Solution Structure of a Bent α-Helix[†,‡]

Li Zhang[§] and Dimitrios Morikis*[,||]

Departments of Chemistry and of Bioengineering, University of California, Riverside, California 92521

Received June 25, 2007; Revised Manuscript Received August 28, 2007

ABSTRACT: We present the solution structure determination of a peptide with sequence Ac-WEAQAR-EALAKEAQARA-NH_2, using NMR data. The peptide has a high population of a stable α-helical structure in the middle with fraying ends. In addition, our data are consistent with the presence of several other transient and interconverting conformers. The peptide sequence was designed using amino acids that have propensity for α-helix specificity. The presence of E−R/K (i, $i + 4$) ion pairs was expected to enhance the stability of the α-helix by introducing favorable electrostatic interactions at the side chain level, in addition to the characteristic backbone (i, $i + 4$) hydrogen bonds. The NMR structural ensemble demonstrates competition between E−R/K (i, $i + 4$) and (i, $i - 1$) ion pair formation, with the (i, $i - 1$) interactions being dominant. The relative topologies of the charged side chains demonstrate flexibility and overall compromised favorable medium/long-range electrostatic interactions. The peptide α-helix is bent despite the lack of an amphipathic sequence. Bending is introduced by a strong (i, $i + 8$) hydrophobic interaction between the side chains of N-terminal tryptophan and leucine at the middle of the peptide sequence.

We examine the role of ion pairs and a hydrophobic pair in the formation of a single α-helix in a 17-residue model peptide with blocked termini. Ion pairs have been observed to stabilize the helical structures of peptides with sequences derived from native proteins or of de novo design (reviewed in ref *1*). In particular, pentapeptide blocks of the type (E/DAAAK/R)$_n$ have been proposed to promote α-helical structures (*2*−*6*). The middle alanine residues promote helix specificities (*7*−*9*), whereas the end-residue pairs E/D−K/R are expected to contribute to the stability of the α-helices through (i, $i + 4$) Coulombic interactions (*1*−*7*). These ion pairs are designed to enhance the contributions of (i, $i + 4$) hydrogen bonds in the stabilities of α-helices.

We have determined the solution structure of the 17-residue model peptide with sequence Ac-W-EAQAR-EALAK-EAQAR-A-NH_2. With the exception of the first and last residues, W1 and A17, the remaining 15 residues form three pentapeptide blocks which are variations of the EAAAR sequence. We have introduced polar residues in the middle of the first and last block by replacing A4 and A14 with Q. The polar residues contribute to the solubility of the peptide. We have also introduced a hydrophobic residue in the center of the middle block by replacing A9 with L. We have also introduced variation in amino acid branching, while maintaining the charge in the middle block, by replacing R11 with K. Because of a lack of branching, lysines are less conformationally restrained than arginines, which contributes to additional flexibility of the middle block. All A, L, E, Q, K, and R residues have preferences for α-helical conformations in peptides and proteins (*10*, *11*). The peptide's first and last blocks are identical, but some asymmetry is introduced by the middle block and the first and last residues that lie outside the end blocks. Overall sequence asymmetry contributes to the spectral resolution of the NMR[1] spectra, which in turn contributes to successful assignments and analysis. The peptide is blocked at the termini to eliminate competing charge−charge or charge−dipole effects from the charged termini (*2*).

A similar peptide, called Q4L9, was shown in an earlier NMR study to possess predominantly a highly α-helical structure, but also other conformations (*6*). This is not unusual, since peptides in solution are dynamic, undergoing conformational interconversions. The Q4L9 peptide has an alanine at position 14, whereas the peptide we study here has a glutamine, the remaining sequence being identical. The present work extends the previous one by performing a complete structure determination using NMR-derived restraints. Our analysis focuses on the competition of (i, $i + 4$) and (i, $i - 1$) E−R/K ion pair formation and their contribution to the peptide helicity. Also, our analysis focuses on the formation of an (i, $i + 8$) hydrophobic pair and its contribution in bending the helical structure. Bending is a typical characteristic of α-helices with amphipathic sequences in aqueous environments, being promoted by clustering of hydrophobic residues on one part of the helix surface (bend interior) and by the presence of polar residues on the

[†] This work was supported by NSF Grant 0427103.

[‡] The coordinates of the NMR ensemble of 10 structures and the average minimized structure and lists of chemical shifts and distance and torsion angle restraints are deposited at the Biological Magnetic Resonance Data Bank (BMRB; http://www.bmrb.wisc.edu/) with Accession No. 11005.

* To whom correspondence should be addressed. Phone: (951) 827-2696. Fax: (951) 827-6416. E-mail: dmorikis@engr.ucr.edu.

[§] Department of Chemistry.

[||] Department of Bioengineering.

[1] Abbreviations: NMR, nuclear magnetic resonance; TOCSY, total correlation spectroscopy; DQF-COSY, double-quantum-filtered correlation spectroscopy; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; PDB, protein data bank; BLAST, basic local alignment search tool.

10.1021/bi701252n CCC: $37.00
Published on Web 10/17/2007

remaining part (bend exterior). Our peptide sequence is not amphipathic, but bending is promoted through the (i, $i + 8$) single hydrophobic pair interaction.

MATERIALS AND METHODS

The peptide Ac-WEAQAREALAKEAQARA-NH_2 was purchased from EZBiolab, Inc. (Westfield, IN). The peptide was shown to have a molecular mass of 1939.9 Da [$M + H^+$] by ESI mass spectrometry. The NMR sample was prepared in 90% H_2O−10% D_2O. The sample concentration was 2−3 mM. The pH was adjusted to ∼5 using microliter quantities of HCl/NaOH. NMR spectra were collected using a 600 MHz Bruker Avance 600 spectrometer. Proton 1D and 2D TOCSY, DQF-COSY, and NOESY spectra were collected at 278 K. The mixing times were 80 ms for the TOCSY spectrum and 200 and 400 ms for the NOESY spectra. The program NMRDraw (*12*) was used for spectral processing. The 400 ms NOESY spectrum was used to extract distance restraints. There was no evidence for unusual spin diffusion effects by comparing the 200 and 400 ms spectra. An earlier study for the similar Q4L9 peptide has assessed possible spin diffusion effects for selected NOE cross-peaks for mixing times in the range of 25−200 ms (*6*). The NMR spectra were calibrated using the H_2O resonance (4.962 ppm at 278 K). The program NMRView (*13*) was used for spectral analysis, the construction of chemical shift and NOE assignment databases, the calculation of NOE peak heights, and the conversion of peak heights to distance restraints. Peak heights were chosen for our analysis because they are more accurate than volumes in cases of overlapping NOE cross-peaks. The bin method of NMRView was used for NOE-to-distance calibration. The DQF-COSY spectrum was collected with high digital resolution in t_2 to extract accurate scalar $^3J(H^N-H^\alpha)$ coupling constant values after processing and analysis. The method of Kim and Prestegard (*14*) was employed to evaluate coupling constants from DQF-COSY cross-peaks, using peak-to-peak separations from the upper and lower lobes of the spectra. In addition, coupling constants were extracted by measuring peak-to-peak separations in 1D spectra for four resonances, where observed peak splittings were present. The $^3J(H^N-H^\alpha)$ coupling constants were converted to φ-torsion angle restraints by solving the Karplus equation with the coefficients of ref *15*. The χ_1 torsions were assigned to one of the three staggered conformations of methylene groups, by analyzing NOE and DQF-COSY cross-peak patterns and intensities. The errors for the torsion angles were ±30° for φ extracted from coupling constants of <7 Hz and ±40° for φ of A17, which exhibited a coupling constant of >7 Hz, and for the χ_1 torsion angles.

