Shibasish Chowdhury
Wei Zhang
Chun Wu
Guoming Xiong
Yong Duan
Department of Chemistry and Biochemistry, Center of Biomedical Research Excellence, University of Delaware, Newark, DE 19716

Breaking Non-Native Hydrophobic Clusters is the Rate-Limiting Step in the Folding of an Alanine-Based Peptide

Abstract: *The formation mechanism of an alanine-based peptide has been studied by all-atom molecular dynamics simulations with a recently developed all-atom point-charge force field and the Generalize Born continuum solvent model at an effective salt concentration of 0.2M. Thirty-two simulations were conducted. Each simulation was performed for 100 ns. A surprisingly complex folding process was observed. The development of the helical content can be divided into three phases with time constants of 0.06–0.08, 1.4–2.3, and 12–13 ns, respectively. Helices initiate extreme rapidly in the first phase similar to that estimated from explicit solvent simulations. Hydrophobic collapse also takes place in this phase. A folding intermediate state develops in the second phase and is unfolded to allow the peptide to reach the transition state in the third phase. The folding intermediate states are characterized by the two-turn short helices and the transition states are helix–turn–helix motifs—both of which are stabilized by hydrophobic clusters. The equilibrium helical content, calculated by both the main-chain Φ–Ψ torsion angles and the main-chain hydrogen bonds, is 64–66%, which is in remarkable agreement with experiments. After corrected for the solvent viscosity effect, an extrapolated folding time of 16–20 ns is obtained that is in qualitative agreement with experiments. Contrary to the prevailing opinion, neither initiation nor growth of the helix is the rate-limiting step. Instead, the rate-limiting step for this peptide is breaking the non-native hydrophobic clusters in order to reach the transition state. The implication to the folding mechanisms of proteins is also discussed.* © 2002 Wiley Periodicals, Inc. Biopolymers 68: 63–75, 2003

Keywords: *hydrophobic clusters; rate limiting step; alanine-based peptide; formation mechanism; all-atom molecular dynamics; point-charge force field; generalized Born continuum solvent model; AMBER; peptide folding; protein folding*

INTRODUCTION

Alanine peptides are among the best-studied peptide fragments. Their simple structures and folding kinetics have served as the model systems to study formation of helical secondary structures.[1] While valuable information has been accumulated from both experimental[2] and theoretical[3,4] studies for over the past half century, interestingly, despite their relatively simple structures and experimentally observed simple kinetics, a detailed description of their folding processes is still lacking. The present prevailing theory attributes helix formation to the formation of main-chain hydrogen bonds in a simple process and neglects the details of side chains. Simulations[5,6] on two small proteins suggested that the time scale of helix formation is about 60 ns, in qualitative agreement with a number of fast kinetic measurements.[2,7,8] This has been challenged by recent experimental studies by Clarke et al.,[9] who observed a ms scale process and suggested that the helix initiation rate is about 5 orders of magnitude slower than the growth rate. At the other extreme, Hummer et al. suggested that helix can initiate within 0.1 ns, based on their studies of short peptides[10] including Ala$_5$.

We have investigated the folding of AK16 [Ac-YG(AAKAA)$_2$AAKA-NH$_2^9$] by all-atom molecular dynamics simulations with Generalized Born solvent model.[11] Starting from a straight chain conformation and with different random velocities, a total of 32 simulations were conducted for 100 ns for an aggregate of 3.2 μs simulations. In this article, we present the results from these simulations.

METHOD

Force Field

AMBER simulation package was used in both simulation and data processing. A recently developed all-atom point-charge force field[12] was used to represent the peptide. The detailed description on the force field will be presented elsewhere.[12] Here we summarize its main features. The new force field is an all-atom point-charge minimalist model developed based on the existing AMBER force field[13] with a new set of charges and main-chain torsion parameters of peptides. The new charge set has been derived based on quantum mechanical calculations using the DFT method, the B3LYP functionals, and the cc-pVTZ basis set that was done in organic solvent with continuum solvent model (ϵ = 4) to mimic the electrostatic environment of proteins. In comparison, the earlier model was developed based on the HF/6-31G* level of theory in the gas phase. The peptide main-chain torsion parameters were obtained by fitting the energy profile of the alanine dipeptide against that calculated quantum mechanically using MP2/cc-pVTZ in organic solvent (ϵ = 4). In addition to the studies discussed here, we also have simulated folding of four β-hairpins and one three-stranded β-sheet. They all successfully folded into their native states. These results strongly suggest that the new force field maintains a reasonable balance between the α-helical and β-sheet conformations. The detail of these simulations will be presented elsewhere.[14]

Simulation

The solvent was represented by a generalized Born solvent model[11] with an effective salt concentration of 0.2*M*. Starting from a straight chain conformation, after initial energy minimization, random velocities were assigned according to Boltzmann's distribution at a temperature of 300 K. A total of 32 simulations were conducted with different random number seeds used to generate the initial random velocities. SHAKE[15] was applied to constrain all bonds connecting hydrogen atoms and a time step of 2.0 fs was used. Born radii were calculated every 5 steps (10 fs) using the method of Bashford and Case.[16] Nonbonded forces were calculated using a two-stage RESPA approach where the forces within a 10 Å radials were updated every time step and those beyond 10 Å were updated every two steps. Temperature was controlled at 300 K using Berendsen's algorithm[17] with a time constant of 2.0 ps. Each simulation was conducted to 100 ns for an aggregate simulation time of 3.2 μs. The trajectories were saved at 10.0 ps intervals and a total of 320,000 snapshots were produced for further analysis.

Clustering Analyses

Simulations of protein folding are typically long time simulations and can produce on the order of 10^5 and to 10^6 sets of coordinates. The large number of snapshots makes it impractical to apply directly those clustering methods that are based on the pairwise comparisons. A semilinear clustering technique was developed earlier and was successfully applied in the analysis of a large pool of structures generated from protein folding simulations.[6] Here we develop the approach further and combine it with a pairwise method of Daura et al.[18]

