Conformational Properties of a Peptide Model for Unfolded α-Helices†

Angela M. Firestine, Veronique M. Chellgren, Shelly J. Rucker, Terrence E. Lester, and Trevor P. Creamer*

Center for Structural Biology, Department of Molecular and Cellular Biochemistry, University of Kentucky, 741 South Limestone Street, Lexington, Kentucky 40536-0509

Received December 19, 2007; Revised Manuscript Received January 23, 2008

ABSTRACT: Models of protein folding often hypothesize that the first step is local secondary structure formation. The assumption is that unfolded polypeptide chains possess an intrinsic propensity to form these local secondary structures. On the basis of this idea, it is tempting to model the local conformational properties of unfolded proteins using well-established residue secondary structure propensities, in particular, α-helix forming propensities. We have used spectroscopic methods to investigate the conformational behavior of a host–guest series of peptides designed to model unfolded α-helices. A suitable peptide model for unfolded α-helices was determined from studies of the length dependence of the conformational properties of alanine-based peptides. The chosen host peptide possessed a small, detectable, α-helix content. Substituting various representative guest residues into the central position of the host peptide at times changed the conformational behavior dramatically, and often in ways that could not be predicted from known α-helix forming propensities. The data presented can be used to rationalize some of these propensities. However, it is clear that secondary structure propensities cannot be used to predict the local conformational properties of unfolded proteins.

Recent years have seen a vast increase in the level of interest in unfolded protein conformational ensembles, and in particular in the presence of local structure (*1–18*). This has been driven in part by the recognition that understanding protein folding will require knowledge of where this process begins (*19*). With this as the goal, a number of groups have focused on unfolded proteins under folding conditions, rather than in the presence of denaturants. The majority of these studies have employed peptide systems that are considered reasonable models for unfolded proteins under folding conditions (*1–10, 12–18*). This work has generally, although not always (*20*), led to the view that unfolded proteins can indeed possess significant, transient local structure. This local structure is not necessarily that adopted by the chain in the final folded form of the protein. In addition, and as might be expected, this local structure is sequence-dependent.

Given the possible presence of transient local structure, it is tempting to employ established secondary structure propensities to generate models for the local conformational behavior of unfolded proteins. α-Helix forming propensities are well-established (*21–26*) and β-sheet propensities less well so (*27–30*). In this work, we explore the conformational properties of a host–guest series of peptides designed to mimic the unfolded ensembles of sequences with tendencies to fold into α-helices. Following a study of the length dependence of α-helix formation by alanine-based peptides, we have selected a host peptide that is too short to form a stable α-helix but long enough to exhibit α-helical tendencies. This host peptide has the sequence Ac-Lys_2-Ala_7-Lys_2-Gly-Tyr-NH_2. The high alanine and lysine content of this peptide imparts a tendency toward α-helix formation (*21*). The central alanine of our host peptide is designated the guest site, into which we have substituted a number of representative residues to determine how the conformational ensemble of the entire peptide is perturbed. Importantly, the guest site is sufficiently far from the ends of the peptide that we believe the behavior of each guest residue is not perturbed by the termini.

We have found that the various guest residues can drastically affect the conformational ensemble adopted by this peptide system. Much of our data can be interpreted in terms of admixtures of right-handed α-helical and left-handed polyproline II (P_{II})[1] helical conformations adopted by individual residues or short stretches of sequence, although other conformations are clearly possible and likely present. Furthermore, the α-helix content of the peptides, as judged by circular dichroism (CD) spectroscopy, does not correlate with the established α-helix forming propensities of the guest residues (*21*). Our data therefore indicate that individual

† This work was funded by grants to T.P.C. from the National Science Foundation (MCB-0444049) and the Petroleum Research Fund administered by the American Chemical Society (43138-AC4). The Jasco J-810 spectrapolarimeter employed in this work was purchased using funds from a NSF-EPSCoR infrastructure grant to the University of Kentucky.

* To whom correspondence should be addressed. Phone: (859) 323-6037. Fax: (859) 257-2283. E-mail: Trevor.Creamer@uky.edu.

[1] Abbreviations: CD, circular dichroism; HPLC, high-pressure liquid chromatography; P_{II}, polyproline II; peptide A5, Ac-Lys_2-Ala_5-Lys_2-Gly-Tyr-NH_2; peptide A7, Ac-Lys_2-Ala_7-Lys_2-Gly-Tyr-NH_2; peptide A9, Ac-Lys_2-Ala_9-Lys_2-Gly-Tyr-NH_2; peptide A11, Ac-Lys_2-Ala_{11}-Lys_2-Gly-Tyr-NH_2; peptide F, Ac-Lys_2-Ala_3-Phe-Ala_3-Lys_2-Gly-Tyr-NH_2; peptide G, Ac-Lys_2-Ala_3-Gly-Ala_3-Lys_2-Gly-Tyr-NH_2; peptide I, Ac-Lys_2-Ala_3-Ile-Ala_3-Lys_2-Gly-Tyr-NH_2; peptide L, Ac-Lys_2-Ala_3-Leu-Ala_3-Lys_2-Gly-Tyr-NH_2; peptide N, Ac-Lys_2-Ala_3-Asn-Ala_3-Lys_2-Gly-Tyr-NH_2; peptide P, Ac-Lys_2-Ala_3-Pro-Ala_3-Lys_2-Gly-Tyr-NH_2; peptide Q, Ac-Lys_2-Ala_3-Gln-Ala_3-Lys_2-Gly-Tyr-NH_2; peptide S, Ac-Lys_2-Ala_3-Ser-Ala_3-Lys_2-Gly-Tyr-NH_2; peptide V, Ac-Lys_2-Ala_3-Val-Ala_3-Lys_2-Gly-Tyr-NH_2.

10.1021/bi702474k CCC: $40.75
Published on Web 02/12/2008

α-helix forming propensities cannot be used to describe unfolded state ensembles. Given the sequence context dependence of β-sheet propensities (*28*), it seems highly likely that these also cannot be used to describe unfolded ensemble behavior.

MATERIALS AND METHODS

All peptides used in this work were obtained from the W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University (New Haven, CT). They were purified to 95% homogeneity by reverse-phase HPLC, and their identities were confirmed using mass spectrometry. Stock solutions were prepared by dissolving peptides in a buffer containing 5 mM potassium phosphate, 5 mM sodium fluoride, and 0.02% sodium azide, with the pH adjusted to 7. Peptide concentrations were determined using the method of Brandts and Kaplan (*31*). Absorbance was measured in a 1.0 cm path length cuvette in a Beckman-Coulter DU 640B spectrophotometer.

Circular dichroism (CD) spectra were measured with a Jasco J-810 spectrapolarimeter employing a 1 mm path length quartz cuvette, with the temperature controlled by a Peltier heating block. Solutions contained approximately 100 μM peptide in 5 mM phosphate buffer (pH 7). Conditions for measuring spectra were as previously described (*1*). Spectra for each peptide were collected at least three times, with errors in the ellipticities estimated to be 3%.