The program CNS (*16*) was used for the structure calculation using molecular dynamics-based simulated annealing (MDSA) in torsion angle space. Other protocols, such as MDSA in Cartesian coordinate space and hybrid distance geometry−MDSA in torsion angle space and Cartesian coordinate space, were used, but without showing significant differences in the calculated NMR ensembles of structures. The default CNS potential energy function and parameters were used. The following bounds were used for the conversion of NOEs to distances: (i) 1.8 Å ≤ r_{ij} ≤ 2.8 Å for strong NOEs; (ii) 1.8 Å ≤ r_{ij} ≤ 3.4 Å for medium NOEs; (iii) 1.8 Å ≤ r_{ij} ≤ 5.0 Å for weak NOEs; (iv) 1.8 Å ≤ r_{ij} ≤ 6.0 Å for very weak NOEs. Corrections of 0.2 and 0.5 Å were added in the upper bounds of distances derived from NOEs involving amides and methyls to account for spin diffusion effects of the amide protons and rotational averaging of the methyl protons. Two β-methylene protons were stereospecifically assigned. The summation method was used for effective distances involving degenerate methylene protons that were not stereospecifically assigned and methyl groups (*17*).

The program PROCHECK-NMR (*18*) was used for structure validations. The programs VMD (*19*) and MOLMOL (*20*) were used for structural analysis. VMD was used for the preparation of molecular graphics. MOLMOL was used to calculate hydrogen bonds. The program CHARMM (*21*) was used to calculate van der Waals (Lennard-Jones 6−12) and electrostatic (Coulombic) interactions. A dielectric constant of 80 was used for the latter. The program WINPEP (*22*) was used to plot the peptide sequence on a helical wheel.

RESULTS

Resonance assignments, spin system identification, and NOE assignments were performed using a combination of 2D TOCSY and NOESY spectra. Figure S1 in the Supporting Information shows the resonance assignments and the identification of each amino acid residue in the 2D TOCSY spectrum. Table S1 in the Supporting Information provides a chemical shift list. Figure S2 in the Supporting Information shows the backbone amide region of the 1D NMR spectrum. The presence of peak splittings for certain amides demonstrates the quality of the 1D spectrum for extraction of $^3J(H^N-H^\alpha)$ coupling constants. Figure S3 in the Supporting Information shows the peptide backbone in a DQF-COSY spectrum. Figure 1A shows the backbone $H^\alpha-H^N$ region of the 2D NOESY spectrum and the $H^\alpha-H^N$ (i, i) and (i, $i + 1$) NOE connectivities. Figure 1B shows the backbone H^N-H^N region of the 2D NOESY spectrum. The peptide helical structure is demonstrated by the presence of medium/strong $H^\alpha-H^N$ and H^N-H^N (i, $i + 1$) and $H^\alpha-H^N$ (i, $i + 3$) NOEs and weak $H^\alpha-H^N$ (i, $i + 2$) and $H^\alpha-H^N$ (i, $i + 4$) NOEs (*23*). Figure 2 summarizes the NOE connectivity patterns. The peptide possesses a helical structure as demonstrated by the presence of $H^\alpha-H^N$, $H^\beta-H^N$, and H^N-H^N (i, $i + 1$) NOEs, $H^\alpha-H^N$ and $H^\alpha-H^\beta$ (i, $i + 3$) NOEs, and $H^\alpha-H^N$ (i, $i + 2$) and $H^\alpha-H^N$ (i, $i + 4$) NOEs (*23*). The uneven distribution of (i, $i + 3$) NOEs along the peptide sequence may be owed to the large number of overlapping cross-peaks, which impedes unambiguous NOE assignments (Figure 2). The presence of less overlapped cross-peaks at the N-terminus is attributed to ring current shifts induced by W1. The presence of both $H^\alpha-H^N$ (i, $i + 4$) and (i, $i + 2$) NOE assignments alone cannot exclude the presence of either or mixed α- and 3_{10}-helices (*23*). A small number of weak or very weak NOEs at the N- and C-termini not consistent with helical structure are also observed (data not shown), suggesting the presence of other conformations with low or very low populations. This is in agreement with an earlier study (*6*).

Figure 3A shows a plot of the differences of the observed chemical shifts for the backbone H^α atoms and their random coil values, as a function of the peptide sequence. The negative difference demonstrates the presence of helical

