Published in final edited form as:
Biopolymers. 2008 September ; 89(9): . doi:10.1002/bip.21004.

Structural insights for designed alanine-rich helices: Comparing NMR helicity measures and conformational ensembles from molecular dynamics simulation

Kun Song[1], **James M. Stewart**[3], **R. Matthew Fesinmeyer**[3], **Niels H. Andersen**[3], and **Carlos Simmerling**[1,2]

Niels H. Andersen: andersen@chem.washington.edu; Carlos Simmerling: carlos.simmerling@stonybrook.edu

[1]Center for Structural Biology, Stony Brook University, Stony Brook, NY 11794

[2]Department of Chemistry, Stony Brook University, Stony Brook, NY 11794

[3]Department of Chemistry, University of Washington, Seattle, Washington 98195

Abstract

The temperature dependence of helical propensities for the peptides Ac-ZGG-$(KAAAA)_3$X-NH_2 (Z = Y or G, X = A, K, and D-Arg) were studied both experimentally and by molecular dynamics simulations. Good agreement is observed in both the absolute helical propensities as well as relative helical content along the sequence; the global minimum on the calculated free energy landscape corresponds to a single α-helical conformation running from K4 – A18 with some terminal fraying, particularly at the C-terminus. Energy component analysis shows that the single helix state has favorable intramolecular electrostatic energy due to hydrogen bonds, and that less-favorable two-helix globular states have favorable solvation energy. The central lysine residues do not appear to increase helicity; however, both experimental and simulation studies show increasing helicity in the series X = Ala → Lys → D-Arg. This C-capping preference was also experimentally confirmed in Ac-$(KAAAA)_3$X-GY-NH_2 and $(KAAAA)_3$X-GY-NH_2 sequences. The roles of the C-capping groups, and of lysines throughout the sequence, in the MD-derived ensembles are analyzed in detail.

INTRODUCTION

Helix-coil theory is one of the most fundamental and intensively studied aspects of biomolecular structure formation[1]. There are several reasons: first, the α-helix is a key secondary structure element in globular protein structure[2]. Second, there are several experimental methods available which can detect the α-helical content. Third, the formation of α-helices is fast (sub-microsecond scale) relative to folding of proteins with multiple secondary structure units. This fast rate permits the all-atom simulations of α-helix formation and also makes helix formation a likely pre-transition state equilibrium event in the folding of proteins with multiple secondary structure units.

Helix formation has long been known to show significant cooperativity, with coupling of hydrogen bonding and φ/ψ propensities between neighboring residues indicated by early theoretical[3–5], and experimental[6, 7] studies. The recent investigations of amyloid fibrils have also found that sequences that adopt α-helices in globular proteins can form β–strands in

Correspondence to: Niels H. Andersen, andersen@chem.washington.edu; Carlos Simmerling, carlos.simmerling@stonybrook.edu.

amyloid aggregates[8, 9]. The quest to understand these secondary structure transitions has, once again, brought helix studies into the spotlight.

Designed alanines-rich helices in which solubility is enhanced through inclusion of polar residues (most commonly Lys and Arg), are the most common short sequences used to study helix formation both experimental[7, 10–15] and in MD simulations[15–18]. Other roles, including solvation changes that increase helicity,[7,18] have been suggested for the Arg and Lys residues in designed helices.[7,17–19] MD simulations of Ala-rich helices of this type have been carried out using both continuum and explicit solvent models. Helix bundles, sequences with two helical segments in close association in a more globular state, have been predicted to be the low energy state for (AAARA)-repeat peptides in continuum solvent models[17], particularly at somewhat warmer temperatures[16]. More typically, ensembles are dominated by single helix structures. Incorrect parameterization of H-bonds has been proposed as the reason for the differences.[17] Sequence differences could also play a role, but the same sequence has yielded both types of structures depending on the solvent model and temperature. With residue-specific measures of helicity these states should be distinguishable.

We now report NMR and molecular dynamics simulation studies on a series of peptides of the general form N-cap-$(KAAAA)_3$-C-cap. One series is very closely related to the parent system (Ac-YGG-$(KAAAA)_3$-K-NH_2) of the series used by the Baldwin group and others to determine the propagation values of X in Ac-YGG-$(KAXAA)_3$-KNH_2 and other helical peptides.[12–14] Contrary to the prior theoretical predictions that peptides of such a short length would not adopt measurable helix content in aqueous solution, Marqusee et al.[10] demonstrated that peptides containing lysine and alanine can be as much as 70% helical in water and this is usually attributed to a uniquely large helix propensity of alanine. Other experimental studies have shown that the α-helix conformation is not favorable for short polyalanine peptides in water at room temperature[7,20,21], and several roles have been suggested for the stabilizing effects of polar side chains, particularly Lys and Arg.[7,17,18] Other simulations, particularly in continuum solvent models, suggest that polar sidechains do not increase helicity,[17] and recent experimental studies have provided propagation values for alanines (1.6) and a non-terminal lysine (0.82) that are in accord with those from the Baldwin group and not indicative of helix stabilization by the lysine side chain when it is not in the C-terminal turn of the helix.[22] Positive charges at the C-terminus of a helix are expected to favor helix formation due to a favorable interaction with the helix macrodipole. D-Arg has been reported to be a particularly favorable C-cap.[23] Explicit solvent simulation has suggested that charged side chains throughout the sequence can favor helix formation by shielding the intramolecular backbone hydrogen bonds from exposure to solvent[18]. As a result, this study also focuses on the contributions of the charged residues to the stabilization of the helix conformation. In this report, the results from all-atom simulations are compared to those obtained primarily from NMR experiments. There is excellent agreement regarding both the distribution of helicity along the sequence and the net helicity that provides: a validation of our simulation protocols and force field parameters, a determination of relative C-capping capabilities, and an atomic-detail model for the structure ensembles that give rise to the experimental observables.

MATERIALS AND METHODS

Peptide synthesis and purification

Peptides were synthesized on an Applied Biosystem 433A synthesizer employing standard Fmoc solid-state peptide synthesis methods and were acetylated while still on-resin. Rink resin provided an amidated C-terminus upon cleavage using 95% trifluoroacetic acid, with 2.5% triisopropylsilane and 2.5% water present. The cleavage products were purified using

reverse-phase HPLC on a Varian C^{18} preparatory-scale column with gradients of water and acetonitrile spiked with 0.1% and 0.085% trifluoroacetic acid, respectively. Collected fractions were lyophilized and their identity and molecular weight confirmed on a Bruker Esquire ion trap mass spectrometer. The following peptides (**r** = D-Arg, A = ^{13}C=O-labeled alanine) were synthesized specifically for this study: (Ac-) indicates that these series of peptides were examined in both the acetylated and non-acetylated form.

(Ac-)KAAAKGY-NH_2	(Ac-)KAAA**r**GY-NH_2	Ac-KGGKAAAAK-NH_2
(Ac-)(KAAAA)$_3$KAAAKGY-NH_2		Ac-GGG(KAAAA)$_3$K-NH_2
(Ac-)(KAAAA)$_3$XGY-NH_2 X = A, K		YGG(KAAAA)$_3$A-NH_2
(Ac-)(KAAAA)$_2$KAAAA**X**GY-NH_2 & (Ac-)YGG-KAAAA(KAAAA)$_2$**X**-NH_2 (X = A, **r**, R)		

The single repeat systems were prepared to serve as random coil controls together with the prior data[24] for Ac-GKAAAKG-NH_2. The same report provided extensive chemical shift deviation (CSD) data for Ac-YGG-KAAAA(KAAAA)$_2$K-NH_2. With the exception of the single repeat systems, several alternate isotopomers were made of each peptide in order to establish secure assignments of the shifts of the different ^{13}C=O labeled sites. The four repeat system, Ac-(KAAAA)$_3$KAAAAKGY-NH_2, was used to determine whether larger structuring shifts could be obtained with a longer, more stable helix. The largest structuring shift observed for a four-repeat peptide to date is +3.41 ppm for GASEDEAAAAK(AAAAK)$_2$GY-NH_2 [James Stewart, unpublished data[25]].

Experimental Measures of Helicity

CD spectroscopy—Circular dichroism stock solutions were prepared by dissolving weighed amounts of peptide in 20 mM aqueous pH 7.0 phosphate buffer to make solutions with a circa 700 μM peptide concentration. The concentration of stock solutions of peptides containing a tyrosine residue was determined by UV based on the absorptivity coefficient, ε = 1420 M^{-1} cm^{-1} at 276 nm due to tyrosine. CD samples were diluted appropriately with buffer to obtain 35 μM peptide solutions. Spectra were recorded on a Jasco J720 spectropolarimeter using 0.10 cm pathlength cells. The calibration of the wavelength and degree ellipticity scales has been described previously[26,27]. Typical spectral accumulation parameters were a scan rate of 100 nm/min and a 0.2 nm step resolution over the range of 190–270 nm with 16 scans averaged for each spectrum. The accumulated average spectra were trimmed at a dynode voltage of 600 V prior to blank subtraction and smoothing, the latter of which used the reverse Fourier transform procedure in the Jasco spectra analysis software. For melting experiments, the temperature was increased from *ca.* 275K to 340K in 5° to 10° increments, depending on the degree of resolution required to elucidate the response. Each step was equilibrated at the target temperature for at least five minutes before data was acquired; larger temperature changes were provided additional time. CD data is reported in residue-molar ellipticity units (deg cm^2·residue·$dmol^{-1}$).

NMR spectroscopy—^{13}C-NMR experiments were routinely performed on Bruker DRX-500 or AV500 spectrometers, with occasional use of a DMX750 to improve resolution when the ^{13}C' signals were badly overlapped. The samples consisted of *ca.* 2 mM peptide in 20mM phosphate buffer at pH 7.0 with 10% D_2O. All samples also contained 0.005mM ^{13}C-labeled urea as a chemical shift reference. Sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was used as the direct or indirect (below) internal chemical shift reference and set to 0 ppm for all conditions independent of temperature, pH, and co-solvent concentration. The method has been described in greater detail in an account of prior

studies of Ac-YGG-(KAXAA)$_3$-K-NH$_2$) with X = A, V.[24] The chemical shift of ^{13}C-urea is given by (165.609 ∓ 0.0056 T ppm, T = °C). Unlike other experimental methods, this method can determine the fractional helicity at individual residue sites.

Determining chemical shift deviations (CSDs) and fractional helicities

The reference chemical shifts for the unstructured state of ^{13}C'-Ala probes was, with two exception, in the range 177.59 – 177.70 ppm at 280K; both the AA<u>A</u>K- and -KA<u>A</u>AA-(D-Arg)- values were downfield, in the 177.91 – 178.05 range depending on the C-terminal sequence. All these resonances shift upfield on warming, 0.007 – 0.012 ppm/°C, a feature that must be taken into account in calculating melting curves. When a full sequence identity was not available in a control peptide, the effects of replacing Lys and/or Gly by Ala was assumed to correspond to the differences in the sequence-dependent corrections reported by Schwarzinger et al.[28] The specific sequence-dependent random coil values employed are available upon request. CSDs are converted to fractional helicities assuming that a 0 – 100% helicity structuring shift at all but the extreme termini corresponds to +3.50 ppm. The following corrections were applied at the termini for Ncap-N1(=K)-N2-N3 - - - C3-C2-C1-Ccap(=K or D-Arg) sequences – +0.25 (N2), +0.10 (N3), −1.15 (C2), −1.3ppm (C1). The N-terminal adjustments reflect helix macrodipole effects; the larger corrections at the C-termini reflect these and the absence of a helical H-bonding interaction. When the C-cap residue does not have a positively charged sidechain to offset helix macrodipole, the Cterminal corrections are larger: −1.25 (C2) and −1.45ppm (C1). The large C-terminal corrections (as much as a 41% reduction in the 0 – 100% helicity structuring shift), are based on both theoretical and experimental measures of this effect. As previously noted[24], a priori calculations of the H-bond effect[29] suggest a 1.4 ppm contribution. Park et al.[30] deconvoluted experimental ^{13}C' melts and arrived at the following %-contributions: 32% backbone geometry, 35% hydrogen bonding by the NH in the amide unit, and 33% by hydrogen bonding to the ^{13}C=O. A recent study of a miniprotein helix suggest that the central residue 100% shift may be as large as 3.7 ppm, with H-bonding responsible for 54% of the shift.[31] The resulting uncertainty in the adjustments introduces additional error into derived fractional helicity measures for the two C-terminal sites in a helix.

