Structural Characterization of α-Helices of Implicitly Solvated Poly-Alanine†

Vernon A. Couch,* Neal Cheng, Krishnan Nambiar, and William Fink*

Department of Chemistry, University of California, Davis, California 95616

Received: September 14, 2005; In Final Form: December 22, 2005

The structural characteristics of α-helices in poly-alanine-based peptides have been investigated via molecular dynamics simulation with the goal of understanding the basic features of peptide simulations within the context of a model system, classical molecular dynamics with generalized Born (GB) solvation, and to shed insight into the formation and stabilization of α-helices in short peptides. The effects of peptide length, terminal charges, proline substitution, and temperature on the α-helical secondary structure have been studied. The simulations have shown that distinct secondary structure begins to develop in peptides with lengths approaching 10 residues while ambiguous structures occur in shorter peptides. The helical content of peptides with lengths ≥10 amino acids is observed to be nearly constant up to $(Ala)_{40}$. Interestingly, terminal charges and proline in the second position from the N-terminus alter the secondary structure locally with little effect on the overall α-helical content of the peptide. The free energy profile of helix formation was also investigated. A large increase in free energy accompanying the formation of helices with more than two consecutive hydrogen bonds in the $(i, i + 4)$ pattern was observed while the free energy increases linearly with additional hydrogen bonds. Values for the change in enthalpy and entropy of helix nucleation and propagation are reported. Additionally the results obtained from the GB model are compared to explicit solvent simulations of two synthetic alanine-based peptides.

Introduction

Since its discovery the α-helix has played an important role in understanding the structure of proteins.[1] The α-helix is a common motif in proteins, and increasingly small α-helical peptides are being utilized as drugs such as HIV fusion inhibitors.[2] The ability to predict secondary and tertiary structure in peptides and proteins is of fundamental importance to unlocking the mysteries of protein structure and function but are also important in the development of smaller, cheaper, and more readily synthesized peptide-based pharmaceuticals. Simple chemical modifications and amino acid substitutions have been shown to dramatically impact the helical content of short alanine-based synthetic peptides in aqueous solution.[3,4] Detailed knowledge of the interactions stabilizing helices in aqueous solution will facilitate the rational design of peptides that fold with predetermined secondary structure.

The purpose of this study is to characterize the α-helical properties of alanine-based peptides within the context of classical generalized Born molecular dynamics simulations,[5] to investigate some of the factors thought to contribute to helix stabilization, and to evaluate the predictive and qualitative value of the model. In particular we have investigated the effect of length on helical content as well as terminal charges and proline substitution in the first turn of the helix. Additionally implicit and explicit solvent simulations of synthetic peptides incorporating these features are compared.

Positive charge, such as a protonated terminal amine, at the N-terminus of α-helices can interact unfavorably with the positive end of the helix dipole destabilizing α-helical structure.[6] Similarly negative charges at the C-terminus are expected to be destabilizing. Experiments in aqueous solution have shown that residues with negatively charged side chains can stabilize the helix when located at the N-terminus.[7−9] Gas-phase experiments have also shown that appropriately placed charges at the termini stabilize helices in a vacuum.[10−12] We have investigated both charged and neutral species to determine the extent of this effect within the context of the model. However, the acetyl and amide groups placed at the N- and C-termini not only neutralize the charges at the termini but can act as additional hydrogen bonding donor and acceptor groups. The regular $(i, i + 4)$ hydrogen bonding pattern leaves the first three backbone amide groups and the last three backbone carbonyls without hydrogen bonding partners. Presta and Rose[13] have suggested that residues that can provide additional hydrogen bonds to the unpaired amide and carbonyl groups of the peptide backbone are important factors in determining the location and stabilization of α-helices. An acetyl group at the N-terminus can act as an N-capping residue and has been shown to increase the helical content of short peptides in solution.[4,7] Hydrogen bonding patterns for the acetyl group have been investigated within the context of the hypothesis of Presta and Rose.

In the body of α-helices, proline is known to be a helix-breaking residue due to the pyrrolidine ring, which sterically hinders adjacent residues, and its lack of an available amide proton for hydrogen bonding.[14] Despite this, proline has been found to have a high affinity for the second position from the N-terminus,[15] and has been successfully utilized in synthetic peptides of high helical content[3]. The structure of the pyrrolidine ring in proline fixes the ϕ dihedral to an α-helical conformation and limits the flexibility of the ψ dihedral.[14] As a result there should be less entropic cost for proline to adopt a helical conformation compared to other amino acids.[15] It has also been suggested that proline occurs in this position not because of its

† Originally submitted for the "Jack Simons Festschrift", published as the December 22, 2005, issue of *J. Phys. Chem. A* (Vol. 109, No. 50).

* Authors to whom correspondence should be addressed. E-mail: vacouch@ucdavis.edu; whfink@ucdavis.edu.

10.1021/jp055209j CCC: $33.50
Published on Web 02/01/2006

helix-initiating propensity but rather as a result of its helix-breaking character where it acts as a helix-demarcating residue.[13,14] We have investigated the effect of proline on the secondary structure in both charged and neutral peptides.

Due to the importance of secondary structure formation in small peptides many experimental and theoretical studies of α-helix formation and stabilization in small peptides have been reported. Many factors play a role in helix stabilization including the propensity of particular amino acids to adopt helical conformations, the position of the amino acid, side chain and termini protonation, and the sequence of amino acids within the peptide. This complexity of interactions makes it difficult to determine a set of parameters for the amino acids that can be fed into an appropriate helix−coil theory to predict the helical characteristics of an arbitrary peptide. Typically the helical content of aqueous peptides is monitored experimentally via circular dichroism measurements at 222 nm characteristic of α-helices,[16] and in the gas-phase ion mobility experiments can discern helical structure.[10] When the circular dichroism is monitored over a range of temperatures a thermal unfolding curve can be generated and fit to the van't Hoff equation to obtain thermodynamic data for the helix−coil transition.[17] Experimental results are often fit to Zimm−Bragg (ZB)[18] or Lifson−Roig (LR)[19] helix−coil theories to obtain thermodynamic and structural information. The effects of various perturbations on the secondary structure of short peptides such as amino acid substitution,[3,4,7−9] temperature jump experiments,[20] and sequence changes[12] have been probed with these and other methods.

Similarly molecular dynamics and Monte Carlo simulations carried out at different temperatures can be fit to the van't Hoff equation[21] or to LR or ZB helix−coil theories[22−27] to obtain thermodynamic data by monitoring helical parameters such as the backbone dihedrals or helical hydrogen bonding. Replica exchange molecular dynamics (REMD) simulations particularly lend themselves to this type of analysis, since an ensemble of structures at various temperatures is simultaneously generated.[23,24] Additional simulation methods such as umbrella sampling have been applied to helix formation in short peptides. Tobias and Brooks[28] utilized umbrella sampling to generate a potential of mean force (PMF) for the helix to extended transition for peptides of the form Ac-Ala_3-NHMe along the coordinate corresponding to the distance between the carbonyl oxygen and amide hydrogen of the first and fifth residues, the (1,5) helical hydrogen bond distance. The extension of this method is difficult since there are numerous $(i, i + 4)$ hydrogen bonding pairs. The peptide growth method is one possible extension where thermodynamic integration is used to calculate the change in free energy as new residues are added to the peptide.[29] We introduce a new method to determine thermodynamic quantities that is similar to both these methods. By generation of a PMF along an abstract coordinate corresponding to sequential helical hydrogen bonds, or sequential helical residues, and performing a two-step linear regression of these PMFs obtained at different temperatures, the enthalpy and entropy of both helix nucleation and propagation can be obtained.

Methods

Simulations. Four sets of model compounds with formulas I, $(Ala)_n$; II, Ac-$(Ala)_n$-NH_2; III, Ala-Pro-$(Ala)_{n-2}$; and IV, Ac-Ala-Pro-$(Ala)_{n-2}$-NH_2 were simulated with $n = 5$, 7, 10, 15, 20, 30, and 40. Groups I and III have charged terminal residues, while groups II and IV have been neutralized via acetylation and amidation at the N- and C-termini, respectively. Molecular dynamics (MD) simulations allow us to study systems that are not accessible to experimentalists due to solubility and synthetic concerns. These systems were chosen to eliminate the ambiguity associated with water soluble peptides containing amino acids with charged side chains that may themselves influence the secondary structure.