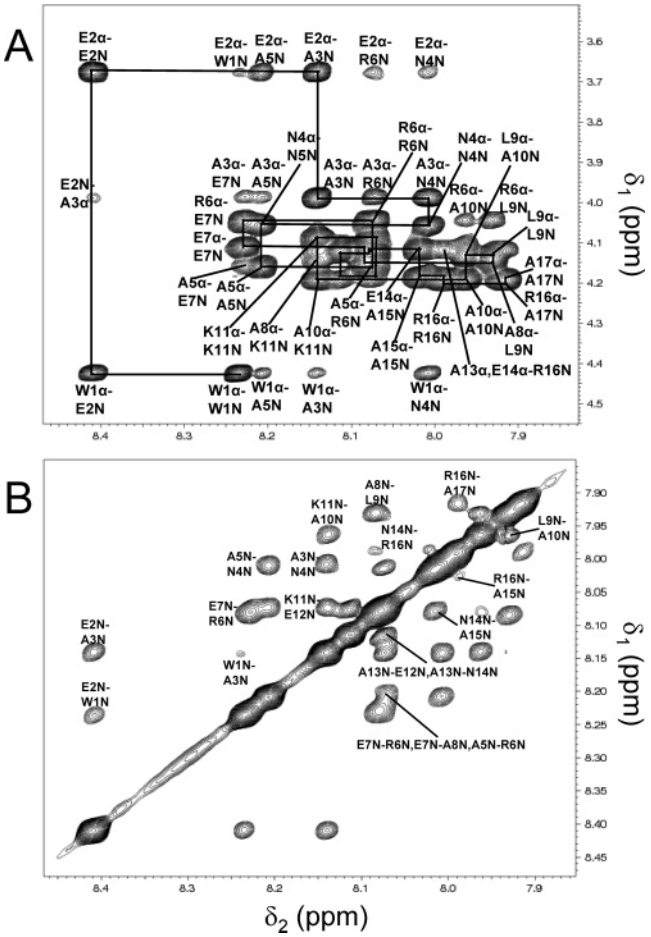

FIGURE 1: Portions of the 2D NOESY spectrum: (A) backbone H^{α} (δ_1)−H^N (δ_2) region, with all NOEs marked and the H^{α}−H^N (*i*, *i*) and (*i*, *i* + 1) NOEs shown; (B) backbone H^N (δ_1)−H^N (δ_2) region, with all NOEs marked once either above or below the diagonal.

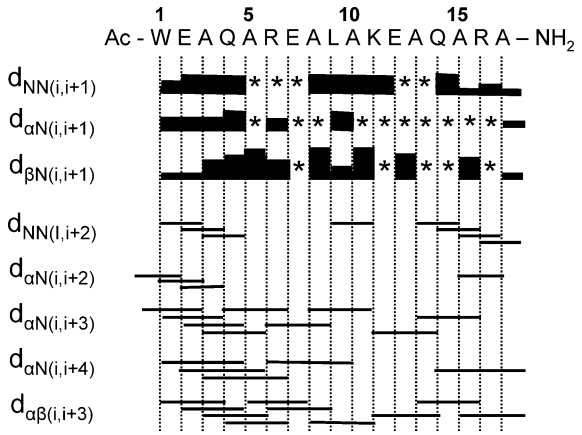

FIGURE 2: NOE connectivity patterns. The asterisks denote ambiguity in the NOE assignment because of cross-peak overlap. The bar height for the (*i*, *i* + 1) NOEs depicts their intensity (very strong, strong, medium, weak). The figure demonstrates NOE evidence for the presence of helical structure.

structure (reviewed in ref *25*). Overall, the N-terminal region (first pentapeptide block) appears to have a higher propensity for helical structure than the rest of the peptide. The mean H^{α} chemical shift difference ± SD for all residues is −0.24 ± 0.12 ppm. For protein helices, the mean H^{α} chemical shift difference is −0.39 ppm and the range is −0.15 to −0.60 ppm (*25*). In the case of helical peptides, smaller negative

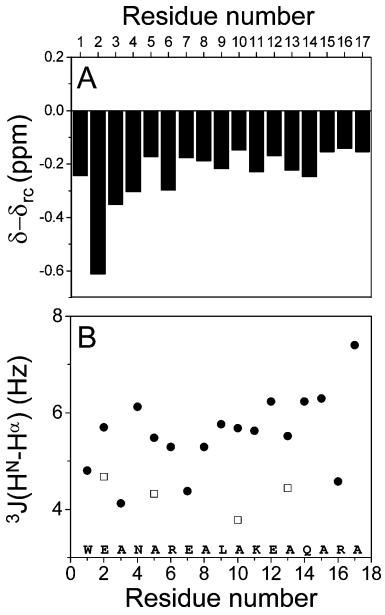

FIGURE 3: (A) Differences between measured chemical shifts and their random coil values against the peptide sequence. The random coil values at the temperature of our experiments (278 K) are from ref *24*. (B) $^3J(H^N-H^{\alpha})$ coupling constant values against the peptide sequence. Solid circles are data derived from the 2D DQF-COSY spectrum, and open squares are data from the 1D spectrum.

differences are not surprising, given the effects of conformational averaging owed to flexibility and the presence of interconverting conformers. We observe similar trends for the backbone amide protons of our peptide with a mean difference ± SD of −0.60 ± 0.14 ppm (Figure S4 in the Supporting Information).

Figure 3B shows a plot of the measured $^3J(H^N-H^{\alpha})$ scalar coupling constants using the DQF-COSY spectrum and, when possible, the 1D spectrum (Figure S2). The use of the 1D spectra slightly underestimates the $^3J(H^N-H^{\alpha})$ coupling constant values, because intensity addition in the overlapping region of the doublet brings the peak maxima closer. The use of the 2D DQF-COSY data overestimates the $^3J(H^N-H^{\alpha})$ coupling constant values because intensity subtraction in the overlapping region of the anti-phase absorptive signals pulls the peak extrema away. These effects are stronger for weak couplings (e.g., helices), where the in-phase or anti-phase peaks are closer to each other, compared to *β*-sheets or extended conformations. The 1D and 2D values set lower and upper bounds, with the actual $^3J(H^N-H^{\alpha})$ coupling constants expected to be closer to the 1D values. The $^3J(H^N-H^{\alpha})$ coupling constant values for residues E2, A5, A10, and A13 are less than 5 Hz using the 1D data and between 5 and 6 Hz using the 2D data. Using the 2D data, 12 out of 17 residues have $^3J(H^N-H^{\alpha})$ coupling constant values of less than 6 Hz and 4 have values between 6 and 6.5 Hz. Taking into account conformational averaging of $^3J(H^N-H^{\alpha})$ coupling constants and errors owed to the limitations of the analysis, these values are close to the theoretical value of 3.9 Hz for helices (*23*). On the contrary, for residues in the *β*-region of the Ramachandran plot, the expected $^3J(H^N-H^{\alpha})$ values should be above 8 Hz. Only the $^3J(H^N-H^{\alpha})$ coupling constant of the C-terminal A17 is in the range of 7−8 Hz.