A series of peptides with the sequence Ac-Lys_2-Ala_n-Lys_2-Gly-Tyr-NH_2 (n = 5, 7, 9, and 11) were examined using circular dichroism (CD) spectroscopy. These peptides were named A5, A7, A9, and A11, respectively. The lysine residues in the peptides were included to maintain solubility. The termini of all peptides were blocked to remove unwanted electrostatic interactions, and the Gly-Tyr pair was included to facilitate concentration determination (*31*). Data obtained from these studies were used to determine a reasonable length to be used as a model for unfolded α-helices.

Host–guest peptides studied in this work were based on the sequence Ac-Lys_2-Ala_3-Xaa-Ala_3-Lys_2-Gly-Tyr-NH_2, with the identity of the guest residue Xaa being varied. The host peptide (A7) has an alanine in the guest position. We examined the guest residues asparagine, glutamine, glycine, valine, isoleucine, leucine, phenylalanine, proline, and serine. Each peptide, except the host, is named on the basis of the one-letter code for the guest residue. For example, the peptide with a leucine guest residue is called peptide L in the host–guest studies.

CD data were analyzed using CDPro (*32*) which includes the CDsstr (*33*), Contin/LL (*34*), and SELCON3 (*35*) programs. Analyses were performed using the SP37A protein reference set since this allows for analysis of α-helix, β-strand, turn, P_{II} helix, and unordered segment content. α-Helix content was also estimated from the per residue molar ellipticity measured at 222 nm, $[\theta]_{222}$, using the equation (*36, 37*)

$$f_\alpha = \frac{[\theta]_{222} - [\theta]_R}{[\theta]_\alpha - [\theta]_R} \quad (1)$$

where f_α is the fractional α-helix content, $[\theta]_R$ is the ellipticity at 222 nm for a disordered chain, and $[\theta]_\alpha$ is the ellipticity at 222 nm for 100% α-helix content for a chain with n backbone amides. $[\theta]_\alpha$ is obtained from an equation derived by Chen et al. (*36*):

$$[\theta]_\alpha = [\theta]_\infty(1 - k/n) \quad (2)$$

where $[\theta]_\infty$ is the per residue molar ellipticity for a chain of infinite length with maximal α-helix content and k is an empirically determined factor interpreted to be the average number of amides in the chain that cannot participate in α-helix formation. There is general agreement for the $[\theta]_\infty$ value of −40000 deg dmol^{-1} cm^2, but the value of k remains in dispute. Estimated values of k have ranged from 2.5 (*21*) to 4.3 (*38*), and even up to 6.3 (*39*). In this work, we have employed the commonly used value of 4 (*37*).

In addition, there is no consensus about the correct value for $[\theta]_R$ in eq 1. For example, $[\theta]_R$ has been estimated to be 0 deg dmol^{-1} cm^2 (*36, 37*), 640 deg dmol^{-1} cm^2 (*40*), and −3000 deg dmol^{-1} cm^2 (*38*). The estimate 0 deg dmol^{-1} cm^2 (*36, 37*) was employed in our analyses.

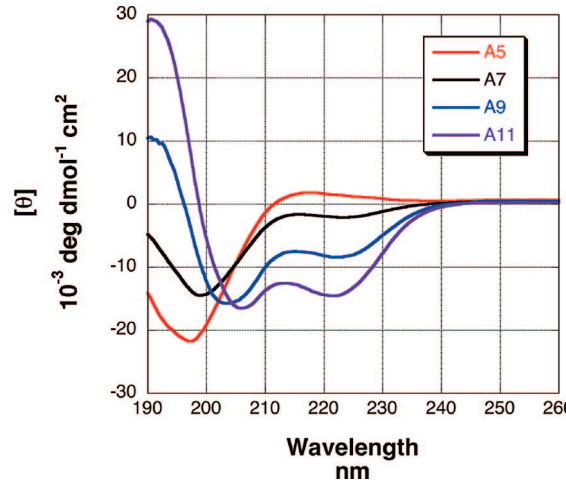

FIGURE 1: CD spectra for peptides A5, A7, A9, and A11 in 5 mM phosphate buffer (pH 7) at 5 °C.

RESULTS

Determination of Host Peptide Length. Peptides of the sequence Ac-Lys_2-Ala_n-Lys_2-Gly-Tyr-NH_2 (n = 5, 7, 9, and 11) were examined to determine a reasonable length for a host peptide. These peptides were named A5, A7, A9, and A11, respectively. Our criteria for the host peptide were that it be too short to form a stable α-helix but be the minimum length necessary to possess a detectable α-helical tendency. We believe these criteria define a reasonable model for an unfolded α-helix.

CD spectra of these peptides in 5 mM phosphate buffer (pH 7) at 5 °C are shown in Figure 1. It is clear from these spectra that there is a growing tendency toward α-helix formation as the number of alanines in the peptides is increased. The spectrum for peptide A5 is suggestive of a high P_{II} helix content. The weak positive band at 218 nm and strong negative band at 195 nm are diagnostic of this conformation (*2, 41*). At the other end of the length scale that was examined, peptide A11 possesses a CD spectrum indicative of significant α-helix content (*37*).

The secondary structure contents of the peptides were determined from the CD spectra at 5 °C using the programs CDsstr (*33*), SELCON3 (*35*), and Contin/LL (*34*) in CDPro (*32*). One should note that these programs attempt to fit

Table 1: Estimated Secondary Structure Contents for Peptides A5, A7, A9, and A11 Determined from the CD Spectra Collected at 5 °C (Figure 1)[a]

peptide	program	α-helix (%)	β-strand (%)	turns (%)	P_{II} (%)	unordered (%)	nrmsd[c]
A5[b]	CDsstr	0	23	15	13	48	0.159
	SELCON3	15	23	12	8	43	−
	Contin/LL	10	18	12	15	55	0.034
A7[b]	CDsstr	4	21	14	13	48	0.099
	SELCON3	12	19	13	10	44	−
	Contin/LL	9	16	11	13	52	0.039
A9	CDsstr	40	16	9	10	25	0.238
	SELCON3	32	8	13	8	36	0.280
	Contin/LL	40	3	10	11	37	0.042
A11	CDsstr	48	9	11	7	25	0.072
	SELCON3	38	7	10	6	39	0.214
	Contin/LL	60	24	0	3	13	0.061

[a] Estimates were obtained from the programs CDsstr (*33*), Contin/LL (*34*), and SELCON3 (*35*) in CDPro (*32*). Analyses were performed using the SP37A protein reference set. [b] The SELCON3 program did not converge for these peptides, and the listed estimates should be considered unreliable at best. [c] The nrmsd is the normalized-root-mean-square deviation of the fit between the calculated CD spectrum and the experimental data.

