Helix Folding of an Alanine-Based Peptide in Explicit Water

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · FEBRUARY 2001

Impact Factor: 3.3 · DOI: 10.1021/jp004048a

2 AUTHORS:

Xiongwu Wu
National Heart, Lung, and Blood Institute
82 PUBLICATIONS **4,425** CITATIONS
SEE PROFILE

Shaomeng Wang
University of Michigan
291 PUBLICATIONS **14,349** CITATIONS
SEE PROFILE

Helix Folding of an Alanine-Based Peptide in Explicit Water

Xiongwu Wu and Shaomeng Wang*

Institute for Cognitive and Computational Sciences and Departments of Oncology and Neuroscience, Georgetown University Medical Center, The New Research Building, EP07, 3970 Reservoir Rd., Washington, D.C. 20007

Received: November 2, 2000; In Final Form: January 2, 2001

Computer simulations using full atomic representations for both the peptide and water molecules were performed to study the folding of a 16-residue alanine-based helical peptide in aqueous solution. Using a recently developed self-guided molecular dynamics (SGMD) method, which was shown to improve the conformational searching efficiency significantly as compared to conventional MD simulation method, reversible folding (folding, unfolding and refolding) of a 16-residue alanine-based synthetic peptide in explicit water at 274 K was successfully accomplished. Consistent with experimental results, the helix was found to be the major secondary structural element in aqueous solution, and among different helix forms, the α-helix is the dominant form. Conformational analysis of our simulation results showed that turns and 3_{10}-helices play an essential role in the folding of α-helix. Interestingly, our results showed that the propagation of a helix segment is more frequent at the C-end than at the N-end. In most helix conformations, the backbone carbonyl groups of the peptide prefer to simultaneously form intramolecular hydrogen bonds with the backbone amide groups of the peptide and intermolecular hydrogen bonds with water molecules, indicating water accessibility to the backbone carbonyl groups is crucial for helix formation in water. Therefore, the helical propensities of amino acids may be related to the water accessibility of their backbone groups in helical conformation. Water molecules also function as hydrogen bonding bridges linking helical residue pairs (i, $i + n$, with $n = 3, 4, 5$), suggesting a role of water bridges in helix folding.

Introduction

The helix is a rudimentary secondary structure in proteins. Helix formation has been implicated to play an important role in protein folding, which remains one of the most challenging and difficult problems in biological science. Elucidation of helix folding mechanism can therefore provide important insights into the protein-folding problem. Alanine-based peptides, which adopt significant populations of helical structure in aqueous solution,[1;2] have served as excellent experimental model systems for the studies of helix folding mechanism. Using alanine-based peptides, Baldwin and his colleagues have performed extensive studies on the thermodynamic and theoretical aspects of helix folding primarily using circular dichroism (CD) and NMR techniques.[3−7] Employing double spin label electron spin resonance (ESR) technique, Millhauser and co-workers have investigated the structural details of alanine-based peptide systems in terms of the 3_{10}- and α-helix populations in solution[8−10] and found that some alanine-based peptides may have significant 3_{10}-helix population in aqueous solution and the 3_{10}-helix may play a crucial role in the folding of the α-helix. In recent years, the time scale of helix folding has also been investigated using alanine-based peptide systems.[11−13]

Computer simulation, referred to as "computer experiment", is highly complementary to experimental and theoretical studies and can provide rich information at various level of resolution. In recent years, this approach has been used to study the helix folding mechanism.[14−23] However, despite extensive theoretical and simulation studies, helix folding mechanism has not been fully understood. A major reason is that helix folding of proteins and peptides at physiological condition remains a rare event in computational studies with current computing power, especially with explicit water molecules.

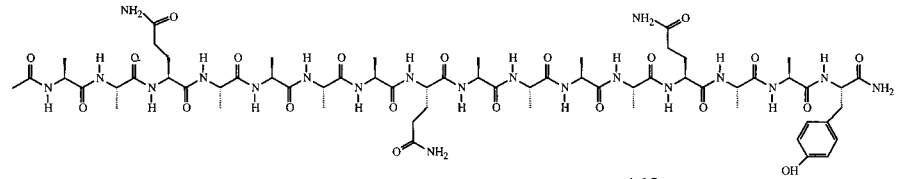
Recently, we developed the self-guided molecular dynamics (SGMD) simulation method to improve the conformational searching efficiency.[24−26] Using model peptide systems, we have demonstrated that while the SGMD can achieve a much-improved conformational searching efficiency as compared to conventional MD method, it does not significantly alter the conformational distribution or other thermodynamic properties with proper parameters.[24] Hence, the SGMD simulation method provides us an unprecedented opportunity to study helix folding with full atomic representation for both the peptide and water molecules.

To be able to study helix folding through simulation, in addition to an efficient simulation method, the force fields used to represent the peptide and the solvent molecules must be fairly accurate to describe the system. Recent successful folding simulations with all atomic detailed model by Daura et al.[22] and by Duan and Kollman,[23] as well as our own simulations,[24] suggest that current force fields such as AMBER,[27] CHARMM,[28] and GROMOS force fields[29] may be accurate enough for the study of helix folding in the presence of explicit solvent molecules.

In this study, we performed SGMD simulations as well as conventional MD simulations for a 16-residue, alanine-based peptide in explicit waters. Reversible helix folding processes were observed in SGMD simulations and enabled us to gain new insights into helix folding.

* To whom correspondence and request of reprints should be sent. Tel: (202) 687−2028. Fax: (202) 687−0617. E-mail: wangs@giccs.goergetown.edu.

10.1021/jp004048a CCC: $20.00
Published on Web 02/17/2001



Ace-Ala-Ala-Gln-Ala-Ala-Ala-Ala-Gln-Ala-Ala-Ala-Ala-Gln-Ala-Ala-Tyr-NH_2

Figure 1. Chemical structure of the 16-mer alanine-based peptide. It consists of three type of residues, alanine (Ala or A), glutamine (Glu or Q), and tyrosine (Tyr or Y), with N-terminal blocked with an acetyl group and C-terminal blocked with an amide group.

Materials and Method

1. Simulation Setup. The alanine-based peptide used in our study has the following sequence Ace-Ala-Ala-Gln-Ala-Ala-Ala-Ala-Gln-Ala-Ala-Ala-Ala-Gln-Ala-Ala-Tyr-NH_2, whose chemical structure is shown in Figure 1. This neutral and water-soluble peptide was first studied in Baldwin's laboratory[30] and was found to have 50% helix content in aqueous solution.