Each peptide was constructed in a fully α-helical conformation using the xleap module of Amber 7.[12] The simulations were run using the sander module with the ff99 force field and the generalized Born solvation model at 300 K, with a dielectric constant of 78.5 corresponding to that of water. Each peptide was equilibrated for 15 ps, followed by 15 ns of production with a 1 fs time step. Additionally, temperature effects were examined via simulations of $(Ala)_{15}$ at temperatures of 200, 250, 300, 350, 400, and 500 K.

To compare the effects of both implicit and explicit solvation, two experimentally studied peptides[3] differing in only one residue, but with very different helical propensities in solution, were chosen. The synthetic peptides Asn-Pro-Glu-$(Ala)_2$-Lys-$(Ala)_3$-Gly-Arg-NH_2 and Gln-Pro-Glu-$(Ala)_2$-Lys-$(Ala)_3$-Gly-Arg-NH_2, denoted p1 and p2, were simulated at 300 K for a total of 15 ns using both generalized Born and explicit solvent. The explicit solvent simulations included 1874 and 2134 TIP3P waters for p1 and p2, respectively.

Analysis. The α-helix can be defined by the regular pattern of hydrogen bonds between the carbonyl of the ith residue and the amide group of the $i + 4$ residue. Similarly, the (ϕ, ψ) backbone dihedrals of the peptide chain, which must lie within a certain range to maintain a helical structure, can also define the α-helix. Both helical parameters, the backbone dihedral angles and the $(i, i + 4)$ hydrogen bonds, were used to monitor the α-helical structure over the MD trajectories. For a given configuration each residue i was considered to be in an α-helical conformation, H_i, if the (ϕ, ψ) dihedrals were within ±25° of the ideal angles of approximately (−60°, −40°). The ideal angles were determined from the statistical analysis of Ramachandran plots to contain the majority of amino acids found in α-helices in proteins.[31] Similarly, a hydrogen bond between the carbonyl O of the ith residue with the amide N of the $i + 4$ residue was considered present if the O−N distance was less than 4 Å.[21] Using these criteria, we define the following counting functions

$$H_i = \begin{cases} 1 & (-85 \le \phi \le -35) \text{ and } (-65 \le \psi \le -15) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$$H_i = \begin{cases} 1 & d(O_i - N_{i+4}) \le 4.0 \text{ Å} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

To calculate the helical content of each configuration the number of helical residues or hydrogen bonds is then summed, divided by the maximum possible (which is given by $(N_r - 2)$ and $(N_r - 4)$, respectively, where N_r is the number of residues), and multiplied by 100 to obtain the percent α-helicity. This gives us

$$\%H = 100 \sum H_i/(N_r - 2) \quad \text{(dihedrals)} \quad (3)$$

$$\%H = 100 \sum H_i/(N_r - 4) \quad \text{(hydrogen bonds)} \quad (4)$$

where the sum is from $i = 2$ to $(N_r - 1)$ residues and $i = 1$ to $(N_r - 4)$ hydrogen bonds, respectively. The percent helicity can be calculated at each time step or averaged over every few time steps to follow the secondary structure dynamically. Alterna-

tively, an average over the entire production phase of the simulation can be used to compare the helical propensities of different peptides. The average percent helicity was calculated from

$$\langle \%H \rangle = \Sigma \, \%H / N_{\text{config}} \tag{5}$$

where the sum is over all configurations, N_{config}.

A probability distribution of the number of hydrogen bonds per configuration was similarly calculated by counting hydrogen bonds, binning according to the number of bonds per configuration, and normalizing. This distribution can be used to obtain the average percent helicity from

$$\langle \%H \rangle = 100 \, \Sigma \, N_{\text{H-bonds}} P(N_{\text{H-bonds}}) / (N_{\text{r}} - 4) \tag{6}$$

where the sum is over all configurations, $N_{\text{H-bonds}}$ is the number of hydrogen bonds present in the configuration, and $P(N_{\text{H-bonds}})$ is the probability of $N_{\text{H-bonds}}$ occurring. The same probability may be used to express a potential of mean force (PMF), W, for the coordinate corresponding to $N_{\text{H-bonds}}$, the total number of hydrogen bonds per configuration,[21] according to

$$W(N_{\text{H-bonds}}) = -RT \ln\{P(N_{\text{H-bonds}})\} \tag{7}$$

Aside from calculating global properties from the two helical parameters it is also possible to determine a helical probability, $\langle \%H_i \rangle$, for each residue or hydrogen bond, i, by simply averaging $\%H_i$ over all configurations. This provides a map of the regions of the peptide that tend to be more or less α-helical, and comparisons of the local structure between different peptides can be examined.

The percent helicity measured from dihedrals tends to overestimate the actual helical content of the peptide, since residues can adopt α-helical configurations without being in an α-helical segment. Similarly, the percent helicity measured according to the hydrogen bonds does not distinguish isolated hydrogen bonds in the $(i, i + 4)$ scheme (turn structures) from those in genuine α-helical segments. To address the issue of α-helix segment length, PMFs were constructed from the distribution of consecutive hydrogen bonds and the distribution of consecutive α-helical residues according to

$$W(N_{\text{s}}) = -RT \ln\{P(N_{\text{s}})\} \tag{8}$$

where $P(N_{\text{s}})$ is the distribution of α-helical segments of length N_{s}. N_{s} can be related to the number of consecutive $(i, i + 4)$ hydrogen bonds or α-helical residues. If we consider a 5 residue peptide in a completely α-helical configuration, then there would be $N_{\text{res}} = 3$ residues in helical conformations (excluding both terminal residues since the (ϕ, ψ) dihedrals cannot be simultaneously defined) corresponding to $N_{\text{H-bond}} = 1$ $(i, i + 4)$ hydrogen bond between the first and fifth residues, leading to the relation

$$N_{\text{s}} = N_{\text{H-bond}} - 1 = N_{\text{res}} - 3 \tag{9}$$

with this initial segment being of length zero. However, the PMF corresponding to the distribution of consecutive hydrogen bonds suggested the relation

$$N_{\text{s}} = N_{\text{H-bond}} - 3 = N_{\text{res}} - 5 \tag{10}$$

Motivated by its form, the PMF, W of eq 8, was then fitted using linear regression to the customary expression from helix−coil theory[19]

$$\Delta G = N_{\text{s}} \Delta G_{\text{s}} + \Delta G_{\text{nuc}} \tag{11}$$

Here we take $\Delta G = W(N_{\text{s}})$, the result calculated by the PMF expression of eq 8. ΔG_{s} is the free energy associated with adding a hydrogen bond to the helix, and ΔG_{nuc} is the free energy required to initiate the α-helix. With the appropriate definition of N_{s}, the resulting linear equation has slope equal to ΔG_{s} and intercept equal to ΔG_{nuc}. Through the use of PMFs from simulations at different temperatures, ΔG_{s} and ΔG_{nuc} can be decomposed into their respective entropic and enthalpic terms. The free energy can thus be expressed as

$$\Delta G = N_{\text{s}} (\Delta H_{\text{s}} - T \Delta S_{\text{s}}) + (\Delta H_{\text{nuc}} - T \Delta S_{\text{nuc}}) \tag{12}$$

When ΔG_{s} versus T is plotted and through the use of linear regression, the PMF slopes can be fitted to a line with slope $-\Delta S_{\text{s}}$ and intercept ΔH_{s}. Similarly, ΔH_{nuc} and ΔS_{nuc} can be determined from a least-squares fit of ΔG_{nuc} versus T. Due to the nature of the PMF rather than plotting ΔG versus $N_{\text{s}} = N_{\text{con}}$, the number of consecutive units satisfying the condition of classification as being α-helical, ΔG versus $N_{\text{s}} = (N_{\text{con}} - 3)$ was used for the hydrogen bonds and $N_{\text{s}} = (N_{\text{con}} - 5)$ was used for the dihedrals (see Discussion and Results sections).

