Characterizing the Rate-Limiting Step of Trp-Cage Folding by All-Atom Molecular Dynamics Simulations

Shibasish Chowdhury, Mathew C. Lee, and Yong Duan*

Department of Chemistry and Biochemistry and Center of Biomedical Research Excellence in Structural and Functional Genomics, University of Delaware, Newark, Delaware 19716

Received: May 17, 2004; In Final Form: June 11, 2004

In this study, the detailed mechanisms of the rapid-folding Trp-cage mini-protein were investigated by extensive all-atom molecular dynamics simulations of both wild-type and mutant proteins using a recently developed point-charge force field within the AMBER simulation package and the generalized Born treatment of solvation. Among the 77 100-ns simulations performed on the wild-type protein, 5 of the simulation trajectories yielded structures with main-chain RMSDs of 1.0−2.0 Å from the native NMR structure. A gradual reduction in the value of the main-chain RMSD distribution was observed during the simulations, which is consistent with the folding funnel theory. The folding time of ∼3 μs based on native tertiary contacts is in reasonable agreement with an experimental value of ∼4 μs. Detailed analysis suggests that packing of the structurally important Trp_{25} side chain is involved in the rate-limiting step and unfolding of the misfolded states and overcoming the additional entropic barrier also contributed to the rate-limiting steps. This is reinforced by the faster folding rate of the W25F mutant. Two putative folding pathways were observed from the simulations, and their folding rates differed by about 200-fold, leading to a 3.2 kcal/mol folding free energy barrier difference. Of this, approximately 2.2 kcal/mol was due to unfolding of the misfolded states, and about 1.0 kcal/mol was due to overcoming the entropic cost to move Trp_{25} side chain into the native orientation. Although formation of the main-chain contacts was not the rate-limiting step, we observed a hierarchical process in which the short-range native contacts formed faster than the long-range ones. These observations are consistent with the contact-order theory.

Introduction

A comprehensive understanding of protein folding mechanisms at both the macroscopic and molecular levels has been a subject of extensive research. Whereas macroscopic information of prototypical systems, such as folding rates and folding free energies, has been accumulating over the past few decades, much needed microscopic information at the molecular level, such as structures of intermediate states and their roles in folding, remains difficult to obtain. Computer simulations at various levels of sophistication, including all-atom molecular dynamics simulations, have played important roles in advancing our understanding of this subject. In the recent decade, we have witnessed exciting developments on both the experimental and theoretical fronts. On the experimental front, methods that are capable of providing (macroscopic) information on the early folding processes[1−3] have been developed. These methods have paved the way for the kinetic studies of fast-folding small proteins and peptides; they have enabled us to directly compare the folding rates from both computational and experimental studies.[4] A growing number of small autonomous-folding proteins, including the villin headpiece subdomain,[5] protein A,[6] $\beta\beta\alpha$,[7,8] and the Trp-cage[9] mini-protein, have been devised as model systems. They have been designed to exhibit properties common to small single-domain proteins including well-packed cores and tertiary contacts and multiple secondary structures.

On the theoretical front, after the folding funnel hypothesis was proposed,[10,11] the focus of the theoretical community has shifted toward understanding the details of the protein folding energy landscapes.[12,13] The discovery of the correlation between the folding rates and their "contact orders"[14] has helped us to better appreciate the entropic contributions to the folding free energy barriers.[15] As computational power has increased, all-atom computer simulations have begun to approach the folding time scales of small proteins.[16−18] With the help of high-level quantum mechanical data,[19] the biomolecular force fields describing the interactions of biomolecules are now capable of capturing the essential molecular interactions, leading to accurate ab initio folding of small proteins. The combinations of these developments have produced realistic simulations on peptides and mini-proteins by which one can now study the protein folding mechanisms of model proteins in atomistic detail. In this paper, we present our effort to study the rate-limiting steps of the Trp-cage mini-protein by applying these advances.

The 20-residue Trp cage[9] ($N_{20}LYIQW_{25}LKDGG_{30}PSSGR_{35}$-$PPPS_{39}$) is the smallest known protein that exhibits many important features common to single-domain proteins. It can fold autonomously in aqueous buffer to a state with an adequate level of thermal stability ($T_m = 42$ °C); it contains multiple secondary structures held together by a tightly packed hydrophobic core. Starting from its N-terminus, the secondary structure of this protein consists of an α-helix formed by residues $L_{21}-K_{27}$ (following the numbering scheme of Neidigh et al.) followed by a 3_{10}-helix ($G_{30}-S_{33}$) and a short stretch of type II polyproline (PPII) helix at the C-terminus. In addition to the presence of a central hydrophobic core around the Trp_{25} side chain, other stabilizing interactions include a salt bridge between

* Corresponding author. Tel.: (302) 831-1099. Fax: (302) 831-6335. E-mail: yduan@udel.edu.