The peptide aqueous solution was constructed by immersing a peptide molecule in its fully extended conformation into a box of TIP3P waters[31] and deleting overlapping water molecules that were within 2.45 Å of the peptide. The simulation system contains 1 peptide and 1945 TIP3P water molecules. The cubic periodic boundary condition was applied in all the simulations to eliminate the boundary effect. The SHAKE algorithm[32] was used to fix all bond lengths so that a large time step of 0.002 ps can be used in the simulations. NTP simulations were performed (T = 274 K and P = 1 atm). Temperature and pressure were controlled using Berendsen's algorithms.[33] The temperature was coupled to a temperature bath with a coupling constant of 1.0 ps, and the pressure coupling constant was set to be 1.0 ps. It is noted that Berendsen's algorithm does not produce a correct thermodynamic ensemble, but it has been frequently used in simulations of biological systems for its gentle constraint behavior. Conventional MD simulations were performed using the AMBER program,[34;35] while SGMD simulations were performed using the SGMD method implemented in the AMBER program.[24;26] Unless otherwise stated, MD has been used to refer to conventional MD simulations and SGMD has been used to refer to the self-guided MD simulations. The AMBER94 all atom force field[27] was used for energy and force calculation. A constant dielectric constant of 1 was used for the electrostatic interaction calculations. The cutoff distance for nonbonded interactions, including Lennard-Jones and electrostatic interaction was set at 10 Å on the basis of the residues.

2. Secondary Structure Analysis. To provide a detailed description of the secondary structure of the peptide during the simulation, the dictionary of secondary structure of protein (DSSP) proposed by Kabsch and Sander[36] was employed. According to the DSSP definition, cooperative secondary structure is recognized as repeats of the elementary hydrogen-bonding patterns "turn" and "bridge". Repeating turns are "helices", repeating bridges are "ladders", and connected ladders are "sheets". Geometric structure is defined in terms of the concepts of torsion and curvature of differential geometry. Local chain "chirality" is the torsion handedness of four consecutive Cα positions and is positive for right-handed helices and negative for ideal twisted β-sheets. Curved pieces are defined as "bends". For the 16-residue alanine-based peptide, the two terminal blocking groups, the acetyl at the N-terminal and the amide at the C-terminal, are counted as separate amino acids when defined the secondary structures of residues 1−16.

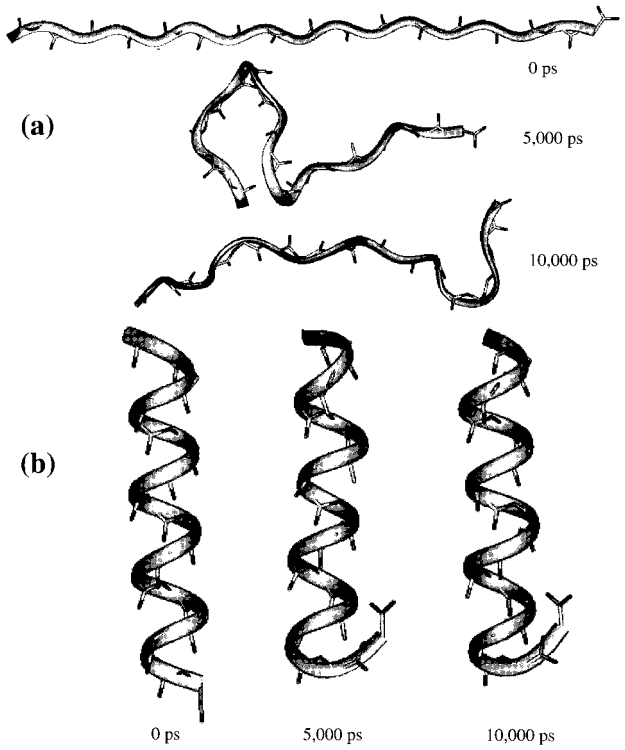


Figure 2. Snapshots of the peptide conformations obtained in the two 10 ns MD simulations: (a) simulation started from a fully extended conformation; (b) simulation started from a complete helix conformation.

Results and Discussion

1. Simulations for the 16-Residue Helical Peptide. We performed 10-ns conventional MD simulations on the peptide system from either a fully extended conformation or a complete helix conformation. Analysis of the simulation trajectories showed that, during the entire 10 ns simulation, no helix segment was formed when starting from a fully extended conformation. When started from a complete helix conformation, the peptide remained in the same complete helix conformation. Some of the snapshots taken from these simulations are shown in Figure 2. These simulation results indicate that a 10 ns MD simulation is not long enough for either helix folding or helix unfolding. A recent study showed that, for a 50 ns MD simulation of a 7 residue β-peptide in methanol at 298 K, helix folding and unfolded occurred only once.[22] In the 1 μs MD simulation performed by Duan and Kollman, substantial helix was formed after 50 ns.[23] It is thus expected that the helix formation of this alanine-based peptide in water might also occur in a similar time scale. Using the conventional MD simulation method, reversible helix folding of this alanine-based peptide would take at least a few hundred nanoseconds to accomplish. Since each 10 ns MD simulation takes about 10 weeks of CPU time on an R10000 SGI workstation, it will require substantial

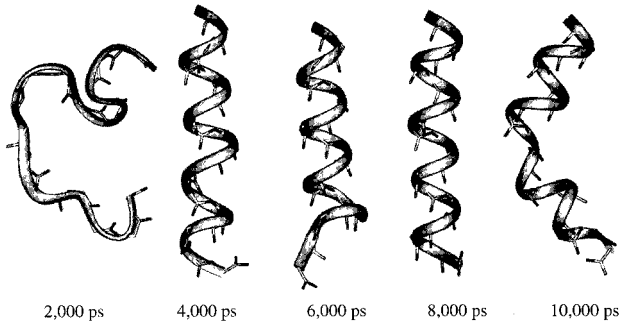


Figure 3. Snapshots of the peptide conformations obtained in the SGMD simulation with $\lambda = 0.3$. The simulation started from a fully extended conformation.

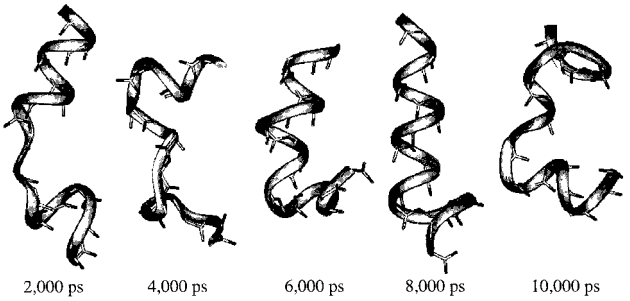


Figure 4. Snapshots of the peptide conformations obtained in the SGMD simulation with $\lambda = 0.4$. The simulation started from a fully extended conformation.

computing resource to perform a few hundred nanoseconds simulations at the present time.

The self-guided molecular dynamics (SGMD) simulation method was developed to enhance the conformational searching ability of a simulation.[24;26] Two parameters, the guiding factor, λ, and the local sampling time, t_l, are used to define the self-guiding effect in a simulation. The local sampling time, t_l, is used to calculate the guiding forces, which are computed as averages of instant forces over a time period of t_l. For simulation efficiency, t_l is chosen to be a relatively short time interval within which no significant conformational changes occur. For simulation of proteins without explicit water, we usually chose a value between 0.05 and 0.5 ps; for simulations of proteins with explicit water molecules, we usually chose a value between 0.2 and 2 ps. The guiding factor, λ, is used to define how much guiding effect will be introduced into a simulation. We usually chose a value between 0.1 and 0.3 for simulations of proteins without explicit water molecules and between 0.2 and 0.5 for simulations of proteins in explicit water solution. When $\lambda = 0$ or $t_l \rightarrow \infty$, a SGMD simulation is identical to a conventional MD simulation. A larger λ will result in faster systematic conformational change (not thermomotion, which depends on temperature). But λ should be small enough to avoid significant conformational change within time interval, t_l.