Our method is a hierarchical approach. In this approach, based on the main-chain root mean square distance (RMSD), each snapshot is compared against the average coordinates of the existing groups after rigid-body alignment.[19] A snapshot may become a member of its closest cluster if the RMSD is smaller than a given cutoff (1.5 Å). Otherwise, a new cluster may form if the minimum RMSD exceeds the cutoff. The clusters were further filtered by removing those structures whose RMSD from the average coordinates of the clusters exceeds the given cutoff. The removed snapshots were then compared to the existing clusters. A total of 3077 clusters were generated from the total of 320,000 snapshots taken from the simulations. A more detailed clustering analysis based on the pairwise
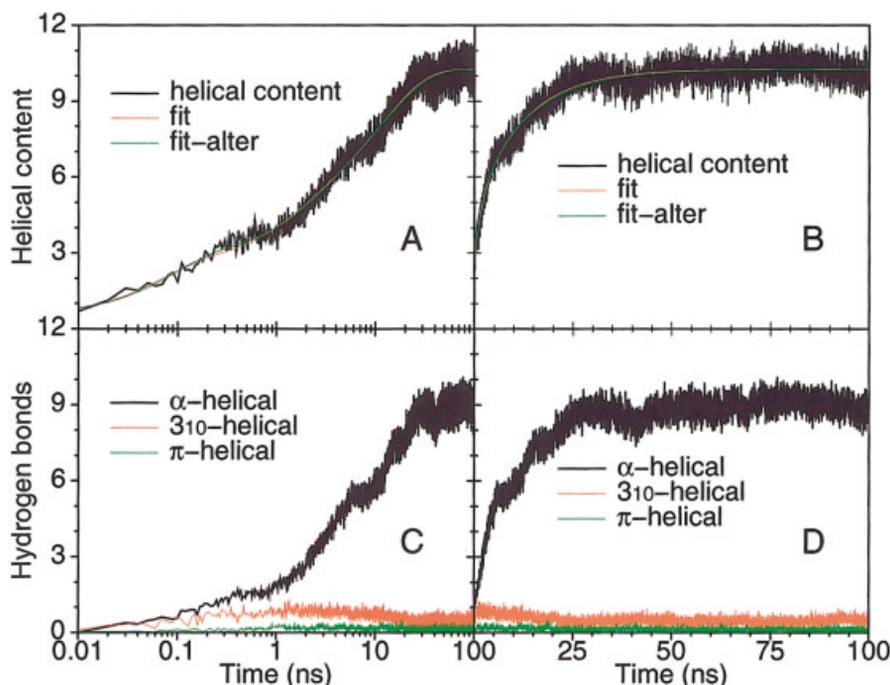
FIGURE 1 Average helical content as measured by main-chain $\Phi - \Psi$ torsion angle (A, B) and main-chain hydrogen bonds (C, D) plotted in logarithmic (A, C) and linear (B, D) time scales. The fitted curves are also shown in A and B and the fitting parameters are given in Table I.

RMSD follows by comparing the representative structures of the cluster using the method of Daura et al.[18] The total number of clusters was reduced to 2259.

RESULTS

Because of its direct analogy to experiments and because it is an effective order parameter, the helical content has been used extensively to gauge the progress of helical peptide folding. However, we shall not limit the scope to the helical content, for such limitation would potentially over simplify the rich kinetics observed in the process. Our analysis is focused on two aspects: main-chain helical content and the diversity of the structures. The former will be measured both by $\Phi-\Psi$ torsion angles and by main-chain helical hydrogen bonds. The later is investigated by analyses of structurally similar clusters.

We defined the helical residue with Φ and Ψ values within an ellipsoid of radii (30°, 25°) centered at the ideal values ($\Phi = -57°$, $\Psi = -47°$). The helical content (number of residues in the helical region) averaged over all 32 simulations is shown in Figure 1. One can clearly identify the three phases of transition that can be fitted by a combination of three exponential functions. The fitting results are summarized in Table I and are plotted in Figure 1A. Two fits

were done. One is the least-square fit and the other is a weighted least square where the weighting factor for the data at time t_i is $w_i = \log(t_{i+1}/t_i)$. Since the least-square fit has a tendency to weigh the long-time trend more favorably, the weighting factor, which weighs the short-time trend more favorably, was used to obtain the short-time trend. Therefore, these two fits allow us to estimate the error margin. Both fits are reasonable with a relative root-mean-square error of 3.7%. The fitted equilibrium helical content is 10.26–10.29, or 64.1–64.4%. These are in remarkable agree-

Table I Summary of Fitting Results by $h[1 - r \exp(-t/\tau)]$

	Least-Square Fit		Weighted Least-Square Fit	
	r	τ (ns)	r	τ (ns)
First phase	0.200	0.062	0.223	0.084
Second phase	0.201	1.41	0.216	2.27
Third phase	0.554	12.05	0.508	13.16
T		19.5		15.6
	$\varepsilon = 3.7\%$,		$\varepsilon = 3.7\%$,	
	$h = 10.26$		$h = 10.29$	

[a] The RMS error is ε.

Table II   Structural Features at the End of Phases I and II (1.0 and 10.0 ns)

	1.0 ns	10.0 ns
Random coils	19	6
One turn	7	0
Two-turn/coil/loop	4	4
Three-turn/coil/loop	1	6
Helix–turn–helix	1	4
Helices	0	12

ment with the experimentally measured helical content by Baldwin and co-workers[20] ranging from 63 to 72% at 0.05–2.5M NaCl and 64% at 0.1M NaCl .

Among the three rates obtained from the simulations, the first two are respectively 0.06–0.08 and 1.4–2.3 ns, depending on the fitting procedure. Both are faster than the typical dead time of experimental apparatus[10] that is on the order of 10 ns. Therefore, the experimentally observed process corresponds the slowest phase observed in our simulations whose rate is the experimental folding rate of the helix. However, since a continuum solvent model is used in our simulations that neglects the solvent viscosity effect, the calculated folding rate is expected to be faster than that obtained when explicit solvent is used. Hence, the extrapolated folding rates are obtained from the rates of the slowest phase multiplied by the ratio between the initiation rates calculated from the explicit solvent simulation[10] and from our simulations. The 15.6 and 19.5 ns extrapolated folding rates agree well with experimental results.[2]

The first phase is an extremely rapid helix initiation phase with a time constant of 0.06–0.08 ns. In this phase, an average of 3.7 residues reach the helical region. One turn of helix and 1.5 main-chain hydrogen bonds form by the end of the phase, which lasts to about 1.0 ns. The growth rate of the helical content is 0.06–0.08 ns. In comparison, earlier molecular dynamics studies on pentapeptides suggested an initia-

tion rate of 0.1 ns.[10] Since our model, in the absence of explicit water, accelerates the kinetics by not counting the solvent viscosity effect, the rate observed in our simulations should be faster than those in explicit water simulations. Thus, we can calibrate the effect by comparing these two types of simulations.

The structural features of the snapshots at 1.0 ns are summarized in Table II. Majority (19 out of 32) of the structures are random coils and are compact. The rest of the structures can be classified into four main categories, including single-turn helix (7 structures), a combination of two-turn helix and coil (4 structures), a three-turn helix with a loop (1 structure), and one helix–turn–helix conformations. The hydrophobic collapse appears to be completed by 1.0 ns, judged by the compactness of the structures. Thus, the helix initiation is concomitant to the hydrophobic collapse, in agreement with earlier simulation results.[6] However, since on average only one turn of helix forms during this phase, our results indicate that the hydrophobic collapse is not the primary driving force for the helix formation, which is not surprising.