10.1021/jp0478920 CCC: $27.50
Published on Web 08/13/2004

the Asp$_{28}$ and Arg$_{35}$ side chains, and hydrogen bonds between NH of Gly$_{30}$ and the backbone carbonyl oxygen of Trp$_{25}$ and from the NH$^{\epsilon 1}$ of Trp$_{25}$ to the backbone carbonyl of Arg$_{35}$. Small size, structural simplicity, and a fast folding rate[20] make this mini-protein an ideal model for computer folding simulation studies.

We and others[18,21−23] have demonstrated that, in the case of Trp-cage mini-protein, molecular dynamics simulations are capable of folding this protein into structures very similar to those obtained from multidimensional NMR spectroscopy.[9] From our previous simulation, we postulated that packing of the Trp$_{25}$ side chain was possibly the rate-limiting step for the folding of this protein. This was a previously unsuspected result. However, this hypothesis was based on the analysis of just one trajectory. To further investigate this hypothesis, we conducted a total of 91 simulations on the wild-type and mutant proteins (77 wild-type, 14 mutant trajectories), each encompassing a 100-ns simulation time. The point mutation is chosen to test our hypotheses about the roles of Trp$_{25}$ in folding. Within a broader context, these simulations also afforded a unique opportunity to account for the detailed folding processes, identify the folding pathways, explore the folding funnels, and reflect upon some of the existing protein folding theories.

Methods

The Duan et al.[19] all-atom point-charge force field was used to represent the Trp-cage protein and its mutant. The AMBER molecular dynamics simulation package was used for both the simulations and data processing throughout this paper.[24] The solvent was represented by the generalized Born solvent model[25] with an effective salt concentration of 0.2 M. A total of 77 independent simulations were performed on the wild-type protein with different initial velocities. Starting from straight chain conformation, after initial energy minimization, molecular dynamics simulations were carried out at 300 K. The SHAKE[26] algorithm was applied to constrain all bonds connecting hydrogen atoms, and an integration time step of 2.0 fs was used. Born radii were calculated every 5 steps (10 fs) using the method of Bashford and Case.[27] Nonbonded forces were calculated using a two-stage RESPA approach, where the forces within a 10-Å radius 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[28] with a time constant of 2.0 ps. All simulations were conducted to 100 ns with an aggregated simulation time of 7.7 μs. The trajectories were saved at 1.0-ps intervals for further analysis. These saved snapshots were clustered by a semilinear clustering technique[29] with a 2.0-Å main-chain RMSD cutoff. The representative structures were those within each cluster closest to their respective cluster averages.

A total of 14 simulations of the single-point mutant W25F were also performed with the aforementioned simulation protocol. The "native" structure of W25F mutant was obtained by replacing the Trp$_{25}$ residue in the wild-type Trp-cage structure with phenylalanine. We assume that the structure remains intact on single point mutation. To assess the validity of the assumption, four 5-ns equilibrium simulations were performed in which the mutant retained the structure close to the wild-type native structure with a main-chain RMSD of approximately 2.0 Å. Thus, in the subsequent discussion, this structure is referred to as the native structure of the W25F mutant.

Results

In this section, we first present the results from the wild-type Trp-cage folding simulations and then the characterizations of the W25F mutant simulations. The folding processes were initially monitored by main-chain root-mean-square deviation (RMSD) from the native structure. Detailed secondary and tertiary structural analyses were performed to characterize the simulations. In addition, the metastable states were identified through clustering analysis of the resulting trajectories.

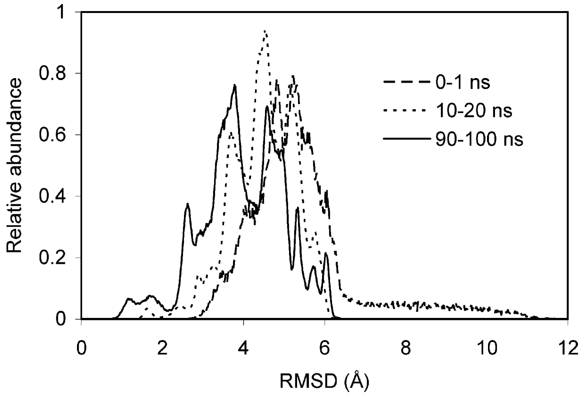

Figure 1. Distribution of the main-chain RMSD during the intervals 0−1 (− − −), 10−20 (• • •), and 90−100 ns (—) of the simulation. The distribution of RMSD during 0−1 ns is scaled up for comparison with the other two distribution plots. To calculate the distribution, bins with a 0.02-Å width were constructed. The main-chain RMSD values (in Å) are shown along the *x* axis, and the numbers of structures falling within a bin are shown along the *y* axis.