Started from the same fully extended conformation, two 10 ns SGMD simulations with $t_l = 2$ ps but $\lambda = 0.3$ and 0.4, respectively, were performed. Several snapshots from these two simulations are shown in Figures 3 and 4, respectively. As can be seen, in both simulations, the peptide folded into conformations with helix segments.

To some extent, gyration radius of the peptide reflects its overall conformation. The gyration radius of the peptide during each of the four simulations is shown in Figure 5. As can be seen, the lowest value of the gyration radius during the MD simulation starting from an extended conformation is around 9 Å, while the gyration radius during the second MD simulation

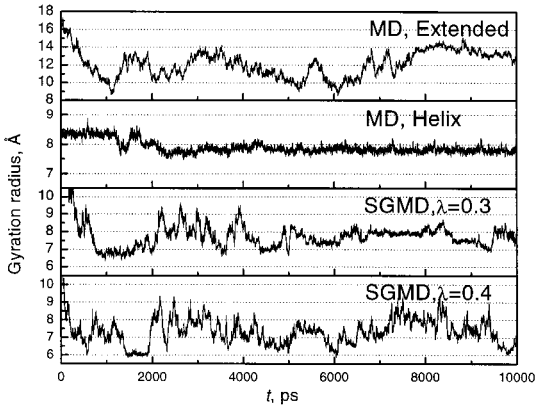


Figure 5. Gyration radius of the peptide during each of the two MD and two SGMD simulations.

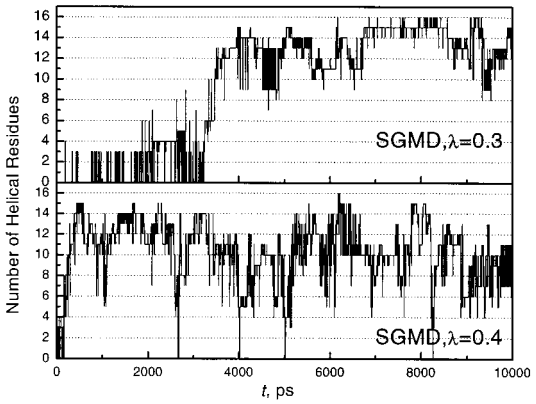


Figure 6. Total number of helical residues observed in the peptide during the two 10 ns SGMD simulations with $\lambda = 0.3$ and 0.4, respectively.

started from a complete helix conformation remains around 8 Å. These results show that the 10 ns MD simulation is not long enough for the formation of helix segments when starting from a fully extended conformation. Furthermore, a 10 ns MD simulation is also not long enough for the peptide to unfold from the complete helix conformation. In contrast, in both the SGMD simulations, the gyration radius decreases from the value of extended conformations to the value even lower than that of a complete helix and fluctuates between 6 and 9 Å. For example, in the 10 ns SGMD simulation with $\lambda = 0.3$, the lowest gyration radius reached was 6.4 Å. In the 10 ns SGMD simulation with $\lambda = 0.4$, the gyration radius of the peptide reached 6.0 Å after about 560 ps. As evident from Figure 5, the conformational change of the peptide during the SGMD simulation with λ of 0.3 is significantly slower than that with λ of 0.4 but much faster than the conformational change in the two MD simulations.

Figure 6 plots the total number of helix residues in each conformation during the two SGMD simulations. As can be seen, while both SGMD simulations reached conformations with high helix ratio (14/16), the simulation with a guiding factor (λ) of 0.4 resulted in a faster folding and unfolding process than the simulation with λ of 0.3. In the SGMD simulation with $\lambda = 0.3$, it took 4000 ps for the peptide to reach the conformations with 90% of helix while it took only 600 ps with $\lambda = 0.4$. Furthermore, during the 10 ns SGMD simulation with λ of 0.4, the folding and unfolding occurred much more frequently than that during the simulation with λ of 0.3. The 10 ns SGMD simulation with λ of 0.3 is not long enough to completely unfold the helix upon the formation of the helix. But the 10 ns SGMD simulation with λ of 0.4 resulted in four times of complete helix

TABLE 1: Percentages of Secondary Structural Elements Observed for Each Amino Acid in the Peptide during the 10 ns SGMD Simulation with $\lambda = 0.4$

residue no.	bend	bridge	ladder	turn	3_{10}-helix	α-helix	π-helix	tot. helix ($\alpha + 3_{10} + \pi$)
1	0	0	0	9.5	0.6	75.0	0.5	76.2
2	9.3	0	0	10.3	0.9	77.4	0.6	78.9
3	5.8	0.2	0.3	8.0	1.0	78.7	0.7	80.5
4	4.0	0	0.3	4.5	0.6	81.9	0.7	83.1
5	2.8	0.1	0	7.6	4.1	82.4	0.7	87.1
6	5.5	0	0	16.3	5.8	69.0	0.3	75.1
7	7.7	0.1	0.3	18.6	6.5	62.3	0.2	69.0
8	10.5	0.2	0.3	20.2	5.4	50.1	0.7	56.2
9	15.0	0	0	7.9	2.6	45.9	0.7	49.1
10	7.4	0	0	10.0	2.2	71.0	0.7	73.9
11	4.5	0	0	22.2	3.9	66.7	0.7	71.3
12	7.4	0	0	23.9	5.3	55.4	0.6	61.3
13	9.9	0	0	19.4	7.1	50.7	0.1	57.9
14	13.5	0	0	25.5	7.7	45.5	0.1	53.3
15	23.2	0	0	25.0	4.6	33.4	0.0	38.0
16	0	0	0	25.5	2.1	10.6	0.0	12.7
av	7.9	0	0.1	15.9	3.8	59.8	0.5	64.0

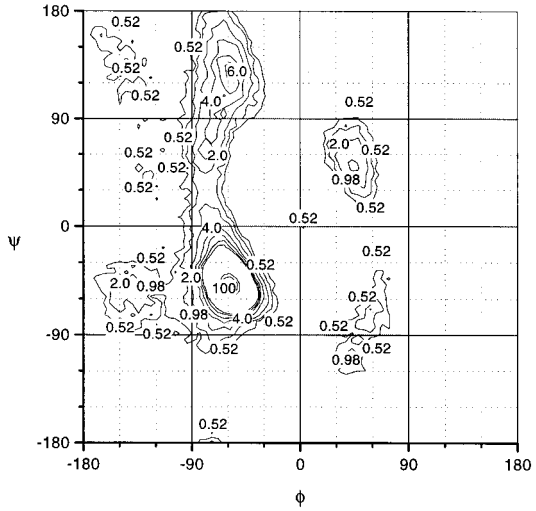


Figure 7. Ramachandran plot of the 16-mer alanine-based peptide obtained in the SGMD simulation with $\lambda = 0.4$. The simulation started from a fully extended conformation.

unfolding, at 2700, 4000, 5000, and 8300 ps, respectively, and subsequent refolding. Therefore, only the 10 ns SGMD simulation with $\lambda = 0.4$ reached a reversible helix folding. Thus, our analyses were focused on the SGMD simulation with $\lambda = 0.4$. It should be noted that even the 10 ns SGMD simulation with $\lambda = 0.4$ is by no means long enough to provide a sufficient sampling of all the important conformational space of this peptide. Nonetheless, the reversible folding of this peptide in explicit waters achieved through the SGMD simulation provided us an opportunity to investigate several important issues concerning the helix folding.