The initiation phase is further studied by the average helical content per residue as measured by their Φ–Ψ torsion angles and is plotted in a colored two-dimensional (2D) map (Figure 2) and summarized in Table III. First of all, the helix was initiated preferentially from the C-terminal part of the peptide. But the degree of preference is marginal. The average helical content of residues Ala_{12}–Lys_{15} is 0.28, which is compared to 0.21 for Ala_3–Ala_{10}. Among the residues, Gly_2 has the lowest average helical content (0.03) in the initiation phase. In fact, its helical content was the lowest throughout the entire simulations. This is not surprising since Gly has been identified as an effective helix stopper.

The intermediate phase has a time constant of 1.4–2.2 ns that lasted to approximately 10 ns. By the end of this phase, an average of 7.5 residues are in the α-helical region, or an additional 3.8 residues from 1.0 to 10 ns, and the total number of hydrogen bonds

■ 0-0.05 ■ 0.05-0.1 ■ 0.1-0.15 ■ 0.15-0.2 ■ 0.2-0.25 ■ 0.25-0.3 ■ 0.3-0.35 ■ 0.35-0.4 ■ 0.4-0.45 ■ 0.45-0.5
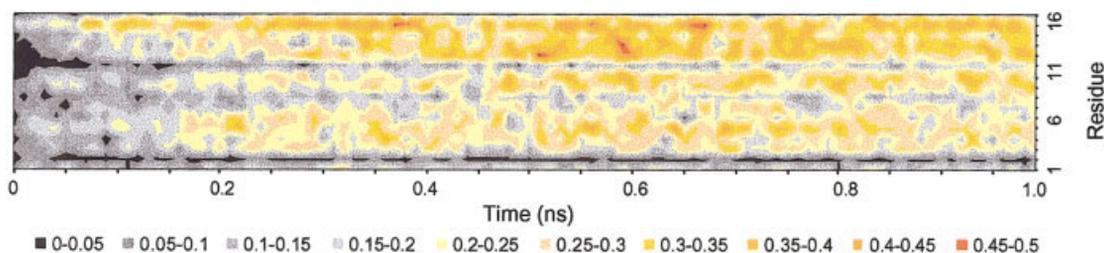
FIGURE 2   Average helical content per residue from 0 to 1 ns. The horizontal axis is time (0–1 ns, left to right) and vertical axis is the residue ID (1–16). The color code is given below the figure.

Table III   Time-Averaged Per-Residue Helical Content

	0–1 ns	5–10 ns	50–100 ns
Tyr_1	0.17	0.25	0.37
Gly_2	0.03	0.06	0.27
Ala_3	0.21	0.39	0.61
Ala_4	0.21	0.43	0.71
Lys_5	0.26	0.45	0.73
Ala_6	0.21	0.44	0.74
Ala_7	0.19	0.52	0.76
Ala_8	0.14	0.50	0.76
Ala_9	0.24	0.63	0.80
Lys_{10}	0.24	0.60	0.80
Ala_{11}	0.12	0.49	0.77
Ala_{12}	0.27	0.57	0.77
Ala_{13}	0.27	0.59	0.73
Ala_{14}	0.24	0.49	0.61
Lys_{15}	0.34	0.42	0.47
Ala_{16}	0.16	0.32	0.38

grows from 1.5 at 1.0 ns to 5.5 at 10 ns. Thus on average the helix grows by a full turn within this period. A substantial number of helical structures forms. In fact, 12 out of 32 trajectories folded to highly native-like structures. Among them, 4 are the full helices and the other 8 are also close to the full helix with only 1–2 residues frayed at the N-terminus and are in the general energy basin of the native full-helix conformation. In addition, there are also 6 trajectories that have three or more turns of helix. Judged by the average helical content, the folding reaches about 73% completion by the end of this phase.

It is noteworthy that there is no single-turn helix species at 10 ns, even though there are 7 trajectories that have single-turn helix conformations at 1.0 ns. Among them, 5 reached the full-helix conformation and the other two form at least three turns of α-helix with disordered N-terminus in loop conformation. This is rather interesting since only 2 of the 4 trajectories that have two turns of α-helix at 1.0 ns reach these states at 10 ns. Among the 12 fully folded trajectories at 10 ns, 4 are basically random coil structures at 1.0 ns without significant presence of helix and 5 are single-turn species at 1.0 ns. Each of the remaining 3 comes from a two-turn helix, a helix–turn–helix, and the helix of three turns. Thus the helical content that immediately follows the helix initiation/hydrophobic collapse phase is not a good indicator of the tendency to reach the full-helix conformation. The absence of single-turn helices at 10 ns also suggests that such a conformation is transient.

Visual inspection on the trajectories that folds at 10 ns further indicates that all of the 12 folded trajectories go through species with helix–turn–helix motif before they reach the full-helix conformation. The lifetime of the helix–turn–helix is rather short, on the order of 1–2 ns, about an order of magnitude shorter than the time constant of the third phase observed in our simulations. Moreover, this suggests that the helix initiation may not trigger the folding.

Three Helical Species

An alternative way to characterize the main-chain conformation is the hydrogen bonding pattern. The most prominent helical form is the α-helix, characterized by both the main-chain torsion angles ($\Phi = -57°$, $\Psi = -47°$) and hydrogen bonds between the carbonyl group of residue i and the amide group of residue $i + 4$. Other helical conformations include the less commonly observed 3_{10}-helix and the rarely observed π-helix, characterized by the $i{:}i + 3$ and the $i{:}i + 5$ hydrogen bonds, respectively.

The main-chain hydrogen bonds are defined when the O:H distance is within 2.8 Å and the C=O:H angle is between 120° and 180°. To assess the robustness of the definition, we also calculated the hydrogen bonds by varying the O:H distance cutoff from 2.5 to 3.0 Å. The overall behavior and the ratio between different species (i.e., 3_{10}, α, and π) remain essentially the same even though, the total number of hydrogen bonds varies somewhat due to the change of the O:H distance cutoff as expected. We will focus our following discussions on the hydrogen bonds calculated using the 2.8 Å cutoff, which is shown in Figure 1 (C and D).