Computational Studies

Replica exchange molecular dynamics simulations—Standard molecular dynamics simulations can become trapped in local minima during computationally affordable simulations. Therefore, replica exchange molecular dynamics (REMD)[32,33] (also known as parallel tempering MD[34]) was used as an enhanced sampling method for this study. We used REMD as implemented in AMBER version 8[35] following procedures we reported previously[36]. REMD employs multiple non-interacting replicas of the system that are simulated independently and simultaneously at several different temperatures. At intervals of 1 ps, exchanges were attempted between conformations at neighboring temperatures based on a Metropolis-type criterion that considers the probability of sampling each conformation at the alternate temperature. The advantage of this method is that the simulations can escape from kinetic traps by "jumping" into alternate local minima being sampled more efficiently at higher temperatures. We employed 12 replicas at 253, 276, 300, 326, 355, 386, 420, 457, 497, 541, 588, and 640 K.

A modified force field based on AMBER parm94[37] was used to represent the peptide. Many AMBER parameter sets have been reported to over stabilize α-helical structure[38,39]. In the parameters that we used, the torsional angle terms were adjusted to reduce this tendency[40]. The generalized Born implicit solvent model[41,42], with intrinsic Born radii taken from Tsui and Case[43] as implemented in AMBER, was used to represent the effects of solvent. All non-bonded interactions were evaluated at every step. The SHAKE algorithm[44] was used to

constrain the bonds involving hydrogen atoms. A 0.002 ps time step was used. Each replica was coupled to a constant-temperature bath using a weak-coupling algorithm[45]. The REMD simulations were carried out for 56 nanoseconds (a total of 672 ns simulation), and 56,000 snapshots were saved for each temperature. The REMD simulations were performed for three sequences: AcGGG-$(KAAAA)_3$-X-NH_2 with X = Lys, D-Arg and Ala.

Analysis of secondary structure content: To compare our results with previous studies, Lifson-Roig helix-coil theory[5,14,46] has been used. In Lifson-Roig theory, there are three states for residues: the "c" state is the residue in random coil conformation, "w" is the residue in helical conformation, and its previous and next neighbor residues are in helical conformation as well, and "v" is a helical residue neighboring with one or two non-helical residues. Overall helical content corresponds to the number of w state residues, with a maximum helix length of N−2 for N residues. Following the work of Garcia and Sanbonmatsu[18] and Sorin and Pande[47], a residue is considered in a helical conformation if $\varphi = -60(\pm 30)°$ and $\psi = -47(\pm 30)°$. The melting curve is generated based on the temperature dependence of the average helical fraction (f_H), calculated using equation 1, where Nw is the number of w-state residues assigned by Lifson-Roig theory and N is the total number of residues. In order to correspond to experimental melting measurements, only the K4 to X19 sequence is considered, setting N to 16 and the maximum helix length to 14.

$$f_H = Nw/(N-2) \quad \text{Equation 1}$$

To complement this approach, we also used DSSP[48] (as implemented in the ptraj module of AMBER) to analyze secondary structure content.

Free energy landscapes: We reconstructed free energy landscapes using principle component analysis (PCA), as implemented in ptraj, to define the reaction coordinates and 2-D histogram analysis to obtain relative free energy for each bin. Due to their high flexibility, Gly residues were not included in this analysis. The alpha carbon coordinates of all other residues were used to generate the covariance matrix. The two largest eigenvalues from the PCA analysis were used to represent order parameters. These two eigenvectors represent only 42% of the total fluctuations in the system, suggesting that more than 2 coordinates are required for a comprehensive view of the free energy landscape even for this short peptide. The relative free energy of each bin was calculated via equation 2, where x is a histogram bin corresponding to a set of PCA eigenvector values, G(x) is the free energy relative to the most populated bin, R is the gas constant, T is the temperature, P(x) is the population in bin x and P(x0) represents the population of the most populated bin.

$$\mathrm{G(x)} = -\mathrm{RTlog(P(x)/P(x0))} \quad \text{Equation 2}$$

Structure analysis: Cluster analysis was used to separate the ensembles into conformational families and representative structures. Moil-view[49] was used to perform the cluster analysis using backbone RMSD for non-Gly residues as a similarity criterion, with average linkage. Clusters were formed with a bottom-up approach using a similarity cutoff of 2.5 Å. The radius of gyration (R_{gyr}) was also used to characterize structural properties of the overall ensemble and individual clusters. The carnal module of AMBER was used for Rgyr calculation. All alpha carbon atoms of non-Gly residues were included in this analysis.

Energy decomposition analysis: To analyze the relative importance of various energy components in different structure families, energy decomposition analysis was carried out using post-processing in the sander module of AMBER. Average total energies were separated into electrostatic, van der Waals and GB solvation terms.

Estimation of simulated data uncertainties: Lower bounds to data uncertainties were estimated as the absolute value of the difference between properties obtained using the entire ensemble sampled and that using only the second half of the data set.

RESULTS AND DISCUSSION

All NMR and CD studies were performed in low ionic strength media (20 mM phosphate, no added salt) in order to retain unscreened Coulombic effects that might be present due to Lys and Arg sidechains as well as, when present, the uncapped N-terminus. CD has been the classic method for accessing net helicity with the $[\theta]_{222}$ value as the standard measure. Baldwin & Luo provided the most commonly used equation for converting CD data to a measure of the net per cent helicity. The $[\theta]_{222}$ value for 100% helicity is given by $\theta_{222} = (1 - 1/Nr)(-44000 + 250\ T)$.[50] For the three repeat systems in the present study, Nr = 16. The thermal gradient of θ_{222} for 100% helicity in this equation is, however, nearly twice as large as the experimental value observed for stable helical proteins and coiled coils well below their melting transitions; this has prompted us to employ a *circa* 0.38%/°C, rather than 0.57%/°C, loss of ellipticity for the 100%-folded baseline in our calculations of CD melts.[51, 52] The temperature dependence of the random coil state for (KAAAA)-repeat peptides bearing a single tyrosine residue separated from the helix N-terminus is $\theta_{222} = +2120 - 64T$.[51] In the present study, CD is used only in a qualitative manner, with data appearing in Table I and full CD melts for several species in the Supporting Materials (Figure S1).

Experimental Helicities from NMR Data

Helix formation results in a downfield shift at $^{13}C{=}O$ sites[53,54] and the resulting structuring shifts melt out. Representative CSD data, and a comparison with CD measures, appear in Table I. CD comparisons between the different series of peptides in this study suggested that the CD value for a comparable degree of helicity at central sites (A8 and A9 in these peptides) were not directly comparable. One rationale for this observation, is that the contribution due to the Tyr sidechain[55] is different when the residue is placed near the C- rather than N-terminus of the helix; the -$(KAAAA)_3GY\text{-}NH_2$ series peptides always display lower CD values than YGG-$(KAAAA)_3$ species with comparable NMR measures of helicity. The Tyr sidechain contribution also appears to be different for Ac-YGG- versus YGG-. Over the systems examined in the present study, the CD data suggested helix termination with D-Arg rather than Lys or Ala, did not uniformly increase the value of $[\theta]_{222}$. The largest increase (21%) was observed in the Ac-$(KAAAA)_3XGY\text{-}NH_2$ comparison (X = **r** versus A) and was not accompanied by a comparable increase in the CSDs at the labeled central alanine sites. As a result, we employ only the NMR data in the present comparisons of experimental measures of helicity and those observed in MD simulations.

Turning to our analysis of the $^{13}C{=}O$ chemical shift data, initially we employed the reference shifts previously derived from Ac-GK<u>AAA</u>KG-NH_2[24] as coil controls for estimate structuring shifts. When this method was used over complete melting curves, chemical shift deviation (CSD) *versus* T, anomalies were observed. For the Cterminal $^{13}C'$-Ala sites of a number of peptides, particularly the less helical ones lacking a positively charged residue at the C-terminus or a capping unit at the N-terminus, the melting data provided "negative" CSDs as helicity melted out at the higher temperatures. The melting data, as CSD *vs* T(°C) for four sites (A3, A8, A13, and A14) of $(KA\underline{AA}A)_3XGY\text{-}NH_2$ (X = K, **r**) appear in Figure 1. The exaggerated downfield shift (0.89 ppm) at A13 for X = **r**, (vs 0.56 ppm for X = K), with little change at the central sites, was initially taken as possible evidence for a D-Arg16 sidechain-NH → O=C-A13 H-bond which would be present in a specific C-capping interaction by D-Arg. However, the melting data A13 in the X = **r** peptide suggests that the

high temperature unfolded limit corresponds to a non-zero (positive) CSD: the CSD at this site, calculated using the prior random coil values, did not melt out as much as central sites in the same peptide and the corresponding site in the Lys-terminated helix. The X = **r** versus K CSD difference at A13 is also retained (1.61 vs 1.39 ppm) in the more helical Ac-(KAAA)$_3$XGY-NH_2 species. These observations prompted us to prepare a new series of more closely matched reference peptides: Ac-KGGKAAAAK-NH_2 and -KAAAAXGY-NH_2 (X = K, **r**) with and without an N-acetyl cap. In KAAAAKGY-NH_2, the two labeled sites were essentially shift coincident (δ^{290} = 177.43 ± 0.02 ppm); acetylation induced downfield shifts (and a small extent of factional helicity) and increased the $\Delta\delta$ for the two probes to 0.045 ppm. In the corresponding $_D$-Arg-GY-NH_2 terminated series, the $\Delta\delta$ for the two labels was 0.38 ± 0.02 ppm independent of whether acetylated or not. A similar $\Delta\delta$ is observed at the melting limit in panel **b** of Figure 1; thus, the downfield shift of A13 cannot be attributed to a helix-capping interaction. Rather it appears to be a random coil shift effect of $_D$-Arg. A new set of position and N-terminus capping specific reference values for 0% helicity was derived based on the expanded set of reference peptides available. The ^{13}C=O shifts in KAAAAXGY-NH_2 (X = K, **r**) display a temperature gradient of − 0.008 ± 0.0005 ppm/°C. In the N-capped reference peptides, the ^{13}C=O shift gradient was − 0.011 ± 0.0014 ppm/°C. These values were used to correct the reference shifts in melting studies.