2. Secondary Structures. The Ramachadran plot is a very useful tool to examine the backbone torsion conformations of peptides and proteins. The Ramachadran plot for this peptide based upon the 5000 conformations recorded during the 10-ns SGMD simulation with λ of 0.4 is shown in Figure 7. As can be seen, the peptide has extensively sampled its backbone torsion space. The most dominant backbone conformation is around the α-helix region (−57°, −47°), followed by the 3_{10}-helix region (−60°, −0°). Interestingly, there is certain population located in the left-handed α-helix region (57°, 47°) and in the β-structure region (−140°, 140°). A significant population is located in a large, random coiled conformational region (−30 to −100°, 0 to 180°).

To provide a detailed description of secondary structure for each residue, we classified the conformations recorded during the simulations according to the dictionary of secondary structure of protein (DSSP).[36] The average ratio of each secondary structural element (H, α-helix; G, 3_{10}-helix; I, π-helix; T, turn; S, bend; B, bridge; E, ladder) for each residue was calculated on the basis of all the 5000 conformations recorded during the SGMD simulation ($\lambda = 0.4$) and listed in Table 1.

As can be seen from Table 1, on average, helix secondary structure, including α-, 3_{10}-, and π-helix, is the dominant form for the peptide and the average helix ratio is 64%, which is comparable to the experimental helix ratio of 50%, as estimated from CD experiments.[30] Among the three different helix elements, α-helix is the dominant form (59.8%) and there is a significant population of the 3_{10}-helix (3.8%). The least populated helix form for this peptide is π-helix (0.5%). Our results thus clearly show that the α-helix conformation is much more populated than the 3_{10}-helix conformation. This is consistent with a previous study for this peptide system, which showed that α-helix is more stable than 3_{10}-helix in aqueous solution.[37]

A number of recent experimental studies using double spin label electron spin resonance (ESR) have suggested that a significant population of the 3_{10}-helix conformation form may exist for short alanine-based peptides.[8−10] On the basis of our simulation studies, it is clear that there is indeed a significant population of the 3_{10}-helix conformation (3.8%) for this peptide in aqueous condition, thus providing direct evidence for the existence of the 3_{10}-helix for alanine-based peptides in water. It is of note that the 3_{10}-helix conformation prefer to form at two locations, residues 5−8 in the middle of the peptide and residues 11−15 near the C-terminus. Using a double spin label ESR technique, Fiori et al. have found that there is a strong evidence of 3_{10}-helix geometry near the C-terminus for 16- and 21-mer alanine-based peptides,[8] and our results are in good agreement with the experimental observations.

Turns are the next most populated secondary structural element (15.9%), which has been implied to play an important role in helix initiation and propagation.[14,38] Indeed, it was found that 87.5% of helix transitions (folding and unfolding) are between helices and turns during the 10 ns SGMD simulation, confirming the important role played by turns in helix folding. We observed a small population of bridges and β-strands, which mainly involve residues 3, 4, 7, and 8. Their low populations suggested that they are transient states under the simulation

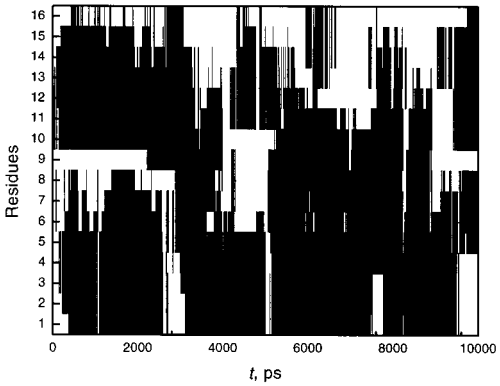


Figure 8. Helix segments observed in the peptide at each conformation during the 10 ns SGMD simulation with $\lambda = 0.4$. The simulation started from a fully extended conformation. The helical residues are marked with a vertical bar at their corresponding position for each conformation.

conditions. Left-handed helix was also observed from 6130 to 6560 ps, located near the C-terminus.

The helix ratio for each residue depends not only on residue types (Ala, Tyr, Gln) but also on the location of the residues. Very low helix ratios were observed for the C-terminal residues 15 and 16, being only 38.0% and 12.7%, respectively. From the N- to the C-terminal, two peaks of the helix ratio were found, located at residues 5 and 10 and separated by residue 9. To further investigate the solution conformation of the peptide during the simulation, for each conformation, each residue was classified as helical (H) or nonhelical (C) on the basis of the DSSP. Residues in α-, 3_{10}-, and π-helices are classified as helical, and all others, as nonhelical. According to this classification, the helix segments of the peptide for each conformation were plotted against the simulation time (Figure 8). The helix segments were formed within the first 100 ps of the simulation. During the 10 ns SGMD simulation, helix folding, unfolding, and refolding occur frequently, suggesting that the peptide structure was not trapped in a particular conformation. As can be seen in Figure 8, the peptide does not simply adopt either a complete helix or random coil conformations but has many rich conformational forms. The conformations of the peptide may be clustered according to the number and location of helix segments in the peptide. Nine major conformational clusters were identified, as shown in Figure 9. As can be seen, the complete helix conformation only accounts for only 0.1% of the total number of the conformations. The coiled conformation (C) only accounts for 0.8% of the conformations. The most populated conformational cluster (HC, 38%) has a helix segment at the N-terminal and a frayed C-terminal. The next two most populated conformational clusters have either one (CHC) or two helix segments (CHCHC) in the middle of the peptide and with frayed N- and C-terminals, accounting for 8.9% and 7.9% of the total conformations, respectively. The next most populated conformation (HCH) has two helix segments at the N- and C-terminal and with a break in the middle of the peptide (5%). The conformation with a C-terminal helix segment and a frayed N-terminal (CH) accounts for 4% of the total conformations. The conformation with a C-terminal helix segment, followed by a coiled segment, a helix segment near the N-terminal, and a frayed N-terminal (CHCH) accounts for about 2% of the total population. Last, the conformation with a helix segment at the N-terminal, followed by a coiled segment, a small helix segment, and a frayed C-terminal (HCHC), accounts for less than 1% of the total conformations. Although these values should be treated with care and are still semiquantitative in nature, our reversible simulation nevertheless showed that this alanine-based peptide

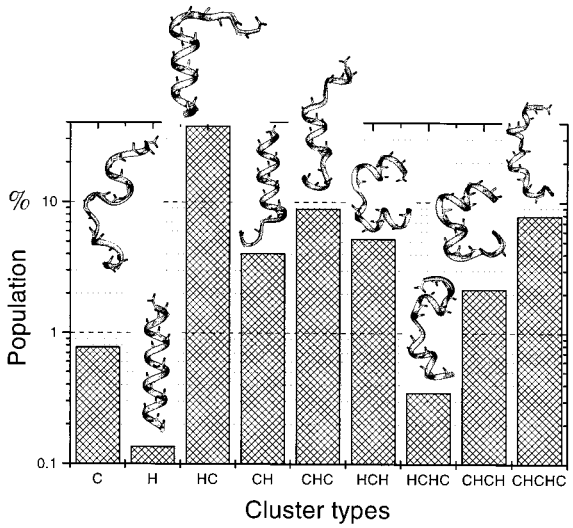


Figure 9. Major secondary structure clusters observed during the 10 ns SGMD simulation with $\lambda = 0.4$. Conformations are clustered on the basis of the number and location of helix segments in the peptide.