Figure 4A,C−F shows various representations of the solution structure of the peptide, using distance restraints

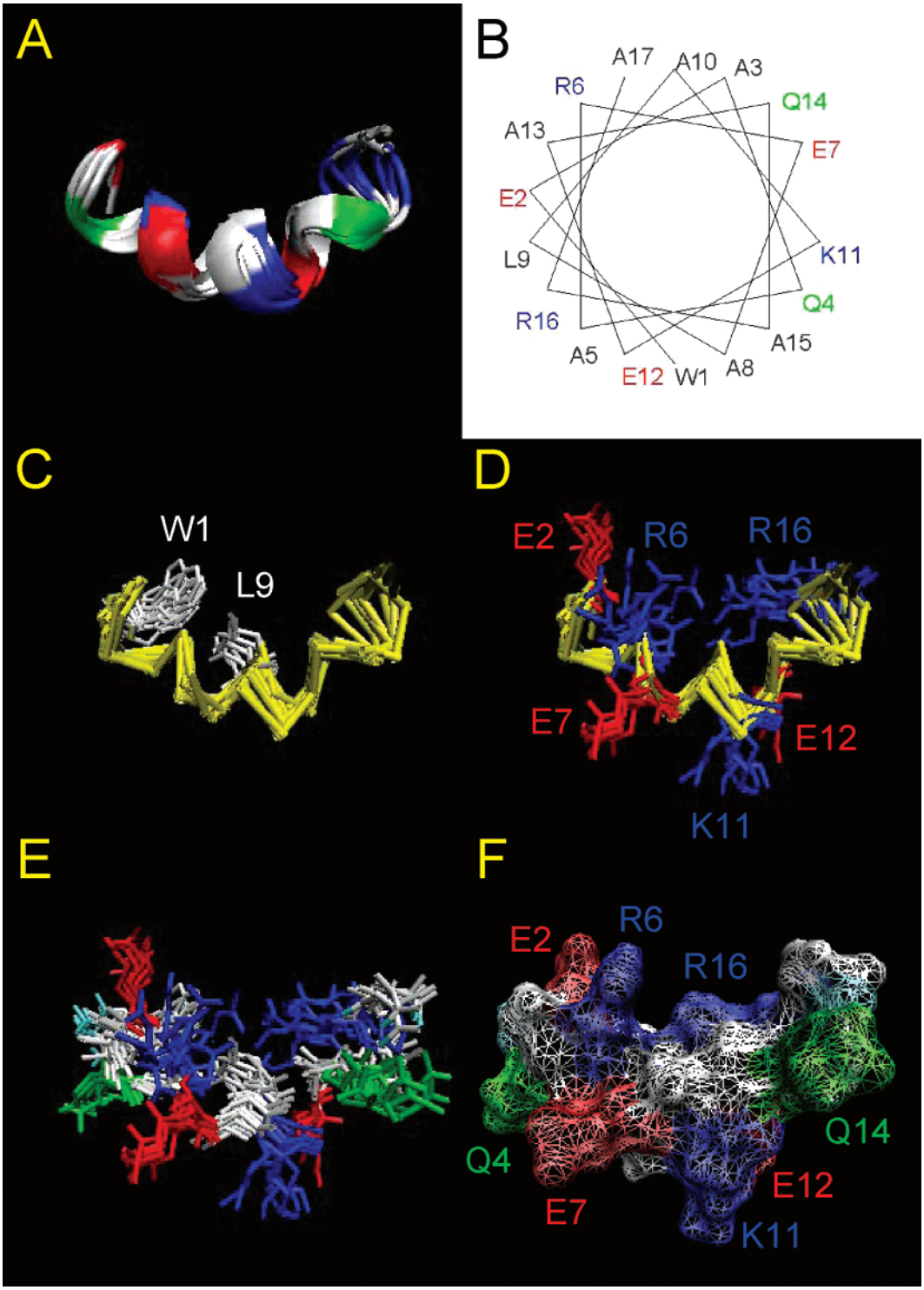

FIGURE 4: (A) Ribbon representation of the NMR ensemble of 10 structures. The thicker ribbons represent α-helical secondary structure. The color code is red for negatively charged amino acids, blue for positively charged amino acids, green for neutral polar amino acids, and white for nonpolar amino acids. (B) Helical wheel plot showing that the peptide is not amphipathic and the proximity of the ion and hydrophobic pairs in a putative perfect α-helix. (C) Demonstration of the W1−L9 hydrophobic pair formation with respect to the backbone C^{α} trace, using the NMR ensemble of structures. The side chains of all other residues have been removed for clarity. (D) Demonstration of the topologies of the ion pairs with respect to each other and the backbone C^{α} trace, using the NMR ensemble of structures. The side chains of all other residues have been removed for clarity. (E) Demonstration of the topologies of all residues, using the NMR ensemble of structures. The color code is red for negatively charged amino acids, blue for positively charged amino acids, green for neutral polar amino acids, white for nonpolar amino acids, and cyan for the blocking groups at the termini. (F) Representation of the peptide surface, using the average minimized structure. The color code is as in (E). In all molecular graphics hydrogen atoms have been removed to simplify the figures. The structure orientations in panels A and C−E are identical, whereas the orientation in panel F is slightly rotated (15° counterclockwise around the horizontal axis) to improve the view of all charged side chains.

derived from NOEs, φ torsion angle restraints derived from $^{3}J(H^{N}-H^{\alpha})$ values, and χ_1 torsion angles derived from combined analysis of DQF-COSY and NOESY data. The numbers and types of restraints are summarized in Table S1A. Table S1B lists a summary of the potential energies for the NMR ensemble of 10 structures, the lowest energy

Table 1: Calculated Secondary Structure

	secondary structure sequence range			
model no.	α-helix	3_{10}-helix	turn	coil
1 (lowest energy)	5−14		1−4	15−17
2	5−12		1−4, 13−16	17
3	5−12		1−4, 13−16	17
4	5−12		1−4, 13−16	17
5	5−12		1−4, 13−16	17
6	5−12		1−4, 13−16	17
7	5−12		1−4, 13−16	17
8	5−12		1−4, 13−16	17
9	5−10		1−4, 11−16	17
10	5−13	13−16	1−4	17
av minimized	5−12		1−4, 13−16	17

[a] Calculated using the algorithm STRIDE (*26*) implemented within the program VMD (*19*).

structure, and the average minimized structure. The various panels in Figure 4A,C−F show visually that the peptide assumes a bent helical structure.