All comparisons with computed ensembles, require a measure of the structuring shift expected for 100% helicity so that the fractional helicity (f_H) can be calculated. In prior studies[24], ^{13}C' CSDs were converted to fractional helicities using equation 3. This equation resulted from cross-calibration ^{13}C'-CSD melts with IR and CD melts. Both of these, however, required making some assumptions about the "100%-folded parameters" for other spectroscopic assays.

$$f_H(T) = (\delta_{obs} - \delta_{ref\ peptide})/(3.040 + 0.00715T(^\circ C)) \quad \text{Equation 3}$$

The NMR study of Ac-(KAAAA)$_3$KAAAAKGY-NH_2 (given, in part in Table 1) yielded structuring shifts greater than the maximal value from equation 3. Helices with a 21-residue span with more effective caps have CSD values as large as 3.41 ppm; since 100% helicity cannot be achieved, as a result, we use equation 4 to convert CSDs to fractional helicities in the present account.

$$f_H(T) = (\delta_{obs} - \delta_{ref\ peptide})/(3.50 + 0.006T(^\circ C)) \quad \text{Equation 4}$$

Given the small remaining error in CSDs, this equation should provide residue-specific fractional helicities that are accurate to 0.03. This degree of accuracy, however, applies only to central portions of helices, and still assumes that there are no position specific structuring shift differences in the fully helical state. Errors are greater, particularly at the C-terminus, due to uncertainty about the contribution of H-bonding to the helical CSD (see Methods). In addition, the current corrections at C-terminal sites may be more problematic for the helices with a DArg C-cap.

For now, however, the data is presented as CSDs rather than fractional helicities. The first question we address is the relative helix stabilizing effects of N-caps. All of the helix ensemble simulations employ Ac-GGG as the N-cap. We wanted to be able to use the more extensive labeling available from the prior studies of Ac-YGG-KAAAAKAAAAKAAAAK-NH_2 and the newly synthesized Ac-(KAAAA)$_3$-XGY-NH_2 peptides in comparisons of sequence position dependence of helicity with simulated

conformational ensembles. The calculated CSDs for a variety of N-capped (and uncapped) species bearing Lys, Ala, and D-Arg (**r**) at the C-terminus appeared in Table I.

Experimental measures of N- and C-capping—With the exception of the long, $(KAAAA)_4K$, helix, N-capping is essential for obtaining systems displaying fractional helicities greater 0.40 (CSD ≥ 1.4) at any position. This is the case even with the D-Arg C-cap species: the maximal central repeat CSD value corresponds to $f_H = 0.29$ for $(KAAAA)_3$ **r**GY-NH_2 and indicates that the best reported[23] C-cap does not lead to greatly enhanced helicity at the center of the helix in these systems. From the NMR data in Table I, it is apparent that the N-capping effects of Ac-GGG, YGG and Ac-YGG are very similar and likely result from the effect of two glycines before a helical segment; further, these capping effects are nearly as large as those observed for N-acetylation. A quantitative treatment of N-capping effects will appear in a subsequent publication; for now our attention is focused on the effects of the C-capping residues.

Figure 2 collects the CSD data for a series of peptides that probe C-capping effects, Ac-$(KAAAA)_3$XGY-NH_2. The Lys- and D-Arg-terminated sequences are distinctly more helical than the Ala-terminated sequence, but still display a loss in structuring shifts near the C-terminus with little loss at the N-capped terminus. Enhanced f_H values at central helical positions for the D-Arg vs Lys C-capped sequences is only evident in the non-acetylated species. While field effects due to the helix macrodipole, positive near the N-terminus and negative at the C-terminus, are expected to change the CSD corresponding to 100% helicity at the termini (see Methods section), the near absence of N-terminal fraying is still evident. N-capping effects are very substantial; the comparisons in Table I provided a number of examples of the N-capping effect on structuring shifts as does the acetylated versus non-acetylated examples in Figure 2. In the absence of an N-capping unit, fraying reappears at the N-terminus.

As previously noted[24], the $^{13}C'$ CSDs towards the C-terminus provide much smaller f_H estimates than those expected based on the AGADIR algorithm[56] and an extended Lifson-Roig statistical mechanical helix/coil model[14]. H-bonding contributions to the $^{13}C'$ CSD are expected and could be a significant contributor to the net structuring shift. Chemical shift calculations based on density functional theory have previously estimated the contribution of hydrogen bond formation to the $^{13}C'$ CSD observed for a central helical residue to be as much as 1.4 ppm[29]; others studies[30, 31] suggest that as much as 60% the experimental CSD for central residues may be due to H-bonding. In the absence of corrections for H-bond contributions, CSD-derived f_H value calculated for C-terminal $^{13}C'$ positions would therefore underestimate the degree of folding as the residues are not capable of forming intramolecular, *i /i+4* or *i/i+3* hydrogen bonds. Corrections applied in comparisons with simulated ensemble statistics are given in the **Methods** section; for example, the CSD required for 100% helicity at the penultimate site in a $(KAAAA)_3A$ helical sequence is only 59% of the value required at a central residue. Even with these substantial corrections, the experimental data for N-capped species still required more C-terminal fraying than that calculated by current helicity prediction methods. This was one of driving forces for computational studies in the present account.

Direct Comparison of Experimental Data and Helical Ensembles from Computational Simulation

Molecular dynamics simulations were performed for AcGGG-$(KAAAA)_3$-X-NH_2 with X = Lys, D-Arg and Ala. Henceforth these are designated as peptides **K19, dR19** and **A19**. For all three peptides, the residues which are near the C-terminus have greatly reduced average helical content corresponding to the substantial helix fraying evident in the ^{13}C-NMR

studies. The GGG units were non-helical ($f_H < 0.04$) with a rapid helicity increase beginning at K4 consistent with a significant N-capping effect by the GGG unit. The helical domain is viewed as K4 – A18/X19 (X19 can be either a helix C-cap or the frayed end of a helix). [Residue locations are given by the residue number, thus the A8/A9 central probes in Table I and Figures 1 and 2, are, henceforth, designated as A11/A12.] The dynamics simulations indicated a greater overall helicity for peptides **K19** and **dR19**, bearing a positively charged side chain at the C-terminus.

The fractional helicities at each residue generated from NMR data and from the simulation ensembles at 275K are compared in Figure 3. Both methods show that the structure ensembles have sequence-dependent helicity measures, and the computational and experimental results are in good agreement, suggesting that further detailed analysis of the simulated structure ensemble is warranted. The relatively high helical content and flat profile across the central portion of these sequences (Figure 3) suggest that the most populated conformation for the peptides is a single helix, rather than isolated smaller helical fragments. Although this is a reasonable expectation and has been observed in an explicit solvent REMD simulation of a (AAARA)-repeat peptide[15,17], some simulations using GB solvation indicated that a single helix structure was not favored[16,17]. This is in disagreement with the present results using GB, but the differences may arise from differences in protein force field model employed in the two studies or possibly sensitivity of the GB model to the identity of the charged residue (D-Arg vs. Lys). Helix-turn-helix conformations have also been reported, as folding intermediates and transition states, in GB AMBER folding trajectory simulations for a (AAAKA)-repeat peptide, but the equilibrated folded state was a single helix with varying degrees of end-fraying.[57]

Although the three peptides show similar helical content profiles in Figure 3, the absolute values are different. The largest differences in the simulated ensembles are in the residues in the middle of the sequence: the peptides with charged residues (**K19** and **dR19**) have noticeably higher fractional helicities than the peptide without a positive side chain at the C-terminus (**A19**). This can be attributed to a favorable sidechain interaction with the helix macrodipole. The helical content profiles of **K19** and **dR19** are similar. (Other simulations, data not shown, reveal that further decreases in the central helicity of in conformational ensembles of **K19** analogs are very modest when the remaining Lys residues are mutated to Ala.) In the simulations shown in Fig. 3, the largest differences are between residues 14–16 in two peptides, with 10% greater helicity for the D-Arg species; this region is not well probed in the experimental study. The experimental probes at A11/A12 do, however, reflect the same trend seen in the simulations. The effects of the C-terminal residue are somewhat greater in the simulations than in the experimental data. The increasing helicity for the C-capping series, Ala → Lys → D-Arg was also experimentally verified for the $(KAAAA)_3$-XGY-NH_2 peptides (see Figure 2). In both the Ac-$(KAAAA)_3$-XGY-NH_2 and Ac-YGG-$(KAAAA)_3$-X-NH_2 series, central repeat Ala ^{13}C=O CSDs were comparably increased for X= D-Arg and L-Arg. In both cases, the NMR data indicates that the N-terminal KAAAA repeat is nearly as folded as the central one, with fraying most notable at the C-terminus; the simulations show nearly the same pattern.

Since the helical profiles from simulation and experiment are in reasonable agreement at low temperature, we examined the stability as a function of increasing temperature (Figure 4). The melting curves for seven Ala residues in an XGG-$(KAAAA)_3$K-span were obtained through both NMR experiments and simulations. These residues were chosen to probe the termini and the central portion of the sequence (which would be affected by equilibrium between single helix and helix-turn-helix conformations). Melting curves for the simulated ensembles were calculated in two ways: from local phi/psi values (panel A) and from the DSSP algorithm (panel B). The fractional helicities calculated from the chemical shift melts

(inset in Figure 4A) are in closer agreement with those from Lifson-Roig theory based on the phi/psi angle in the MD simulated ensembles. The melting curves for the central and N-terminal repeat sites were very similar, with enhanced melting observed in the less structured C-terminal repeat, particularly in the experimental data. In general, melting curves generated by the simulations are flatter than the ones from the NMR experiments, likely due to the lack of temperature dependence in the continuum solvent model that we employed.

Features of the Computed Helical Ensembles—Figure 5 shows the melting profile for average overall helical content in the peptide as assigned by Lifson-Roig theory. Two interesting observations can be drawn. One is that the D-Arg C-capped sequence displays enhanced stability in the lower temperature simulations. One possible rationale is that only a particular D-Arg conformation can stabilize the helix structure; as D-Arg becomes more disordered, the Lys and D-Arg capped sequences have similar helicities at temperatures higher than 300 K.

The predicted Tm for **K19** is 282K with $f_H = 0.58$ at 273K. There are a number of experimental determinations for comparison, all reported for Ac-YGG- (rather than Ac-GGG-) capped system. Baldwin's estimate, based on CD, was $f_H = 0.676$ at 273K[12]. Given the adjustment in the 100% helicity value (3.50 *versus* 3.04 ppm, in CSD terms) this is excellent agreement. FT-IR monitored T-jump experiments indicate that $k_F = k_U$ at 285K[58].

Number and length of helical segments: As previously noted, the high α-helical content at 275 K and flat profile across the central portion of the sequence (Figure 3) implies the predominant population of a conformation consisting of a single α-helix. To verify this feature of the simulations, we calculated the population of structures in the ensemble containing different number of continuous α-helix segments using the Lifson-Roig model, in which a helical segment is defined as 3 or more continuous residues adopting helical backbone conformation. We note that the resulting helix conformation populations (Figure 6) only take into account the number of helices, not their length (discussed below). At 275 K, more than half of structures (64(±0.5)% for **A19**, 69(±1)% for **K19**, and 71(±1.5)% for **dR19**) contain only a single continuous helix; this is also consistent with the general practice of applying the single sequence approximation in helix/coil treatments of short peptides. The population of conformations with a single helix decreases with increasing temperature slightly faster than the two helix conformation populations (helix-turn-helix bundle motifs).

Free energy landscape and structure families: To illustrate the structural ensembles sampled in our simulation, we constructed the free energy landscape using principle component analysis (PCA) of the ensemble sampled at 275K. Previous analyses showed that the three peptides have similar overall structural ensembles, therefore the peptide **K19** was chosen as a representative for the PCA landscape analysis. As described in Methods, the largest eigenvectors from PCA were used to define reaction coordinates for the landscapes, and relative free energies along these coordinates were calculated from histogram populations. We complemented these landscapes with cluster analysis of the ensemble of structures. Following cluster analysis, eigenvalues were calculated for samples of structures in each of the six most populated clusters (which together represent 78% of the ensemble), and these were projected onto the landscape in order to clarify the structural properties of the different basins. The results are shown in Figure 7, along with representative conformations which illustrate the general features adopted by structures in each basin. The arrows indicate their positions on the free energy landscape. In order to more fully characterize the extent of secondary structure formation in structural families, we calculated average helical content for each residue, similar to Figure 3, except that each cluster was considered separately. The results are shown in Figure 8, which provides insight into basis for the end-fraying seen in

Figure 3. The diminished net helicity in the early portion of the N-terminal repeat represents the contribution of two single helix conformers that begin later in the sequence and the lesser helicity in the N-terminal vs C-terminal helices of the two helix bundle conformations. All clusters within the ensemble display a similar progression of helix fraying at the C-terminus.