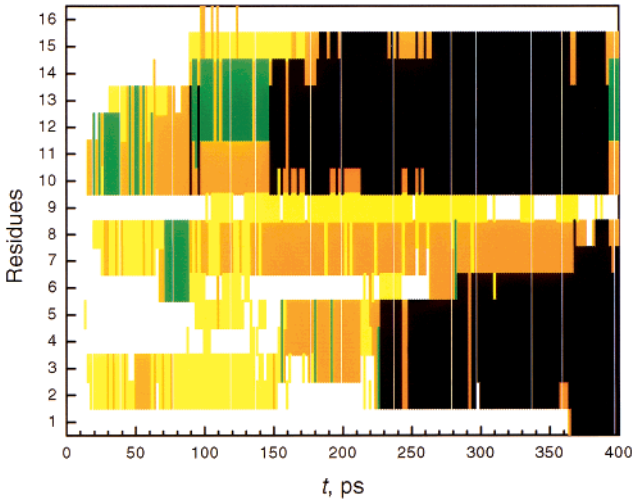


Figure 10. Secondary structural segments observed during the first 400 ps of the SGMD simulation with $\lambda = 0.4$. The simulation started from a fully extended conformation. Secondary structures are defined by following DSSP convention. For each conformation, secondary structures of residues are indicated by colored bars at their corresponding positions: yellow, bend; red, turns; green, 3_{10}-helix; black, α-helix.

adopts multiple helical conformations and the complete helix conformation only accounts for a very small percentage of total conformations. Accordingly, the experimentally observed 50% helix content for this alanine-based peptide must represent the average of various conformational forms.

3. Helix Initiation and Propagation. Two major issues in helix folding are helix initiation and propagation. Since helix folding and unfolding occurred repeatedly during the 10 ns SGMD simulation with λ of 0.4, we had the opportunity to investigate the mechanism of helix initiation and propagation.

We first examined the helix initiation starting from the fully extended conformation. In Figure 10, the evolvement of the secondary structures during the first 400 ps of the simulation was shown. As can be seen, starting from the fully extended conformation, the peptide first formed bends and turns (e.g., on residues A10 and A11 at 16 ps, on A7 and A8 at 28 ps, and A2 and Q3 at 30 ps). Thereafter, some turn structures evolved into 3_{10}-helix segments (e.g., on residues A10-A12 at 20 ps, and on A6-Q8 at 72 ps), which can either transfer back to turns (e.g., on A11−A13 at 38 ps and on A6−Q8 at 88 ps) or further

TABLE 2: Number of Transitions between Different Secondary Structures Found in the 5000 Conformations Recorded during the SGMD Simulation with $\lambda = 0.4$[a]

from/to	coil	bend	turn	3_{10}-helix	α-helix	π-helix
coil				0	0	0
bend				3 (0)	0 (0)	0 (0)
turn				131 (89)	74 (58)	8 (17)
3_{10}-helix	0	6 (5)	122 (87)		58 (20)	0 (0)
α-helix	0	0 (0)	76 (54)	63 (21)		10 (2)
π-helix	0	0 (0)	9 (19)	0 (0)	8 (2)	

[a] The data in parentheses are the numbers found in the 5000 conformations recorded during the SGMD simulation with $\lambda = 0.3$.

evolve to form α-helices (e.g., on A9−A12 at 96 ps and on A2−A5 at 228 ps). α-Helices were also observed to transfer back to turns (e.g., A2−A5 at 244 ps) and 3_{10}-helices (e.g., on A12−A14 at 98 ps and on A12−A14 at 392 ps). Turns are also observed to transfer to α-helices directly (e.g., on A10−Q13 at 92 ps and on A2−A5 at 248 ps).

To gain a more quantitative description on the helix initiation, we analyzed the number of direct transitions between different secondary structural elements. The numbers are shown in Table 2. During the 10 ns simulation with λ of 0.4, there were a total of 253 transitions between 3_{10}-helix segments and turns, 150 transitions between α-helix segments and turns, 17 transitions between π-helix segments and turns, and 121 transitions between 3_{10}-helix and α-helix. There were only 9 transitions between bends and 3_{10}-helix segments. These results thus showed that turns play an important role in helix initiation (both α- and 3_{10}-helices). It is of note that turns were implicated to play a crucial role in helix folding in several previous studies.[14;38;39] Furthermore, our results indicated that 3_{10}-helices play an important role in α-helices initiation, even though a significant number of α-helices initiation are directly from turns.

To more quantitatively assess the relative rates of transitions between different secondary structural elements, we made an assumption that the transition rate may be related to the average population of secondary structures:

$$\upsilon_{SS1\rightarrow SS2} = k_{SS1\rightarrow SS2}[SS1] \quad (1)$$

Here, SS1 and SS2 represent the secondary structures before and after the transition, $\upsilon_{SS1\rightarrow SS2}$ and $k_{SS1\rightarrow SS2}$ are the transition rate and kinetic constants, respectively, and [SS1] represents the population of the secondary structure, SS1. We assumed that the average secondary structure percentages listed in Table 1 as their "equilibrium" population and obtained the following relative kinetic constants:

$$\frac{k_{T\rightarrow G}}{k_{T\rightarrow H}} = 1.77\ (1.53);$$

$$\frac{k_{G\rightarrow H}}{k_{T\rightarrow H}} = 3.28\ (4.56);$$

$$\frac{k_{H\rightarrow T}}{k_{H\rightarrow G}} = 1.20\ (2.57)$$

T, G, and H represent turn, 3_{10}-helix, and α-helix secondary structural elements, respectively. The numbers in parentheses are the results for the SGMD simulation with $\lambda = 0.3$. These data suggest that turns are more like to become 3_{10}-helices than to α-helices, and α-helices can be more easily to be initiated from 3_{10}-helix than from turns. However, α-helix is more likely to unfold to turns than to 3_{10}-helix. It is interesting to note that although the 10 ns SGMD simulation with λ of 0.3 is not long enough to achieve reversible folding, the kinetic constants qualitatively agree with that from the SGMD simulation with λ of 0.4.