Figure 4A shows the secondary structure of the NMR ensemble of 10 structures, calculated using the implementation of the algorithm STRIDE (*26*) within the program VMD (*19*). According to STRIDE, the peptide forms predominantly α-helical secondary structure, which is more stable in the middle and fraying at the ends (Figure 4A, Table 1). The α-helix starts at residue 5 and ends at residues 10−14, with 7 structures out of 10 and the average minimized structure ending at residue 12 (Table 1). Turn structures are observed at the termini. The C-terminus is slightly more disordered than the N-terminus (Figure 4A). In only one structure there is a short segment of a 3_{10}-helix toward the C-terminus (Table 1). The STRIDE assignment is somehow in disagreement with the presence of (i, $i + 3$) NOE connectivities (Figure 2), which indicate an α-helix extending all the way from the N-terminus to the C-terminus. This may be a genuine attribute of α-helical secondary structures derived directly from NOE connectivity patterns, compared to stringent calculations based on combined torsion angle and hydrogen bond analyses. The disagreement may be owed to the fact that STRIDE is optimized on the basis of knowledge of protein (as opposed to peptide) secondary structures for α-helices and β-sheets, using sets of high-resolution structures deposited at the Protein Data Bank (PDB) (*27*). STRIDE uses hydrogen bond energy criteria and statistically derived backbone torsion angle information for secondary structure determination (*26*). It should be noted that STRIDE suggests turn structures at residue segments 1−4 and 11/13−16 (Table 1), which can be seen as less defined extensions of the strictly defined α-helix in the peptide center. Ramachandran plots and putative hydrogen bond lists (Figures S5 and S6 and Table S3 in the Supporting Information) support a highly flexible helical structure throughout the peptide. Flexibility is indicated by the nonuniform distribution of putative hydrogen bonds, throughout each structure and across different structures of the NMR ensemble, and the presence of some nonstandard hydrogen bonds for helical structures (Table S3). The latter may be related to the bending of the structure (Figure 4A). The presence of several C═O···H−N (i, $i + 3$) and (i, $i + 4$) hydrogen bonds suggests possible competing transient α- and 3_{10}-helices. The only three residues not participating in hydrogen bonds are residues E2, A3, and Q14 (Table S3), supporting higher flexibility at the termini. It is possible that the "knowledge-based" secondary structure assignment of STRIDE is more rigorous than the NOE connectivities of Figure 2, which are based on distance-dependent dipolar couplings, without the consideration of subtler conformational (geometric) parameters between hydrogen bond donor and acceptor atoms. The NOE connectivity patterns also reflect the presence of multiple and interconverting conformers. The chemical shift differences (Figure 3A) suggest more helicity toward the N-terminus. However, these differences may be owed to side chain effects from Trp1 (*6*). A holistic view of all the measured and calculated data suggests the presence of a highly flexible and bent α-helix, which is more stable in the middle of the peptide sequence, with several other transient and interconverting conformers. This type of flexibility is expected for peptides and is known to contribute to the averaging of the measured NMR parameters (*6*).

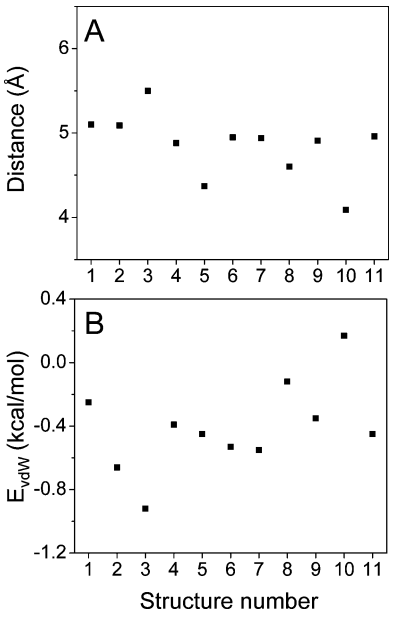

FIGURE 5: (A) Calculated distance for the W1−L9 hydrophobic pair, using the NMR ensemble of structures (nos. 1−10) and the average minimized structure (no. 11). The distance is from the geometric center of the phenyl ring of W1 and the C^{γ} atom of L9. (B) Calculated pairwise van der Waals potential energy term for the W1−L9 hydrophobic pair, using the NMR ensemble of structures (nos. 1−10) and the average minimized structure (no. 11).

Despite the lack of amphipathicity, shown in the helical wheel of Figure 4B, the peptide secondary structure is bent (Figure 4A,C−F). This is because of the formation of a strong (i, $i + 8$) hydrophobic interaction between W1 and L9. Figure 4C depicts the spatial arrangement of the side chains of W1 and L9 with respect to each other and the C^{α} trace of the peptide backbone. Figure 5A quantifies the distance between the geometric center of the phenyl ring of W1 and the C^{γ} atom of L9 for the NMR ensemble of structures and the average minimized structure. This distance is in the range of 4−5.5 Å (average ± SD = 4.84 ± 0.40 Å, using the NMR ensemble, 4.96 Å in the average minimized structure). Figure 5B shows the calculated Lennard-Jones 6−12 potential energy term for the side chains of W1−L9, using the NMR ensemble of structures and the average

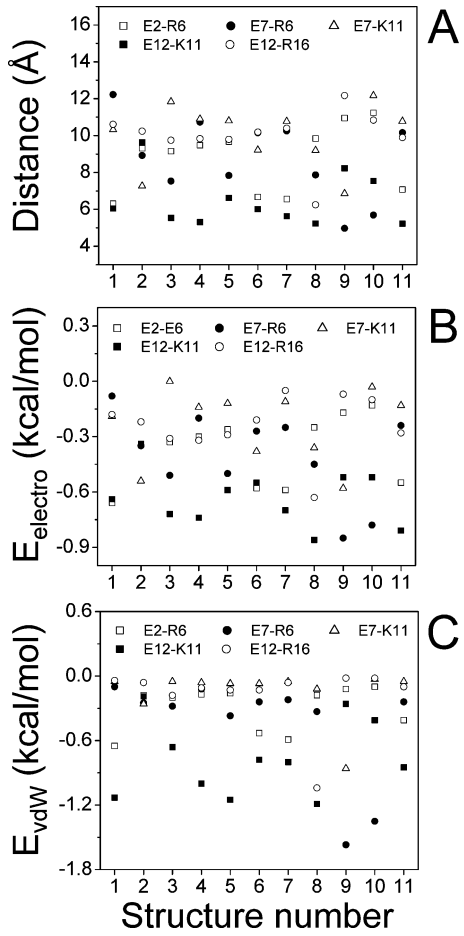

FIGURE 6: (A) Calculated ion pair distances, using the NMR ensemble of structures (nos. 1−10) and the average minimized structure (no. 11). The ion pairs and their corresponding symbols are shown in the legend. The distances are from the C^{δ} atoms of E2, E7, and E12 and the C^{ζ} atoms of R6 and R16 or the N^{ζ} atom of K11. (B) Calculated pairwise electrostatic potential energy term for the ion pairs of panel A (shown in the legend). (C) Calculated pairwise van der Waals potential energy term for the ion pairs of panel A (shown in the legend).

minimized structure. With the exception of one structure (no. 10), the remaining show stabilizing Lennard-Jones contributions to the internal energy of the peptide in the range of −1 to slightly less than 0 kcal/mol.