Overall Assessment of the Wild-Type Trp-Cage Folding. The distribution of the main-chain RMSD during the initial period of the simulations (0.0−1.0 ns) varied between 3.0 and 6.5 Å (Figure 1). The long tail beyond 6.5 Å corresponds to the initial straight chain and the rapid collapse. Two closely spaced peaks at around 4.75 and 5.25 Å suggested the existence of partially folded structures in the denatured state and the unfolded-state ensemble. These are clearly transient states in the folding process. As the folding progressed, the overall ensemble of the structures moved toward the native structure indicated by the reduction of the main-chain RMSD. The main peak shifted to around 4.5 Å between 10 and 20 ns (dotted line) and was accompanied by three smaller peaks near 3.75 and 5.25 Å. During the later stage of folding (90−100 ns), the main peak at 4.5 Å diminished substantially, and the two small peaks near 5.25 Å almost disappeared, suggesting that they were early-stage intermediate states. This was concurrent with an increase at 3.75 Å, indicating that this peak represents a late-stage species. A small peak below 2.0 Å appeared during the interval between 20 and 30 ns. This region represents a basin of protein conformations similar to the native structure and will hence forth be referred to as the native basin. The distribution of the main-chain RMSD in this region increased during the final 10 ns of the simulations, indicating increased sampling of the native basin during the late stage of folding. Taken together, the progression of the simulations led to an overall shift of structures toward the native state and suggested that the equilibrium structure of the simulations resided close to the native basin. The existence of numerous peaks on the RMSD distributions indicated a rugged energy landscape on which folding took place.

There were five trajectories whose main-chain structures were within 2.0-Å RMSD from the native structure in the final 20 ns of the simulations. Under the assumption of a simple two-state kinetics and given the 4.0-μs experimental folding time,[20] one might expect approximately 2 of the total 77 trajectories to reach the native basin by 100 ns. The higher-than-expected number of trajectories observed in our simulations was most likely due to the lack of solvent viscosity in the continuum

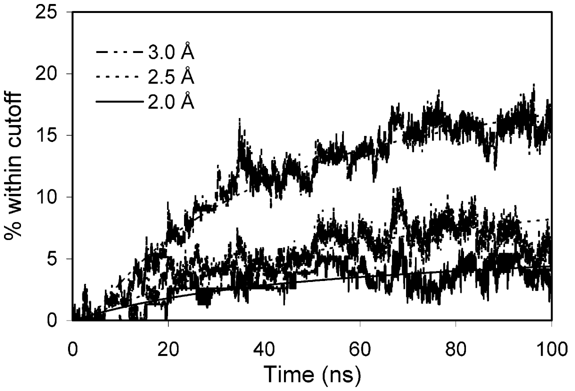

Figure 2. Time development of the percentage growth of Trp-cage molecules that fell into the native structural basin with main-chain RMSD cutoffs of 2.0, 2.5, and 3.0 Å. They were fitted by the expression $[1 - r_1 \exp(-t/\tau_1) - r_2 \exp(-t/\tau_2)]$, and the results are summarized in Table 1.

TABLE 1: Summary of Fitting Results from the Time-Dependent Population of Folded Molecules in the Ensembles of 77 Trajectories

	RMSD cutoff		
	2.0 Å	2.5 Å	3 Å
r_1	0.03	0.08	0.15
r_2	0.97	0.92	0.85
τ_1 (ns)	28	51	27
τ_2 (ns)	6154	6073	1921
rms error	0.88	1.25	1.16
τ_2/τ_1	220.0	119.0	71.5
ΔG_f (kcal/mol)	3.2	2.9	
$\Delta G'_f$ (kcal/mol)	2.2	1.9	

solvent model employed, resulting in increased conformational sampling relative to that of explicit solvent simulations.

Figure 2 shows the percentages of trajectories that fell within 2.0, 2.5, and 3.0 Å of main-chain RMSD from the native structure. All of these curves were best fitted to a combination of two exponential functions, indicating that the folding of Trp cage took two different paths, with one of them being faster by 2 orders of magnitude. The fitting results are summarized in Table 1. The fast folding rates varied from ~25 to ~60 ns, depending on the cutoff used, whereas the slower rates were between ~2 and ~6 μs; the latter were in qualitative agreement with the 4.0-μs experimental folding time of Trp cage.[20] Note that only 3−15% of all trajectories followed the fast folding pathways, which implies that the slow path was dominant. Given the agreement between the folding rates, these simulation results suggest that the experimentally measured rate corresponds to the slow refolding pathway.

Clustering Analysis. A total of 2532 clusters were identified using the main-chain RMSD as the clustering criterion. Not surprisingly, most of these clusters were too poorly populated to be of interest and were excluded from further analyses. Among the more populated ones, there were 36 clusters that each contained more than 70000 structures (i.e., more than 0.9% of total saved snapshots). In our analyses, we focused our attention on these well-populated clusters.