TABLE 3: α-Helical Segment Propagation Events Observed during the Two SGMD Simulations

		λ	
posn	direction	0.3	0.4
N-end	fold	36	121
	unfold	28	119
C-end	fold	298	582
	unfold	311	586

Previously, Millhauser proposed a helix folding model on the basis of the experimental data obtained from X-ray crystallography, CD, NMR, and ESR, as well as theoretical and simulations studies.[10] In Millhauser's model, nascent helices were formed first from random coiled conformations. The nascent helix further evolved into 3_{10}-helix, which can subsequently form the α-helix. Our simulation results support this model but also suggest additional possible extensions to the model. The nascent helix is essentially a type of turn conformations according to the DSSP. Thus, our simulation results and Millhauser's model both suggest that the nascent helices (turns) and 3_{10}-helices play important roles in helix folding. But our simulation results indicate that although turns are more readily to fold into 3_{10}-helices, a significant percentage of turns conformations can fold into α-helices without going through 3_{10}-helices as the intermediate. In addition, we observed that α-helices can be directly initiated from π-helices, and π-helices can be initiated directly from turns, although these events were relatively rare.

Comparison of the two SGMD simulations with different guiding parameters showed that with respect to the helix initiation, although the absolute number of transitions between turns, α-helices, 3_{10}-helices, and π-helices differ from each other, the general conclusions from the two simulations are essentially the same. For example, both simulations clearly indicated that turns play a dominant role in the formation of helices (α-, 3_{10}-, and π-helices). Moreover, both simulations showed that turns most likely transfer to 3_{10}-helices and least likely transfer to π-helices among different forms of helices. While α-helices can be formed directly from turns, a significant percentage of α-helices are initiated from 3_{10}-helices. Furthermore, both simulations indicated that helix is the dominant conformational form of this alanine-based peptide in aqueous solution.

Helix propagation is another major issue in helix folding. On the basis of the secondary structures of simulation conformations, we can analyze helix propagation. As can be seen from Figure 10, during the first 400 ps of the SGMD simulation with $\lambda = 0.4$, helix propagation occurred in each of the two major α-helix segments. The first stable α-helix segment formed at 148 ps from residue A10 to residue A13. It propagated to A14 at 150 ps and to A15 at 184 ps. From time to time, we observed fluctuations at both N- and C-ends of the helix segment. Another stable α-helix segment formed at 228 ps from A2 to A5. It propagated to A6 at 284 ps, to A1 at 366 ps, and to A8 at 368 ps. Thus, folding (growth) and unfolding occurred frequently at both ends of the helix segments.

To more quantitatively examine helix propagation, we analyzed the folding and unfolding of helix segments at both of its N- and C-ends in more detail on the basis of the two 10 ns SGMD simulations. The results are summarized in Table 3. Interestingly, the folding and unfolding of helix segments are much more frequent at the C-end than at the N-end of a helix

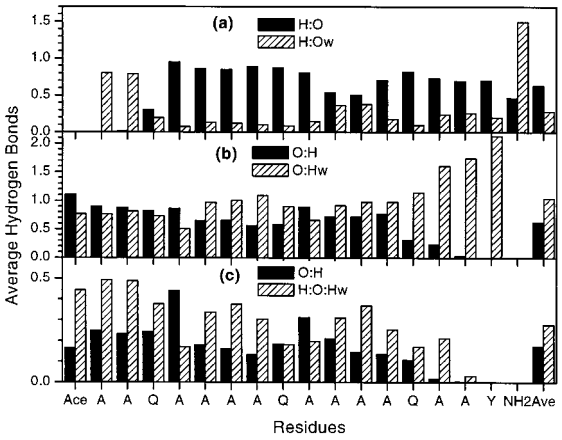


Figure 11. Average number of hydrogen bonds of the peptide found in the 5000 conformations recorded during the 10 ns SGMD simulation with $\lambda = 0.4$. Hydrogen bonds are defined as oxygen−hydrogen pairs with distance less than 2.5 Å. (a) Hydrogen bonds between the backbone amide hydrogens and the backbone carbonyl oxygens (H:O) and between backbone amide hydrogens and water (H:Ow). The C-terminal blocking group NH_2 has two hydrogens and can form two hydrogen bonds with water. (b) Hydrogen bonds between the backbone carbonyl oxygens and the backbone amide hydrogens (O:H) and between backbone carbonyl oxygens and water (O:Hw). (c) Hydrogen bonds for the backbone carbonyl oxygens either interacting with the backbone amide hydrogens alone (O:H) or interacting with both the backbone amide hydrogens and water (H:O:Hw).

segment. This was observed in both SGMD simulations with different guiding parameters. In fact, we found that this difference in propagation frequency at the N- and C-ends may be explained from the different capability of forming hydrogen bonds with water molecules, which is discussed in the next section.

4. Peptide−Water Interaction. In aqueous solution, a peptide can form intermolecular hydrogen bonds with water molecules and intramolecular hydrogen bonds within the peptide chain. The competition between these two types of hydrogen bonds probably plays an important role for peptide and protein folding. In a fully unfolded state, a peptide is in a fully hydrated state where most hydrogen bonding groups of the peptide form hydrogen bonds with water molecules. In folded states, some intermolecular peptide−water hydrogen bonds are now replaced by intrapeptide hydrogen bonds. We analyzed the hydrogen-bonding interaction of each amino acid in the peptide during the 10 ns SGMD simulation with $\lambda = 0.4$, and the results are shown in Figure 11a−c.

As can be seen from Figure 11a, on average the amide hydrogen of each residue can form only one hydrogen bond, primarily with a carbonyl group in the peptide, except residues near the very end of N-terminal region where no backbone carbonyl groups are available for intramolecular hydrogen-bonding formation. In contrast, the carbonyl oxygen can form two hydrogen bonds because it has two electron lone pairs. As can be seen from Figure 11b, the backbone carbonyl oxygens form approximately the same number of hydrogen bonds with the amide groups in the peptide and with water molecules, except those residues near the C-terminus.

From a fully hydrated coil conformation to a complete helix conformation, the amide groups change from intermolecular hydrogen bonding with water molecules to intramolecular hydrogen bonding with the carbonyl groups in the peptide. Thus, there is no change in the total number of hydrogen bonds formed for the amide groups. In a fully unfolded state, each backbone carbonyl group in the peptide can form two optimal hydrogen bonds with water molecules but can only form one intramolecular optimal hydrogen bond in a helical conformation. Thus, it may be essential for each of the peptide carbonyl groups to form an additional hydrogen bond with a water molecule simply for the stability of the helix. We therefore calculated the average hydrogen bonds of the backbone carbonyl oxygen atoms with the amide group alone, or simultaneously with both amide group and water, as shown in Figure 11c. We can see that most of the backbone carbonyl oxygens prefer to simultaneously hydrogen bond with water and with the backbone amide groups. The carbonyl group of only two residues, A4 and A9, prefers to form a single hydrogen bond with the backbone amide groups. The reason for this is that the side chain of residues Q8 and Q13 partially block the access of waters to the backbone carbonyl oxygens of A4 and A9 in a helix conformation. Therefore, helix hydration is important for helix folding of short peptides in water. In fact, helix hydration has been often observed in crystal structures of proteins.[39]

Helix hydration of a short peptide requires its backbone carbonyl groups to have water accessibility in a helical conformation. Obviously, the major obstacle for water to access the backbone carbonyl groups comes from the side chains of the peptide. Branched side chains are more likely to block the water accessibility and, therefore, are not good for helix folding in water. One major reason that many alanine-based peptides can form helix in water is that the side chains of alanine residues are small, which allow the backbone carbonyl groups to remain hydrated even in helical conformations.