The majority (93.9%) of the main-chain hydrogen bonds is α-helical (residue i to $i + 4$), which shows the same overall trend as the helical content measured by the Φ–Ψ torsion angles. When averaged over the last 50 ns of the simulations (equilibrium phase), the total number of α-helical hydrogen bonds is 9.0. In comparison, the number of hydrogen bonds in 3_{10}- and π-helix conformations are 0.46 (4.8%) and 0.13 (1.3%), respectively, lower than those found in other simulations[21,22] (to be discussed later). The combined total number of non-α-helical hydrogen bonds is 0.59. Among which, 0.36 (or 61%) are bifurcated. The overwhelming majority (99.5%) of the bifurcation is from the carbonyl group (i.e., one carbonyl group forms hydrogen bonds with two amide groups). Without the bifurcated bonds, only 0.23 (or 2.4%) hydrogen bonds are truly non-α-helical. These strongly suggest that both the 3_{10}-helical and π-helical hydrogen bonds are transient and are formed primarily due
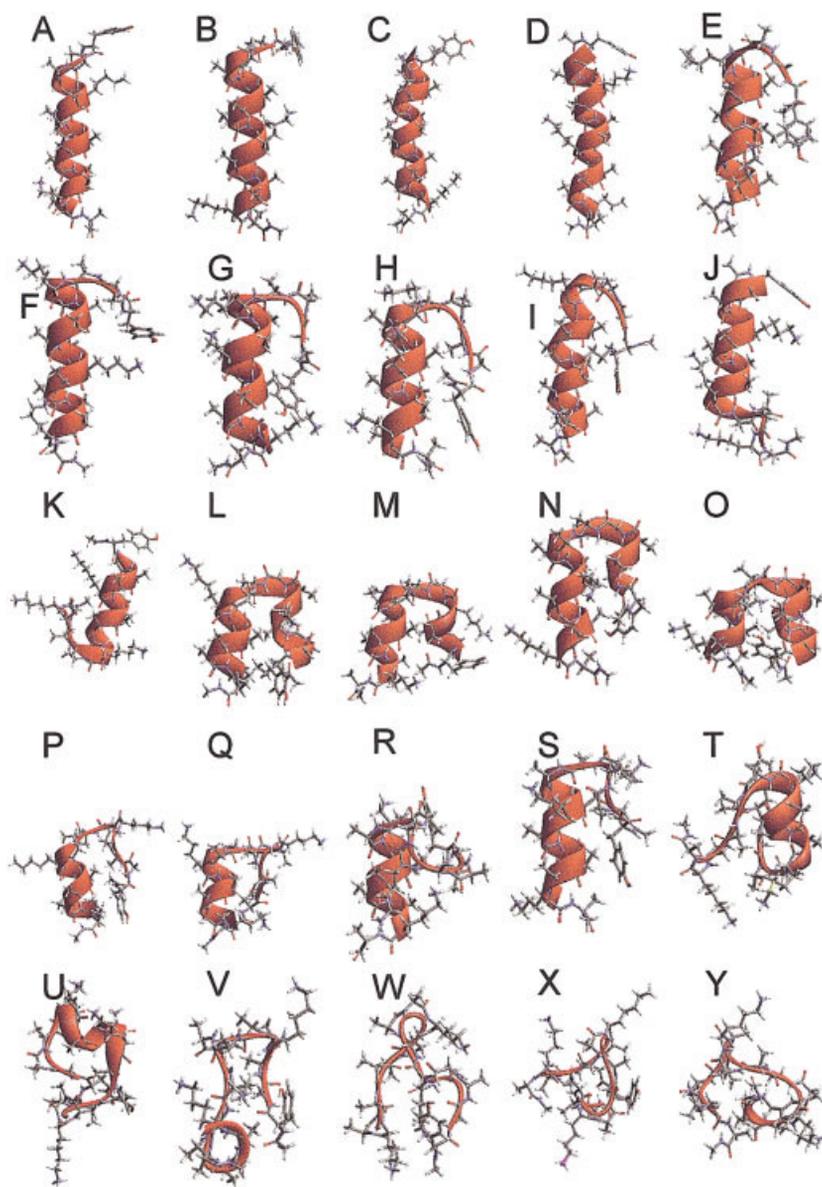
FIGURE 3 Representative structures of the most populated clusters. They are classified into six categories, namely, full helix (A–D), three-turn helix with a loop or coil (E–K), helix–turn–helix (L–O), two-turn helix with loop/coil (P–U), single-turn helix (V), and random coil (W–Y).

to thermal fluctuation. Thus, neither 3_{10}-helix nor π-helix appears to be the transition state of the α-helix folding as suggested by earlier simulations on short helices. This can be further demonstrated by their development over the course of the simulations. Indeed, at around 0.2–0.3 ns, the number of 3_{10} hydrogen bonds is at its highest (0.74) and the ratio between 3_{10}- and α-helical species reaches maximum (about 1:2). This is the time when short helices start to form (initiation) and the most simulations are still in the random-coil state. Thus, a high degree of fluctuation is expected. After that, the number of 3_{10} hydrogen

bonds decreases continuously and reaches an average of 0.46 in the last 50 ns. Nevertheless, the number of 3_{10} hydrogen bonds is always less than the α hydrogen bonds, and there is no significant accumulation of 3_{10} conformations. Thus, we conclude that 3_{10}-helix is not a transition state of α-helix folding.

Since there were 14 sites that can form α-helical hydrogen bonds, the average 9.0 hydrogen bonds in the α-helical region would correspond 64% helical content or 66% when measured by the total number of hydrogen bonds in the helical region (sum of all three species subtracting the bifurcated bonds). These are in

Table IV **Summary of Clusters at the Time Intervals 0–10, 10–20, and 0–100 ns**

	0–10 ns		10–20 ns		0–100 ns	
	Clusters	Frames	Clusters	Frames	Clusters	Frames
Full helix	1	4556 (14%)	2	15,116 (47%)	4	191,228 (60%)
Three-turn helix	5	3364 (11%)	4	4075 (13%)	7	47,442 (15%)
Helix–turn–helix	5	2559 (8%)	3	1471 (5%)	4	20,344 (6%)
Two-turn helix	2	965 (3%)	7	4566 (14%)	6	8868 (3%)
Single-turn helix	7	3648 (11%)	1	355 (1%)	1	5856 (2%)
Random	8	3608 (11%)	4	3510 (11%)	3	4422 (1%)
''uncounted''		13,300 (42%)		2907 (9%)		43657 (14%)

good agreement with that measured by the Φ–Ψ torsion angles and are also in agreement with the experiments of Padmanabhan et al.,[20] which is 64% at 0.1M salt concentration.

Clustering Analysis

A hierarchical approach is applied to group the closely related structures into a smaller number of clusters as described in the Method section. All structures were superimposed using main-chain least-square superimposition. A 1.5 Å cutoff was used for the RMSD. A total of 2259 clusters were identified by this clustering technique. However, since most clusters are poorly populated and do not represent the main features of the folding processes, we choose to focus our discussions on the most populated clusters. Representative structure from each of the 25 most populated clusters is shown in Figure 3A–Y. Finally, by visual inspection, it is possible to classify these most populated clusters into 6 categories according to the features of their main-chain structures. They are summarized in Table IV. The first cluster contains the full helix structure (Figure 3A). This is by far the most populated cluster and about 53% of all saved structures fall into this cluster. There are three additional clusters that are populated by the full helix (Figure 3C and 3D). Thus, an aggregate total of 60% of all saved structures is in the full helix conformation, whereas 47,442 (15%) of saved snapshots fall in seven clusters that contain the helix with frayed ends (Figure 3E–K). Among them, only two have the N-terminal helix (J, K) with a combined population of 2175 (0.7%) snapshots and the remaining five clusters have the C-terminal helix (Figure 3E–I, 45267 snapshots). Thus, the peptide has a high-degree of preference to form a partial helix at the C-terminus with a ratio of about 20:1. The third category contains the helix–turn–helix motifs. Four clusters and a combined total of 20,344

(6.4%) snapshots are in this category. Other categories that contain partial helix include six clusters of two-turn helix (8868 snapshots, 2.8%) and a cluster containing single-turn helix (5856 snapshots, 1.8%). Three remaining clusters (4422, 1.4%) do not show significant structural feature and are classified as the random coil. There are also 43,657 (13.6%) snapshots that fall into less populated clusters and are not accounted. These clusters are typically featureless and are mostly random coils. Thus, a total of 48,079 snapshots (15%) are in the random coil states.