Table 2: α-Helix Contents Estimated from eqs 1 and 2 for Peptides A7, A9, and A11 Determined from the Per Residue Molar Ellipticity at 222 nm from CD Spectra Collected at 5 °C (Figure 1)

peptide	$[\theta]_{222}$ (deg $dmol^{-1}$ cm^2)	$[\theta]_\alpha$[a] (deg $dmol^{-1}$ cm^2)	n[b]	α-helix (%)
A7	-2092 ± 65	−27700	13	8
A9	-8432 ± 253	−29300	15	29
A11	-14572 ± 437	−30600	17	48

[a] Estimated from eq 2 using a $[\theta]_\infty$ value of −40000 deg $dmol^{-1}$ cm^2 and a k of 4 and rounded to the nearest 100. [b] Number of amides in the peptide backbone, including acetyl and amide blocking groups.

protein CD data to the input experimental data. Thus, they may not provide accurate secondary structure estimates for peptide CD data. The results of our analyses are listed in Table 1. The SELCON3 analyses for peptides A5 and A7 did not converge, so the estimated secondary structure contents for these should be considered unreliable at best. The three CD analysis programs provided somewhat variable results, with the Contin/LL estimates generally giving the best fits to the experimental data, as judged by the normalized-root-mean-square deviations [nrmsd (Table 1)]. General trends appear to be that α-helix content increases with peptide length, β-strand, P_{II} helix, and unordered contents decrease by varying amounts, and the estimated turn content remains somewhat constant. Notably, there is a large increase in the estimated α-helix content between peptides A7 and A9, consistent with the significant strengthening of the negative band at 222 nm seen in Figure 1.

We also estimated the α-helix content of peptides A7, A9, and A11 from $[\theta]_{222}$ using eqs 1 and 2 (Table 2). Peptide A5 possesses significant positive ellipticity at 222 nm and could not be analyzed using this approach. In this analysis, it is assumed that the only contribution to the negative ellipticity observed at 222 nm is from any α-helix present in the peptide. The obtained estimates agree reasonably with those obtained with CDPro (Table 1). Again, we estimate a significant increase in α-helix content between peptides A7 and A9.

The CD spectrum for peptide A7 in dilute phosphate buffer at 5 °C is shown in Figure 2a along with a difference

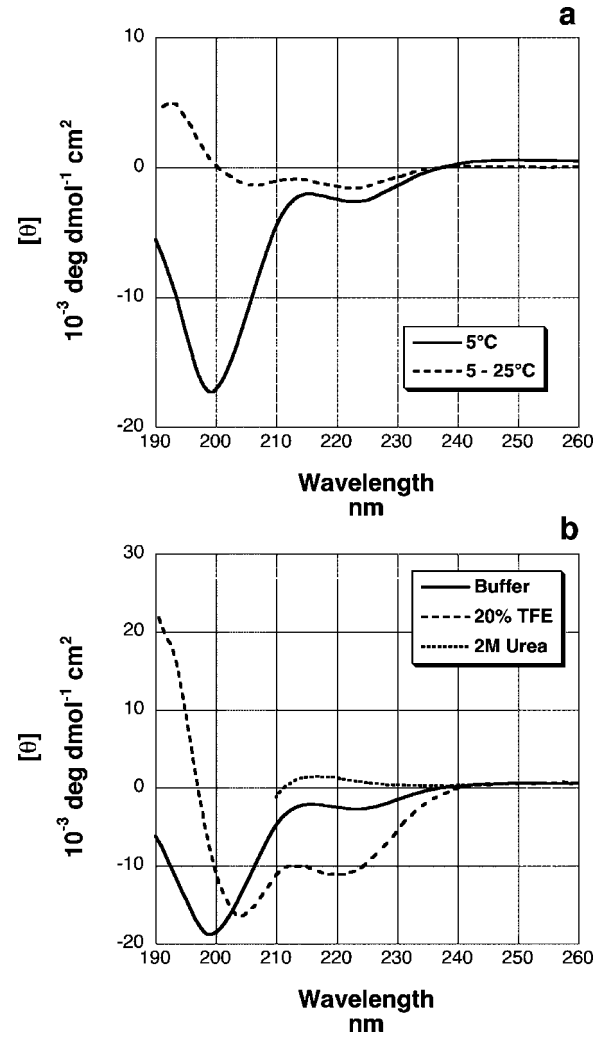

FIGURE 2: (a) CD spectrum collected for peptide A7 in 5 mM phosphate buffer at 5 °C (—) and difference spectrum generated by subtracting a 25 °C spectrum from the 5 °C spectrum (---). (b) CD spectra collected for peptide A7 at 5 °C in phosphate buffer (—), 20% (w/v) TFE (---), and 2 M urea (•••).

spectrum generated by subtracting a spectrum collected at 25 °C from that collected at 5 °C. The difference spectrum in Figure 2a is clearly that of an α-helix with characteristic negative bands at 222 and 208 nm. This confirms that peptide A7 does indeed possess detectable α-helix content. By increasing the temperature, we have altered the conformational manifold adopted by peptide A7 in such a way that local α-helical conformations are disfavored.

To further probe the conformational properties of peptide A7, the cosolvent 2,2,2-trifluoroethanol (TFE) was added at a concentration of 20% (w/v). The resulting CD spectrum is shown in Figure 2b. The increase in strength of the weak negative band at 222 nm, and the shift in wavelength of the stronger negative band from 198 to 204 nm, is indicative of increased α-helix content. The addition of 2 M urea has a very different effect upon peptide A7 (Figure 2b). The negative band at 222 nm is abolished in favor of a weak positive band at 217 nm. This weak positive band is due to the presence of left-handed polyproline II (P_{II}) helix content, which is favored by the presence of urea (*1*).

Host–Guest Peptide Studies. On the basis of the results of our peptide length studies, host–guest peptides of the sequence Ac-Lys_2-Ala_3-Xaa-Ala_3-Lys_2-Gly-Tyr-NH_2, with

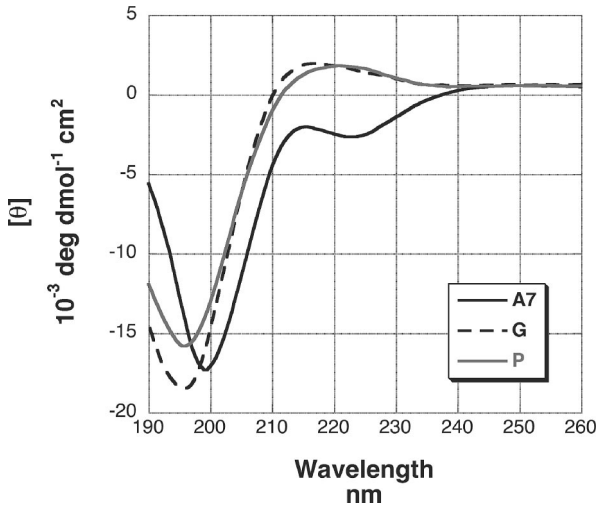

FIGURE 3: CD spectra collected for peptides A7, G, and P in 5 mM phosphate buffer at 5 °C.

Xaa denoting the guest site, were employed as a model system for the unfolded states of a polypeptide sequence that is destined to fold into an α-helix. The host peptide, peptide A7, corresponds to the sequence with an alanine in the guest position. This peptide has a tendency toward α-helix formation but is too short to form a stable helix. Importantly, it is long enough that the behavior of the central guest residue is unlikely to be perturbed by the peptide termini. Host–guest peptides are named using the single-letter code for the guest residue, with the exception of the host peptide A7.

