modified simulation described in the text. The band below 1600 cm^{-1} which overlaps with the amide I' in the IR spectrum is assignable to the antisymmetric COO^- stretching vibration.

signal of the protonated dipeptide. In a first step, we simulated the amide I' profiles with the distribution of protonated GDG (DdP). This yielded the solid lines in Figure 7. While the IR profile is well reproduced, the calculation slightly underestimates the VCD signal. The dashed line is a modified simulation for which the calculated VCD agrees with the experimental profile within the limits of spectral noise. We calculated a $^3J(H^N H^{\alpha})$ value of 7.54 Hz. That is slightly below the values observed for ionized GDⁱG (7.69 Hz) but reproduces the values obtained for DdP (7.5 Hz) and for the blocked $G_2D^iG_2$ (7.5 Hz).⁹

The corresponding Ramachandran plot is shown in Figure S7. A comparison of mole statistical weights in Figure S3 reveals that DⁱdP has a slightly higher pPII population. The most decisive difference is the absence of any population of $\beta I/II_{i+2}$. This makes sense because the hydrogen bond between terminal groups that stabilizes this conformation is absent in DⁱdP. α -turns are still populated. This is not surprising since the aspartate side chain is perfectly capable to stabilize α -turns in both protonation states because its functional group serves as a hydrogen bonding donor. Altogether, our results indicate that influence of the side chain charge of DdP is even less pronounced than it is in GDG and DDD. We conclude from this observation that it is mostly the interaction between the terminal groups which causes the difference between the distributions of D in the protonated and ionized state of the side chain of GDG. Apparently, this interaction becomes less relevant for longer oligopeptides where $\beta I/II_{i+2}$ is more likely to be stabilized by hydrogen bonds between peptide groups.

SUMMARY AND CONCLUSIONS

Aspartic acid residues are peculiar because of their frequent occurrence in turn-like conformations. The earlier work of Dyson and co-workers seems to indicate that this reflects a high

intrinsic propensity of this amino acid residue,⁷ but a determination of its intrinsic conformational preferences was only undertaken by Hagarman et al. and subsequently by Rybka et al.^{14,22} Both found that cationic GDG exhibits an above average population of turns in its Ramachandran plot. Adopted structures include α and to a lesser extent $\beta I/II_{i+2}$ conformations. The sampling of these conformations and the low pPII content of its Ramachandran plot distinguish D from aliphatic and aromatic residues and to a lesser extent from residues with other hydrogen bond accepting side chains.¹⁵ In this paper we show that the propensities of D can be significantly modified by its like neighbors. We observe that the latter increases the population of $\beta I/II_{i+2}$ mostly at the expense of α -turns. For D dimers in GDDG, the pPII content of the D2 gets significantly reduced further from its already low value in GDG. The presence of a phenylalanine neighbor reduces the population of turns. Only the D2 residue of GFDG was found to still exhibit a recognizable I/II' - β turn population. In line with the effects caused by downstream aliphatic neighbors F increases the pPII propensity of D. The concomitant increase of the pPII population of F is even more interesting and in fact very surprising. The pPII fraction of 0.81 obtained for F1 in GFDG puts this residue on the same footing with alanine, which is totally unexpected. The results for GDFG and GFDG together with the earlier ones for GDyG peptides shows that the capability of D to initiate and adopt turn conformations in peptides and proteins depends heavily on the character of its neighbors. Figure 8 shows a selection of the most probable conformations of the investigated peptides, i.e., α - β and pPII- β for GDDG, pPII- $\beta I/II_{i+2}$ - β for GDDDG, and pPII-pPII for GDFG.

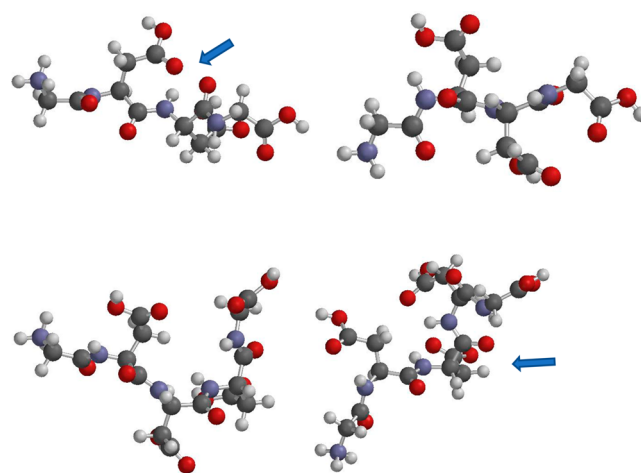


Figure 8. Most probable structures of GDDG, GDDDG, and GDFG in water inferred from the conformational analysis in the text. Upper panel: (left) GDDG (α - β), (right) GDDG (pPII- β). Lower panel: (right) GDFG (pPII-pPII), (left) GDDDG (pPII- $\beta I/II_{i+2}$ - β). The arrows indicate the positions of the turn structures.