To investigate further the differences between the dihedral and hydrogen bonding measurements of the percent helicity for the shorter peptides a clustering analysis was performed on the peptides of length 5 using the NMRClust program.[32] Clustering was also performed on the peptides of length 15 to obtain a set of structures representative of the ensemble obtained from the simulation. Of the 150 000 structures collected over the MD trajectory for each peptide, 1500 structures were selected sequentially (i.e., every 100th structure) for clustering. The population of each cluster represents the weight, w_i, of that cluster within the ensemble. Along with the secondary structure assignments of the representative structures obtained from the dssp program,[17] the average percent helicity, $\langle \%H \rangle_{\text{dssp}}$, was calculated according to

$$\langle \%H \rangle_{\text{dssp}} = 100(\Sigma \, w_i f_{\text{H}}) / (\Sigma \, w_i) \tag{13}$$

where w_i is the weight of the ith cluster and f_{H} is the helix fraction. The summation in eq 13 is ideally taken over all clusters but in practice was truncated to the first 10 clusters. The $\langle \%H \rangle_{\text{dssp}}$ was compared to the $\langle \%H \rangle$ values obtained from eq 5 to verify that the clusters accurately represent the total ensemble.

Results and Discussion

Percent Helicity. The percent helicity as a function of length for each of the four sets of peptides is illustrated in Figure 1. In each case the behavior of the longer peptides, 10 or more amino acids, is similar reaching an equilibrium value of approximately 40% and 30% for the dihedral and hydrogen bond measures, respectively, as calculated from eqs 3 and 4. The different substitutions show little global effect on the overall helical content of the longer peptides. The 15 residue peptides $(Ala)_{15}$, Ac-$(Ala)_{15}$-NH_2, Ala-Pro-$(Ala)_{13}$, and Ac-Ala-Pro-$(Ala)_{13}$-NH_2 resulted in $\langle \%H \rangle$ values of 37.9, 43.4, 40.9, and 43.4 from dihedral measures and 33.8, 36.1, 35.1, and 37.7 from hydrogen bond measures, respectively. The percent helicity of various alanine-based peptides have been determined experimentally from circular dichroism data. The 15 residue peptide CH_3CO-$A_5QA_4QA_2GY$-NH_2 was reported to be 42% helical, in good agreement with the dihedral measure obtained for Ac-$(Ala)_{15}$-NH_2.[8] However, the peptide NH_2-X-$AKA_4KA_4KA_2GY$-$CONH_2$

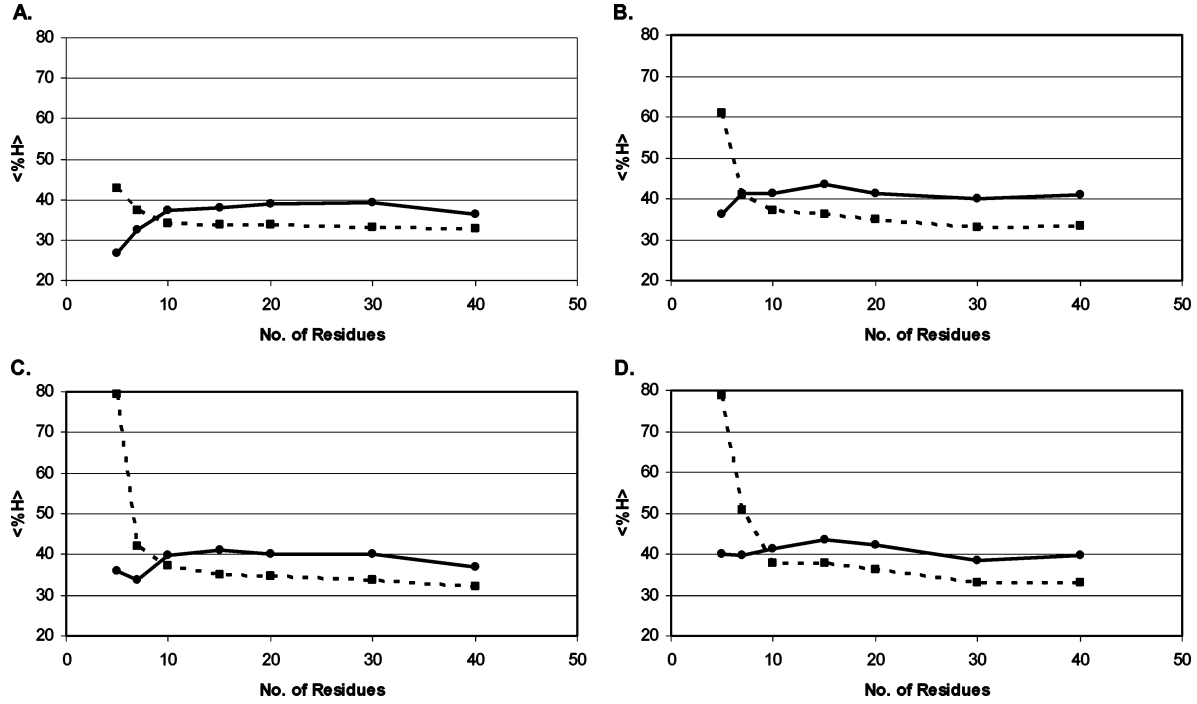

Figure 1. Average percent helicity vs peptide length for (A) model I, $(Ala)_n$ (zwitterion), (B) model II, Ac-$(Ala)_n$-NH_2 (neutral), (C) model III, Ala-Pro-$(Ala)_{n-2}$ (proline containing zwitterion), and (D) model IV, Ac-Ala-Pro-$(Ala)_{n-2}$-NH_2 (proline containing neutral). Solid lines and dashed lines indicate dihedral and hydrogen bond measures, respectively.

was found to be 23% percent helical when X = Ala and the N-terminus is charged, and when X = acetyl the percent helicity is reported as 58%.[7] These values are roughly 10% below the values obtained for the zwitterionic peptide, model I, and 10% above the value obtained for the acetylated peptide, model II. The peptide X-HN-$A_4EA_3KA_4YR$-$CONH_2$ was found to be 17% helical when X = hydrogen, and 32% helical when X = acetyl.[4] Finally, the peptide $APAEA_2KA_3GR$-$CONH_2$ was reported to be only 2.4% helical, which is an order of magnitude smaller than the value observed for model III peptides of similar lengths.[3] Gas-phase ion mobility experiments on the peptide Ac-$A_3G_{12}K$ were observed to be 26% helical at 213 K and 38% at 391 K.[12] The experimental results suggest that the values obtained from the simulations are reasonable, although the generalized Born (GB) model appears less sensitive to the substitutions than experiment.

In contrast the 5 residue peptides, which correspond to a single helical turn with one hydrogen bond, exhibit vastly differing behaviors among the set. Model I (Figure 1A) has divergent percent helicity measurements in this length regime. Although the single (1, 5) hydrogen bond occurs with a frequency of 45%, the dihedral measurement shows a very low α-helical content of 27%. This indicates the presence of predominantly turn structures and that the residues in a single turn have more torsional degrees of freedom than residues within an extended helix. The other peptides show a significant increase in the frequency of the (1, 5) hydrogen bond (61% for model II and 80% for models III and IV) with moderate increases in the dihedral measurements (36% for models II and III and 40% for model IV). From this we see that removing charges from the termini increases the propensity to form the (1, 5) bond, increasing the likelihood of helix initiation, that proline reduces the conformational degrees of freedom in the first loop of the helix, helping to secure the fairly mobile N-terminus to an α-helical structure, and that both substitutions (i.e., model IV) can work together to further increase the helical propensity. That the difference persists for the very short peptides indicates that the substitutions have a short range of action with little global effect on the secondary structure of the larger peptides.