We limited our studies to the guest residues asparagine, glutamine, glycine, valine, isoleucine, leucine, phenylalanine, proline, and serine. We believe these residues are representative of the noncharged residues. We decided not to study charged guest residues because they would interact strongly with the flanking lysine residues. We also opted to exclude tyrosine and tryptophan from our study because of their well-known effects upon CD spectra that complicate subsequent analysis (*42*).

In previous work, we had demonstrated that the programs included in CDPro do not provide accurate estimates of P_{II} helix content for peptides (*43*). Indeed, an oligopeptide consisting of 14 prolines, which will predominantly form P_{II} helix (*44–47*), was estimated to contain just 14% P_{II} helix. The positive band around 217 nm for peptide A5 suggests that it possesses significant P_{II} helical content, yet the estimates in Table 1 range from just 8 to 15%. The inability of these programs to accurately estimate P_{II} helix content is most likely a result of the use of protein CD data to fit the input peptide data. The accuracy of the estimates of other secondary structures when using CDPro to analyze peptide data is also unclear. In addition, α-helix content estimates obtained from eqs 1 and 2 are made assuming that no other conformations contribute significantly to the measured ellipticity at 222 nm. As will be seen, this assumption is likely incorrect for the peptides studied here. Consequently, we have not attempted to estimate secondary structure contents for the host–guest peptides using either CDPro or eqs 1 and 2.

Figure 3 depicts the CD spectra collected at 5 °C for the peptides with the strongest α-helix breakers as guests, peptides G and P. As one might expect from their respective α-helix forming propensities (*21*), the spectra for peptides G and P show no evidence for α-helix content. Notably, these two peptides appear to have significant P_{II} helix contents as evidenced by positive bands at 216 and 221 nm and strong negative bands at 195 and 196 nm, respectively. These are the hallmarks of P_{II} helix formation by sequences with little or no proline content (*1, 41*). Note that the positions of these bands are influenced by the presence of one or more prolines (*41, 48*).

CD spectra collected at 5 °C for the peptides with apolar guest residues, A7, L, I, V, and F, are shown in Figure 4a. Notably, peptides A7, L, I, and V all possess a weak negative band around 222 nm indicative of α-helix content. In contrast, peptide F possesses a weak positive band at 217 nm, suggesting the presence of the P_{II} helical conformation. The corresponding 5–25 °C difference spectra given in Figure 4b confirm the presence of α-helix in peptides A7, L, I, and to a lesser extent V. The *β*-branched residues have low α-helix propensities, valine having the lowest after glycine and proline (*21*). Nonetheless, peptide V shows evidence of some α-helix content even in this host peptide system that is a weak α-helix former as seen by the difference spectrum given in Figure 4b. Remarkably, peptide I appears to have a similar, possibly even higher, α-helix content compared to that of peptide L (Figure 4b).

The spectra collected for peptides with polar guest residues, S, N, and Q, are given in Figure 4c, with the 5 °C – 25 °C difference spectra given in Figure 4d. The spectrum for peptide A7 is included for reference. Peptide Q possesses a very weak negative band at ∼222 nm (Figure 4c), indicating the possible presence of α-helix. However, inspection of the difference spectrum reveals more complex behavior (Figure 4d). In fact, the difference spectrum for peptide Q appears to be similar to the 5 °C spectrum, suggesting that the various conformations populated by this peptide at the lower temperature are altered in a somewhat uniform manner when the temperature is increased to 25 °C. On the basis of the addition of TFE or urea (data not shown), and earlier work in which we have demonstrated that glutamine has a very high P_{II} helix forming propensity (*43, 49, 50*), we conclude that peptide Q differs from peptide A7 in that it appears to have a higher P_{II} content and a lower α-helix content.

Peptides N and S possess somewhat surprising CD spectra (Figure 4c). The peptide N spectrum has a broad positive band centered at 220 nm and a strong negative band at 195 nm, suggesting significant P_{II} helix content. This is surprising because we had previously demonstrated that asparagine has a very low P_{II} helix forming propensity (*43, 49, 50*). It is notable that asparagine and glutamine guests lead to very dissimilar peptide behavior (Figure 4c,d), despite the fact that these two residues differ by a single methylene in their side chains. The peptide S spectrum also suggests P_{II} helix content (Figure 4c). The magnitude of the P_{II} helix-like signal decreases with an increase in temperature (Figure 4d). Serine has a relatively low P_{II} helix forming propensity (*50*).

On the basis of the observed difference spectra (Figure 4), and assuming the α-helix contents are decreased at similar rates upon heating from 5 to 25 °C, we would predict our peptides have apparent α-helix contents in the following order: A7 > I and L > V and Q > F and S > G and P > N. The corresponding α-helix propensities for the guest residues are in the following order: alanine > leucine >