The capability to populate $\beta I/II_{i+2}$ conformations of aspartic acid residues can be understood in terms of side chain and backbone solvation. As shown by Avbelj, residues with short side chains reduce the free energy cost of right-handed α -conformation population by up to 6.27 kJ/mol compared with residues with more branched side chains.⁶² The close proximity of $\beta I/II_{i+2}$ and right-handed α -conformations suggests that the number will not be much different for the latter. Our reasoning is

strongly supported by the coil library Ramachandran plot of D in Figure S1, which is actually dominated by $\beta\text{I}/\text{II}_{i+2}$ sampling. An investigation of the crystal structures of a series of blocked tBuCO-PD(N)-NH-Me peptides revealed that both, ionized and protonated D residues have a preference for either type I/II' β -turn ($i+2$) and type II β -turn ($i+2$) conformation with a slightly higher propensity of the ionized residue.⁶³ This observations corroborates the notion that these turn supporting conformations are populated irrespective of the ionization state of the side chain.

Besides functioning as initiation sites for protein folding, aspartic acid residues can be relevant for promoting the formation of compact and thus of residual structures in intrinsically disordered proteins. Turn formations involving aspartic acid were reported for several segments between residue 252 and 348 (called D segments in the following) of the disordered tau protein.⁶⁴ NMR experiments with the disordered 130 residue fragment of *Staphylococcus aureus* fibronectin-binding protein revealed that $^3\text{J}(\text{H}^{\text{H}}\text{H}^{\alpha})$ -values of E, D, and T containing peptides are way lower than the respective GxG- and G₂xG₂-values which could well result from $\beta\text{I}/\text{II}_{i+2}$ population stabilized by hydrogen bonding and nonlocal interaction in compact domains of the protein.^{24,47} The list of short linear motifs (SLIM) by Davey et al. contains a comparatively large number of D, L and K residues.³⁵ Interestingly, most of them have an aliphatic neighbor which can be expected to curtail the turn forming tendencies of D. Only the β -1 subunit of the AF-3 complex contains a D₃ segment, a DD segment appears in the SLIM of sorting nexin-13. Neduva et al. reported the evolution of SLIM segments of three proteins which all contain D residues.³⁴ The D residues of the AP-2 binding sites form the clathrin coat assembly protein are highly evolutionary conserved. Their aliphatic neighbors (I, V) can be expected to eliminate turns and increase pPII propensity, while the conserved GD pair should facilitate turn formation. Another important D-containing segment is the DxxDxxx motif in interaction partners of yeast protein phosphatase 1.³⁴ Motifs differ heavily regarding occupation of the x-residues, these segments can contain aliphatic and hydrophilic residues, but there is quite an impressive number of motifs, where the x-gaps are filled with D residues so that the respective segments contain D₄, D₃, and D₂ segments. We hypothesize that the conformational flexibility and thus the binding affinities of these motifs are heavily regulated by the neighbors of the aspartic acid residues. This subject does certainly deserve further investigations.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.1c06472>.

H bonding connections, Ramachandran plots of aspartic acid residues taken from the coil library of Sosnick and co-workers, E.COSY NMR and NMR spectra of GDDG, mole fractions of Gaussian subdistributions of GDDG, GDDDG, GDFG, and GFDG, UVCD spectra of GDVG measured as a function of temperature, temperature dependence of chemical shifts and $^3\text{J}(\text{H}^{\text{N}}\text{H}^{\alpha})$, and the Ramachandran plots (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Reinhard Schweitzer-Stenner – Department of Chemistry, Drexel University, Philadelphia, Pennsylvania 19026, United States; orcid.org/0000-0001-5616-0722; Phone: 215-895-2268; Email: rs344@drexel.edu

Authors

Bridget Milorey – Department of Chemistry, Drexel University, Philadelphia, Pennsylvania 19026, United States; Present Address: Department of Chemistry and Chemical Biology, Rutgers University, New Brunswick, New Jersey 08854, United States

Harald Schwalbe – Institut für Organische Chemie und Chemische Biologie, Johann Wolfgang Goethe Universität, 60438 Frankfurt, Germany; orcid.org/0000-0001-5693-7909

Nichole O'Neill – Department of Chemistry, Drexel University, Philadelphia, Pennsylvania 19026, United States

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jpcb.1c06472>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Raghad Kurbaj for recording the UVCD spectra of GDVG and Prof. Brigita Urbanc for a critical reading of the manuscript. This project was supported by a grant from the National Science Foundation (MCB-1817650). B.M. acknowledges a travel grant from the International Office of Drexel University to conduct research in Frankfurt, Germany. Work at BMRZ is supported by the state of Hesse.