To investigate the changes in population among different types of structures at different stages of folding, structures saved from 0 to 10 ns and from 10 to 20 ns are clustered separately. The results are summarized in Table IV. Not surprisingly, about 53% the structures in the time interval of 0–10 ns are random coils which is reduced to about 20% in 10–20 ns, comparable to the percentage of the random coil structures in the entire simulations. A significant increase from the first 10 ns to the second 10 ns is the number of snapshots in the full helix category. There are 15,116 (47%) snapshots that have the full helix structure in the second 10 ns period, which is also comparable to the percentage full helix structure in the entire simulations. In comparison, only 4556 (14%) are the full helix in the period of 0–10 ns. Interestingly, the number of snapshots both in the three-turn helix and in the helix–turn–helix categories remained essentially constant in both 0–10 and 10–20 ns. Their percentage populations also remain relatively constant in the 0–100 ns clustering analysis. These suggest that they may be linked to the folding transition (to be discussed later).

On the other hand, there is substantial decrease in the single-turn helix structures (3648, 11.4%, to 355, 1.1%) from the first to the second 10 ns periods. In comparison, only 1.8% of the snapshots are single-turn helix when the complete trajectory sets are clus-

tered. Thus, the single-turn helix is a transient structure. It exists with substantial presence only at the very early stages of folding.

In contrast, the number of 2-turn helix structures increases substantially from 965 (3%) to 4566 (14.3%) for the same two 10 ns periods. However, it decreases to 2.8% in the clusters of the complete trajectories. This suggests that the 2-turn helix species resemble the folding intermediates. This is reinforced by the fact that the population increases at a time scale close to the folding time scale of the peptide.

DISCUSSION

The formation mechanisms of α-helices have been studied for quite some time and substantial amount of knowledge has been accumulated. The recent resurgence of interests in the simple helical peptides has been fueled largely by the advancement in two important and complementary areas. One is the advent of the ultrafast experiments that are capable of detecting folding events at the nanosecond time scale. The other development is the significantly improved computer simulation methods that allow detailed all-atom simulations, such as the simulations presented in this article, to reach tens of nanoseconds or longer time scales. Thus, detailed and direct comparisons can be made. Therefore, in this section, we focus on the comparison with experiments and we try to explain some of the experimental observations.

Folding Transition States

3_{10}-Helices have been repeatedly observed as the transition states in helix formation in earlier simulations. Some even observed that 3_{10}-helices could have relatively long lifetimes. This was not observed in our simulations. We observed that about 4.8% of the helical hydrogen bonds were in the 3_{10}-helix conformation. This is lower than that obtained from the analysis of protein structures that indicates that 12% of the helices found in proteins are 3_{10}-helices.[23] We think the different environment may be responsible for the difference. This is because 3_{10}-helices form more readily in low-dielectric environment and our simulations on the peptides were conducted in high-dielectric environment. Furthermore, since 61% of these non-α-helical hydrogen bonds are bifurcated, we suggest that the 3_{10}-helix hydrogen bonds are due to thermal fluctuation, in agreement with Ferrara et al.[21] This is also supported by the fact that their presence is appreciable only at the beginning of the simulations when the peptide is highly labile.

Instead in almost all of our simulations we observed that the formation of the native state was proceeded by a helix–turn–helix motif. This motif has a typical lifetime of a nanosecond, much shorter than the observed folding time. Thus, we may attribute it as the folding transition state. This is supported by the clustering analysis results.

There are two types of species that have essentially constant abundance throughout the simulations, namely the three-turn helix and the helix–turn–helix motif, suggesting their role as the transition states. From a kinetic perspective, a characteristic pattern of a transition state is that (1) it should have relatively short lifetime, (2) it should show some accumulation at the beginning of folding due to the time needed to reach the exit, (3) the relative abundance should decrease at the intermediate period comparable to the folding time scale, and (4) their relative abundance should then increase when the peptide reaches equilibrium due to reverse transition. The helix–turn–helix motif is the only one that fits the profile. Its lifetime is an order of magnitude shorter than the folding time. Its abundance increases to an average of 8% in the first 10 ns, decreases to 5% in the second 10 ns, and reaches an average of 6.4% in the last 80 ns.

Since this transition state has a well-defined structure, we anticipate that the chain entropy contribution would not be important. This, however, may not be generalized to proteins in which chain entropy can play a significant role. Such a role has been clearly demonstrated by the observation that protein folding rates are linked to their contact orders.[24] Yet the enthalpic component cannot be ignored in general. Both terms have to be taken into account for a complete description of the transition state since free energy barriers comprise both enthalpic and entropic components. Thus, our observation complements the contact order theory.

Inspection of the folding trajectories indicates that an alternative pathway exists. In this pathway, the three-turn helix species form first that lead to the formation of an asymmetric helix–turn–helix with the single-turn helix at the N-terminus. Analysis of transitions between these well-populated species indicates (data not shown) that there are substantial (on the order of 10^2) direct transitions between the three-turn and the full helix conformations. In comparison, transitions between other species take place primarily by going through other species of lower population. This suggests that the free energy barrier separating these two conformations is rather low. Thus, we conclude that the three-turn species are in the broad energy basin of the full helix.

Intermediate State in Helix Folding

The very notion that helix folding may involve intermediate states can spark lively debate among the researchers in the field. However, we believe our simulations demonstrate the existence of such states.

An important characteristics of folding intermediate state is that it should have a lifetime comparable to the folding time. Kinetically, this leads to the increased accumulation at the intervening time period and substantially reduced population at the equilibrium. Since the simulated time of helix formation is 12–13 ns, an increased accumulation in the second 10 ns period would be expected. Indeed, the two-turn helix species showed just that. There are only 3% snapshots that are two-turn helices in the first 10 ns period (0–10 ns). The accumulation increased substantially to 14% in the time interval of 10–20 ns. However, its population reduces to 3337 (or about 1.3%) in 20–100 ns. We thus conclude that the two-turn species are folding intermediates. Moreover, the increase in the helical content in the second phase is largely attributable to the formation of these species. Thus, the characteristic time to reach this state is 1.4–2.3 ns.