First, we classified these 36 most-populated clusters into two broad categories: the native and nonnative clusters. Clusters with representative structures within the deviation range of the NMR structures were identified as native clusters. All other clusters were considered nonnative. The native cluster was the most populated, comprising a total of ~0.28 million snapshots (or ~3.6% of all saved snapshots), which is slightly more than the expected 2.5% to fall into the native ensemble given a folding rate of 4.0 μs and simple two-state kinetics.

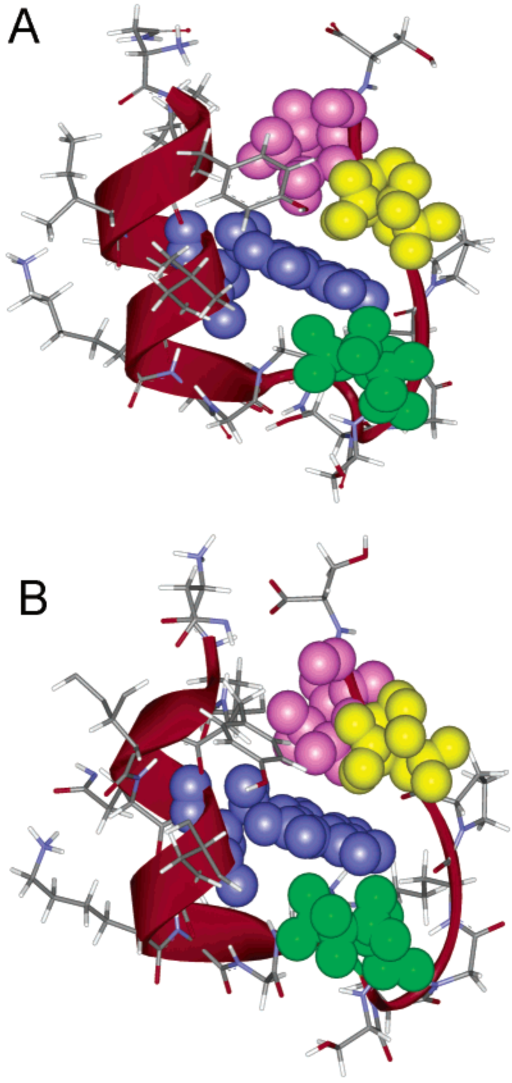

Figure 3. (A) One of the 38 NMR structures. (B) Representative structure of the most populated cluster. The main chains are shown in ribbon, and structurally important Trp_{25} (purple), Pro_{31} (green), Pro_{37} (yellow), and Pro_{38} (pink) residues are shown in space-filling models, while all other side chains are shown in sticks.

Figure 3 shows a direct comparison of the NMR structure and the representative structure of the most populated native cluster. The high degree of resemblance between these two structures is readily apparent. Some of the more salient features include a well-packed hydrophobic core and the well-formed native secondary structural units (i.e., α-, 3_{10}-, and PPII helices). Other important features include the packing of the Trp_{25} side chain between Pro_{31} and Pro_{39} (β face toward Pro_{31} and α face toward Pro_{38}); the closed Tyr_{22} side chain, completing the native Trp_{25} packing; and the stable native hydrophobic core with Leu_{21}, Trp_{25}, Leu_{26}, Pro_{31}, and Pro_{36-38} residues. All of these structural features are in excellent agreement with those observed from the NMR structures. In fact, the range of RMSD values between the simulated and NMR structures is only 1.1−1.7 Å. In comparison, the pairwise RMSD of the 38 models in the NMR ensemble ranges from 0.3 to 1.4 Å. Thus, this native cluster was within the range of experimental uncertainties.

Because the simulation time was only 100 ns, about 40 times shorter than the experimental folding time, most clusters are