Hydrogen Bonds per Conformation. The probability distribution of the number of helical hydrogen bonds per configuration (Figure 2A) is increasingly Gaussian in form as peptide length increases and is centered about the constant helix fraction of 0.33 or 33%, consistent with the average percent helicity calculated from eq 4. The associated PMF, W of eq 7, illustrates the deviation of the shorter peptides to a higher helix fraction (Figure 2B). As the peptide length increases the probability of becoming fully helical diminishes, the PMF becomes increasingly quadratic, and the system maintains with high probability approximately one-third of the possible helical hydrogen bonds. The various substitutions studied showed little effect on the distribution of hydrogen bonds for the longer peptides, further confirming the local nature of the effect of the substitutions on the overall helical content. Figure 2C shows the probability distribution of the number of hydrogen bonds per configuration for each model of length 40 residues. The maximum of each distribution occurs at the same helix fraction with little variation in the standard deviation. The probability distribution for model I (Figure 2D) of length 40 was fit to a normalized Gaussian distribution[34] of the form

$$P(x;\mu,\sigma) = (2\pi)^{-1/2} \exp\{-0.5*(x-\mu)^2/\sigma^2\} \qquad (14)$$

where x is, in this case, the helix fraction, σ is the standard deviation, and μ is the mean. μ was taken to be 0.33 as determined from the percent helicity, and σ was determined from the full-width at half-maximum to be 0.06.

Helical Probability. The helical probability per residue or per hydrogen bond is a measure of the frequency with which particular residues or hydrogen bonds are in α-helical conformations. This allows us to locate within the peptide those regions more likely to be α-helical and allows us to see the effect of the substitutions on the local structure of the peptide. The N-terminus of the model I peptides is significantly less helical

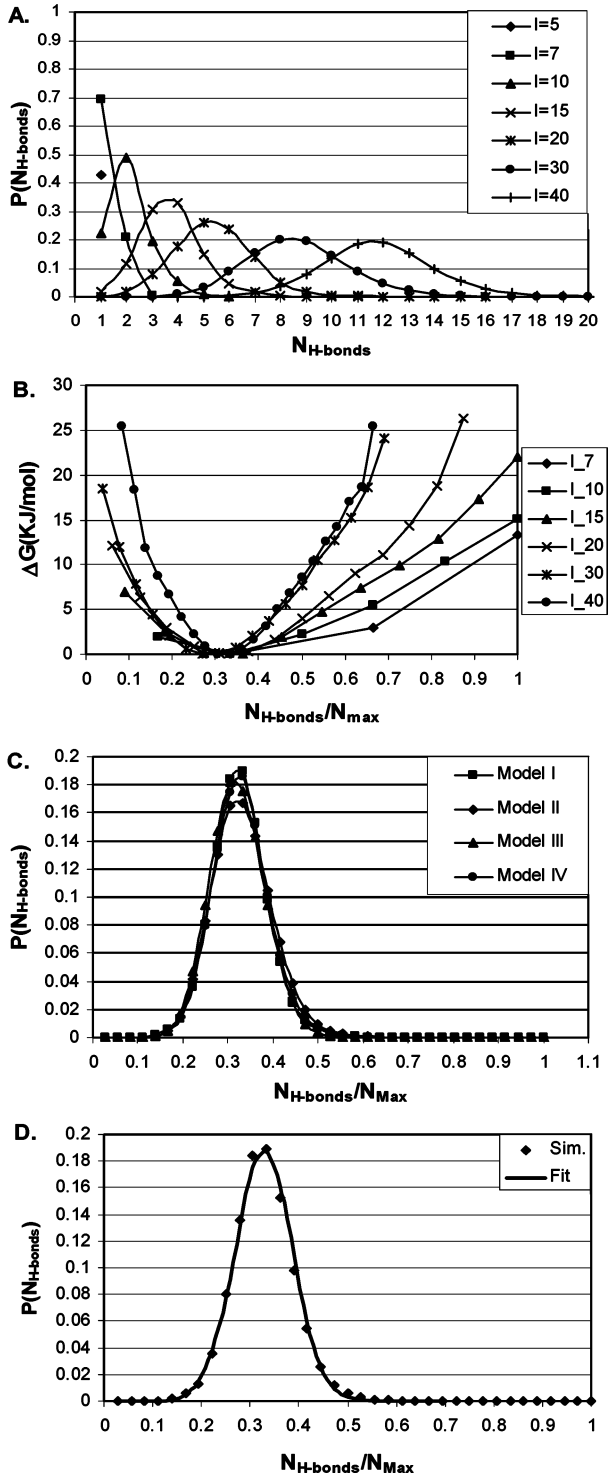

Figure 2. (A) Normalized distribution of the number of hydrogen bonds per configuration for the model I peptides of various lengths. (B) Potential of mean force generated from the distributions in part A. (C) The distribution of the number of hydrogen bonds for each model of length 40 residues. (D) Gaussian fit to the distribution for model I, again for 40 residues. In parts B−D the *x*-axis has been scaled according to the maximum number of hydrogen bonds to represent the helix fraction.

than the remaining residues as seen from the hydrogen bonding profile shown in Figure 3A. The C-terminus appears to be less affected by the charge but shows a small decrease in helical probability. The plot for model I is binodal, with a loss of helicity in the central region of the peptide, indicating that hydrogen bonds are most likely to occur in the regions adjacent to the termini and away from the center of the peptide. This behavior is also observed in the larger peptides but with increasingly irregular patterns of high and low helical probability as peptide length increases. This differentiation into distinct domains with different hydrogen bonding patterns indicates the emergence of tertiary structure. Figure 3B shows that the model II peptides, with neutral termini, exhibit the opposite behavior as in the case of model I. Here an increase in helicity in the terminal regions compared to the residues in the body of the peptide is observed. Unlike model I both the N- and C-termini are affected to a similar extent, and there is a nearly constant helical probability in the body of the peptide.

When proline is substituted in the second position of the zwitterion form, model III, as shown in Figure 3C, there is an increase in the helicity of the N-terminus while the C-terminus remains unenhanced as in the case of model I. Interestingly, enhancement is observed for the second, rather than the first, (i, $i + 4$) hydrogen bond for all peptide lengths. Additionally, the helical probability per residue shows a distinct increase in helicity for Pro-2 and Ala-3, indicating more rigid backbone dihedrals near the N-terminus (Figures 3B and 3F). Again we see a lower probability to form hydrogen bonds in the center of the peptide, an apparent characteristic of the zwitterion peptides that is absent from both neutral models. Model IV, shown in Figure 3D, with neutral termini and proline in the second position, exhibits the greatest enhancement at the N-terminus among all models. The enhancement of the first and second (i, $i + 4$) hydrogen bonds is similar to that observed in models II and III, respectively, illustrating how both substitutions work. Proline with reduced torsional degrees of freedom promotes the second (i, $i + 4$) hydrogen bond in the helix while the acetylation promotes the first. The C-terminus of the model IV peptides exhibit very similar behavior to the model II peptides.

The helical probability per residue (Figures 3E−H) shows that the residues immediately following the N-terminus are most likely to have α-helical dihedral values, and consequently the N-terminus is the most likely position for helix nucleation. Models I and II have decreasing helical probability as we progress toward the C-terminus. Conversely, models II and IV show a constant helical probability throughout the body of the peptide. In all cases the two residues preceding the C-terminus show a significant decrease in helical probability, indicating more flexibility in the C-terminal region.