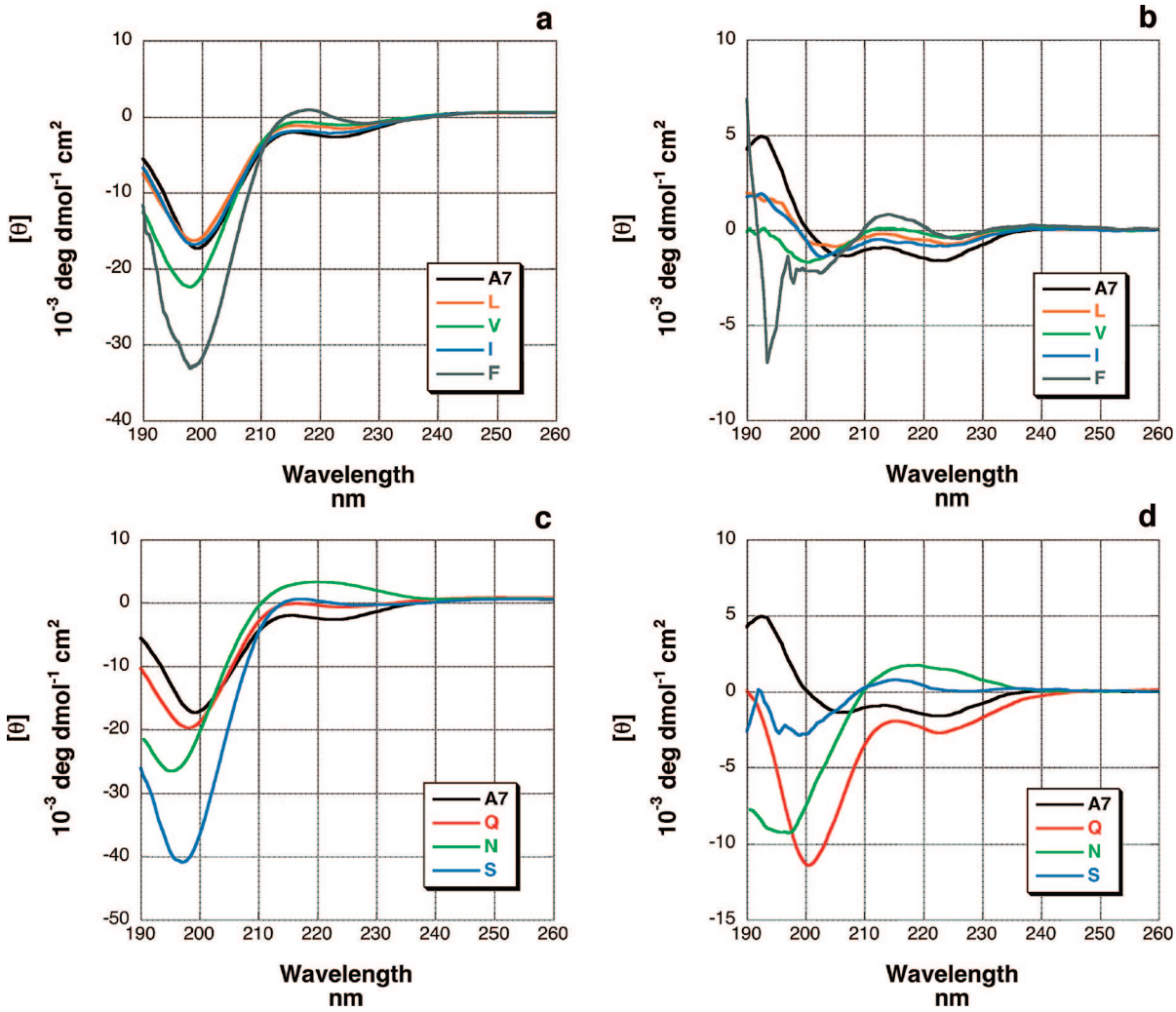

FIGURE 4: (a) CD spectra collected for peptides A7, L, V, I, and F at 5 °C. (b) Difference (5 °C − 25 °C) spectra for peptides A7, L, V, I, and F. (c) CD spectra collected for peptides A7, Q, N, and S at 5 °C. (d) Difference (5 °C − 25 °C) spectra for peptides A7, Q, N, and S.

glutamine > isoleucine > serine > asparagine > phenylalanine > valine > glycine ≫ proline (*21*). The correlation between known α-helix propensities and observed behavior of our peptide system appears weak at best.

DISCUSSION

We have chosen to study a model for unfolded α-helices because the α-helix forming propensities of residues are well-determined as a result of extensive peptide (*21–24*) and protein studies (*25, 26*). We limited our studies to the guest residues asparagine, glutamine, glycine, valine, isoleucine, leucine, phenylalanine, proline, and serine because we consider these to be representative of the noncharged residues. We did not examine charged guest residues because they would interact strongly with the flanking lysine residues used to solubilize the peptides. Furthermore, we opted to exclude tyrosine and tryptophan from our study because of their well-known effects upon CD spectra that complicate subsequent analysis (*42*).

Choice and Characterization of the Host Peptide. We chose peptide A7 as our host peptide on the basis of the length dependence of α-helix formation shown in Figure 1 and Tables 1 and 2. Given the significant α-helix contents estimated for A9 and A11 (Tables 1 and 2), we felt that these peptides would not make good models for unfolded α-helices. Peptide A5, on the other hand, has positive ellipticity around 222 nm, suggesting significant P_{II} helix content and very low α-helix content (Figure 1 and Tables 1 and 2). Peptide A7 displays negative ellipticity around 222 nm, indicating a low but detectable α-helix content (Figure 1), making it a more suitable model for an unfolded α-helix under folding conditions.

The criteria for choosing peptide A7 as the host peptide were that it should be too short to be capable of folding into a stable α-helix while at the same time be the minimum length necessary to be able to detect the α-helical tendency. The CD data shown in Figure 2 confirm that the detectable conformational properties of peptide A7 meet these criteria. This peptide is also sufficiently long that the behavior of the central residue should be similar to that of a residue in a much longer polypeptide chain. If the C-terminal Gly-Tyr pair is not counted, there are five residues on each side of the guest site, which should be sufficient to account for most local intrachain interactions.

It is notable that peptide A7 appears to have detectable P_{II} helix content in addition to a weak α-helix tendency (Figure 2) (*1*). It has been shown previously that alanine and, to some extent, lysine have significant P_{II} helix forming

propensities (*2, 3, 5, 49, 51*). These propensities appear to arise from a combination of favorable backbone solvation and minimization of backbone steric clashes (*8, 9, 52*). These are precisely the conditions one would expect for an unfolded protein under folding conditions. We feel then that our host peptide A7 is a good model system for unfolded protein sequences that are destined to fold into α-helices.

Pappu and co-workers (*53, 54*) have recently published calculations that indicate local conformational preferences are confined to stretches of approximately seven to eight residues. This length estimate does not appear to vary with sequence. Our host peptide, A7, is a total of 13 residues in length. The influence of the C-terminal Gly-Tyr pair upon the conformational properties of this peptide is likely very small, if not negligible. This leaves us with an "effective" length of 11 residues, approximately 1.5 times times the local conformational preference length of Pappu and co-workers (*53, 54*). This suggests that our choice of peptide length is appropriate. It is long enough that we should capture local interactions yet short enough that we will not detect longer range interactions that could lead to cooperative structure formation.

Our host peptide A7 is similar to the XAO peptide studied by Kallenbach and co-workers (*3*) but would be expected to have a higher α-helix content because the flanking lysines are better α-helix formers than the shorter flanking aminobutyric acid and ornithine residues in XAO (*1, 55*). An alternative, or perhaps complementary, explanation comes from the work of Scheraga and co-workers (*56*) and Garcia and co-workers (*57*). These groups performed calculations that indicate the bulky charged lysine side chain can desolvate the backbone of a peptide, leading to enhanced α-helix formation.

Note that although we believe the α-helix and P_{II} helix conformations to be the most prevalent for each residue in peptide A7 in dilute phosphate buffer, we are neither suggesting that the entire peptide is in a two-state equilibrium between these conformations nor ruling out the presence of other conformations. Rather, our data would appear to indicate that these are the two most populated conformations for individual residues and perhaps short stretches of adjacent residues. It has been demonstrated that transitions into and out of the P_{II} conformation are not cooperative (*58*), so the existence of short stretches of P_{II} helix is very unlikely. On the other hand, α-helix formation is well-known to be highly cooperative, so short stretches of α-helix are more probable. Peptide A7 exists in an ensemble of conformational states, a detectable fraction of which appear to contain short, frayed α-helices, with other fractions consisting of conformations with a number of, not necessarily consecutive, residues in the P_{II} conformation. Clearly, these fractions of the ensemble are not mutually exclusive or totally inclusive. Some conformations could possess residues in a short α-helix and other residues in the P_{II} conformation, and yet others will be completely devoid of both the α-helical or P_{II} conformations.

Host–Guest Peptide Studies. Why does the correlation between the known α-helix forming propensities and the apparent α-helix contents of our peptides appear so weak? This is because we are only looking at one side of the folding process. The α-helix propensities are derived from differences between folded and unfolded ensembles. In other words, they are a measure of how energetically favorable it is for individual residues to be part of a folded α-helix versus an unfolded ensemble. The peptides studied here model just the unfolded ensemble. The propensities cannot be used to predict the conformational behavior of our peptides. Indeed, it is clear from our data that interactions between residues separated in sequence by a few residues can lead to large differences in the observed conformational behavior. Such interactions could not be predicted using single-residue propensities. However, the observed conformational properties of our peptides can be used in part to rationalize the α-helix forming propensities.