Figure 4D depicts the spatial arrangement of the charged side chains with respect to each other and the C^{α} trace of the peptide backbone. Apparently, there is competition between E−R/K (i, $i + 4$) and E−R/K (i, $i - 1$) ion pair formations. Figure 6A quantifies the distances between the C^{δ} atoms of glutamic acids and the C^{ζ} atoms of arginines or the N^{ζ} atom of lysine for the NMR ensemble of structures and the average minimized structure. There are no ion pairs within 3.5 Å that would suggest the presence of a salt bridge or up to 4.5−5 Å that would suggest the presence of a weak salt bridge. The calculated distances are within 5−12 Å, suggesting that all charged residues participate in medium/long-range (i, $i + 4$) or (i, $i - 1$) electrostatic interactions. In the average minimized structure (no. 11, Figure 6A), the ion pairs for E2−R6 and E12−K11 are within 5−8.5 Å, whereas the ion pairs for E7−R6, E7−K11, and E12−R16 are found within 9.5−12.5 Å. In the NMR ensemble (structure nos. 1−10, Figure 6A), the following ion pairs are found within 5−8.5 Å: E2−R6 (three times), E7−R6 (five times), E7−K11 (two times), E12−K11 (nine times), and E12−R16 (one time). Statistically, the E12−K11 ion pair interaction is dominant, followed by E7−R6, both of which are E−K/R (i, $i - 1$) interactions. The three possible E−K/R (i, $i + 4$) ion pair interactions, E2−R6, E7−K11, and E12−R16, are always present, albeit at an overall weaker strength than the two possible (i, $i - 1$) ion pair interactions. Figure 6B,C shows the calculated Coulombic and Lennard-Jones 6−12 potential energy terms for the side chains of the E−R/K ion pairs, using the NMR ensemble of structures and the average minimized structure. All Coulombic interactions are stabilizing in the range of −0.9 to 0 kcal/mol. All Lennard-Jones interactions are stabilizing in the range of −1.6 to 0 kcal/mol.

Figure 4E depicts the spatial arrangement of all side chains in the NMR ensemble of structures. Figure 4F shows the surface representation of the average minimized structure of the peptide, colored according to the amino acid type. The figure depicts the distributions of the charged (E, R, K), neutral polar (Q), and nonpolar (W, A, L) amino acids to the peptide surface and demonstrates the presence of polar residues in both the interior and exterior bends of the α-helix.

DISCUSSION

We have studied a 17-residue peptide of de novo design that forms a single element of secondary structure. The peptide consists of three pentapeptide blocks, spanned by a tryptophan, W1, at the N-terminus and an alanine, A17, at the C-terminus. The first and third pentapeptide blocks are identical with sequence EAQAR, and the middle pentapeptide block has the sequence EALAK. The high alanine content of the peptide promotes a helical structure, and so does the presence of glutamic acids, glutamines, arginines, and lysine (*10*, *11*). The rationale in incorporating the capability for the formation of (i, $i + 4$) ion pair electrostatic interactions in the side chains is to promote the formation of α-helices, which are stabilized by backbone (i, $i + 4$) hydrogen bonds. This idea has been proposed by Baldwin and co-workers (*2*) and has been successfully tested in a number of studies (*2*−*6*). In solution, α-helices are typically amphipathic with a bent structure, which is promoted by clustering of hydrophobic side chains in an interior surface and the presence of polar side chains in a solvent-exposed surface. Our 17-residue peptide has the above characteristics to promote an α-helical structure, except for an amphipathic sequence. However, it does assume a bent structure because of an (i, $i + 8$) hydrophobic interaction between the side chains of the N-terminal residue and the middle residue of the peptide. This observation was made in a previous study for the Q4L9 peptide, using NMR data (*6*). The current study presents a complete structure determination of a similar peptide but symmetric in the sequences of the first and last pentapeptide blocks, using NMR data and NMR-derived restraints. The only difference in the sequences of the two peptides is that Q4L9 has an alanine at position 14 (*6*) whereas the peptide studied here has a glutamine.

Peptides consisting of three E/DAAAK/R pentapeptide blocks were designed to be monomeric by uniformly spacing charged residues, which are solvent-accessible in helical

conformation (*1*−*6*). However, the introduction of W at position 1 and L at position 9 raises the possibility of aggregation (*6*). Aggregation could occur with concurrent unfolding, in which case the peptide would exhibit a circular dichroism spectrum with $[\theta]_{222}$ ellipticity inconsistent with helical structure and also NOE connectivity patterns incompatible with helices. Aggregation could also occur without unfolding, in which case the circular dichroism spectrum and the NOE connectivity patterns would still be characteristic of a helix. If the aggregation is not massive, it would not affect the NMR line widths. The possibility of peptide aggregation was tested in the study of Q4L9, using circular dichroism spectroscopy at different concentrations and temperatures, analytical ultracentrifugation at different concentrations, and NMR spectroscopy (*6*). The elipticity $[\theta]_{222}$ for Q4L9 was found to be independent of peptide concentration in the range of 2−1800 μM (at temperatures of 5 and 60 °C). However, these data do not exclude possible aggregation in the helical state. Analytical ultracentrifugation experiments, using dilute and concentrated Q4L9 solutions of 0.1 and 5 mM, were consistent with the monomeric peptide state (*6*). Our NMR data and those for Q4L9 do not show evidence for massive aggregation because there is no obvious line broadening for the NMR resonances. We have also been concerned with the possibility of dimer-only formation in an antiparalell arrangement of two helices, where the following two (i, $i + 8$) pairs of hydrophobic interactions could be possible: W1(monomer 1)−L9-(monomer 2) and W1(monomer 2)−L9(monomer 1). (Such an arrangement excludes higher order multimers with the same mechanism.) This arrangement would be consistent with the absence of significant line broadening and would explain the observed W1−L9 NOEs, but would also be responsible for E2−A8 and other NOEs, which are not observed. Despite the above discussion, the presence of low-population aggregates that do not significantly affect the NMR data cannot be excluded.