The calculated free energy landscape demonstrates that this peptide has a well-defined and deep global free energy minimum at 275K, approximately 1.5–2.0 kcal/mol lower than alternate local minima. Consistent with the Lifson-Roig analysis, the global free energy minimum corresponds to a family of conformations with a single long helix (Figure 7) with large fractional helicities for residues 5–17 (Figure 8). The population of this cluster is ~63%; the next two most populated clusters (clusters 3 and 2, with 4–5% population) are α-helix bundle structures, with each corresponding to a different local minimum on the free energy landscape. Both have turns near the middle of the sequence, as shown in Figure 8, but the turns occur at different points. This diversity in the location of the turn or coil in structures with 2 helical segments is also reflected in the observation that structures with 2 helical segments comprise ~20% of the low-temperature ensemble (Figure 6), yet none of the individual conformation clusters have a population of more than 5%.

In order to investigate the specific interactions that determine the relative stabilities of these conformation types, we performed energy and hydrogen bond analysis on the structures in the six largest clusters. Average values for number of intramolecular hydrogen bonds and energy components are shown in Table 2. The single long helix cluster (cluster 1) has an average of 2–3 more intramolecular hydrogen bonds than the other conformations, and as a result, a more favorable electrostatic energy. This is partially offset by a less favorable solvation energy (EGB) (>10 kcal/mol higher than other clusters). Several of the less populated clusters, including the two-helix bundle states, display a more favorable intramolecular van der Waals energy.

The radius of gyration at different temperatures: The radius of gyration is used to characterize the effective size and shape of the molecule. Figure 9 shows the radius of gyration (R_{gyr}) distributions at seven temperatures. To clarify the representation, the data are divided into groups according to temperature. Four temperatures, 253 K, 276 K, 300 K and 326 K are shown in Figure 9A. At these four temperatures, there are two peaks in the R_{gyr} distribution. One is located at 7.5 Å, which represents the single helix structures (cluster 1 in Figure 7). The other peak is near 6 Å, which represents more compact structures such as partial helices, helical bundles, and collapsed coils. The 7.5Å peak is most populated at temperatures ≤ 300 K. With increasing temperatures, the height of the single helix peak is reduced, and the height and the width of the globular state peak are increased slightly. At 354 K (Figure 9B), the peak for the single helix structure has disappeared; only the globular collapsed coil conformations remain. With increasing temperature, the peak shifts to higher values. At temperatures above 450K, the ensemble is composed of random coil structures, with a broad Gaussian distribution of R_{gyr} centered about 9.5 Å.

The Role of the Lysine Sidechains in Helices—MD simulations of, and theoretical conformation searches for, (AAAXA)-repeat peptides (X = R or K)[18,59] and some experimental studies[7, 60, 61] have shown that amino acids with charged sidechains can significantly enhance α–helix formation. Other experimental data indicate that central lysines in helices have a net destabilization effect (versus alanine at the same position).[12, 22, 31, 62, 63] The stabilization of helical content by the charged side chains, when observed in computational studies, arises from shielding of the α-helical intramolecular backbone hydrogen bonds from the solvation due to sequestration with the solvent localized at the sidechain terminus.[7, 18, 59] However, some previous GB simulations indicated that

helical conformations were actually destabilized by the presence of ionizable sidechains, due to competition between α-helical i-i+4 backbone hydrogen bonds and those involving the arginine guanidinium groups and backbone carbonyl oxygen atoms[17]. These backbone-sidechain interactions were not observed in similar simulations using explicit water molecules. The current argument concerning the effects of central polar residues in Ala-rich helices appears to focused on whether or not sidechain-induced backbone desolvation increases net helicity. Evidence has been presented that branched non-polar sidechains owe a portion of their helix destabilization to an enthalpic effect (backbone desolvation)[64] rather than only the classical unfolding entropy expectation. The present experimental studies only confirm a helix-favoring effect of *C-terminal* Lys/Arg residues, which likely reflects a Coulombic interaction with the helix macrodipole[65] rather than backbone desolvation or an H-bonding interaction.

These continuing issues prompted us to examine the role of lysine sidechains in our simulations. To determine the extent of Lys sidechain interaction with the backbone, we calculated hydrogen bond populations in the 275K simulated ensembles of peptide **K19**. The Lys9 site was selected as a representative because its location in the central region of the sequence would allow interaction with backbone residues on either the N-terminal or C-terminal side. The data shows that no significant sidechain hydrogen bonding is present in the ensemble, with all possible Lys9 Nζ - backbone carbonyl hydrogen bonds having less than 1.5% population (Figure S2). A similar situation applies to Lys14, but hydrogen bonding involving the sidechain amino group of Lys19 reaches populations approaching 5%. It is, however, unclear whether this is a statistically significant difference.

Since the controversy concerning lysine's potential α-helix stabilizing effects centers on Lys residues in the central helical repeat, we calculated the correlation between helix formation and Lys9's sidechain position. To quantify the position of the sidechain, two distances were measured: one is the distance between Ala5 Cα and Lys9 Nζ, the other is the distance between Ala13 Cα and Lys9 Nζ. The preference of Lys to interact with the N- and C-terminal end of the chain is indicated by the difference in the two distances (Figure S3). At low radius of gyration values corresponding to compact coil and helical bundle conformations, a broad, flat distribution of Lys distances is observed with no apparent preference for N-terminal or C-terminal direction: the slight bias toward the N-terminus may reflect the partial formation of helices in that region. For the larger radius of gyration values that correspond to extended helical conformations, the amino group is 5 Å closer to the Ala5 than to Ala13; in a significant conformational peak. For further insights, structures belonging to the single helix cluster were reclustered based on the Lys9 side chain conformation (Figure S4). In one cluster the methylene groups of the Lys9 sidechain shields the carbonyl oxygen of the upstream i-4 residue (A5) from solvent. In the other, the two distances are same: the sidechain is perpendicular to the axis of the helix. Interestingly, two highly similar conformations were reported in a previously explicit solvent simulations of a (AAARA)-repeat peptide[18]. Our implicit water simulations exhibit a similar preference for a configuration with protection of backbone hydrogen bonds by the nonpolar region of the ionizable sidechain, with the charged amino group interacting with solvent. This type of conformation for Lys sidechains in a helix is also supported by NMR experiments done by Groebke et al[60]: medium intensity NOEs were observed between the (i-3) Ala α-H and the β-, γ-, δ-H protons of the Lys, with a weak interaction with the ε-H. These NOEs were absent in the random coil model. Our simulation data (Figure S5) are consistent with these observations.

Conclusion

We have studied the temperature dependence of helical propensities for the peptides Ac-GGG-$(KAAAA)_3X$-NH_2 (X = A, K, and D-Arg) using replica exchange molecular dynamics simulations and the generalized Born continuum solvent model. In contrast to prior simulations of (AAARA)-repeat peptides with continuum solvent models, which showed that the structures that were most populated were two helix bundles[16,17], we found a large population (62%) of single-helix structures at 275 K, more than 10 times the size of any other conformation family. For X = Lys, the calculated Tm was 282 K, fully consistent with T-jump rate studies[58] of the corresponding Ac-YGG capped species. We did find some helix bundle structures (clusters 2 and 3, in Figure 7), and their contribution to the helical states increases somewhat at higher temperatures (reflected in the increasing peak at R_{gyr} = 6 Å in Figure 9). Compact structures that are helix bundles never account (Figure 6) for more than 20% of the total conformations (*and* ≤ 20% of the helical states) at ≤ 326 K. Above 326 K, the bundle conformations melt out more rapidly than the single helix state. Our continuum solvent simulations are, thus, in better agreement with explicit water simulations of (AAARA)-repeat peptides and support the common use of the single sequence approximation for helix-coil equilibrium calculations. The apparent discrepancy between our and prior GB helix simulations could be the result of differences in backbone parameters employed in the studies, although the high population of single helix conformation observed by the same group using explicit solvent[17] suggests that this is not the cause if the explicit solvent data is well converged. Of course, sequence differences in the two peptides are another possible reason for the discrepancy: peptide **K19** has solubilizing Lys residues rather than the Arg residues. We have shown that this GB model performs less well for Arg than Lys, leading to insufficient desolvation penalty for the Arg guanidinium group upon formation of intramolecular hydrogen bonds[66]. Peptide **K19** is, among the helical peptides that have been studied in MD folding simulations, most closely analogous to AK16, Ac-$YGAA(KAAAA)_2KA$-NH_2, which was found to be 64% helical at equilibrium with only minor contributions due to two-helix bundle conformations.[57] In the AK16 folding trajectories, helix-turn-helix conformations were viewed as folding transition states. The equilibrium folded state, with enhanced helicity starting abruptly after the Gly linker, appears late in the AK16 trajectory with the breaking of non-native hydrophobic interactions with the tyrosine viewed as the slow step in extended single helix formation. Similar observations in the initial stages of the present study prompted us to use Ac-$GGG(KAAAA)_3X$-NH_2 rather than Ac-$YGG(KAAAA)_3X$-NH_2 (X = A, K, and D-Arg) for the MD simulations herein.

Our simulation results are compared with data obtained from NMR chemical shifts of -$ZGG(KAAAA)_3X$-NH_2 and Ac-$(KAAAA)_3XGY$-NH_2 sequences (X = A, K, and D-Arg) and good agreement is found with both the absolute helical propensities as well as relative helical content along the sequence. Since the melting temperatures are also in reasonable agreement with the experimental results, detailed structural analysis of the simulation ensembles was performed. Cluster analysis showed that the global minimum on the calculated free energy landscape corresponds to a nearly full-length α-helical conformation. Helical bundle conformations were populated, but these local minima were 1.5–2.0 kcal/mol higher in free energy. Energy component analysis shows that the single helix state has favorable intramolecular electrostatic energy due to hydrogen bonds, while the globular states have favorable solvation energy due to the exposure of the polar atoms. The agreement in helicity profiles for the sequences examined in experimental and molecular dynamics ensembles indicates that MD will continue to be a useful method for examining the details of polypeptide secondary structuring.

The analysis of lysine sidechain conformations suggests that polar residues could favor helix formation by protecting the backbone hydrogen bonds; but the effect does not appear to offset the presumed entropic advantage of unfolding; Lys effects a net increase in helicity (versus Ala) only at the C-terminal position. This is in agreement with the view of Baldwin[64] rather than Scheraga[7,59] and Kemp[60] regarding the relative helix propensities of Lys and Ala and the consequences of backbone desolvation for Ala-rich helices. Here it worth noting that a more recent report[67] from the Kemp laboratory places the propagation value for alanine in an (Ala)[15] segment at 1.5, essentially the value we used to reproduce the experimental fractional helicities observed in the present study. A structural analysis of C-terminal ending group D-Arg in the simulated helical ensembles shows that it can from H-bonds with Ala16, which may provide some additional stability to helical structures. Both experimental and simulation studies show increasing helicity in the series X = Ala → Lys → D-Arg; however, none of the data sets show a specific DArg effect as large as that quoted in the initial reports (1.2 kcal/mol helix stabilization)[23] on this C-capping function: the largest net helical stabilization observed in the present NMR studies for a C-terminal Ala → D-Arg equates to less than a 0.4 kcal/mol helicity increment. A similar increase in helicity was observed in the MD studies. A modest increase in helicity due to placing Lys, Arg or D-Arg at the C-terminal position can be rationalized as an interaction with the helix macrodipole.[65] A resolution of the questions concerning the specific Ccapping effects of positively charged sidechains, particularly of the D-Arg function, will require studies of additional peptides with both L- and D-configured terminal groups and other polar functions. Here also, the comparison of experimental data and computational simulations should be particularly useful.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Studies at UW were supported by NSF grants (CHE-0315361 and -0650318) to NHA. Studies at SBU were supported by the National Institutes of Health (GM61678 to CS) and the National Computational Science Alliance Grant MCA02N028 (C.S.), which provided computational resources at NCSA.