For the individual flanking alanine, and perhaps lysine, residues in our peptides, the most populated states are α-helix-like and P_{II} helix-like, although β-strand-like conformations are clearly possible. The level of α-helix and P_{II} helix content is modulated in large part by the properties of the central guest residue. When the guest residue is an alanine, the peptide is capable of forming two to three consecutive turns of α-helix in a cooperative manner. This is not sufficient for stable α-helix formation but clearly results in a sufficiently high α-helix population that it can be detected via CD spectroscopy. Note that this is in contrast to the recent suggestion that low levels of α-helix content in short peptides are not discernible using CD spectroscopy (*38*).

The "α-helix-breaking" residues proline and glycine prevent detectable α-helix formation, leading the flanking residues to rearrange their conformational ensembles (Figure 3). This rearrangement appears to involve a significant increase in P_{II} helix content concomitant with the decrease in α-helix content. The conformational properties of the individual guest residues are not sufficient to explain the observation of apparent high P_{II} helix contents for these two peptides. Although proline has the highest P_{II} helix forming propensity (*3, 48, 50*), this one residue could not contribute all of the P_{II} signal observed for peptide P (Figure 3). In the case of peptide G, it seems highly unlikely that this highly flexible residue significantly populates the P_{II} conformation. In addition, the achiral glycine guest in peptide G would contribute little to the CD signal. For these two peptides, the majority of the P_{II} helix signal must then arise from the flanking sequences. The flanking alanines and lysines have high α-helix forming propensities (*21*), but the sets of five residues on either side of the guest position are not long enough to form stable α-helices. Notably, alanine and, to a lesser extent, lysine also possess significant P_{II} helix forming propensities (*2, 3, 5, 50*). The flanking sequences appear then to default, at least in part, to the P_{II} helical conformation, leading to the conformational behavior indicated in Figure 3.

Although peptides P and G possess similar overall CD signals, arising predominantly from the flanking sequences, the unfolded ensemble contributions to the α-helix propensities for these two residues are clearly very different. The imino ring of proline restricts its backbone conformational behavior and, additionally, sterically restricts the immediately preceding residue to the β-region of $\varphi-\Psi$ space (*47, 59*). The entropic cost of folding the proline into an α-helix would be low given that proline can populate only either a P_{II} helix-like or α-helix-like conformation. The unique major unfolded ensemble contribution to proline's α-helix propensity may then be the steric penalty paid for folding the restricted

preceding residue into an α-helical conformation. Glycine, on the other hand, is by far the most flexible of the residues, presumably leading to a high entropic penalty to its α-helix propensity. However, the entropic cost of restraining a glycine has recently been challenged by Daggett and co-workers (*60*).

The observed behavior of peptides with guest residue aliphatic or aromatic side chains could not all be predicted by the established α-helix forming propensities. Peptide L does have detectable α-helix content, at a lower level than peptide A7 (Figure 4). This is as might be expected given that leucine has an α-helix forming propensity lower than that of alanine, but high enough for it to be considered an α-helix favoring residue (*21*). Interpretation of the spectra for peptide F is complicated by the interference caused by the phenylalanine residue's absorbance in the far UV (*42*). However, phenylalanine is known to have a relatively weak α-helix forming propensity (*21*), so the lack of detectable α-helix content may not be surprising.

Peptides I and V, however, have significantly higher α-helix contents (Figure 3) than what might be expected given the unfavorable α-helix propensities of isoleucine and valine (*21*). This is particularly the case for peptide I which has a similar, or even higher, apparent α-helix content than peptide L. This suggests that a significant portion of the unfavorable components of the propensities for *β*-branched residues occur in the folded ensemble rather than the unfolded ensemble. Indeed, it has been observed that the *β*-branched residues lose significant side chain conformational entropy upon folding into an α-helix (*61, 62*). This entropic loss is due to unfavorable steric interactions between the side chain and residues in the preceding turn of α-helix. The peptides examined here could form at most 1.5 turns of α-helix prior to the guest site. Given that the ends of any α-helix formed by these peptides are likely to be highly frayed, the formation of a complete turn of α-helix prior to the guest residue is likely a rare event, irrespective of the chemical identity of the guest residue. The *β*-branched guest residue is not precluded from being part of a short α-helix, but it is unlikely that such a helix would be long enough for the side chain to sterically interfere with propagation. The *β*-branched isoleucine and valine residues then do not prevent but rather reduce the level of α-helix formation through unfavorable steric interactions in our peptides.

The apparent lack of correlation between α-helix content in the peptides containing polar guest residues (Figure 4) and the corresponding α-helix forming propensities can also be explained by considering contributions to the folded and unfolded ensembles separately. Serine and asparagine both have middling α-helix propensities (*21*), and low P_{II} helix propensities (*48, 50*), yet peptides S and N apparently have significant P_{II} helix content and no detectable α-helix content (Figure 4). These may be situations in which solvation plays a significant role. Both the serine and asparagine side chain polar groups will possess significant solvation shells. These are relatively short side chains, so in the folded α-helix, the associated solvation shells will be close to the backbone atoms of the residues in the preceding and/or following turns. It is possible that the side chain solvation shells are not "compatible" with solvation of the backbone polar atoms, leading to some level of unfavorable desolvation. Such side chain-induced desolvation would be less likely to occur in the unfolded ensemble because the backbone could adopt conformations in which favorable solvent−backbone interactions are maximized. In this way, the serine and asparagine could disfavor the folded α-helical conformation, leading the flanking alanines and lysines to default to the P_{II} helix conformation.

An additional contribution to the observed conformational behavior of peptide N could be asparagine's propensity to adopt the left-handed α conformation (*63, 64*). This would be compatible with the broad positive band observed for this peptide (Figure 4) since this conformation would lead to a CD spectrum that is the mirror image of the usual right-handed α-helix, with positive bands at 222 and 208 nm.

Although glutamine has a significant α-helix forming propensity (*21*), we detect little or no α-helix content in peptide Q (Figure 4d). Notably, we do not see significant P_{II} helix content either, despite glutamine having a very high propensity to adopt this conformation (*43, 48, 50*). We know from previous work that a relatively low concentration of urea can induce significant P_{II} helix content in this peptide (*1*). These data suggest that peptide Q does possess some P_{II} helix content, but we do not as yet have an explanation for either the shape of its CD spectrum or the observed temperature behavior (Figure 4). It is also notable that peptides Q and N display such dissimilar behavior, despite the side chains of the guest residues differing by just a single methylene. This observation suggests that glutamine to asparagine substitutions, as well as the reverse, are not necessarily conservative when considering the unfolded ensemble.