■ REFERENCES

- (1) Hutchinson, E. G.; Thornton, J. M. A Revised Set of Potentials for B-turn Formation in Proteins. *Protein Sci.* **1994**, *3* (12), 2207–2216.
- (2) Duddy, W. J.; Nissink, J. W. M.; Allen, F. H.; Milner-White, E. J. Mimicry by Asx- and ST-Turns of the Four Main Types of β -Turn in Proteins. *Protein Sci.* **2004**, *13* (11), 3051–3055.
- (3) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963**, *7* (1), 95–99.
- (4) Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. E. Folding of Immunogenic Peptide Fragments of Proteins in Water Solution. I. Sequence Requirements for the Formation of a Reverse Turn. *J. Mol. Biol.* **1988**, *201* (1), 161–200.
- (5) Dyson, H. J.; Wright, P. E. Peptide Conformation and Protein Folding. *Curr. Opin. Struct. Biol.* **1993**, *3* (1), 60–65.
- (6) Dyson, H. J.; Wright, P. E. Insights into the Structure and Dynamics of Unfolded Proteins from Nuclear Magnetic Resonance. *Adv. Protein Chem.* **2002**, *62*, 311–340.
- (7) Chandrasekhar, K.; Dyson, H. J.; Profy, A. T. Solution Conformational Preferences of Immunogenic Peptides Derived from the Principal Neutralizing Determinant of the HIV-1 Envelope Glycoprotein Gp120. *Biochemistry* **1991**, *30* (38), 9187–9194.
- (8) Rose, G. D. Prediction of Chain Turns in Globular Proteins on a Hydrophobic Basis. *Nature* **1978**, *272* (5654), 586–590.
- (9) Shi, Z.; Chen, K.; Liu, Z.; Ng, A.; Bracken, W. C.; Kallenbach, N. R. Polyproline II Propensities from GGXGG Peptides Reveal an Anticorrelation with β -Sheet Scales. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (50), 17964–17968.
- (10) Shi, Z.; Woody, R. W.; Kallenbach, N. R. Is Polyproline II a Major Backbone Conformation in Unfolded Proteins? *Adv. Protein Chem.* **2002**, *62*, 163–240.