α-Helix Segment Length. The fact that hydrogen bonds must occur in succession to form α-helices is not satisfactorily addressed by simply monitoring the number of hydrogen bonds occurring over the molecular dynamics (MD) trajectory. The construction of a potential of mean force (PMF) along the coordinate corresponding to the number of consecutive hydrogen bonds or consecutive helical residues based upon the criteria of eqs 1 and 2 has yielded very interesting results into the stability of α-helices (Figure 4). It is clear that although small changes in the percent helicity and helical probability per hydrogen bond were observed, the PMF is essentially the same for each peptide regardless of length or substitution. The primary effect of the substitutions is to reduce the free energy of helix nucleation. This is consistent with our findings that the substitutions have rather local effects, at least concerning secondary structural content, by increasing the probability of helix nucleation particularly near the N-terminus. The observed $\Delta\Delta G_{nuc}$, compared to model I, are approximately −0.3, −0.1, and −0.3 kcal/mol for models II−IV, respectively. It appears that the vast majority of the hydrogen bonds occur alone or in

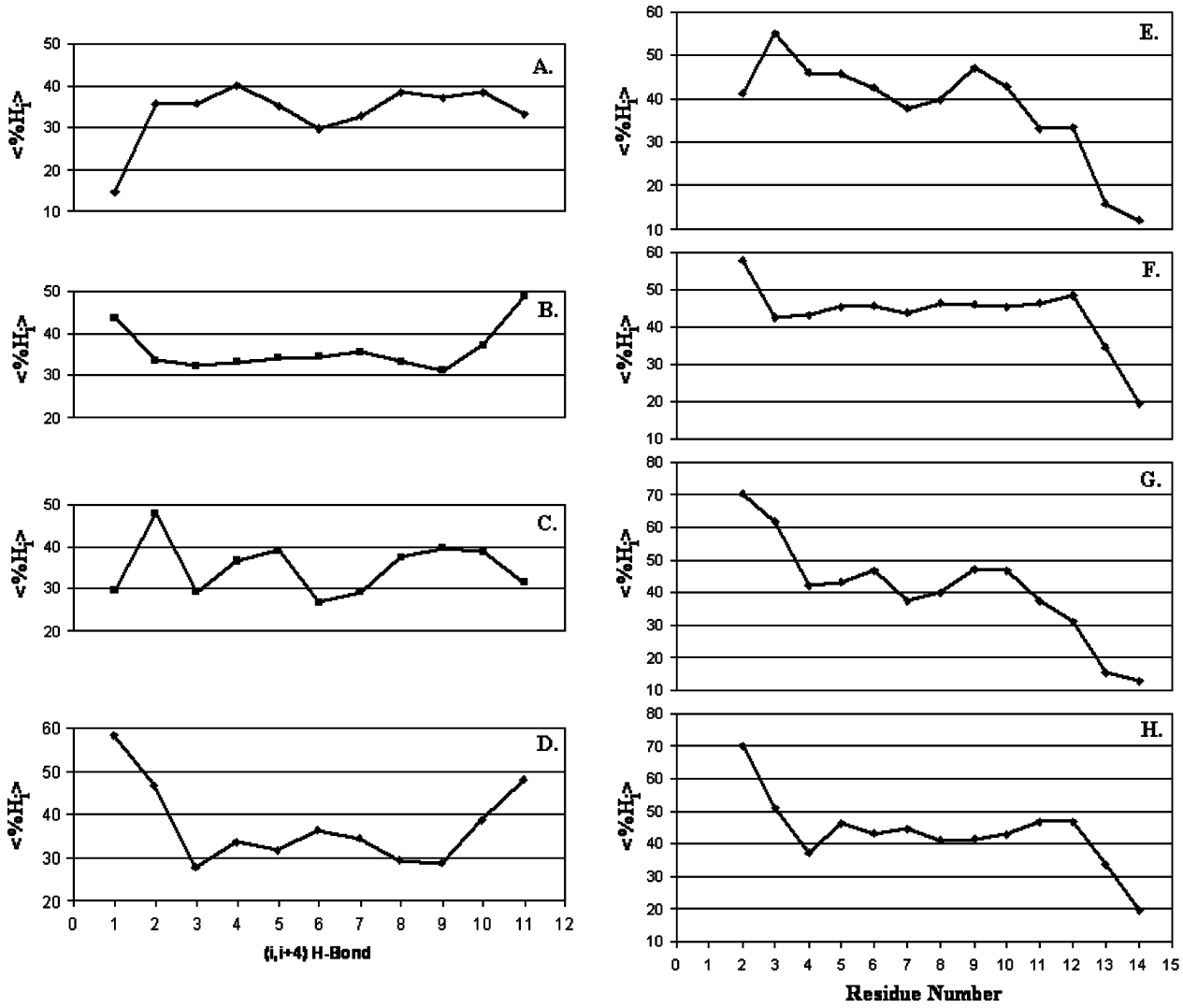

Figure 3. Helical probability, $\langle \%H_i \rangle$, for models I−IV of length 15. Parts A−D represent the helical probability per (i, $i + 4$) hydrogen bond for models I−IV respectively, and parts E−H represent the corresponding helical probability per residue.

groups of 2, as indicated by the jump in the PMF at three consecutive bonds (Figures 4A and 4B), followed by a linear region out to large segment lengths where sampling becomes relatively poor. This seems to indicate that there is minimal energetic or entropic cost to the formation of individual (i, $i +$ 4) hydrogen bonds, but the formation of extended α-helices is energetically unstable at room temperature.

The PMF of the number of consecutive (i, $i + 4$) hydrogen bonds at different temperatures (Figure 5A) were used to determine the enthalpy and entropy of helix formation. The multiple regression of PMFs from different temperature simulations yielded $\Delta H_{nuc} = -0.53 \pm 0.40$ kcal/mol and $\Delta S_{nuc} = -8.1 \pm 1.0$ cal/(mol K), and $\Delta H_s = -1.21 \pm 0.08$ kcal/mol per hydrogen bond and $\Delta S_s = -5.26 \pm 0.22$ cal/(mol K) per hydrogen bond. Figure 5 illustrates ΔG obtained from eq 12 overlaid with the PMFs used in the fit. The values of ΔH_s and ΔS_s correspond to a melting temperature $T_m = 230$ K for the infinite chain. The poor sampling of short helical segments at temperatures of 200 and 250 K corresponds to the freezing of the helix as it becomes thermodynamically stable. This leads to large errors in the estimate of ΔG_{nuc}, and simulations at these temperatures were dropped from the fit.

To validate this approach the percent helicity calculated from eq 4 was fit to the van't Hoff equation. Assuming a two-state model the helix−coil equilibrium constant can be approximated as $K_{eq} = f_H/(1 - f_H)$ where f_H is the helix fraction and equals $\langle \%H \rangle/100$. At equilibrium we can write $\ln K_{eq} = (-\Delta H/RT) + (\Delta S/R)$. A plot of $\ln K_{Eq}$ vs $1/T$ results in a line with slope $-\Delta H/R$ and intercept $\Delta S/R$.[24] This analysis resulted in $\Delta H = -1.18 \pm 0.17$ kcal/mol per residue and $\Delta S = -5.15 \pm 0.47$ cal/(mol K) per residue, which agrees with ΔH_s and ΔS_s to within 2%. Not only do we see good agreement between the values of ΔH and ΔS, this comparison illustrates the connection between the percent helicity measurements and the PMFs obtained from eq 8.

Interestingly, nucleation appears to occur with three consecutive hydrogen bonds as seen from the large increase in the PMF corresponding to ΔG_{nuc}. The PMFs of consecutive hydrogen bonds and consecutive helical residues (based on dihedral criterion) show nearly exact correspondence when the x-axis is shifted by three hydrogen bonds and five residues, respectively, illustrating the relationship between the number of hydrogen bonds and the number of helical residues. From this we see that ΔH_s and ΔS_s also represent the enthalpic and entropic cost of adding an additional helical residue and that the jump in the PMF seen with the hydrogen bonds can be related to the per residue PMF by $\Delta G_{nuc} = 5^*\Delta G_s$ (Figure 5B).