Rate-Limiting Step in Helix Formation

An important feature of α-helices is their highly ordered main-chain conformation in which the main-chain carbonyl group of residue i forms hydrogen bond with the amide group of residue $i + 4$. This has inspired many to suggest that the main-chain hydrogen bonds could play key role in the helix formation. This can be exemplified by the prevailing opinion that recognizes helix initiation as the rate-limiting step at which the peptide forms the first turn and the first main-chain hydrogen bond. This is consistent with the classical helix–coil transition theory, which argues that the first turn of helix would be more difficult to form in comparison to the subsequent helix growth due to the additional entropic cost to bring four residues into the α-helical region. On the other hand, the available multiple initiation sites can compensate partially such entropic cost since the first turn may form anywhere along the helix and the number of possible initiation sites grows linearly to the chain length. Thus, one would expect that the folding rate would be qualitatively proportional to the number of residues in the helix if the initiation were the rate-limiting step. So far, there has been no experimental evidence that indicates such proportionality.

Our simulation results indicate that the rate-limiting step is not the initiation. We observed that the helix could form its first turn rapidly within nanosecond time scale, corresponding to the rapid phase observed in the simulations. More importantly, the initiation does not trigger the quick formation of the entire helix, as envisioned by the "initiation" model. In fact, the initiation is at least an order of magnitude faster than the completion of the helix. Our simulations indicate that the completion of the helix has a time constant of 12–13 ns, whereas the initiation phase has a time constant of 0.06–0.08 ns. The later is in qualitative agreement with all-atom explicit solvent simulations on short peptides.[10] In our opinion, the two orders of magnitude disparity in these two important rates rules out the possibility that the helix initiation would play any significant role in the peptide folding. It is not the rate-limiting step. Our data further suggests that the initiation is not linked to the rate-limiting step. In fact, the single-turn helix, though can form very rapidly, is unstable, judging from its lack of significant presence after 10 ns. Its abundance dropped from 11% in 0–10 ns to 1% in 10–20 ns. Presumably, some of these single-turn helix species develop into multiturn species, and others dissipate and go back to the random-coil states. In a qualitative sense, this is consistent with the classical helix–coil transition theory.

Is the helix growth the rate-limiting step? Our data suggests that the answer is no. Indeed, the completion of the helix is a slow process that has a time constant of 12–13 ns (or 16–20 ns extrapolated time constant). However, helix growth itself is not the reason that causes the reduction in rates. This becomes evident when the various helical species are compared in different time period. The relative abundance of 3-turn helix and the helix–turn–helix species remains basically constant in all three periods under consideration (0–10 ns, 10–20 ns, and 0–100 ns) both of which have multiple turns of helix. This contradicts the suggestion that helix growth is the rate-limiting step. Under the "growth" scenario, one would expect a significant change of abundance in these two highly helical species due to the perceived barrier leading to the full helix conformation. The relative constant abundance of these two species shows that immediately after formation of these species, the peptide soon adopts the full helical conformation. Thus, the helix growth itself is not the rate-limiting step.

To understand the kinetic process, one must go beyond the helix itself. We propose that the rate-limiting step is breaking the physical interactions responsible to stabilize the intermediate state—namely, the two-turn helix species.

There are six clusters that have two-turn helix motif (Figure 3P–U). Not surprisingly, most of them

are also present in the time period of 10–20 ns with similar populations. Four of them have the C-terminal helix and the other two form the helices in the middle. A common feature of all six clusters is their compactness. In fact, they are among the most compact structures, more compact than the full helix. Another important common feature is that the helical and nonhelical segments are held together by well-packed hydrophobic clusters involving the Tyr side chain. Interestingly, the Tyr side chain is always in contact with solvent. The interior of the hydrophobic cluster is formed mainly by Ala side chains. We conclude that the hydrophobic interaction is mainly responsible to the stability of these structures. Thus, the rate-limiting step is to break the hydrophobic cluster.

In a related work, Eaton and co-workers[7] studied the kinetics of helix formation using the MABA-labeled peptide [MABA–$(A)_5$–$(AAARA)_3$–ANH_2] as the model system. The unfolding rates were measured by monitoring the fluorescence of MABA in series of laser T-jump experiments. The burial and exposure of the N-terminal aromatic group were seen in the folding/unfolding transitions of the helical peptide. Therefore, our observation that the N-terminal Tyr side chain can be partially buried is in agreement with those experiments.

The native state of a typical well-folded protein is usually dominated by an ensemble of well-ordered structures. In the case of marginally stable peptides and proteins, the native state can comprise disordered structures. Baldwin and co-workers[20] showed that depending upon the salt concentration, the native state of AK16 only has an equilibrium helical content of 63–73%. Our simulations further indicate that the native state represents a dynamic equilibrium between the full α-helix and partial helices with frayed ends. Interestingly, the side chains of the frayed ends form small hydrophobic cluster with the aliphatic portion of the Lys side chains (Figure 3E–I).

The renewed interests in helix–coil transition is largely attributable to the relatively new ultrafast experimental methods such as the laser T-jump experiments pioneered by Gruebele and co-workers.[25] In the study of the MABA-labeled peptide, Eaton and co-workers found that the fluorescence intensity relaxed with the time constants of 8, 20, and 12 ns[7] in three temperature jumps (265 + 10 K, 273 + 20 K, and 303 + 16 K, respectively)**.** Although the varying relaxation rates were linked to the temperature dependence of the fundamental kinetic rates, additional measurements seem desirable to rule out the trivial correlation to the T-jump sizes. Nevertheless, the relaxation was attributed to the breaking of the hydrogen bond between MABA carbonyl and the amide groups. Eaton

and co-workers then applied a kinetic version of Schellman model,[26] and concluded that the time scale of helix propagation is on the order of 10 ns.

We think an alternative explanation is plausible and appears to be more convincing in light of the simulation results notwithstanding the difference between the two peptides. A related observation in our simulation was that the native state comprises both fully and partially folded helices where the termini of the later were in random-coil states. Thus it is likely what was observed in the experiments of Eaton and co-workers was the propagation of the random-coil segments when the temperature was raised. Indeed, the observed fluorescence decrease due to temperature jump may reflect lose of hydrogen bonds between the carbonyl and the amide groups. A critical difference lies in the consideration of the side chains. In the Schellman theory, only the hydrogen bonds were considered. The side chains were not taken into account. Our simulations indicate that small hydrophobic clusters can form by the side chains. The aromatic group of the N-terminal Tyr residue can be partially buried when the residue is in the coiled state. Since the coiled state comprises multiple conformations, the temperature jump effectively shifts the distribution from one conformation to another. The relaxation rates would be dictated by the stability of the conformations, particularly the hydrophobic clusters. Thus, the observed relaxation in the laser temperature jump experiments may be the process of breaking the cluster that is required to shift the equilibrium. The relaxation time may be the lifetime of the hydrophobic cluster formed by the aromatic side chain. As a direct consequence, the observed relaxation rates should be similar to the folding rate of the peptide since the later is also dictated by the lifetime of hydrophobic cluster. Indeed, Williams et al. found that the folding time of suc-Fs peptide is 16 ns.[2] In comparison, the relaxation times of MABA-labeled peptide were 8, 20, and 12 ns. The two peptides have identical sequence with different N-terminal groups. The similarity between these two types of rates is difficult to explain under the "initiation-growth" scenario. Another consequence is that since multiple conformations are involved it is possible that the relaxation rates may depend on the temperature jump size in addition to its dependence on the temperature.