Our study shows a competition between the E−R/K (i, $i + 4$) ion pair interactions. These medium/long-range electrostatic interactions occur at distances of 4−12.5 Å between the central heavy atoms of the ionization sites. No ionic interaction within 3.5 Å is observed that would suggest the presence of a strong salt bridge. The end result is a compromised conformation for the charged side chains, with the middle ones (R6, E7, K11, E12) participating in both next-neighbor and four-residue-apart interactions. Overall, the E−R/K (i, $i - 1$) ion pair interactions are more dominant than those at (i, $i + 4$). It is possible that the helix bending promotes the (i, $i - 1$) interactions. The observed medium/long-range electrostatic interactions contribute to the stability of the peptide, which has a sequence with helix specificity. The unusual characteristic of our peptide is the strength of the (i, $i + 8$) hydrophobic interaction between the phenyl ring of W1 and the branched methyl groups of L9. This hydrophobic pair interaction is responsible for bending the structure of the peptide, without compromising its α-helical character. The peptide retains a helical structure throughout its sequence, with a strong α-helix in its middle and fraying ends, as expected since the middle residues participate in both forward and backward hydrogen bonds. Small peptides in solution are dynamic, fluctuating within the same structure and interchanging their structures among several conformers. The calculated ion pair distances and interaction energies demonstrate differences of the competing ionic interactions in the NMR ensemble of structures. Because of the dynamic character of the peptide, we expect that in real time there is fluctuation of the competing interactions. There is also fluctuation in the distance and energy of interaction for the single (i, $i + 8$) W1−L9 hydrophobic pair, with possible contributions from alanines or the hydrophobic portions of the polar side chains. Local fluctuations in proteins and peptides are responsible for the formation and deformation of hydrogen bonds in α-helical structures, as is often shown by site-specific hydrogen/deuterium exchange and NMR spectroscopy for all-helical proteins (e.g., ref *28*). The dual hydrogen bond participation of the middle helical residues dampens the local fluctuations with the effect of a better definition of the local secondary structure. Further dampening also occurs in proteins because of side chain−side chain interactions, involving side chains from different elements of secondary structure (*28*), which is not the case here.

The dynamic character of small peptides in solution that demonstrate a predominant high-population structure has been shown by standard NMR spectroscopy (reviewed in ref *29*). This was also evidenced by nonstandard NMR spectroscopy for a partially folded peptide with significant helical content, whose sequence was derived from a protein (*30*). In agreement with an earlier NMR study (*6*), our peptide flexibility and the presence of other conformers with lower populations are demonstrated by averaging of NMR parameters, such as scalar coupling constants, the presence of weak or very weak NOE cross-peaks not consistent with α-helices, and the presence of other than (i, $i + 4$) backbone hydrogen bonds. In particular, the large content of (i, $i + 3$) putative hydrogen bonds suggests transitional 3_{10}-helices. Previous spectroscopic studies have suggested the presence of 3_{10}-helicity in alanine-based peptides (*31*, *32*). The lower content of rigorously defined α-helical structure at the termini, calculated using STRIDE (*26*), supports the notion that helix unfolding starts at the termini and gradually proceeds toward its center. This is in agreement with site-specific hydrogen/deuterium exchange studies of an all-helical protein by NMR which show higher stability (larger protection factors) in the middle of the α-helices compared to their termini (*28*).

As for the amino acid content, A, R, K, and L residues (in that rank order) have the highest α-helical tendencies, measured by the relative stabilization of α-helical conformation in a peptide undergoing α-helical dimer−random coil monomer transition (*10*). In addition, W, Q, and E are in the top 10 in the same rank order list. All, E, A, L, Q, K, and R residues (in that rank order) show preferences for location in protein α-helices, compared to β-strands or reverse turns, with A having the highest frequency of occurrence in the middle of protein α-helices, followed by L (summarized in ref *11*). K, R, E, and Q residues demonstrate lower than those of A and L, but significant frequencies of occurrence in the middle of protein α-helices (*11*). In addition, acidic residues (mainly E) show preferences toward the N-terminus and basic residues (mainly K) toward the C-terminus, which is attributed to favorable interactions with the helix dipole (*11*). The peptide studies demonstrate the local sequence effects that contribute to α-helix nucleation. The protein studies incorporate long-range contribu-

tions, involving side chains from different elements of secondary structure.

We performed a BLAST (*36*) search to determine the importance of E/DAAAK/R sequences in proteins, by measuring the sequence occurrences in the PDB (*27*) and in particular in helical structures. The BLAST search has revealed the following "sequences, hits (number)" in the PDB, where the number in parentheses denotes sequence hits in helices: EAAAK, 15 (11); EAAAR, 21 (18); DAAAK, 11 (8); DAAAR, 18 (17); KAAAE, 32 (26); RAAAE, 30 (17); KAAAD 33 (26); RAAAD, 10 (10). Overall, 133 out of 170 hits (78.2%) involve helices. From the sequence hits in parentheses, 13 out of 133 were located partially in helices. Only one sequence hit was located fully in a β-sheet, and seven were located partially in a β-sheet. When we allowed variation in the middle alanine position, the BLAST search revealed the following "sequence, hits": E/DAXAK/R, 50; K/RAXAE/D, 26. In these hits X represents any amino acid excluding E, D, A, K, and R. In earlier studies of proteins, we have identified the presence of charge−charge (i, $i \pm 4$) interactions within helices, where the middle residues are not necessarily alanines (*33*, *34*). Another study of a predominantly helical protein has identified numerous (i, $i + 4$) salt bridges, which were proposed to be stabilizing factors (*35*). The data discussed here demonstrate the significance of (i, $i + 4$) ion pair interactions in helix promotion.

In summary, we present the NMR-derived solution structure of a predominantly α-helical element of secondary structure. This structure is a bent α-helix, despite the absence of an amphipathic sequence. The bend is smooth as opposed to typical kinks introduced by prolines disturbing helical structures. We critically discuss the content and stability of secondary structure and the origins of α-helical structure formation. Bending is promoted by a strong (i, $i + 8$) hydrophobic interaction between the side chains of the N-terminal and middle residues. The amino acids that comprise the remaining sequence are helix promoters. The strategic incorporation of charge−charge (i, $i \pm 4$) and (i, $i \pm 1$) interactions also contributes to the stability of the α-helix, although they are not the defining factors for its structure.