The value of ΔH_s is in excellent agreement with both computationally and experimentally derived values ranging from −0.6 to −1.3 kcal/mol per residue, with the most probable and

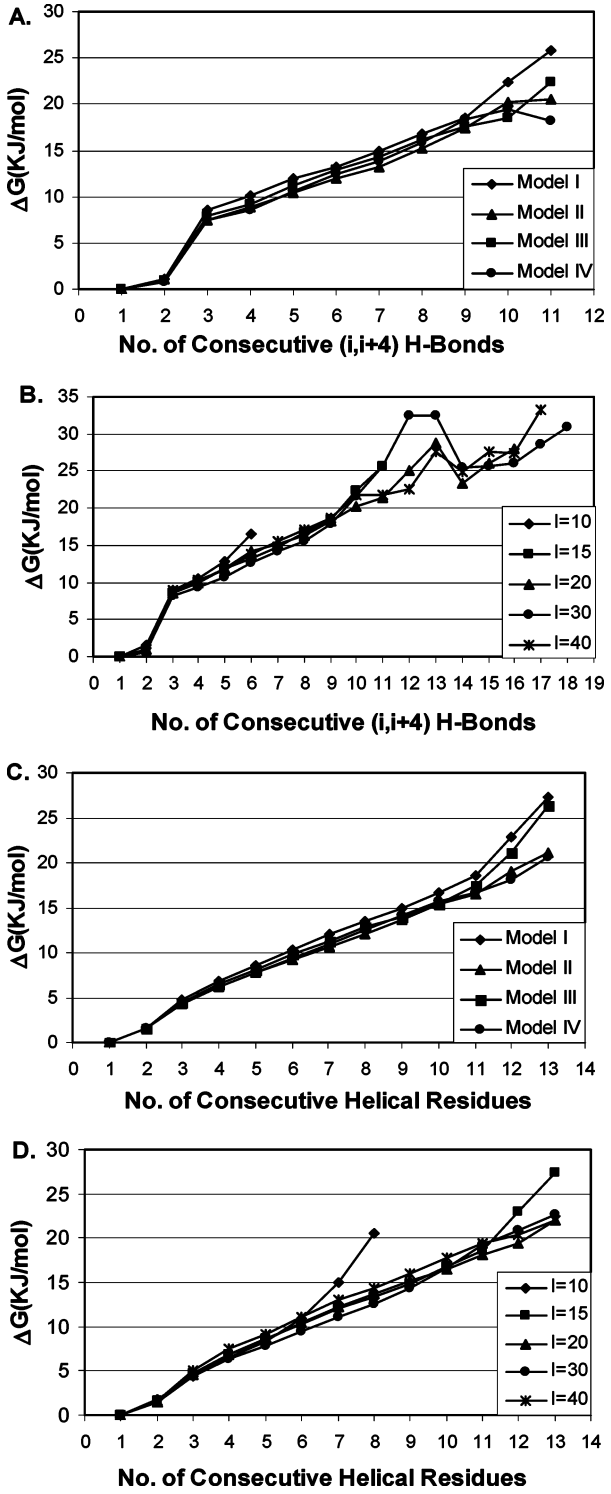

Figure 4. PMF of the number of consecutive hydrogen bonds at $T = 300$ K calculated from eq 8. (A) The PMF of each model with peptide length 15 residues along an abstract coordinate defining successive hydrogen bonds using the criterion of eq 2. (B) The same PMF as in part A for model I peptides from length 10 to 40 residues. (C) The PMF of each model with peptide length 15 residues along an abstract coordinate defining successive helical residues based upon the criterion of eq 1. (D) The same PMF as in part C for model I peptides from length 10 to 40 residues.

accepted value near −1 kcal/mol per residue.[17,21,22,23,28] Recent replica exchange molecular dynamics (REMD) simulations and thermal unfolding experiments have estimated ΔS_s at −2 cal/(mol K) per residue,[21,23] less than half of our value. To compare ΔG_{nuc} we calculated the Zimm−Bragg helix nucleation parameter σ. From our value of ΔG_{nuc} we obtain a value for σ of 0.042 using the expression $\sigma = \exp(-\Delta G_{nuc}/RT)$.[19] This value is an order of magnitude larger than the typical value range reported for σ of 0.001−0.004,[20] leading to an underestimate of the free energy of nucleation by a factor of 2. Although the free energy of nucleation is underestimated according to accepted values of σ, the exact meaning of σ within the particular interpretation of Zimm−Bragg helix−coil theory is ambiguous at best. Larger values of σ are often encountered in peptide simulations[22−24,26,28] and are required to interpret experimental data such as temperature jump experiments.[20] One reason ΔG_{nuc} may be underestimated is the differing definition of the reference state, which in the case of the present work is not an extended coil state but rather a turn structure containing $(i, i + 4)$ hydrogen bonds prior to complete nucleation.

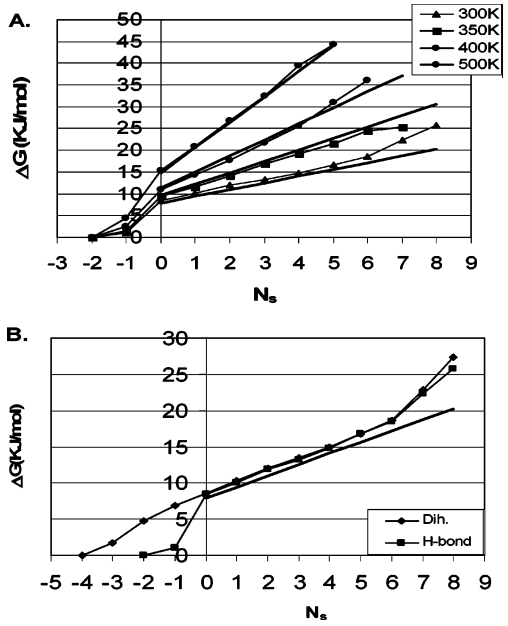

Figure 5. (A) PMF of α-helix segment length based on hydrogen bonds, $W(N_s)$ of eq 8, for $(Ala)_{15}$ at each temperature used in the fitting procedure. (B) PMF of α-helix segment length as a function of consecutive helical dihedral angles and hydrogen bonds. The solid lines in parts A and B indicate the values predicted from eq 12.

Comparing Models. Having discussed the results for the generalized poly-alanine chain within the context of generalized Born molecular dynamics simulations, it is of interest to consider how implicitly and explicitly solvated models differ. The question of which features of short aqueous peptides are correctly represented by GB simulations is unclear, and we take up this subject now. Although asparagine and glutamine differ by only one side chain CH_2 group, circular dichroism (CD) experiments have shown that p1, with Asn at N1, is 5 times more helical than p2, with Gln at N1, with $\%H$ values of 24.4% and 4.8% respectively[3]. Explicit simulations of p1 result in $\langle \%H \rangle$ values of 46% for both dihedral and hydrogen bond criterion, while p2 results in $\langle \%H \rangle$ of 29% and 35% for dihedral and hydrogen bond criteria, respectively. Although explicitly solvated simulations of both peptides exhibit a much greater helical content than determined from CD data, the difference between the $\langle \%H \rangle$ values obtained for p1 and p2 are similar to experiment with $\Delta(\%H) = \langle \%H_{p1} \rangle - \langle \%H_{p2} \rangle = 20$ (experiment), 17 (dihedrals), 11 (hydrogen bonds). Generalized Born simulations, however, have very similar $\langle \%H \rangle$ values of 47% (dihedrals) and 46% (hydrogen bonds) for p1, and $\langle \%H \rangle$ values of 46% (dihedrals) and 43% (hydrogen bonds) for p2. The generalized

Born simulations seem unable to differentiate p1 and p2 and result in percent helicities very similar to that of p1 in explicit solvent.

To further understand the apparent shortcomings of GB, we can examine PMFs, $W(N_s)$, obtained from eq 8 using the distributions of consecutive helical residues as determined by the hydrogen bond criterion and the dihedral angle criterion for p1 and p2 with explicit and implicit solvation. $W(N_s)$ calculated from eq 8 for p1 (Figures 6A and 6C) with similar features and relatively small deviations is quite similar for both GB and explicit solvent simulations. $W(N_s)$ for p2 (Figures 6B and 6D), in contrast, exhibits appreciable deviations between the two solvation models. These findings suggest that the GB solvation model adequately describes p1 but not p2. However, this result is very interesting because we can use both solvation models together to better understand the behavior of various peptides and to discern the interactions stabilizing or destabilizing secondary structure. Although GB screens the interaction of charges according to the dielectric of the medium, in the absence of explicit peptide−solvent interactions, one would expect electrostatic interactions to dominate. Hence, the generalized Born method sees little difference in the electrostatic properties of Asn and Gln at the N-terminus, resulting in similar helical propensities. In explicit solvent the peptide−solvent interactions apparently involving the Gln side chain result in a strong destabilization of the second, third, and fourth hydrogen bonds. A destabilization absent in the GB simulations as can be seen from the helical probability illustrated in Figure 7, where part A shows the results for p1 and part B for p2.

From this we can conclude that the generalized Born simulation can adequately describe some aqueous peptides but certainly not all. In cases where GB fails to give an adequate description, the results can be useful in differentiating the behavior observed in more realistic simulations. A full analysis of the helical propensities and interactions of Asn versus Gln at the N1 position is the subject of a future report.