- (11) Grdadolnik, J.; Mohacek-Grosec, V.; Baldwin, R. L.; Avbelj, F. Populations of the Three Major Backbone Conformations in 19 Amino Acid Dipeptides. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (5), 1794–1798.
- (12) Eker, F.; Griebenow, K.; Cao, X.; Nafie, L. A.; Schweitzer-Stenner, R. Preferred Peptide Backbone Conformations in the Unfolded State Revealed by the Structure Analysis of Alanine-Based (AXA) Tripeptides in Aqueous Solution. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (27), 10054–10059.
- (13) Hagarman, A.; Measey, T. J.; Mathieu, D.; Schwalbe, H.; Schweitzer-Stenner, R. Intrinsic Propensities of Amino Acid Residues in GxG Peptides Inferred from Amide I' Band Profiles and NMR Scalar Coupling Constants. *J. Am. Chem. Soc.* **2010**, *132* (2), 540–551.
- (14) Hagarman, A.; Mathieu, D.; Toal, S.; Measey, T. J.; Schwalbe, H.; Schweitzer-Stenner, R. Amino Acids with Hydrogen-Bonding Side Chains Have an Intrinsic Tendency to Sample Various Turn Conformations in Aqueous Solution. *Chem. - Eur. J.* **2011**, *17* (24), 6789–6797.
- (15) Hagarman, A.; Mathieu, D.; Toal, S.; Measey, T. J.; Schwalbe, H.; Schweitzer-Stenner, R. Amino Acids with Hydrogen-Bonding Side Chains Have an Intrinsic Tendency to Sample Various Turn Conformations in Aqueous Solution. *Chem. - Eur. J.* **2011**, *17* (24), 6789–6797.
- (16) Ding, L.; Chen, K.; Santini, P. A.; Shi, Z.; Kallenbach, N. R. The Pentapeptide GGAGG Has PII Conformation. *J. Am. Chem. Soc.* **2003**, *125* (27), 8092–8093.
- (17) Shi, Z.; Anders Olson, C.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. Polyproline II Structure in a Sequence of Seven Alanine Residues. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (14), 9190–9195.
- (18) Schweitzer-Stenner, R.; Measey, T. J. The Alanine-Rich XAO Peptide Adopts a Heterogeneous Population, Including Turn-like and Polyproline II Conformations. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (16), 6649–6654.
- (19) Garcia, A. E. Characterization of Non-Alpha Helical Conformations in Ala Peptides. *Polymer* **2004**, *45* (2), 669–676.
- (20) Gnanakaran, S.; Garcia, A. E. Validation of an All-Atom Protein Force Field: From Dipeptides to Larger Peptides. *J. Phys. Chem. B* **2003**, *107* (46), 12555–12557.
- (21) Andrews, B.; Zhang, S.; Schweitzer-Stenner, R.; Urbanc, B. Glycine in Water Favors the Polyproline II State. *Biomolecules* **2020**, *10* (8), 1121.
- (22) Rybka, K.; Toal, S. E.; Verbaro, D. J.; Mathieu, D.; Schwalbe, H.; Schweitzer-Stenner, R. Disorder and Order in Unfolded and Disordered Peptides and Proteins: A View Derived from Tripeptide Conformational Analysis. II. Tripeptides with Short Side Chains Populating Asx and β -Type like Turn Conformations. *Proteins: Struct., Funct., Genet.* **2013**, *81* (6), 968–983.
- (23) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley & Sons: New York, 1970; Vol. 23.
- (24) Penkett, C. J.; Redfield, C.; Dodd, I.; Hubbard, J.; McBay, D. L.; Mossakowska, D. E.; Smith, R. A. G.; Dobson, C. M.; Smith, L. J. NMR Analysis of Main-Chain Conformational Preferences in an Unfolded Fibronectin-Binding Protein. *J. Mol. Biol.* **1997**, *274* (2), 152–159.
- (25) Serrano, L. Comparison between the φ Distribution of the Amino Acids in the Protein Database and NMR Data Indicates That Amino Acids Have Various φ Propensities in the Random Coil Conformation. *J. Mol. Biol.* **1995**, *254* (2), 322–333.
- (26) Avbelj, F.; Baldwin, R. L. Origin of the Neighboring Residue Effect on Peptide Backbone Conformation. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 10967.
- (27) Jha, A. K.; Colubri, A.; Zaman, M. H.; Koide, S.; Sosnick, T. R.; Freed, K. F. Helix, Sheet, and Polyproline II Frequencies and Strong Nearest Neighbor Effects in a Restricted Coil Library. *Biochemistry* **2005**, *44* (28), 9691–9702.
- (28) Pappu, R. V.; Srinivasan, R.; Rose, G. D. The Flory Isolated-Pair Hypothesis Is Not Valid for Polypeptide Chains: Implications for Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (23), 12565–12570.