TABLE 2: Summary of the Most Populated Clusters

RMSD[a]	total energy[b]	no. of snapshots	E_{el}[b,c]	E_{GB}[b,d]	vdW[b,e]	bonding energy[b,f]	features[g]
1.05	0.0	278915	0.0	0.0	0.0	0.0	native
2.61	5.0	80413	26.3	−24.3	−6.5	9.5	nativelike
3.43	6.6	79190	133.5	−102.2	−3.6	−21.0	nativelike
2.82	7.0	191646	44.8	−34.4	−2.3	−1.1	nativelike
3.99	9.3	123545	34.4	−26.2	−3.1	4.2	full helix
3.81	10.6	108106	151.0	−135.6	1.2	−6.0	partial helix
4.52	10.8	99788	65.5	−50.2	0.5	−5.1	random coil
4.43	11.2	96323	49.0	−28.9	3.2	−12.1	random coil
4.51	11.4	84398	226.9	−191.3	−6.2	−18.0	hairpin with partial helix
4.9	11.5	95606	119.5	−109.5	−7.4	8.9	random coil
2.55	13.1	99056	74.8	−74.9	−1.6	14.9	partial helix
3.59	13.8	94004	231.0	−210.4	−2.8	−4.0	hairpin with partial helix
5.38	14.8	99661	20.1	−3.3	3.0	−5.0	hairpin
3.75	15.1	87137	189.1	−150.8	−4.6	−18.6	partial helix
4.31	15.4	75201	130.4	−103.1	−6.4	−5.6	random coil
4.41	16.0	98876	66.2	−52.9	−0.3	3.0	random coil
4.7	16.1	122768	155.6	−133.8	−4.0	−1.7	random coil
2.95	16.9	113955	120.2	−87.5	−0.4	−15.4	nativelike
5.51	16.9	80266	255.5	−208.1	−3.8	−26.8	random coil
3.76	17.7	93565	157.0	−136.6	4.0	−6.7	hairpin
3.83	18.1	78994	41.3	−15.7	−5.2	−2.4	random coil
4.95	18.5	112693	156.0	−127.1	−6.5	−4.0	partial helix
4.27	18.7	103741	91.4	−61.2	−3.5	−7.9	random coil
2.74	18.8	93207	92.3	−63.4	−4.5	−5.6	random coil
4.57	19.2	147199	7.3	15.1	8.9	−12.2	hairpin
5.25	20.3	74309	198.0	−155.5	−8.3	−13.9	random coil
4.51	23.0	75837	204.8	−160.4	−8.9	−12.6	random coil
4.01	23.6	86958	227.5	−199.1	1.1	−5.8	random coil
5.18	24.2	91893	236.7	−205.3	3.9	−11.1	random coil
3.6	25.0	156909	203.6	−169.3	4.8	−14.1	random coil
4.59	25.3	81923	244.4	−225.7	3.3	3.3	partial helix
4.72	26.0	82805	200.6	−168.5	−1.0	−5.1	random coil
4.24	26.1	94718	125.0	−102.0	−5.6	8.7	hairpin
4.52	26.5	82443	217.9	−175.6	−1.0	−14.8	random coil
5.44	28.6	99862	136.5	−90.4	−0.6	−16.8	random coil
5.97	29.4	87845	243.7	−203.9	−4.3	−6.1	random coil

[a] RMSD is for the main-chain structures relative to the NMR structures in Å. [b] Energies were obtained after energy minimization and are in kcal/mol. [c] E_{el} is the electrostatic energy. [d] E_{GB} is the solvation free energy calculated by the generalized Born model. [e] vdW is the van der Waals energy. [f] Bonding energies include bond, bond angle, dihedral angle, and 1−4 energy terms. [g] Key structural features.

expected to be nonnative and to represent the unfolded-state ensemble as demonstrated by Snow et al.[22] By visual inspection, we classified the nonnative clusters into five broad categories (Table 2) whose representative structures are shown in Figure 4A−E. Figure 4A represents four nativelike clusters (∼6.0% of snapshots) in which the overall architecture was close to the native structure. However, the 3_{10}-helix as well as the stacking of Trp_{25} with Pro_{31} were not observed. The salt bridge between Asp_{28} and Arg_{35} was also absent. More importantly, the Trp_{25} side chain remained outside the cage and was blocked by part of the collapsed PPII helix. The second type of clusters had only the α-helical segment without the native topology (Figure 4B), which comprised six clusters and 612460 (∼8.0%) snapshots. The third type of clusters contained hairpin structures (Figure 4C) with the native contacts between Trp_{25} and Pro_{36-39} (four clusters and 435143 or ∼5.6% snapshots). The fourth type of clusters represented hairpin structures with partial native helical segment (Figure 4D) [two clusters and 178402 (∼2.3%) snapshots]. The last category of clusters contained random coil structures (Figure 4E) in which no significant native structural feature was observed. They account for 1783631 (23%) snapshots.

The relative stabilities of the 36 most populated clusters were analyzed by minimizing the energies of the representative structures (Table 2). The structure with the lowest backbone RMSD from the native structure had the lowest energy; it was between 5 and 32 kcal/mol lower than the other 35 structures. The largest contribution to conformational stability came from the electrostatic interaction energy, which was partially compensated by the solvation free energy. The nativelike clusters with Trp_{25} outside the cage had more favorable van der Waals (vdW) energies (∼2−7 kcal/mol) than the native structure. This is understandable as close contacts between the Trp_{25} side chain and the "cage" generated unfavorable vdW energy. However, the more favorable electrostatic energy of the native structure more than compensated for the unfavorable vdW energies, which, in turn, made the native structures more energetically stable.