Clustering. The clustering analysis of models I−IV of length 5 residues yielded structures consistent with the helical measurements. The most populous clusters are shown in Figure 8, and each is a remarkably similar turn structure with 1 (i, $i + 4$) and 1 (i, $i + 3$) hydrogen bond. Most interesting are the relative populations of each structure. Model II with proline and charged termini has a relative population of 53% for this structure while model IV with proline and neutral termini has a population of only 11%, indicating that salt bridging between the termini is a major factor in structural determination at such short lengths. The dssp program requires at least four nonterminal residues to classify α-helical secondary structure, making a comparison to the $\langle \%H \rangle$ measurements via eq 13 irrelevant.

The clustering of model peptides I−IV of length 15 residues yielded much more informative results than that of the 5 residue peptides. Figure 9 illustrates the representative structures of the most populous clusters.

Zwitterionic peptides, models I and III, exhibit a turn in the center of the helix resulting from the interaction of the charged termini, which causes the peptide to fold to bring the termini together in space. The first three most populous clusters of model I exhibit this structural characteristic, which accounts for 26% of the 1500 structures analyzed. However, extended helices are also observed but with a much lower frequency of occurrence. These findings are in agreement with the helical profile (Figure 3) where a drop in helicity in the center of the peptide was observed for models I and III. In the case of model I the N-terminus is extended, directing the charged amino group away from the short helical segment immediately following the turn near the center of the peptide. In contrast the most populous structure for model III has the N-terminus much more closely associated with the preceding helical segment while the C-terminal regions of both models I and III are very similar.

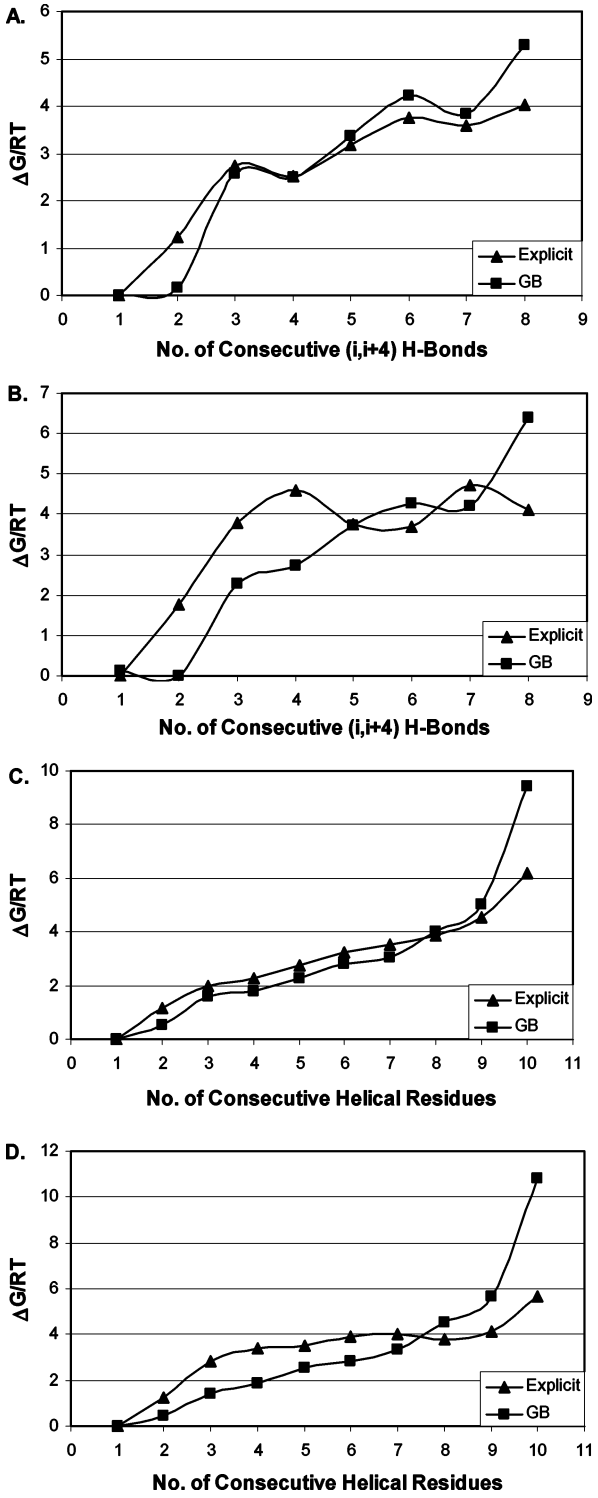

Figure 6. Comparison of explicit solvent and generalized Born simulations. PMF of eq 8 generated from the distribution of consecutive (i, $i + 4$) hydrogen bonds for (A) p1, Asn-Pro-Ala-Glu-$(Ala)_2$-Lys-$(Ala)_3$-Gly-Arg-NH_2, and (B) p2, Gln-Pro-Ala-Glu-$(Ala)_2$-Lys-$(Ala)_3$-Gly-Arg-NH_2. The same PMF generated from the distribution of consecutive helical residues for (C) p1 and (D) p2. Triangles indicate explicit solvent results, and squares represent the GB results.

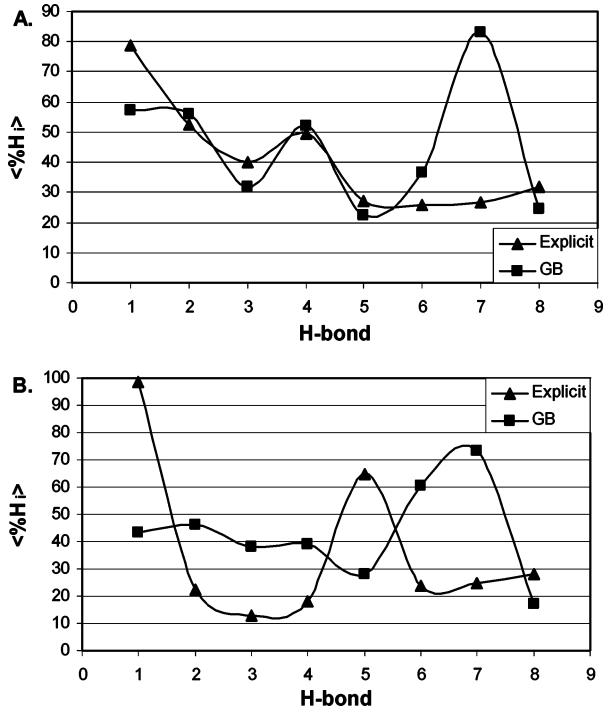

Figure 7. Helical probability per hydrogen bond, $\langle \%H_i \rangle$, for (A) p1 and (B) p2 from implicitly solvated (squares) and explicitly solvated (triangles) simulations.

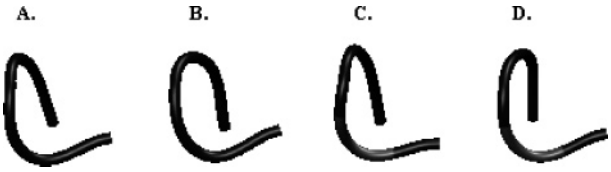

Figure 8. Cartoon representations of the most populous cluster for models I−IV (parts A−D) from simulations of peptides of length 5 residues.

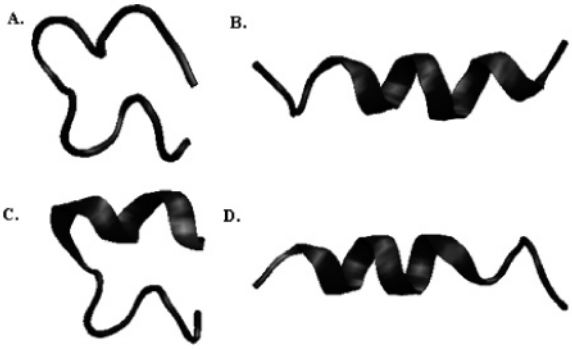

Figure 9. Cartoon representations of the most populous cluster for models I−IV (parts A−D) from simulations of peptides of length 15 residues.