ACKNOWLEDGMENT

We thank Dr. Gene Merutka for very insightful discussions. We also thank Dr. Dan Borchardt of the UCR Analytical Chemistry Instrumentation Facility for the use of the NMR spectrometer.

SUPPORTING INFORMATION AVAILABLE

(I) Table of chemical shifts, (II) tables of structure determination statistics, NMR-derived restraints and potential energies, (III) TOCSY spectrum, amide/aromatic-side chain region, (IV) 1D NMR spectrum, amide region, (V) DQF-COSY spectrum, backbone region, (VI) amide proton chemical shift differences for observed and random coil values, (VII) Ramachandran plots for the NMR ensemble of structures and the average minimized structure, and (VIII) tables of hydrogen bonds for the NMR ensemble of structures and the average minimized structure. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Scholtz, J. M., and Baldwin, R. L. (1992) The mechanism of α-Helix formation by peptides, *Annu. Rev. Biophys. Biomol. Struct. 21*, 95−118.
2. Marqusee, S., and Baldwin, R. L. (1987) Helix stabilization by Glu- ... Lys+ salt bridges in short peptides of *de novo* design, *Proc. Natl. Acad. Sci. U.S.A. 84*, 8898−8902.
3. Merutka, G., Lipton, W., Shalongo, W., Park, S. H., and Stellwagen, E. (1990) Effect of central-residue replacements on the helical stability of a monomeric peptide, *Biochemistry 29*, 7511−7515.
4. Merutka, G., Shalongo, W., and Stellwagen, E. (1991) A model peptide with enhanced helicity, *Biochemistry 30*, 4245−4248.
5. Merutka, G., and Stellwagen, E. (1991) Effect of amino acid ion pairs on peptide helicity, *Biochemistry 30*, 1591−1594.
6. Merutka, G., Morikis, D., Bruschweiler, R., and Wright, P. E. (1993) NMR evidence for multiple conformations in a highly helical model peptide, *Biochemistry 32*, 13089−13097.
7. Marqusee, S., Robbins, V. H., and Baldwin, R. L. (1989) Unusually stable helix formation in short alanine-based peptides, *Proc. Natl. Acad. Sci. U.S.A. 86*, 5286−5290.
8. Scholtz, J. M., York, E. J., Stewart, J. M., and Baldwin, R. L. (1991) A neutral, water-soluble, α-helical peptide: the effect of ionic strength on the helix-coil equilibrium, *J. Am. Chem. Soc. 113*, 5102−5104.
9. Lyu, P. C., Liff, M. I., Marky, L. A., and Kallenbach, N. R. (1990) Side chain contributions to the stability of alpha-helical structure in peptides, *Science 250*, 669−673.
10. 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.
11. Creighton, T. E. (1993) *Proteins, Structures and Molecular Properties*, W. H. Freeman and Co., New York.
12. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277−293.
13. Johnson, B. A., and Blevins, R. A. (1994) NMRView: a computer program for the visualization and analysis of NMR data, *J. Biomol. NMR 4*, 603−614.
14. Kim, Y. M., and Prestegard, J. H. (1989) Measurement of vicinal couplings from cross peaks in COSY spectra, *J. Magn. Reson. 84*, 9−13.
15. Wang, A. C., and Bax, A. (1995) Reparametrization of the Karplus relation for $^{3}J(H^{\alpha}-N)$ and $^{3}J(H^{N}-C')$ in peptides from uniformly $^{13}C/^{15}N$-enriched human ubiquitin, *J. Am. Chem. Soc. 117*, 1810−1813.
16. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 905−921.
17. Nilges, M. (1993) A calculation strategy for the structure determination of symmetrical dimers by ^{1}H NMR, *Proteins: Struct., Funct., Genet. 17*, 297−309.
18. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR 8*, 477−486.
19. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J. Mol. Graphics 14*, 33−38.
20. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51−55.
21. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem. 4*, 187−217.
22. Hennig, L. (1999) WinGene/WinPep: User-friendly software for the analysis of amino acid sequences, *Biotechniques 26*, 1170−1172.
23. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
24. Merutka, G., Dyson, H. J., and Wright, P. E. (1995) Random coil ^{1}H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG, *J. Biomol. NMR 5*, 14−24.

25. Szilagyi, L. (1995) Chemical shifts in proteins come of age, *Prog. Nucl. Magn. Reson. Spectrosc. 27*, 325−443.
26. Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment, *Proteins: Struct., Funct., Genet. 23*, 566−579.
27. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The protein data bank, *Nucleic Acids Res. 28*, 235−242.
28. Morikis, D., and Wright, P. E. (1996) Hydrogen exchange in the carbon monoxide complex of soybean leghemoglobin, *Eur. J. Biochem. 237*, 212−220.
29. Dyson, H. J., and Wright, P. E. (1991) Defining solution conformations of small linear peptides, *Annu. Rev. Biophys. Biophys. Chem. 20*, 519−538.
30. Bruschweiler, R., Morikis, D., and Wright, P. E. (1995) Hydration of the partially folded peptide RN-24 studied by multidimensional NMR, *J. Biomol. NMR 5*, 353−356.
31. Miick, S. M., Martinez, G. V., Fiori, W. R., Todd, A. P., and Millhauser, G. L. (1992) Short alanine-based peptides may form 3_{10}-helices and not α-helices in aqueous solution, *Nature 359*, 653−655.
32. Fiori, W. R., Miick, S. M., and Millhauser, G. L. (1993) Increasing sequence length favors α-helix over 3_{10}-helix in alanine-based peptides: evidence for a length-dependent structural transition, *Biochemistry 32*, 11957−11962.
33. Morikis, D., Roy, M., Newlon, M. G., Scott, J. D., and Jennings, P. A. (2002) Electrostatic properties of the structure of the docking and dimerization domain of protein kinase A IIα, *Eur. J. Biochem. 269*, 2040−2051.
34. Morikis, D., Elcock, A. H., Jennings, P. A., and McCammon, J. A. (2001) Native-state conformational dynamics of GART: A regulatory pH-dependent coil-helix transition examined by electrostatic calculations, *Protein Sci. 10*, 2363−2378.
35. Wang, C. L. A., Chalovich, J. M., Graceffa, P., Lu, R. C., Mabuchi, K., and Stafford, W. F. (1991) A long helix from the central region of smooth-muscle caldesmon, *J. Biol. Chem. 266*, 13958−13963.
36. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389−3402.

BI701252N