In the last 50 years, many studies have been done to investigate the helical propensities of amino acids but some controversies still remain in the field. When the α-helix was proposed by Pauling in 1951,[40] it was believed that the major driving force for helix formation was due to the backbone hydrogen bonding. Later studies, including the empirical studies[41] and the host−guest experimental studies,[42−44] led to the idea that helix-forming propensities of residues with chemically equivalent backbones are due, at least in part, to conformational restrictions imposed upon the side-chain by the helix itself. But a more recent calculation on the side-chain entropy provided opposing evidence to the idea that helix-forming propensities of residues are due to conformational restrictions imposed upon the side chain by the helix itself[45] and suggested that the idea by Pauling[40] is in fact correct. Our simulation results on backbone hydration suggest that the helical propensities of amino acids are related to the water accessibilities of the backbone groups, especially the carbonyl groups, in helical conformation. Therefore, the helical propensities of amino acids indeed originate from the backbone but are modulated by the side chain of the amino acids.

Helix folding is marked by the formation of hydrogen bonds between residues i and $i + 4$ (i and $i + 3$ for 3_{10} helix, or i and $i + 5$ for π helix). Figure 12 shows the backbone hydrogen bonding of the peptide observed during the 10 ns SGMD simulation with $\lambda = 0.4$. Consistent with our DSSP analysis, (i, $i + 4$) hydrogen bonds, which characterize the α-helix, are the dominant in the hydrogen bond distribution, followed by (i, $i + 2$) hydrogen bonds for turns, (i, $i + 3$) hydrogen bonds for 3_{10} helices, and (i, $i + 5$) hydrogen bonds for π helices. Trace amount of other hydrogen bonds are observed, e.g., at (7, 4), (8, 5), and (6, 18).

One important role water molecules may play is to bridge two hydrogen bonding groups within a peptide to induce a more structured conformation. We examined one type of water bridges

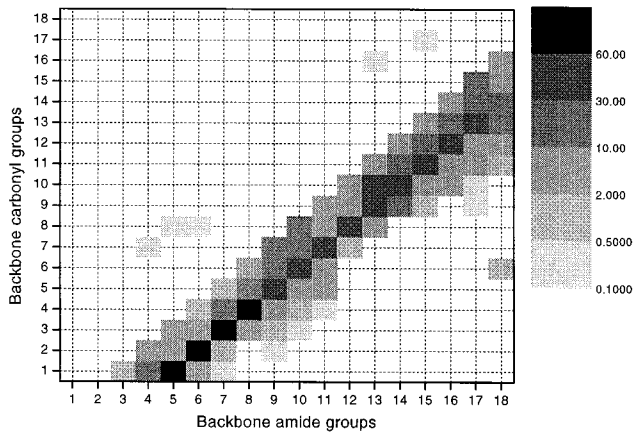


Figure 12. Distribution of hydrogen bonds between the peptide backbone amide and carbonyl groups found in the 5000 conformations recorded during the 10 ns SGMD simulation with $\lambda = 0.4$. Note that the N-terminal acetyl group is counted as residue 1 and the C-terminal amide group is counted as residue 18.

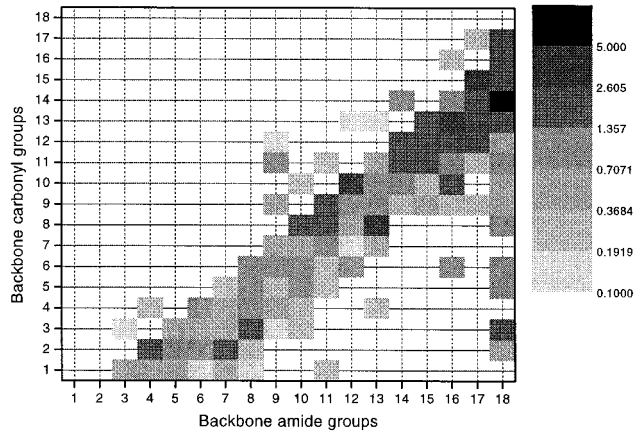


Figure 13. Distribution of water bridges between residues in the peptide found in the 5000 conformations recorded during the 10 ns SGMD simulation with $\lambda = 0.4$. Note that the N-terminal acetyl group is counted as residue 1 and the C-terminal amido group is counted as residue 18.

which hydrogen bond with a backbone amide group and a backbone carbonyl group of the peptide. The distributions of water bridges between different residues of the peptide are shown in Figure 13. Interestingly, water bridges are concentrated on residue pairs (i, $i + n$), ($n = 3$, 4, and 5), which would be hydrogen bonded to each other in helical conformations. This suggests that the water bridges may play a role in helix folding.

Since water bridges may play a role in helix initiation and propagation, we analyzed the water bridges between different secondary structures, as shown in Figure 14. As can be seen, a large portion of water bridges form between helix and coil structures and between helix and turn structures, which may be important for helix propagation. Also, we observed a significant number of water bridges between turns, which suggest a role in helix initiation. Because 3_{10} helices have low population, the number of water bridges involving 3_{10} helices is limited.

Among the water bridges involving helix segments, we found residues at the C-end form much more water bridges than those at the N-end (Figure 15). One reason could be that, at the C-end, residues expose their carbonyl oxygens to water, which can form more hydrogen bonds than the amide hydrogens exposed to water at the N-end. More water bridges at the C-end may also account for the higher activity in helix propagation at the C-end of a helix segment during the simulation (Table 3).

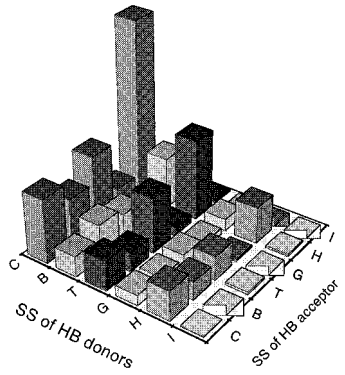


Figure 14. Possibilities of water bridges between different types of secondary structures found in the 5000 conformations recorded during the 10 ns SGMD simulation with $\lambda = 0.4$.