The clustering analysis of individual trajectory, which produced structures within 2 Å of native structure at any point of the simulation, revealed that folding of these trajectories went through some common intermediate states. The initial common intermediate state of the folding was the formation of the hairpin structure in which Trp_{25} side chain was in contact with C-terminal proline side chains. This was followed by helix formation at the N-terminus. There were two scenarios in this state. In the first case, we observed that the orientation of Trp_{25} side chain (χ angle) was close to that in the native structure. It was noticed that the formation of native main-chain structure from this state was stable throughout the simulation. In the other case, the Trp_{25} side chain was flipped by 180° from that of native orientation (discussed later), and formation of the native structure (main chain) from this state was less stable. In the latter case, frequent transitions from nativelike main-chain structure to

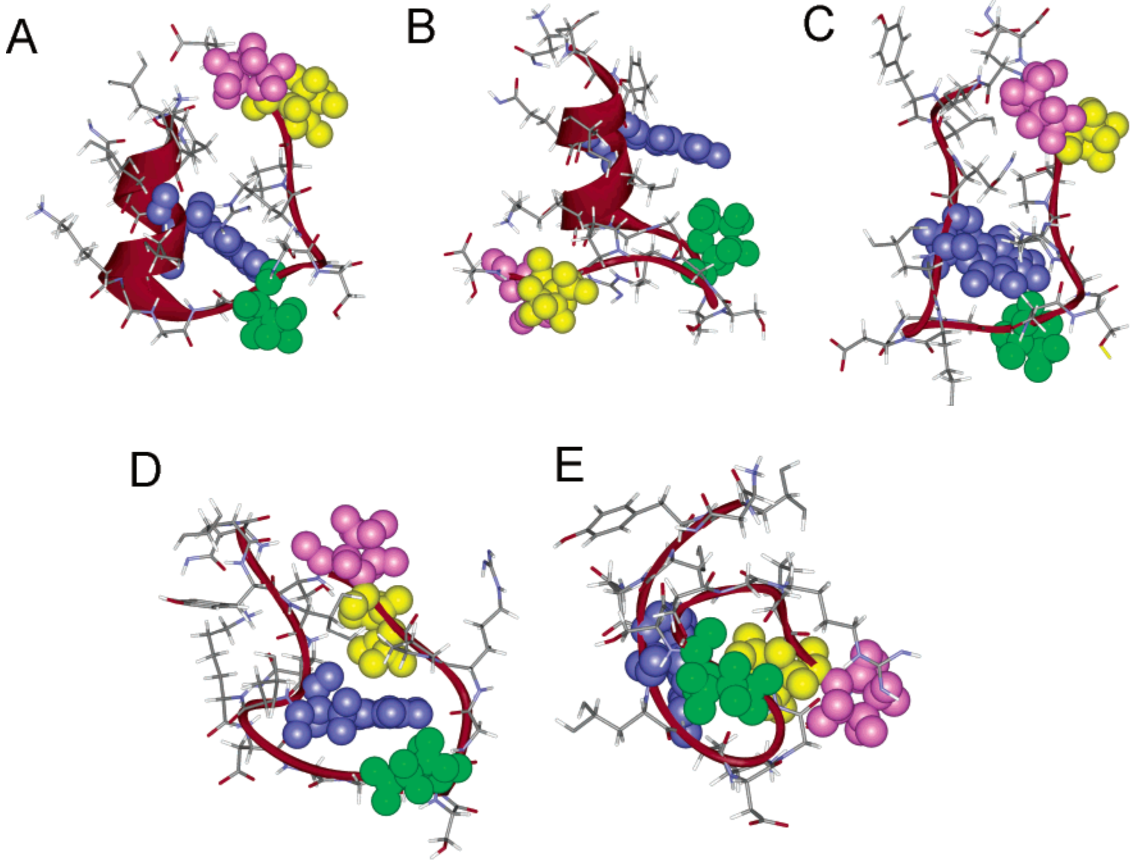

Figure 4. Representative structure of five clusters under the broad category of nonnative clusters. Color and side chain models are the same as in Figure 3.