REFERENCES

1. Doig AJ. Biophysical Chemistry. 2002; 101–102:281–293.
2. Barlow DJ, Thornton JM. J Mol Biol. 1988; 201:601–619. [PubMed: 3418712]
3. Schellman JA. J Phys Chem. 1958; 62:1485–1494.
4. Zimm BH, Bragg JK. J Chem Phys. 1959; 31:526–535.
5. Lifson S, Roig A. J Chem Phys. 1961; 34:1963–1974.
6. Doty P, Yang JT. J Am Chem Soc. 1956; 78:498–500.
7. Scheraga HA, Vila JA, Ripoll DR. Biophysical Chemistry. 2002; 101:255–265. [PubMed: 12488006]
8. Pan KM, Baldwin M, Nguyen J, Gasset M, Serban A, Groth D, Mehlhorn I, Huang ZW, Fletterick RJ, Cohen FE, Prusiner SB. Proc Natl Acad Sci U S A. 1993; 90:10962–10966. [PubMed: 7902575]
9. Jayawickrama D, Zink S, Vandervelde D, Effiong RI, Larive CK. J Biomol Struct Dyn. 1995; 13:229–244.
10. Marqusee S, Robbins VH, Baldwin RL. Proc Natl Acad Sci U S A. 1989; 86:5286–5290. [PubMed: 2748584]
11. Huang CY, Klemke JW, Getahun Z, DeGrado WF, Gai F. J Am Chem Soc. 2001; 123:9235–9238. [PubMed: 11562202]
12. Doig AJ, Baldwin RL. Protein Science. 1995; 4:1325–1336. [PubMed: 7670375]

13. Rohl CA, Chakrabartty A, Baldwin RL. Protein Science. 1996; 5:2623–2637. [PubMed: 8976571]
14. Andersen NH, Tong H. Protein Science. 1997; 6:1920–1936. [PubMed: 9300492]
15. Jas GS, Kuczera K. Biophys J. 2004; 87:3786–3798. [PubMed: 15339816]
16. Zhang W, Lei HX, Chowdhury S, Duan Y. J Phys Chem B. 2004; 108:7479–7489.
17. Nymeyer H, Garcia AE. Proc Natl Acad Sci U S A. 2003; 100:13934–13939. [PubMed: 14617775]
18. Garcia AE, Sanbonmatsu KY. Proc Natl Acad Sci U S A. 2002; 99:2782–2787. [PubMed: 11867710]
19. Baldwin RL. J Biol Chem. 2003; 278:17581–17588. [PubMed: 12582164]
20. Miller JS, Kennedy RJ, Kemp DS. Biochemistry. 2001; 40:305–309. [PubMed: 11148022]
21. Miller JS, Kennedy RJ, Kemp DS. J Am Chem Soc. 2002; 124:945–962. [PubMed: 11829602]
22. Lin JC, Barua B, Andersen NH. J Am Chem Soc. 2004; 126:13679–13684. [PubMed: 15493925]
23. Schneider JP, DeGrado WF. J Am Chem Soc. 1998; 120:2764–2767.
24. Fesinmeyer RM, Peterson ES, Dyer RB, Andersen NH. Protein Science. 2005; 14:2324–2332. [PubMed: 16131660]
25. Stewart JM, Lin JC, Andersen NH. manuscript submitted to Chem Comm.
26. Andersen NH, Brodsky Y, Neidigh JW, Prickett KS. Bioorganic & Medicinal Chemistry. 2002; 10:79–85. [PubMed: 11738609]
27. Andersen NH, Liu ZH, Prickett KS. FEBS Letters. 1996; 399:47–52. [PubMed: 8980117]
28. Schwarzinger S, Kroon GJA, Foss TR, Chung J, Wright PE, Dyson HJ. J Am Chem Soc. 2001; 123:2970–2978. [PubMed: 11457007]
29. Xu XP, Case DA. Biopolymers. 2002; 65:408–423. [PubMed: 12434429]
30. Park SH, Shalongo W, Stellwagen E. Proteins Struct Funct Genet. 1998; 33:167–176. [PubMed: 9779786]
31. Lin, J. PhD thesis. Chemistry Department, University of Washington; 2007.
32. Sugita Y, Okamoto Y. Chem Phys Lett. 1999; 314:141–151.
33. Sugita Y, Okamoto Y. Progress of Theoretical Physics Supplement. 2000:402–403.
34. Schug A, Wenzel W, Hansmann UHE. J Chem Phys. 2005; 122 194711/1-7.
35. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. J Computational Chemistry. 2005; 26:1668–1688.
36. Roe DR, Hornak V, Simmerling C. J Mol Biol. 2005; 352:370–381. [PubMed: 16095612]
37. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. J Am Chem Soc. 1995; 117:5179–5197.
38. Okur A, Strockbine B, Hornak V, Simmerling C. J Computational Chemistry. 2003; 24:21–31.
39. Zaman MH, Shen MY, Berry RS, Freed KF, Sosnick TR. J Mol Biol. 2003; 331:693–711. [PubMed: 12899838]
40. Simmerling C, Strockbine B, Roitberg AE. J Am Chem Soc. 2002; 124:11258–11259. [PubMed: 12236726]
41. Hawkins GD, Cramer CJ, Truhlar DG. Chem Phys Lett. 1995; 246:122–129.
42. Hawkins GD, Cramer CJ, Truhlar DG. J Phys Chem. 1996; 100:19824–19839.
43. Tsui V, Case DA. Biopolymers. 2000; 56:275–291. [PubMed: 11754341]
44. Ryckaert JP, Ciccotti G, Berendsen HJC. J Computational Physics. 1977; 23:327–341.
45. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. J Chem Phys. 1984; 81:3684–3690.
46. Hong Q, Schellman JA. J Phys Chem. 1992; 96:3987–3994.
47. Zagrovic B, Pande VS. Nature Structural Biology. 2003; 10:955–961.
48. Kabsch W, Sander C. Biopolymers. 1983; 22:2577–2637. [PubMed: 6667333]
49. Simmerling C, Kollman P. Abstracts of Papers of the American Chemical Society. 1996; 211 92-Comp.
50. Luo PZ, Baldwin RL. Biochemistry. 1997; 36:8413–8421. [PubMed: 9204889]
51. Andersen NH, Cort JR, Liu ZH, Sjoberg SJ, Tong H. J Am Chem Soc. 1996; 118:10309–10310.

52. Neidigh JW, Fesinmeyer RM, Andersen NH. Nature Structural Biology. 2002; 9:425–430.
53. Andersen NH, Dyer RB, Fesinmeyer RM, Gai F, Liu ZH, Neidigh JW, Tong H. J Am Chem Soc. 1999; 121:9879–9880.
54. Wishart DS, Sykes BD. Methods Enzymol. 1994; 239:363–392. [PubMed: 7830591]
55. Chakrabartty A, Kortemme T, Padmanabhan S, Baldwin RL. Biochemistry. 1993; 32:5560–5565. [PubMed: 8504077]
56. Munoz V, Serrano L. Biopolymers. 1997; 41:495–509. [PubMed: 9095674]
57. Chowdhury S, Zhang W, Wu C, Xiong G, Duan Y. Biopolymers. 2003; 68:63–75. [PubMed: 12579580]
58. Werner JH, Dyer RB, Fesinmeyer RM, Andersen NH. J Phys Chem B. 2002; 106:487–494.
59. Vila JA, Ripoll DR, Scheraga HA. Proc Natl Acad Sci U S A. 2000; 97:13075–13079. [PubMed: 11078529]
60. Groebke K, Renold P, Tsang KY, Allen TJ, McClure KF, Kemp DS. Proc Natl Acad Sci U S A. 1996; 93:4025–4029. [PubMed: 8633010]
61. Williams L, Kather K, Kemp DS. J Am Chem Soc. 1998; 120:11033–11043.
62. Chakrabartty A, Kortemme T, Baldwin RL. Protein Science. 1994; 3:843–852. [PubMed: 8061613]
63. Spek EJ, Olson CA, Shi Z, Kallenbach NR. J Am Chem Soc. 1999; 121:5571–5572.
64. Luo PZ, Baldwin RL. Proc Natl Acad Sci U S A. 1999; 96:4930–4935. [PubMed: 10220396]
65. Couch VA, Cheng N, Nambiar K, Fink W. J Phys Chem B. 2006; 110:3410–3419. [PubMed: 16494355]
66. Geney R, Layten M, Gomperts R, Hornak V, Simmerling C. J Chemical Theory and Computation. 2006; 2:115–127.
67. Job GE, Kennedy RJ, Heitmann B, Miller JS, Walker SM, Kemp DS. J Am Chem Soc. 2006; 128:8227–8233. [PubMed: 16787087]

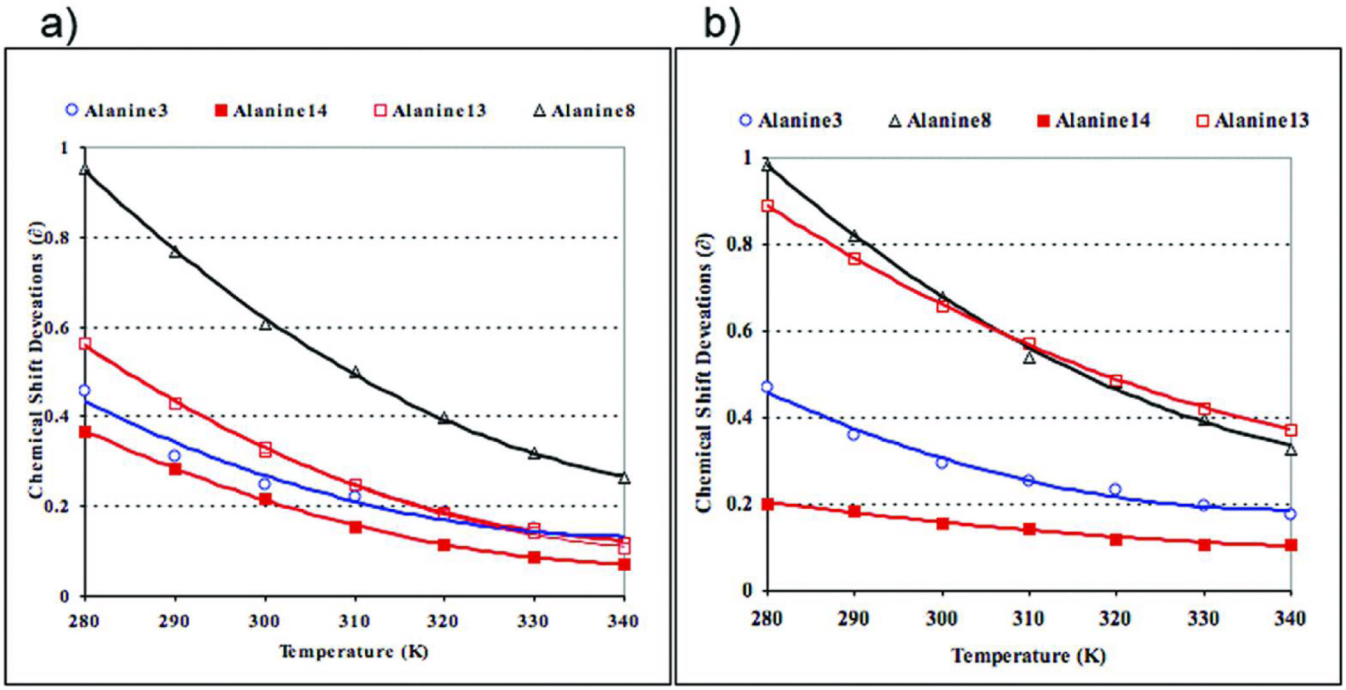

FIGURE 1.