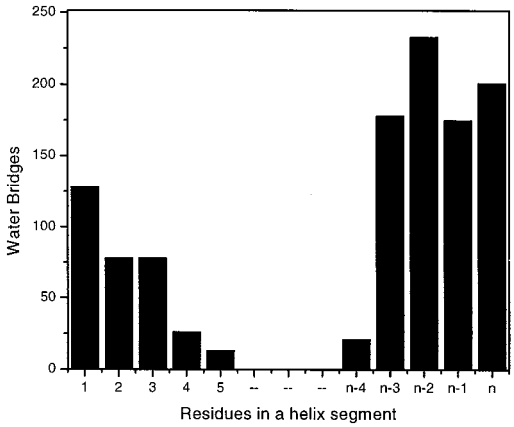


Figure 15. Number of water bridges at each residue in a helix segment found in the 5000 conformations recorded during the 10 ns SGMD simulation with $\lambda = 0.4$.

Conclusions

Reversible folding of a 16-residue alanine-based synthetic peptide in explicit water was successfully accomplished through a SGMD simulation. Analysis of the simulation trajectories showed that this alanine-based peptide adopts multiple major conformational clusters with one or more helix segments in the peptide. The helix is the major secondary structural elements in aqueous solution, and among different helix forms, the α-helix is the dominant one, followed by the 3_{10}-helix. Our simulation results showed that turns play an important role for the helix initiation and are more readily transferred to 3_{10}-helix segments than to α-helices. Although the α-helix can be initiated directly from turns, frequent transitions occurred between 3_{10}-helix and α-helix segments, suggesting that 3_{10}-helices play an important role for the formation of α-helices. Interestingly, our simulation results show that helix propagation is more active at the C-end of a helix segment. Water−peptide hydrogen bonds are widely observed in both helical and nonhelical structures. In helical structures, the carbonyl oxygens prefer to form hydrogen bonds simultaneously with both amide hydrogen and water hydrogen. These hydrogen-bonding interactions may be energetically essential for helix folding in water. To form a helix in water, side chains of a peptide should not block the water access to carbonyl oxygens in helix conformation. Therefore, the helical propensities of amino acids may be related to the water accessibility of their backbone groups in helical conformation. Water bridges are observed frequently during the simulation. It is observed that water bridges are concentrated on residue pairs (i, $i + n$), ($n = 3$, 4, and 5), which would be hydrogen bonded

to each other in helical conformations, and suggests that the water bridges may play a role in helix folding. Our observation that many water bridges link helical and nonhelical residues indicates the importance of water bridges in helix propagation. More water bridges are formed in the C-end of a helix segment may be the reason the C-end propagation is more active. The successful reversible helix folding of this alanine-based peptide paves the way for the study of the folding of other peptides and small proteins which adopt either defined secondary structures, such as turns, helices, and β-hairpins, or tertiary structures.

Acknowledgment. The financial support from the National Institutes of Health to this work is highly appreciated (Grant No. GM59188).

References and Notes

(1) Scholtz, J. M.; Qian, H.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers* **1991**, *31*, 1463−1470.
(2) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95−118.
(3) Baldwin, R. L. *Biophys. Chem.* **1995**, *55*, 127−135.
(4) Bousquet, J. A.; Garbay, C.; Roques, B. P.; Mely, Y. *Biochemistry* **2000**, *39*, 7722−7735.
(5) Odaert, B.; Baleux, F.; Huynh-Dinh, T.; Neumann, J. M.; Sanson, A. *Biochemistry* **1995**, *34*, 12820−12829.
(6) Sasidhar, Y. U.; Ramakrishna, V. *Indian J. Biochem. Biophys.* **2000**, *37*, 34−44.
(7) Bolin, K. A.; Hanson, P.; Wright, S. J.; Millhauser, G. L. *J. Magn Reson.* **1998**, *131*, 248−253.
(8) Fiori, W. R.; Millhauser, G. L. *Biochemistry* **1993**, *32*, 11957−11962.
(9) Miick, S. M.; Marttinez, G. V.; Fiori, W. R.; Todd, A. P.; Millhauser, G. L. *Nature* **1992**, *359*, 653−655.
(10) Millhauser, G. L. *Biochemistry* **1995**, *34*, 3873−3877.
(11) Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. *Biochemistry* **1996**, *35*, 691−697.
(12) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. *Biochemistry* **1997**, *36*, 9200−9210.
(13) Riley, M. L.; Wallace, B. A.; Flitsch, S. L.; Booth, P. J. *Biochemistry* **1997**, *36*, 192−196.
(14) Daggett, V.; Levitt, M. *J. Mol. Biol.* **1992**, *223*, 1121−1138.
(15) Tirado-Rives, J.; Jorgensen, W. L. *Biochemistry* **1993**, *32*, 4175−4184.
(16) Smythe, M. L.; Huston, S. E.; Marshall, G. R. *J. Am. Chem. Soc.* **1993**, *115*, 11594−11595.
(17) DiCapua, F. M.; Swaminathan, S.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 6145−6155.
(18) Sung, S.; Wu, X. W. *Biopolymers* **1997**, *42*, 633−644.
(19) Sung, S.; Wu, X. W. *Proteins* **1996**, *25*, 202−214.
(20) Hirst, J. D.; Brooks, C. L., III. *Biochemistry* **1995**, *34*, 7614−7621.
(21) Young, W. S.; Brooks III, C. L. *J. Mol. Biol.* **1996**, *259*, 558−572.
(22) Daura, X.; Jaun, B.; Seebach, D.; van Gunstern, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925−932.
(23) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740−744.
(24) Wu, X.-W.; Wang, S. *J. Phys. Chem. B* **1998**, *102*, 7238−7250.
(25) Wu, X.-W.; Wang, S. *J. Phys. Chem. B* **2000**, *104*, 8023−8034.
(26) Wu, X.-W.; Wang, S. *J. Chem. Phys.* **1999**, *110*, 9401−9410.
(27) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(28) MacKerell, A. D. Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fisher, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(29) van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen molecular simulation (GROMOS) library manual (Biomos)*; Nijenborgh 4, 9747 AG Gronigen, The Netherlands, 1987.
(30) Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *J. Am. Chem. Soc.* **1991**, *113*, 5102−5104.
(31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(32) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(33) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(34) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S. Jr.; Weiner, P. A. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.
(35) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230−252.
(36) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.
(37) Tirado-Rives, J.; Maxwell, D. S.; Jorgesen, W. L. *J. Am. Chem. Soc.* **1993**, *115*, 11590−11593.
(38) Tobias, D. J.; Brooks, C. L., III. *Biochemistry* **1991**, *30*, 6059−6070.
(39) Sundaralingam, M.; Sekharudu, Y. C. *Science* **1989**, *244*, 1333−1337.
(40) Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acd. Sci. USA* **1951**, *37*, 205−210.
(41) Chou, P. Y.; Fasman, G. D. *Adv. Enzymol.* **1978**, *47*, 45−148.
(42) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. *Nature* **1990**, *344*, 268−270.
(43) Lyu, P. C.; Liff, M. I.; Marky, L. A.; Kallenbach, N. R. *Science* **1990**, *250*, 669−673.
(44) O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250*, 646−651.
(45) Creamer, T. P.; Rose, G. D. *Proc. Natl. Acd. Sci. USA* **1992**, *89*, 5937−5941.