The peptides with neutral termini exhibit α-helical structures that extend the length of the peptide, consistent with the nearly constant helical profile for the interior residues of models II and III (Figure 3). Additionally, the N- and C-termini are participating in the helix with both the N-capping acetyl group and the C-terminal amide forming hydrogen bonds to unpaired backbone amide and carbonyl groups in accord with the predictions of Presta and Rose.[13] Figure 10 illustrates the probability distribution of distances of the acetyl oxygen and backbone amide nitrogens and shows a high probability of hydrogen bonding with residues 2, 3, and 4. The stabilization of such hydrogen bonds, at least within the context of these

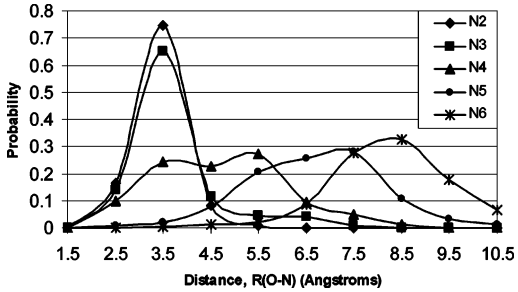

Figure 10. Probability distribution of the acetyl oxygen and backbone amide nitrogen distance for the residues 2 through 6, where residue 1 is taken as the first alanine not the acetyl group.

simulations, is unclear, since we have determined that ΔG is positive for the addition of hydrogen bonds to the helix. However, the determining factor in the sign of ΔG is the entropic cost of adopting an α-helical configuration upon bond formation. At the termini it is not necessary for the terminal or capping groups to maintain a strict helical conformation, since these residues must be helix-breaking simply due to the fact that they are located at the termini. If we consider the majority of helix-initiating events to occur at the N-terminus and that there is very little entropic penalty for the N-cap to form a hydrogen bond, then there should be a reduction in ΔG_{nuc} nearly equal to ΔH_s. In fact we see a $\Delta\Delta G_{\text{nuc}}$ of only one-quarter of ΔH_s, indicating the predominant effect to be due to the favorable interaction of the acetyl group with the helix dipole opposed to the additional hydrogen bond.

Conclusion

This study has shown that within the context of classical generalized Born molecular dynamics simulations the α-helix is not inherently stable at room temperature, although it is a common secondary structural motif in equilibrium with turn and extended structures. In all cases of peptide length greater than 10 residues the percent helicity was observed to be approximately 30% with an increasingly Gaussian distribution of helical hydrogen bonds centered about this mean as peptide length increases. This result coupled with the ambiguous percent helicity measurements for peptides of lengths less than 10 residues and the new definition of the helix nucleus leads to the conclusion that true α-helical secondary structure is not likely to occur in peptides less than 10 residues with an absolute minimum length of 7 residues necessary for nucleation.

The helix-initiating step occurs via a turn structure intermediate with the formation of both helical hydrogen bonds and the transition of several residues to an α-helical conformation. The relatively large frequency with which one and two consecutive (i, $i + 4$) hydrogen bonds are observed implies the rapid formation and dissociation of pre-helix structures with relatively few nucleating events. From examining the PMF corresponding to sequential helical hydrogen bonds, $W(N_s)$ of eq 8, we have observed that helix initiation occurs when three consecutive hydrogen bonds form and that this corresponds to five consecutive residues adopting α-helical dihedral conformations. It is typical to assume that a single (i, $i + 4$) hydrogen bond with the three nonterminal residues in helical dihedral conformations is a sufficient definition of helix nucleation; however, the current simulations suggest that up to two helical hydrogen bonds can form without the entropic loss associated with the residues forming a highly regular structure.

The substitutions studied were found to have little effect on the overall form of the PMF with primary influence on the free

energy of nucleation. This lack of sensitivity to substitution is also present in the percent helicity measurements, which exhibited the greatest difference between models for short peptides, with little effect on larger peptides, indicating a short range of action. However, the helical profile of each model was significantly altered by the substitutions with primary effect on the N-terminus and, to a lesser extent, the residues in the body of the peptide. Incorporation of proline into the second position with an acetyl N-capping group showed the greatest increase among the models studied in the frequency with which the first two helical hydrogen bonds form, increasing the likelihood of helix formation at the N-terminus. We conclude that proline in the second position supports helical structure by limiting the dihedral conformational degrees of freedom, as observed from the helical probability per residue. The removal of charge from the termini has a profound effect on the α-helical secondary structure and the tertiary structure observed from the clustering analysis. The fact that the replacement of positive charge with a polar group at the N-terminus has a limited range of action is in agreement with the conclusions of Åqvist[6] that the effect of the helix dipole is primarily localized in the last, or in this case the first, turn of the helix.

Acknowledgment. We thank April Voss for her help organizing this paper.

References and Notes

(1) Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 205.
(2) Wild, C. T.; Shugars, D. C.; Greenwell, T. K.; McDanal, C. B.; Matthews, T. J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9770.
(3) Forood, B.; Feliciano, E. J.; Nambiar, K. P. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 838.
(4) Forood, B.: Reddy, H. K.; Nambiar, K. P. *J. Am. Chem. Soc.* **1994**, *116*, 6935.
(5) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.
(6) Åqvist, J.; Luecke, H.; Quiocho, F. A.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2026.
(7) Doig, A. J.; Baldwin, R. L. *Protein Sci.* **1995**, *4*, 1336.
(8) Cochran, D. A. E.; Penel, S.; Doig, A. J. *Protein Sci.* **2001**, *10*, 463.
(9) Cochran, D. A. E.; Penel, S.; Doig, A. J. *Protein Sci.* **2001**, *10*, 1305.
(10) Jarrold, M. F. *Annu. Rev. Phys. Chem.* **2000**, *51*, 179.
(11) Kinnear, B. S.; Hartings M. R.; Jarrold, M. F. *J Am. Chem. Soc.* **2001**, *123*, 5660.
(12) Breaux, G. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **2003**, *125*, 10740.
(13) Presta, L. G.; Rose, G. D. *Science* **1988**, *240*, 1632.
(14) Polinsky, A.; Goodman, M.; Williams, K. A.; Deber, C. M. *Biopolymers* **1992**, *32*, 399.
(15) Richardson, J. S.; Richardson, D. C. *Science* **1988**, *240*, 1648.
(16) Hirst, J. D.; Brooks, C. L, III. *J. Mol. Biol.* **1994**, *243*, 173.
(17) Scholtz, J. M.; Marquesee, S.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Santoro, M.; Bolen, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2854.
(18) Zimm, B. H.; Bragg, J. K. *J. Phys. Chem.* **1959**, *31*, 526.
(19) Qian, H.; Schellman, J. A. *J. Phys. Chem.* **1992**, *96*, 3987.
(20) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **1997**, *36*, 9200.
(21) Jas, G. S.; Kuczera, K. *Biophys. J.* **2004**, *87*, 3786.
(22) Sung, S.; Wu, X. *Proteins: Struct., Funct., Genet.* **1996**, *25*, 202.
(23) Ohkubo, Y. Z.; Brooks, C. L., III. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 24, 13916.
(24) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472.
(25) Mitsutake, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *112*, 10638.
(26) Shental-Bechor, D.; Kirca, S.; Ben-Tal, N.; Haliloglu, T. *Biophys. J.* **2005**, *88*, 2391.
(27) Young, W. S.; Brooks, C. A., III. *J. Mol. Biol.* **1996**, *259*, 560.
(28) Tobias, D. J.; Brooks, C. L., III. *Biochemistry* **1991**, *30*, 6059.
(29) Wang, L.; O'Connell, T.; Tropsha, A.; Hermans, J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10924.
(30) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner P. K.; Kollman, P. A. *AMBER 7*; University of California: San Francisco, CA, 2002.
(31) Hovmoller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr., Sect. D* **2002**, *58*, 768.
(32) Kelley, L. A.; Gardner, S. P.; Sutcliffe, M. *J. Protein Eng.* **1996**, *9*, 1063.
(33) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(34) Bevington, P. R.; Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill: San Fransisco, CA, 1992.