Structuring shift melts for the non-acetylated $(KAAAA)_3X$ peptides with X = Lys (a) and D-Arg (b). Random coil shifts were derived from AcGKAAAKG-NH_2 and thus assume no differences in sequence effect for Lys and D-Arg. The more remote probes, A3 and A8, are, presumably, not influenced by the C-terminal substitution; these indicate that the D-Arg species is slightly more helical and resistant to melting.

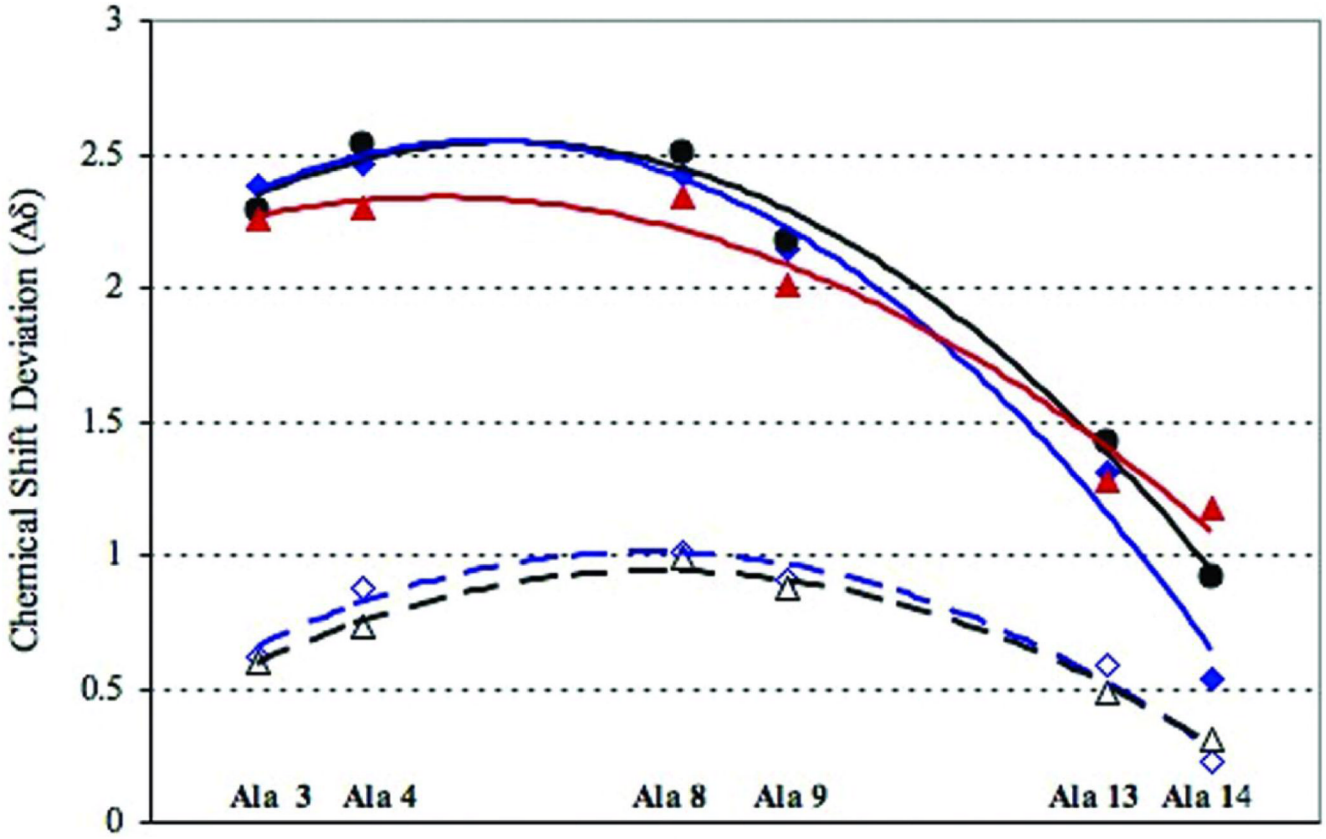

FIGURE 2.

The CSDs observed for Ac-$(KAAAA)_3XGY-NH_2$ (X = A, K, D-Arg) at the probed helix positions (initial K = #1) at 280K are shown as filled symbols. The corresponding points for $KAAAA(KAAAA)_2$-X-$GYNH_2$ (X = K, D-Arg) are shown as open symbols. The lines are trendlines drawn through the data to guide the eye and have no theoretical significance. Throughout C-caps are color-coded: X = D-Arg points and lines are in blue, Lys in black, and Ala in red. The same color-coding is used in Figure 3.

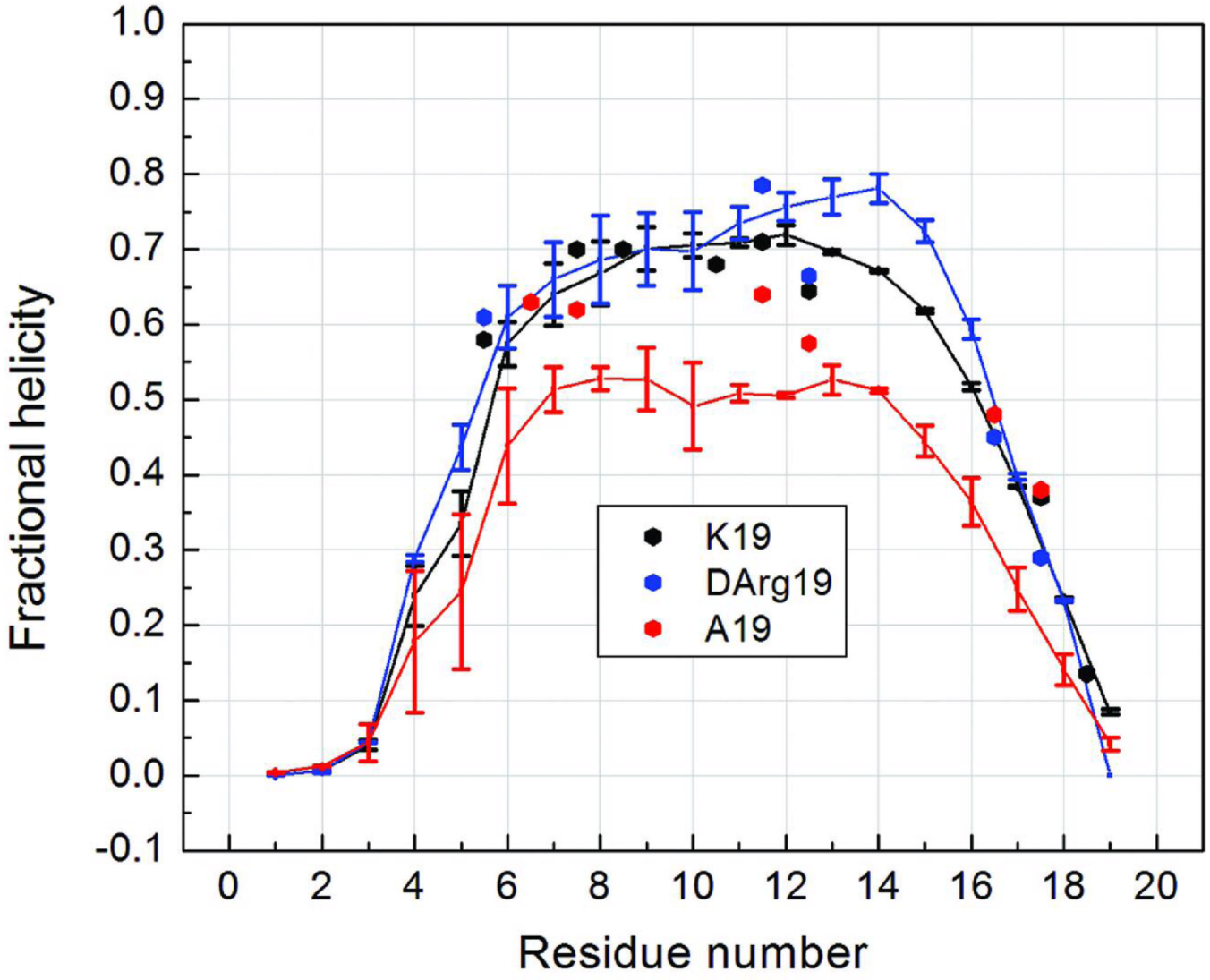

FIGURE 3.

Fractional helicities at each residue of peptides **A19, K19** and **dR19**. The continuous lines connect perresidue values obtained from REMD ensembles at 275 K. The experimental values (points) are from ^{13}C-NMR for the corresponding Ac-YGG-capped peptides; the **A19** data was for the peptide lacking the N-terminal acetyl. Since the NMR data reports on the amide linkage between residue i and i+1, we place the experimental point based on the ^{13}C=O CSD of residue i halfway between i and i+1 on this plot.

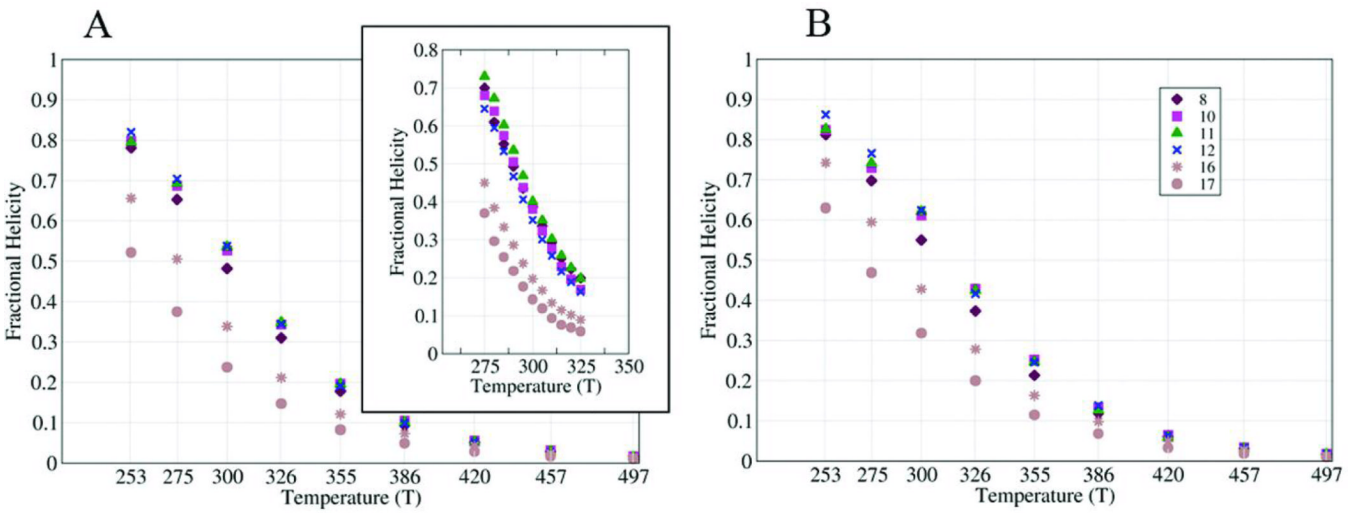

FIGURE 4.