Recently, Pappu and co-workers (*53, 54*) used computational methods to explore the conformational properties of host–guest peptide models for unfolded proteins. They found residues Ala, Ile, Leu, Gln, Phe, and Ser displayed very small, or no, differences in the extent to which they populated the α-region of the Ramachandran plot but that there were significant differences in the various populations of the *β*-strand and P_{II} helix regions. The differences in α-helix contents described in this work are small, although for the most part larger than errors in the CD measurements (Figures 3 and 4). This apparent disagreement between our results and those of Pappu and co-workers (*53, 54*) is quite likely due to the use of a potential with no attractive terms in the computer simulations. The lack of attractive interactions such as hydrogen bonds would significantly reduce the probability of formation of α-helices. Nonetheless, the significant range of conformational behavior exhibited by our host–guest peptides (Figures 3 and 4) is likely to be largely due to differences in the populations of the extended conformations, in good agreement with the calculations of Pappu and co-workers (*53, 54*).

The findings described here demonstrate that residue propensities determined for structures that form in highly cooperative manners such as the α-helix cannot be used to describe the behavior of the unfolded conformational ensembles populated by peptides and proteins. Interactions between residues spaced a few residues apart in sequence appear to be able to dominate the conformational properties of the peptide models for unfolded ensembles examined here. *β*-Sheet propensities are known to be strongly context dependent (*28*), so it seems highly likely that these also cannot be used to describe unfolded ensemble behavior.

Perhaps the only propensities that can be used are for the P_{II} helical conformation, a result of the noncooperative nature of transitions into and out of this conformation (*58*).

ACKNOWLEDGMENT

We thank Sara Manning for useful discussions. We also thank Dr. Rick McCann for use of his spectrophotometer.

REFERENCES

1. Whittington, S. J., Chellgren, B. W., Hermann, V. M., and Creamer, T. P. (2005) Urea promotes polyproline II helix formation: Implications for protein denatured states. *Biochemistry 44*, 6269–6275.
2. Rucker, A. L., and Creamer, T. P. (2002) Polyproline II helical structure in protein unfolded states: Lysine peptides revisited. *Protein Sci. 11*, 980–985.
3. Shi, Z., Chen, K., Liu, Z., Ng, A., Bracken, W. C., and Kallenbach, N. R. (2005) Polyproline II propensities from GGXGG peptides reveal an anticorrelation with β-sheet scales. *Proc. Natl. Acad. Sci. U.S.A. 102*, 17964–17968.
4. Shi, Z., Olson, C. A., Rose, G. D., Baldwin, R. L., and Kallenbach, N. R. (2002) Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. U.S.A. 99*, 9190–9195.
5. Eker, F., Griebenow, K., Cao, X., Nafie, L. A., and Schweitzer-Stenner, R. (2004) 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. 101*, 10054–10059.
6. Mezei, M., Fleming, P. J., Srinivasan, R., and Rose, G. D. (2004) Polyproline II helix is the preferred conformation for unfolded polyalanine in water. *Proteins 55*, 502–507.
7. Fitzkee, N. C., and Rose, G. D. (2004) Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. U.S.A. 101*, 12497–12502.
8. Pappu, R. V., and Rose, G. D. (2002) A simple model for polyproline II structure in unfolded states of alanine-based peptides. *Protein Sci. 11*, 2437–2455.
9. Drozdov, A. N., Grossfield, A., and Pappu, R. V. (2004) Role of Solvent in Determining Conformational Preferences of Alanine Dipeptide in Water. *J. Am. Chem. Soc. 126*, 2574–2581.
10. Garcia, A. E. (2004) Characterization of non-α helical conformations in Ala peptides. *Polymer 45*, 669–676.
11. Shortle, D. (2002) The expanded denatured state: An ensemble of conformations trapped in a locally encoded topological space. *Adv. Protein Chem. 62*, 1–23.
12. Barron, L. D., Blanch, E. W., and Hecht, L. (2002) Unfolded proteins studied by Raman optical activity. *Adv. Protein Chem. 62*, 51–90.
13. Keiderling, T. A., and Xu, Q. (2002) Unfolded peptides and proteins studied with infrared absorption and vibrational circular dichroism spectra. *Adv. Protein Chem. 62*, 111–161.
14. Fleming, P. J., Fitzkee, N. C., Mezei, M., Srinivasan, R., and Rose, G. D. (2005) A novel method reveals that solvent water favors polyproline II over β-strand conformation in peptides and unfolded proteins: Conditional hydrophobic accessible surface area (CHASA). *Protein Sci. 14*, 111–118.
15. Anil, B., Li, Y., Cho, J. H., and Raleigh, D. P. (2006) The unfolded state of NTL9 is compact in the absence of denaturant. *Biochemistry 45*, 10110–10116.
16. Wickstrom, L., Okur, A., Song, K., Hornak, V., Raleigh, D. P., and Simmerling, C. L. (2006) The unfolded state of the villin headpiece helical subdomain: Computational studies of the role of locally stabilized structure. *J. Mol. Biol. 360*, 1094–1107.
17. Tang, Y., Goger, M. J., and Raleigh, D. P. (2006) NMR characterization of a peptide model provides evidence for significant structure in the unfolded state of the villin headpiece helical subdomain. *Biochemistry 45*, 6940–6946.
18. Li, Y., Picart, F., and Raleigh, D. P. (2005) Direct characterization of the folded, unfolded and urea-denatured states of the C-terminal domain of the ribosomal protein L9. *J. Mol. Biol. 349*, 839–846.
19. van Gunsteren, W. F., Bürgi, R., Peter, C., and Daura, X. (2001) The key to solving the protein-folding problem lies in an accurate description of the denatured state. *Angew. Chem., Int. Ed. 40*, 352–355.
20. Makowska, J., Rodziewicz-Motowidlo, S., Baginska, K., Makowski, M., Vila, J. A., Liwo, A., Chmurzynski, L., and Scheraga, H. A. (2007) Further evidence for the absence of stretches of polyproline II conformation in the XAO peptide. *Biophys. J. 92*, 2904–2917.
21. Chakrabartty, A., Kortemme, T., and Baldwin, R. L. (1994) Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci. 3*, 843–852.
22. O'Neil, K. T., and DeGrado, W. F. (1990) A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science 250*, 646–651.
23. Lyu, P. C., Liff, M. I., Marky, L. A., and Kallenbach, N. R. (1990) Side chain contributions to the stability of α-helical structure in peptides. *Science 250*, 669–673.
24. Park, S. H., Shalongo, W., and Stellwagen, E. (1993) Residue helix parameters obtained from dichroic analysis of peptides of defined sequence. *Biochemistry 32*, 7048–7053.
25. Blaber, M., Zhang, X. J., and Matthews, B. W. (1993) Structural basis of amino acid α-helix propensity. *Science 260*, 1637–1640.
26. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1997) A direct comparison of helix propensity in proteins and peptides. *Proc. Natl. Acad. Sci. U.S.A. 94*, 2833–2837.
27. Smith, C. K., Withka, J. M., and Regan, L. (1994) A thermodynamic scale for the β-sheet forming tendencies of the amino acids. *Biochemistry 33*, 5510–5517.
28. Minor, D. L., Jr., and Kim, P. S. (1994) Context is a major determinant of β-sheet propensity. *Nature 371*, 264–267.
29. Minor, D. L., Jr., and Kim, P. S. (1994) Measurement of the β-sheet forming propensities of the amino acids. *Nature 367*, 660–663.
30. Kim, C. A., and Berg, J. M. (1993) Thermodynamic β-sheet propensities measured using a zinc-finger host peptide. *Nature 362*, 267–270.
31. Brandts, J. F., and Kaplan, L. J. (1973) Derivative spectroscopy applied to tyrosyl chromophores. Studies on ribonuclease, lima bean inhibitors, insulin, and pancreatic trypsin inhibitor. *Biochemistry 12*, 2011–2024.
32. Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem. 287*, 252–260.
33. Compton, L. A., and Johnson, W. C., Jr. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem. 155*, 155–167.
34. Provencher, S. W., and Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry 20*, 33–37.
35. Sreerama, N., and Woody, R. W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem. 209*, 32–44.
36. Chen, Y. H., Yang, J. T., and Chau, K. H. (1974) Determination of the helix and β form of proteins in aqueous solution by circular dichroism. *Biochemistry 13*, 3350–3359.
37. Kallenbach, N. R., Lyu, P., and Zhou, H. (1996) CD spectroscopy and the helix-coil transition in peptides and polypeptides, in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G. D., Ed.) pp 201–259, Plenum Press, New York.
38. Ozdemir, A., Lednev, I. K., and Asher, S. A. (2002) Comparison between UV Raman and circular dichroism detection of short α helices in bombolitin III. *Biochemistry 41*, 1893–1896.
39. Gans, P. J., Lyu, P. C., Manning, M. C., Woody, R. W., and Kallenbach, N. R. (1991) The helix-coil transition in heterogeneous peptides with specific side-chain interactions: Theory and comparison with CD spectral data. *Biopolymers 31*, 1605–1614.
40. Padmanabhan, S., York, E. J., Gera, L., Stewart, J. M., and Baldwin, R. L. (1994) Helix-forming tendencies of amino acids in short (hydroxybutyl)-L-glutamine peptides: An evaluation of the contradictory results from host-guest studies and short alanine-based peptides. *Biochemistry 33*, 8604–8609.
41. Woody, R. W. (1992) Circular dichroism and conformation of unordered polypeptides. *Adv. Biophys. Chem. 2*, 37–79.
42. Woody, R. W., and Dunker, A. K. (1996) Aromatic and cystine side-chain circular dichroism in proteins, in *Circular dichroism and the conformational analysis of biomolecules* (Fasman, G. D., Ed.) pp 109–158, Plenum Press, New York.
43. Chellgren, B. W., Miller, A. F., and Creamer, T. P. (2006) Evidence for polyproline II helical structure in short polyglutamine tracts. *J. Mol. Biol. 361*, 362–371.