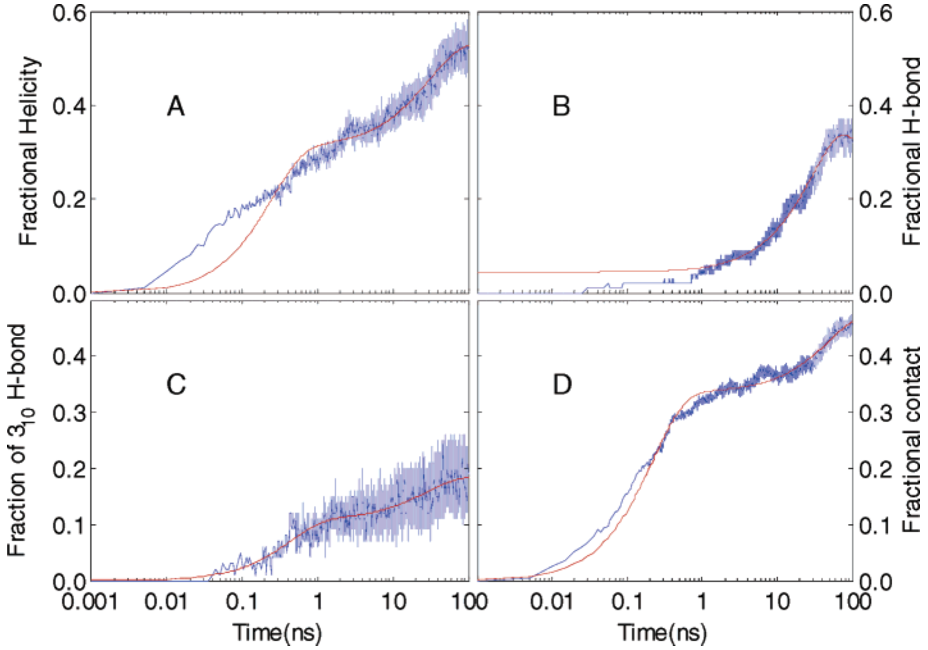

Figure 5. Fractional helicity of the α-helix measured by (A) the main-chain ϕ and ψ torsion angles and (B) the i to $i + 4$ hydrogen bonds. (C) Fraction of 3_{10}-helix hydrogen bonds. There are an average of 3.97 i to $i + 4$ hydrogen bonds and 2.13 i to $i + 3$ hydrogen bonds in the native NMR structures. (D) Fraction of native tertiary contacts. The residual contacts within 5 Å are considered only for those residues that are at least five residues (i to $i + 5$) apart along the primary chain. Three phase fitting curves are also shown (red line).

nonnative main-chain structure were observed (also confirmed by RMSD). The kinetics of folding event was subsequently affected by this unfolding event.

Secondary and Tertiary Structures. Figure 5A shows the fractional helicity averaged over the 77 trajectories as measured by the main-chain torsion angles ($\Phi = -57° \pm 40°$ and $\Psi = -47° \pm 40°$). The α-helix started to form very early (about 30% by 1.0 ns). In contrast, the helix propagated gradually in the later stages and reached 40% mark at around 20 ns and 55% at 100 ns. The helicity was fitted to a combination of three exponential functions (with an rms error of 1.5%). The rates of the first two phases were 0.24 and 30.0 ns, respectively, corresponding to the initiation and propagation of the α-helix. The slowest third phase had a rate constant of ∼1.0 μs, comparable to the folding time of the protein as measured by the growth rate of native structures. The *i* to *i* + 4 main-chain hydrogen bonds, the defining feature of α-helix, also indicated three phases with similar rates where the slowest phase was also ∼1.0 μs (Figure 5B). Thus, there was substantial α-helix formation before folding was completed, and α-helix formation appears unrelated to the rate-limiting step. However, completion of the α-helix (rate constant of 1.0 μs) was still substantially slower than the isolated α-helix and was linked to folding of the entire protein. This was consistent with earlier studies[16,17] on other small proteins and with the notion that the secondary structures in proteins are determined by both local sequence propensity and tertiary contacts. Because helices can propagate at much higher rates than the folding of proteins, the lower rate observed for the completion of helical secondary structure in the Trp cage indicates that it must have formed cooperatively with the overall folding of the proteins.

Formation of the 3_{10}-helix was monitored by the *i* to *i* + 3 main-chain hydrogen bonds and is shown in Figure 5C. About 10% of the 3_{10}-helix formed by 1.0 ns. Thus, the 3_{10}-helix also initiated early. Thereafter, the growth of the 3_{10}-helix slowed considerably and reached 12% by 10 ns and only 20% by the end of the simulation. Given its architectural role, this was not surprising. In the native structure, the 3_{10}-helix bridges between two secondary structures (α- and PPII helices) and is in contact with the Trp_{25} side chain; its formation is closely linked to the folding of the entire protein. Because short helices are unstable,[32] they require stabilizing interactions in proteins, often in the form of tertiary contacts. In this case, the short 3_{10}-helix requires direct contacts with the hydrophobic core to be stable. Because the formation of 3_{10}-helix progressed much faster than the folding of the entire protein (20% vs <10%), formation of this fragment was not the rate-limiting step. Furthermore, because of the requirement for stabilizing tertiary interactions, its formation was closely linked to the folding of the protein and was thus cooperative.