Average α-helical propensities of representative residues at different temperatures. Different symbols are used for each residue examined. Helical propensity was calculated using local backbone conformation of the residue (left) and DSSP (right). The inset in A shows the NMR shift melts.

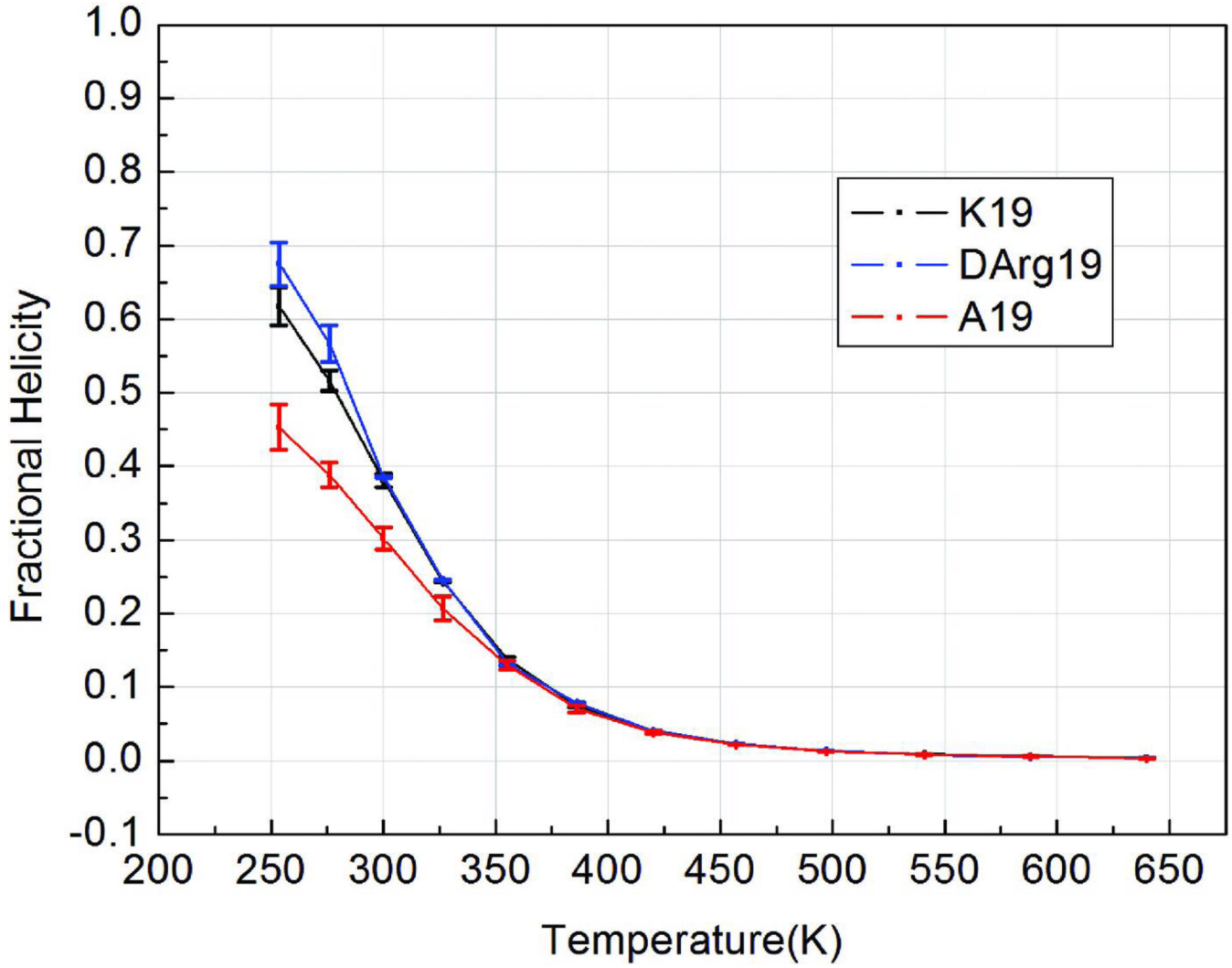

FIGURE 5.
Lifson-Roig based melting curves for the helical segment (K4 through X19, normalized for this region of the three peptides.

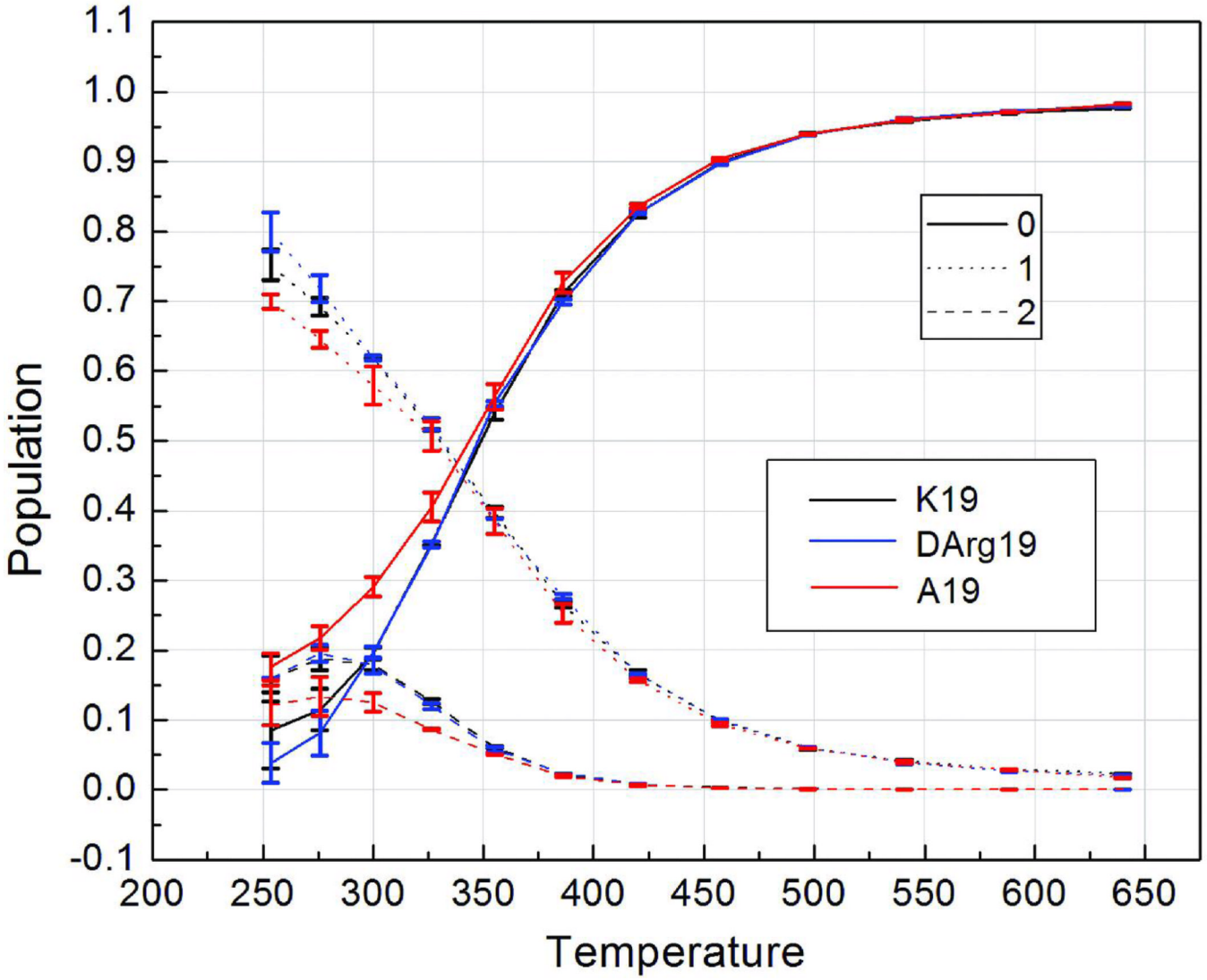

FIGURE 6.

Temperature-dependent populations of structures containing different number of continuous helical segments (shown using different line styles) sampled in the REMD simulation ensembles. The numbers in the legend refer to the number of continuous helical segments. Different peptide sequences are indicated using different colors.

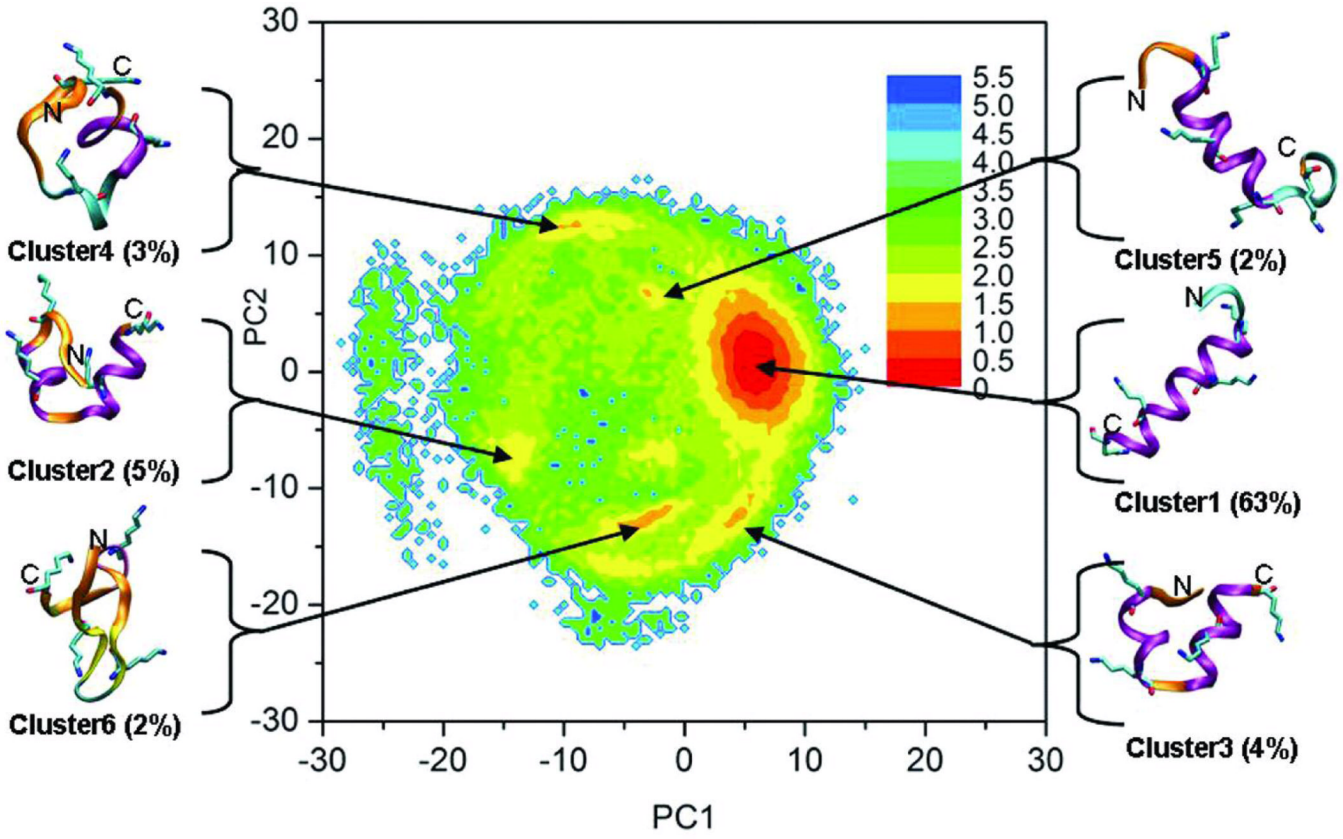

FIGURE 7.