Implication to the Folding Mechanisms of Proteins

Being perhaps the simplest peptides with well-defined structures, the alanine peptides exhibit surprisingly rich folding features. Despite the obvious and pro-

found differences, the folding of a single helix and that of proteins bear resemblances including some of the important features. The most surprising feature in helix formation is perhaps the formation of the non-native loop stabilized by the hydrophobic cluster, much like what has been observed in proteins as commonly referred to as the hydrophobic collapse.[6] Though its native (full helix) structure does not have tertiary contacts, the folding of helix was significantly affected by the non-native tertiary contacts. So much so, the folding rate was almost exclusively determined by the lifetime of the non-native tertiary contacts. In proteins, though much has yet to be discovered, the formation of the native contacts and breaking of the non-native ones are intertwined. The question is which one of the two would be dominant at the rate-limiting step. We would like to suggest that breaking of non-native states should be important for some proteins. If so, the folding rate would be dictated largely by the unfolding of non-native marginally stable states in those proteins.

The formation of the hydrophobic cluster, even in the simple α-helical peptides reinforces the notion that hydrophobic interaction plays a significant role in protein folding. In this regard, our simulation results are in line with the hydrophobic collapse model that envisions a rapid collapse into compact conformations when proteins start to fold from extended conformations. Interestingly, the collapsed conformations were even more compact than the native conformation. Thus, a possible avenue to validate our results would be a direct measurement of the size of the peptide during folding. Care should be taken, though, to ensure that the starting conformations are extended. This becomes an issue since the non-native state is highly disordered and its structural features are poorly characterized. Since our simulations were started from the fully extended (straight chain) conformation, which is not typically accessible to experiments, some of the early folding events, such as helix initiation, may also be inaccessible to experiments.

Because of the biological consequences of protein folding, it is a common practice to put the features observed during folding into the context of biology. Furthermore, because biological systems have been under evolutionary pressure, it is thought that the existing features in biological systems have been "optimized." Though we are not in a position to test these directly, our results do not seem to suggest the "optimized" mechanisms that should result in the fastest folding pathway. In the case of the alanine-based peptide, the fastest pathway is the simple initiation-growth mechanism characterized by the absence of the marginally stable non-native state. Our data seem

to suggest that the hydrophobic collapse did not assist folding since the collapsed marginally stable conformation resulted in a delay of the completion of folding. Because the hydrophobic collapse usually results in nonspecific conformations that do not resemble the structures of the native state, it is difficult to justify how a positive role it may play during folding. Indeed, the hydrophobic collapse can greatly reduce the conformational space. But the hydrophobic force also stabilizes the non-native states. The net effect is a free energy surface of significantly increased roughness. Consequently, folding rates are reduced.

Chain entropy has long been recognized as one of the important components constituting the free energy barrier. A recent example was the work by Plaxco and Baker, who found correlation between the "contact orders" and folding rates.[24] The simple helix-forming peptides, however, may be at the other end of the spectrum. In the absence of tertiary contacts at its native state, the chain-entropy contribution during folding is almost negligible. Instead, the folding rate is dictated almost entirely by the lifetime of the non-native states. Given the similarity between the observed folding processes of this short peptide and those of small protein,[6] that both undergo hydrophobic collapse and both show marginally stable non-native states, we further postulate that the lifetime of non-native states can significantly affect the folding rates of proteins.

Comparisons with Existing Computational Studies

There have been numerals simulation studies on the alanine-based peptides. We focus our comparisons to the direct folding simulations using atomic-level models. One of the early atomic-level simulations was done by Sung,[3,27] who studied folding of both $(Ala)_{16}^3$ and Ala and Lys mixed peptides[27] using the Weiner et al. force field[28] and a mean-field treatment of solvent. Later, Sung and Wu[29] simulated folding of AQ16 with Cornell et al. force field.[13] In these studies, Sung found a significant portion of the peptides in 3_{10}-helical conformation. Loop and other non-native conformations were found before the peptides became helices. The observation of non-native states is consistent with our results.

Ferrara et al.[21] studied folding of AQ15 [Ace-$(AAQAA)_3$-NHCH$_3$] using CHARMM19 force field with a simple distance-dependent dielectric treatment ($\epsilon = 2r$) to mimic solvation electrostatic screening effect. In addition to the α-helical conformation, they also found substantial presence of π- (7%) and 3_{10}-helices (6.5%), consistent with earlier equilibrium

simulations of Shirley and Brooks.[30] In the work of Ferrara et al.,[21] a large number of simulations were conducted that started from different random conformations. Each of the simulations was stopped when the simulated structure was close to the native helical conformation as judged by its RMSD from the full helix conformation. Therefore, the simulation time would be the "first passage" time. On a qualitative basis, the first passage time should be similar to the folding time. But it would be difficult to make quantitative comparisons with experiments due to the difficulty to measure such a quantity from experiments. It would also be difficult to obtain simple quantities such as equilibrium helical content that is frequently used for critical comparisons. The substantial presence of π-helix is also a concern.

Hummer et al. studied three short peptides, including Ace-;Ala$_5$-Nme, Ace-Ala$_2$-Gly-Ala$_2$-Nme, and Ace-Gly$_5$-Nme,[10] with the Cornell et al. force field and explicit solvent. Hummer et al. estimated that the helix initiation rate was 0.1 ns. In comparison, the initiation rate estimated from our simulations is about 0.06–0.08 ns. Thus, by neglecting solvent viscosity effect, the continuum solvent model appears to accelerate the process marginally by a factor of 1.25–1.7. Since the absence of viscosity in our continuum model should also make the unfolding of the intermediate states faster, one may need to correct the time constant. Assuming the same ratio holds true for the unfolding of the intermediate states, we can then extrapolate the folding time by multiplying the correcting factor obtained above. Indeed, the underlying assumption is that the process is not diffusion driven. Given the agreement, we think it is likely that neither initiation nor unfolding of the intermediate states is driven by diffusion. Visual inspection confirms that. The peptide quickly adopts the helix–turn–helix conformation when it unfolds from the intermediate states. We did not observe large scale movement (i.e., unfolding to extended conformations) during that process. Such a simple extrapolation may likely be invalid in the folding of other peptides and proteins in which large scale movement may play important role.