- (29) Ting, D.; Wang, G.; Shapovalov, M.; Mitra, R.; Jordan, M. I.; Dunbrack, R. L. Neighbor-Dependent Ramachandran Probability Distributions of Amino Acids Developed from a Hierarchical Dirichlet Process Model. *PLoS Comput. Biol.* **2010**, *6* (4), e1000763.
- (30) Toal, S. E.; Kubatova, N.; Richter, C.; Linhard, V.; Schwalbe, H.; Schweitzer-Stenner, R. Corrigendum: Randomizing the Unfolded State of Peptides (and Proteins) by Nearest Neighbor Interactions between Unlike Residues. *Chem. - Eur. J.* **2017**, *23* (71), 18084–18087.
- (31) Milorey, B.; Schweitzer-Stenner, R.; Andrews, B.; Schwalbe, H.; Urbanc, B. Short Peptides as Predictors for the Structure of Polyarginine Sequences in Disordered Proteins. *Biophys. J.* **2021**, *120*, 662–676.
- (32) Sosnick, T. R. Sampling library. <http://godzilla.uchicago.edu/cgi-bin/rama.cgi>.
- (33) Neduva, V.; Russell, R. B. Linear Motifs: Evolutionary Interaction Switches. *FEBS Lett.* **2005**, *579* (15), 3342–3345.
- (34) Neduva, V.; Linding, R.; Su-Angrand, I.; Stark, A.; De Masi, F.; Gibson, T. J.; Lewis, J.; Serrano, L.; Russell, R. B. Systematic Discovery of New Recognition Peptides Mediating Protein Interaction Networks. *PLoS Biol.* **2005**, *3* (12), e405.
- (35) Davey, N. E.; Cowan, J. L.; Shields, D. C.; Gibson, T. J.; Coldwell, M. J.; Edwards, R. J. SLiMPrints: Conservation-Based Discovery of Functional Motif Fingerprints in Intrinsically Disordered Protein Regions. *Nucleic Acids Res.* **2012**, *40* (21), 10628–10641.
- (36) Toal, S. E.; Kubatova, N.; Richter, C.; Linhard, V.; Schwalbe, H.; Schweitzer-Stenner, R. Randomizing the Unfolded State of Peptides (and Proteins) by Nearest Neighbor Interactions between Unlike Residues. *Chem. - Eur. J.* **2015**, *21*, 5173–5192.
- (37) Schweitzer-Stenner, R.; Eker, F.; Huang, Q.; Griebenow, K. Dihedral Angles of Trialanine in D₂O Determined by Combining FTIR and Polarized Visible Raman Spectroscopy. *J. Am. Chem. Soc.* **2001**, *123* (39), 9628–9633.
- (38) Toal, S.; Meral, D.; Verbaro, D.; Urbanc, B.; Schweitzer-Stenner, R. PH-Independence of Trialanine and the Effects of Termini Blocking in Short Peptides: A Combined Vibrational, NMR, UVCD, and Molecular Dynamics Study. *J. Phys. Chem. B* **2013**, *117* (14), 3689–3706.
- (39) Schweitzer-Stenner, R.; Hagarman, A.; Toal, S.; Mathieu, D.; Schwalbe, H. Disorder and Order in Unfolded and Disordered Peptides and Proteins: A View Derived from Tripeptide Conformational Analysis. I. Tripeptides with Long and Predominantly Hydrophobic Side Chains. *Proteins: Struct., Funct., Genet.* **2013**, *81* (6), 955–967.
- (40) Schweitzer-Stenner, R. Distribution of Conformations Sampled by the Central Amino Acid Residue in Tripeptides Inferred from Amide I Band Profiles and NMR Scalar Coupling Constants. *J. Phys. Chem. B* **2009**, *113* (9), 2922–2932.
- (41) Ding, K.; Gronenborn, A. M. Protein Backbone ¹HN-¹³C α and ¹⁵N-¹³C α Residual Dipolar and J Couplings: New Constraints for NMR Structure Determination. *J. Am. Chem. Soc.* **2004**, *126* (20), 6232–6233.
- (42) Wirmer, J.; Schwalbe, H. Angular Dependence of ¹J(Ni,C α i) and ²J(Ni,C α (i-1)) Couplings Constants Measured in J-Modulated HSQCs. *J. Biomol. NMR* **2002**, *23* (1), 47–55.
- (43) Zhang, S.; Schweitzer-Stenner, R.; Urbanc, B. Do Molecular Dynamics Force Fields Capture Conformational Dynamics of Alanine in Water? *J. Chem. Theory Comput.* **2020**, *16* (1), 510–527.
- (44) Krimm, S.; Bandekar, J. Vibrational Spectroscopy of Peptides and Proteins. *Adv. Protein Chem.* **1986**, *38*, 181–364.
- (45) Schweitzer-Stenner, R. Secondary Structure Analysis of Polypeptides Based on an Excitonic Coupling Model to Describe the Band Profile of Amide I' of IR, Raman, and Vibrational Circular Dichroism Spectra. *J. Phys. Chem. B* **2004**, *108* (43), 16965–16975.
- (46) Schweitzer-Stenner, R.; Hagarman, A.; Toal, S.; Mathieu, D.; Schwalbe, H. Disorder and Order in Unfolded and Disordered Peptides and Proteins: A View Derived from Tripeptide Conformational Analysis. I. Tripeptides with Long and Predominantly Hydrophobic Side Chains. *Proteins: Struct., Funct., Genet.* **2013**, *81* (6), 955–967.
- (47) Schweitzer-Stenner, R.; Toal, S. E. Construction and Comparison of the Statistical Coil States of Unfolded and Intrinsically