44. Helbecque, N., and Loucheux-Lefebvre, M. H. (1982) Critical chain length for polyproline-II structure formation in H-Gly-(Pro) $_n$-OH. *Int. J. Pept. Protein Res. 19*, 94–101.
45. Helbecque, N., and Loucheux-Lefebvre, M. H. (1978) Synthesis and circular dichroism studies of two polypeptides H-[Gly-(Pro)$_3$]$_n$-OH and H-[Gly-(Pro)$_4$]$_n$-OH. *Int. J. Pept. Protein Res. 11*, 353–362.
46. Chellgren, B. W., and Creamer, T. P. (2006) Side-chain entropy effects on protein secondary structure formation. *Proteins 62*, 411–420.
47. Creamer, T. P. (1998) Left-handed polyproline II helix formation is (very) locally driven. *Proteins 33*, 218–226.
48. Kelly, M., Chellgren, B. W., Rucker, A. L., Troutman, J. M., Fried, M. G., Miller, A.-F., and Creamer, T. P. (2001) Host-guest study of left-handed polyproline II helix formation. *Biochemistry 40*, 14376–14383.
49. Chellgren, B. W., and Creamer, T. P. (2004) Short Sequences of Non-Proline Residues Can Adopt the Polyproline II Helical Conformation. *Biochemistry 43*, 5864–5869.
50. Rucker, A. L., Pager, C. T., Campbell, M. N., Qualls, J. E., and Creamer, T. P. (2003) Host-guest scale of left-handed polyproline II helix formation. *Proteins 53*, 68–75.
51. Ding, L., Chen, K., Santini, P. A., Shi, Z., and Kallenbach, N. R. (2003) The pentapeptide GGAGG has PII conformation. *J. Am. Chem. Soc. 125*, 8092–8093.
52. Chellgren, B. W., and Creamer, T. P. (2004) Effects of H_2O and D_2O on polyproline II helical structure. *J. Am. Chem. Soc. 126*, 14734–14735.
53. Tran, H. T., and Pappu, R. V. (2006) Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions. *Biophys. J. 91*, 1868–1886.
54. Tran, H. T., Wang, X., and Pappu, R. V. (2005) Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins. *Biochemistry 44*, 11369–11380.
55. Padmanabhan, S., and Baldwin, R. L. (1991) Straight-chain nonpolar amino acids are good helix-formers in water. *J. Mol. Biol. 219*, 135–137.
56. Vila, J., Williams, R. L., Grant, J. A., Wojcik, J., and Scheraga, H. A. (1992) The intrinsic helix-forming tendency of L-alanine. *Proc. Natl. Acad. Sci. U.S.A. 89*, 7821–7825.
57. Ghosh, T., Garde, S., and Garcia, A. E. (2003) Role of backbone hydration and salt-bridge formation in stability of α helix in solution. *Biophys. J. 85*, 3187–3193.
58. Chen, K., Liu, Z., and Kallenbach, N. R. (2004) The polyproline II conformation in short alanine peptides is noncooperative. *Proc. Natl. Acad. Sci. U.S.A. 101*, 15352–15357.
59. MacArthur, M. W., and Thornton, J. M. (1991) Influence of proline residues on protein conformation. *J. Mol. Biol. 218*, 397–412.
60. Scott, K. A., Alonso, D. O., Sato, S., Fersht, A. R., and Daggett, V. (2007) Conformational entropy of alanine versus glycine in protein denatured states. *Proc. Natl. Acad. Sci. U.S.A. 104*, 2661–2666.
61. Creamer, T. P., and Rose, G. D. (1994) α-Helix-forming propensities in peptides and proteins. *Proteins 19*, 85–97.
62. Creamer, T. P., and Rose, G. D. (1992) Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix-forming propensities. *Proc. Natl. Acad. Sci. U.S.A. 89*, 5937–5941.
63. Srinivasan, N., Anuradha, V. S., Ramakrishnan, C., Sowdhamini, R., and Balaram, P. (1994) Conformational characteristics of asparaginyl residues in proteins. *Int. J. Pept. Protein Res. 44*, 112–122.
64. Deane, C. M., Allen, F. H., Taylor, R., and Blundell, T. L. (1999) Carbonyl-carbonyl interactions stabilize the partially allowed Ramachandran conformations of asparagine and aspartic acid. *Protein Eng. 12*, 1025–1028.

BI702474K