Wu and Wang[22] recently studied folding of AQ16 [Ac-(AAQAA)$_3$-YNH$_2$] in aqueous solution with explicit representation of solvent and Cornell et al. force field. The self-guided MD[31] approach was used to achieve reversible folding within relatively short (10 ns) simulation time. In this approach, the spatially and temporally averaged force was added to the molecular mechanical force. Two simulations were conducted with different coupling constants for the average force. Although encouraging, the average equilibrium helicity seemed to be dependent on the choice of the coupling constant. More simulations are desirable to rule out such dependency. Nevertheless, they found that the average abundance of α-, 3_{10}-, and π-helical species were respectively 59.8, 3.8, and 0.5%.

Garcia and Sanbonmatsu[32] recently studied the Fs peptide [Ace–A$_5$–(AAARA)$_3$–A–Nme] using Cornell et al. force field and a replica-exchange algorithm[33] for accelerated sampling efficiency. Two sets of simulations were conducted with the poly-Ala as the control. The Fs peptide appears to be more stable and more helical. The increased stability and helicity was attributed to the screening effect of Arg side chains.

Regardless how simulations were conducted, these studies have focused almost exclusively on the characteristics of main-chain conformations, even though marginally stable non-native states were frequently observed. These observations are consistent with our results. An exception was the work of Elmer and Pande,[34] who studied the folding of a 12-mer poly-phenylacetylene. The solvent effect was modeled by additional Lenard–Jones interaction between ring carbon atoms and the depth of the potential was adjustable to reflect ring-stacking interactions. An interesting correlation between the strengths of ring-stacking interaction and the folding rate was observed, consistent with the notion that the side-chain contacts in the non-native states determine the folding rates. The results are also in qualitative agreement with ours.

CONCLUDING REMARKS

Our simulations demonstrate that a peptide as simple as the alanine peptide can exhibit remarkably complex folding kinetics. The detailed atomic-level modeling allows us to capture the main features of the folding process. Much like what has been observed in the folding of small proteins, the folding of the peptide begins with a hydrophobic collapse that is concomitant to the formation of transient helical elements. The development of the helical content shows characteristics of three phases. Helices initiate extreme rapidly with a time constant of 0.06–0.08 ns, similar to that estimated from explicit solvent simulations. The second phase is the development of a folding intermediate state that has a time constant of 1.4–2.3 ns. The folding intermediate state is characterized by the two-turn short helices formed either in the middle of the peptide or at the C-terminus and stabilized by the hydrophobic cluster. The peptide reaches the equilibrium state in the third phase with a characteristic time constant of 12–13 ns. A mixture of full and partial helices is found at equilibrium. The calculated equilibrium helicity of 64–66% is in remarkable agree-

ment with experiments. By comparing the helix initiation time against explicit solvent simulation results, we are able to obtain a factor to take into account solvent viscosity effect that is absent in our continuum solvent treatment. An extrapolated folding time of 16–20ns is obtained that is in qualitative agreement with experiments, notwithstanding the differences in the peptides. Contrary to the prevailing opinion, neither initiation nor progression of the helix is the rate-limiting step. Instead, the rate-limiting step for this peptide is breaking the non-native hydrophobic clusters in which the peptide reaches the transition state characterized by the helix–turn–helix motif.

REFERENCES

1. Chakrabartty, A.; Baldwin, R. L. Adv Protein Chem 1995, 46, 141–176.
2. Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. Biochemistry 1996, 35, 691–697.
3. Sung, S. S. Biophys J 1994, 66, 1796–1803.
4. Wu, X. W.; Sung, S. S. Proteins 1999, 34, 295–302.
5. Duan, Y.; Wang, L.; Kollman, P. A. Proc Natl Acad Sci USA 1998, 95, 9897–9902.
6. Duan, Y.; Kollman, P. A. Science 1998, 282, 740–744.
7. Thompson, P. A.; Eaton, W. A.; Hofrichter, J. Biochemistry 1997, 36, 9200–9210.
8. Gilmanshin, R.; Williams, S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. Proc Natl Acad Sci USA 1997, 94, 3709–3713.
9. Clarke, D. T.; Doig, A. J.; Stapley, B. J.; Jones, G. R. Proc Natl Acad Sci USA 1999, 96, 7232–7237.
10. Hummer, G.; Garcia, A. E.; Garde, S. Proteins 2001, 42, 77–84.
11. Tsui, V.; Case, D. A. J Am Chem Soc 2000, 122, 2489–2498.
12. Duan, Y.; et al. 2002, submitted.
13. Cornell, W. D.; et al. J Am Chem Soc 1995, 117, 5179–5197.
14. Wu, C.; Chowdhury, S.; Duan, Y. 2002, in preparation.
15. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. J Comp Phys 1977, 23, 327–341.
16. Bashford, D.; Case, D. A. Ann Rev Phys Chem 2000, 51, 129–152.
17. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. J Comp Phys 1984, 81, 3684–3690.
18. Daura, X.; van Gunsteren, W. F.; Mark, A. E. Proteins 1999, 34, 269–280.
19. Kabsch, W. Acta Cryst 1976, A32, 922–923.
20. Padmanabhan, S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. J Mol Biol 1996, 257, 726–734.
21. Ferrara, P.; Apostolakis, J.; Caflisch, A. J Phys Chem B 2000, 104, 5000–5010.
22. Wu, X.; Wang, S. J Phys Chem B 2001, 105, 2227–2235.
23. Andersen, C. A.; Bohr, H.; Brunak, S. FEBS Lett 2001, 507, 6–10.
24. Plaxco, K. W.; Simons, K. T.; Baker, D. J Mol Biol 1998, 277, 985–994.
25. Ballew, R. M.; Sabelko, J.; Gruebele, M. Proc Natl Acad Sci 1996, 93, 5759–5764.
26. Schellman, J. A. J Phys Chem 1959, 62, 1485–1494.
27. Sung, S. S. Biophys J 1995, 68, 826–834.
28. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. J Comp Chem 1986, 7, 230–252.
29. Sung, S. S.; Wu, X. W. Proteins 1996, 25, 202–214.
30. Shirley, W. A.; Brooks, C. L., III Proteins 1997, 28, 59–71.
31. Wu, X. W.; Wang, S. M. J Phys Chem B 1998, 102, 7238–7250.
32. Garcia, A. E.; Sanbonmatsu, K. Y. Proc Natl Acad Sci 2002, 99, 2782–2787.
33. Sugita, Y.; Okamoto, Y. Chem Phys Lett 1999, 314, 141.
34. Elmer, S.; Pande, V. S. J Phys Chem B 2001, 105, 482–485.