Disordered Proteins from Nearest-Neighbor Corrected Conformational Propensities of Short Peptides. *Mol. BioSyst.* **2016**, *12* (11), 3294–3306.

(48) Baxa, M. C.; Haddadian, E. J.; Jha, A. K.; Freed, K. F.; Sosnick, T. R. Context and Force Field Dependence of the Loss of Protein Backbone Entropy upon Folding Using Realistic Denatured and Native State Ensembles. *J. Am. Chem. Soc.* **2012**, *134* (38), 15929–15936.

(49) Duitch, L.; Toal, S.; Measey, T. J.; Schweitzer-Stenner, R. Triaspartate: A Model System for Conformationally Flexible DDD Motifs in Proteins. *J. Phys. Chem. B* **2012**, *116* (17), 5160–5171.

(50) Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill: New York, 1969; Vol. 67.

(51) Toal, S. E.; Verbaro, D. J.; Schweitzer-Stenner, R. Role of Enthalpy-Entropy Compensation Interactions in Determining the Conformational Propensities of Amino Acid Residues in Unfolded Peptides. *J. Phys. Chem. B* **2014**, *118* (5), 1309–1318.

(52) Hernández, B.; Coïc, Y. M.; Kruglik, S. G.; Carelli, C.; Cohen, R.; Ghomi, M. Octreotide Used for Probing the Type-II' β -Turn CD and Raman Markers. *J. Phys. Chem. B* **2012**, *116* (31), 9337–9345.

(53) Schweitzer-Stenner, R.; Eker, F.; Griebenow, K.; Cao, X.; Nafie, L. A. The Conformation of Tetraalanine in Water Determined by Polarized Raman, FT-IR, and VCD Spectroscopy. *J. Am. Chem. Soc.* **2004**, *126* (9), 2768–2776.

(54) Merutka, G.; Jane Dyson, H.; Wright, P. E. Random Coil ^1H Chemical Shifts Obtained as a Function of Temperature and Trifluoroethanol Concentration for the Peptide Series GGXGG. *J. Biomol. NMR* **1995**, *5* (1), 14–24.

(55) Porter, L. L.; Rose, G. D. Redrawing the Ramachandran Plot after Inclusion of Hydrogen-Bonding Constraints. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (1), 109–113.

(56) Avbelj, F.; Baldwin, R. L. Origin of the Neighboring Residue Effect on Peptide Backbone Conformation. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (30), 10967–10972.

(57) Woody, R. W. Aromatic Side-chain Contributions to the Far Ultraviolet Circular Dichroism of Peptides and Proteins. *Biopolymers* **1978**, *17* (6), 1451–1467.

(58) Verbaro, D.; Ghosh, I.; Nau, W. M.; Schweitzer-Stenner, R. Discrepancies between Conformational Distributions of a Polyalanine Peptide in Solution Obtained from Molecular Dynamics Force Fields and Amide I' Band Profiles. *J. Phys. Chem. B* **2010**, *114* (51), 17201–17208.

(59) McColl, I. H.; Blanch, E. W.; Hecht, L.; Kallenbach, N. R.; Barron, L. D. Vibrational Raman Optical Activity Characterization of Poly(L-Proline) II Helix in Alanine Oligopeptides. *J. Am. Chem. Soc.* **2004**, *126* (16), 5076–5077.

(60) Schweitzer-Stenner, R.; Toal, S. E. Anticooperative Nearest-Neighbor Interactions between Residues in Unfolded Peptides and Proteins. *Biophys. J.* **2018**, *114* (5), 1046–1057.

(61) Bastida, A.; Zúñiga, J.; Requena, A.; Miguel, B.; Cerezo, J. On the Role of Entropy in the Stabilization of α -Helices. *J. Chem. Inf. Model.* **2020**, *60* (12), 6523–6531.

(62) Avbelj, F. Amino Acid Conformational Preferences and Solvation of Polar Backbone Atoms in Peptides and Proteins. *J. Mol. Biol.* **2000**, *300* (5), 1335–1359.

(63) Mcharfi, M.; Aubry, A.; Boussard, G.; Marraud, M. Backbone Side-Chain Interactions in Peptides - IV. β -Turn Conformations of Asp and Asn-Containing Dipeptides in Solute and Solid States. *Eur. Biophys. J.* **1986**, *14* (1), 43–51.

(64) Mukrasch, M. D.; Markwick, P.; Biernat, J.; Von Bergen, M.; Bernadó, P.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. Highly Populated Turn Conformations in Natively Unfolded Tau Protein Identified from Residual Dipolar Couplings and Molecular Simulation. *J. Am. Chem. Soc.* **2007**, *129* (16), 5235–5243.