The free energy landscape for peptide **K19** at 275 K, along with representative structures obtained from cluster analysis of the ensemble. X and Y axes represent the first 2 principle components. Relative free energy values (in kcal/mol) are represented by color as indicated by the legend. Representative structures for each basin are shown. The colors of the structures reflect secondary structure type: alpha helix – purple, extended beta – yellow, turn – cyan, coil – orange. Only lysine side chains are shown. The populations of the clusters are shown in parentheses.

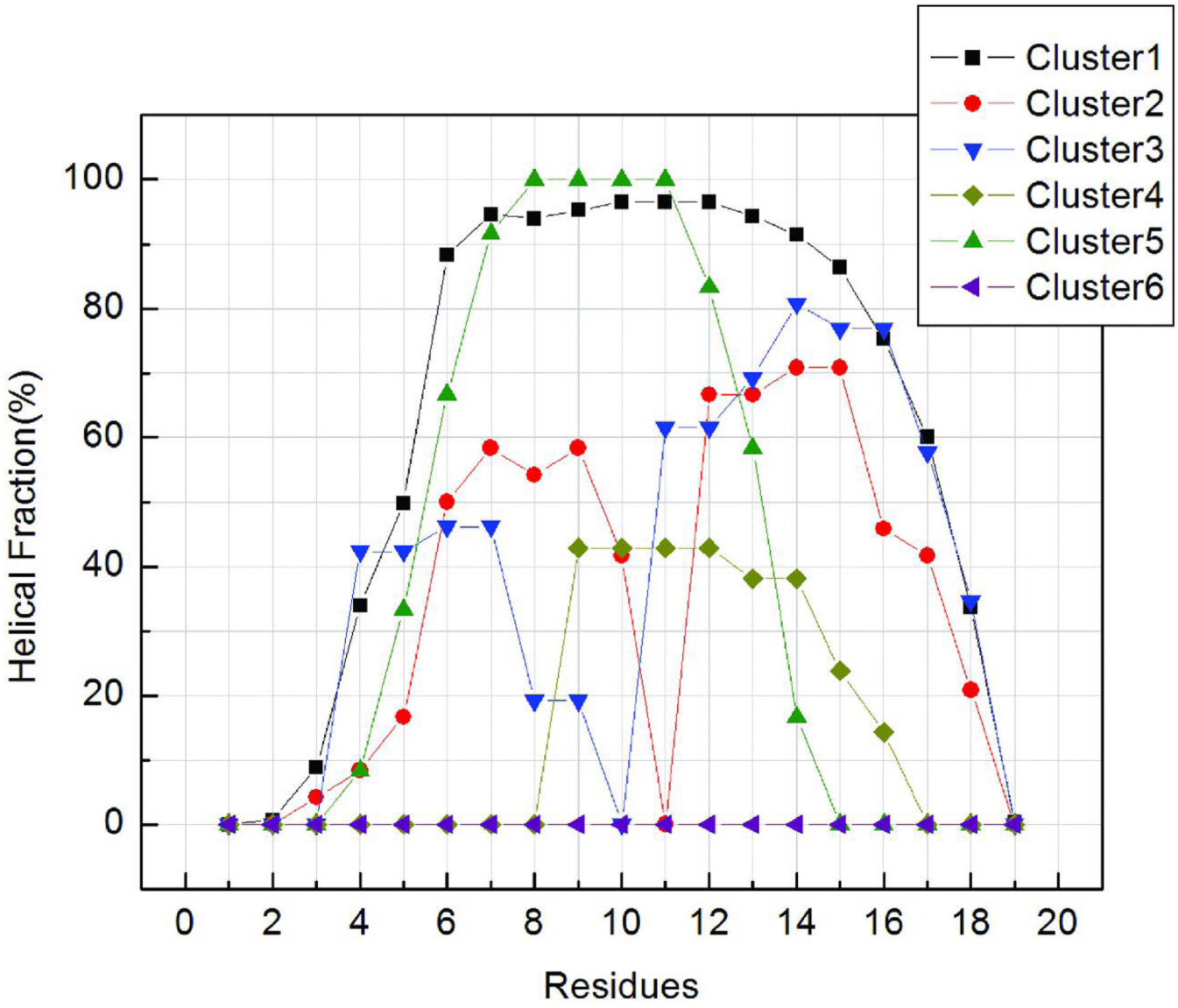

FIGURE 8.

The normalized average helical propensity as a function of sequence for each cluster shown in Figure 7. Similar to the overall ensemble shown in Figure 3, helical content is reduced at the termini of most clusters. Most clusters also have nearly flat profiles along the middle of the sequence, with the exception of clusters 2 and 3 which correspond to helix-turn-helix motifs. Cluster 6 shows no helical content.

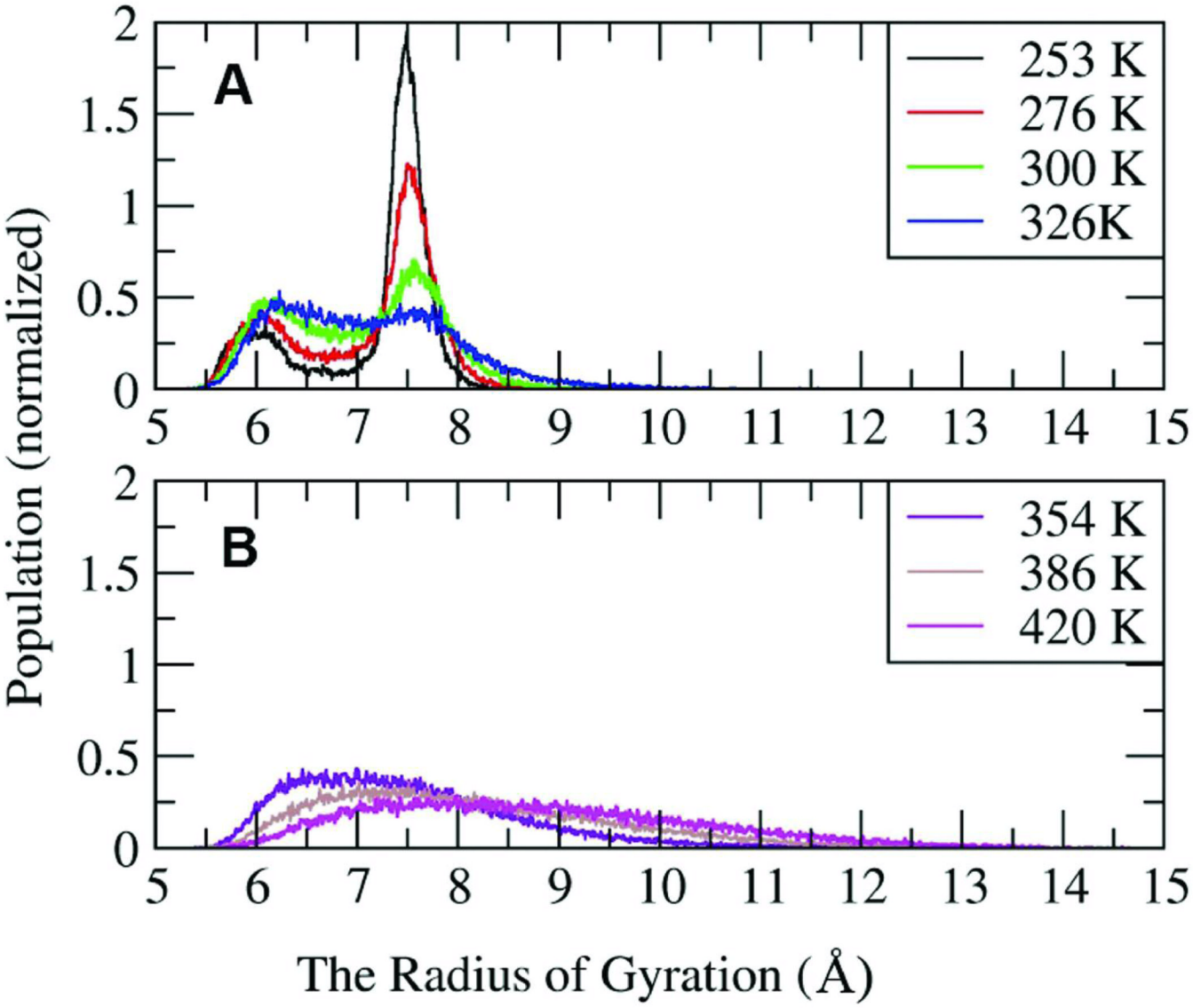

FIGURE 9.
The distributions of the radius of gyration for peptide **K19** at different temperatures.

TABLE I

NMR Structuring Shifts and CD Measures of Helicity for Peptides with Different N- and C-Capping Motifs

Peptide Sequence	^{13}C=O structuring shifts at 280K			CD, 278K
	A4	A8	A9	$-[\theta]_{222}$
Ac-GGG(KAAAA)$_3$K-NH$_2$	2.37	2.41	n. d.	n.d.
YGG(KAAAA)$_3$A-NH$_2$	1.98	2.07	1.82	18,700
Ac-YGG(KAAAA)$_3$A-NH$_2$	2.29	2.28	2.01	16,700
Ac-YGG(KAAAA)$_3$K-NH$_2$	2.22	2.33	1.98	21,400[a]
Ac-YGG(KAAAA)$_3$R-NH$_2$	n. d.	2.51	n. d.	19,300
Ac-YGG(KAAAA)$_3$r-NH$_2$	n. d.	2.45	2.05	18,700[b]
(KAAAA)$_3$rGY-NH$_2$	0.87	1.01	0.90	2,900
Ac-(KAAAA)$_3$AGY-NH$_2$	2.30	2.34	2.01	11,900
Ac-(KAAAA)$_3$KGY-NH$_2$	2.54	2.51	2.18	12,200
Ac-(KAAAA)$_3$rGY-NH$_2$	2.47	2.43	2.16	14,300
Ac-(KAAAA)$_3$KAAAAKGY-NH$_2$	2.94	3.13	2.85	18,700
(KAAAA)$_3$KAAAAKGY-NH$_2$	1.54	2.18	2.08	n.d.

[a] Prior value from a graph reported by Andersen et al.[51]

[b] Huang et al.[11] reported $[\theta]_{222}$. = −18,300°.

Table 2

The average number of hydrogen bonds and average values of nonbonded energy components (in kcal/mol) for the six structural clusters shown in Figure 7. Electrostatic is the intramolecular electrostatic energy. VDW is the intramolecular van der Waals energy. EGB is the electrostatic component of solvation energy as calculated by the generalized Born model. The free energy relative to cluster #1 is calculated based on the population of the clusters. The total energy is lowest for cluster 1, which has the highest population, and similar for all of the other clusters except #6, which has no helical content.

	# Intramolecular H-bonds	Electrostatic	VDW	EGB	Sum	Free energy
Cluster1	11.3	142.8	−29.4	−488.3	−374.8	0
Cluster2	8.3	168.3	−33.4	−504.9	−369.9	1.43
Cluster3	9.1	160.7	−30.9	−498.8	−369.1	1.45
Cluster4	8.8	167.2	−31.2	−504.7	−368.8	1.70
Cluster5	8.3	162.5	−26.4	−503.7	−367.6	1.78
Cluster6	6.4	170.6	−32.3	−503.6	−365.3	1.83