On the other hand, the PPII helix ($\phi = -79° \pm 30°$, $\psi = 149° \pm 30°$) formed very early and independently from the rest of the protein. Approximately 90% of the PPII helix formed within 100 ps (data not shown). The C-terminal proline residues remained in the same conformation as that observed immediately after the initial collapse of the straight conformation.

Thus, our simulations have delineated the chronological order of secondary structure formation. PPII helix formed first and right after the initial collapse of straight chain and remained stable throughout the simulations, suggesting substantial presence of PPII helix in the denatured state. Initiation of the α-helix and 3_{10}-helix started at around the same time. However, the α-helix fragment folds faster than the 3_{10}-helix. We also observed greater fluctuation in the 3_{10}-helicity than the α-helicity, indicating that the α-helix in the mini-protein was more stable than the 3_{10}-helix.

TABLE 3: Fitting Results[a] of C_β−C_β Contacts between the Residue Pairs

contacting residues	r_1	r_2	τ_1 (ns)	τ_2 (ns)
L_{21}−P_{38}	0.51	0.49	0.35	499
Y_{22}−P_{38}	0.51	0.49	0.32	481
W_{25}−P_{38}	0.82	0.18	0.31	68
D_{28}−R_{35}	0.84	0.16	0.34	18

[a] Basic fitting equation was $[1 - r_1 \exp(-t/\tau_1) - r_2 \exp(-t/\tau_2)]$.

Early formation of the native tertiary contacts is often believed to be the main reason for the fast folding of proteins. To examine this hypothesis, we considered the contacts between residues that were at least five residues apart (*i* and *i* + 5) along the primary chain to exclude the short-range contacts of the helical secondary structures. A contact was defined if the distance of any atom pairs between two residues was less than 5.0 Å. The fraction of the native contacts averaged over all simulations is shown in Figure 5D. The contact plot can be best fitted to a combination of three exponential functions. In this analysis, we observed that a significant portion of the native contacts developed very early. About 32% of the native tertiary contacts formed within the first nanosecond with a subnanosecond rate constant, suggesting that the protein contained a substantial number of native tertiary contacts in the denatured state. This observation suggested that the denatured state of Trp-cage protein was very compact. Thus, we would expect that internal friction of the protein could contribute significantly toward the rate of folding. However, only about 48% of the native tertiary contacts were formed by the end of the simulations (100 ns), indicating considerably slower processes in the later stages. The later two phases had rate constants of ∼50 ns and ∼3 μs, respectively, which qualitatively agree with the folding rates based on the backbone RMSD. The ∼3-μs value was also similar to that observed from folding experiments. These data further strengthened our earlier conclusion pointing to the existence of two different folding pathways. It also confirmed that formation of the native tertiary contacts was slower than formation of the α-helix. Despite the small size of the Trp-cage protein, our observation was consistent with the theory proposed by Plaxco and Baker in which they postulated that the free energy barrier of folding was attributable to the chain entropy and could be characterized by the contact order.[14]

Long-range tertiary contacts were identified from the NMR structures by partitioning the protein into two segments, N_{20}−G_{30} and P_{31}−S_{39}, each corresponding to one side of the protein. We counted the residue pairs between the two segments that were separated by at least five residues that had their C_β atoms within 7 Å of each other. Four such pairs were found; they were L_{21}−P_{38}, Y_{22}−P_{38}, W_{25}−P_{38}, and D_{28}−S_{33}. Their time evolution averaged over the simulations was monitored and was fitted to combinations of two exponential functions. The rate constants (Table 3) of the fast phase were 0.31−0.35 ns and were independent of contacting residual pairs. This time scale corresponded to the initial collapse process, suggesting that the initial collapse was nonspecific. In comparison, the second phase showed a clear trend. The rates of L_{21}−P_{38} (∼500 ns) and Y_{22}−P_{38} (∼480 ns) contacts were about an order of magnitude lower than those of W_{25}−P_{38} (∼70 ns) and D_{28}−S_{33} (∼20 ns). Thus, the rates of the slow phase were correlated with the separation of the residues along the primary chain: contacts of protein residues that were further apart in the primary sequence formed more slowly than those of nearer residues.

Trp_{25}: Contacts, Burial, and Side Chain Orientation. The average contacts between Trp_{25} and the rest of the protein were analyzed and are shown in Figure 6. There were extensive contacts between Trp_{25} and the C-terminal residues in the early stages (between 0 and 2 ns), especially Pro_{36}. During this period, contacts from Trp_{25} to Ile_{23} and Lys_{27} were also observed. These contacts were mainly hydrophobic in nature, suggesting that Trp_{25} formed native contacts with proline residues during the initial hydrophobic collapse. The contacts between Gly_{30} and Trp_{25} developed at around 10 ns of the simulation. At around 20 ns of the simulation, Trp_{25} began to form contacts with the

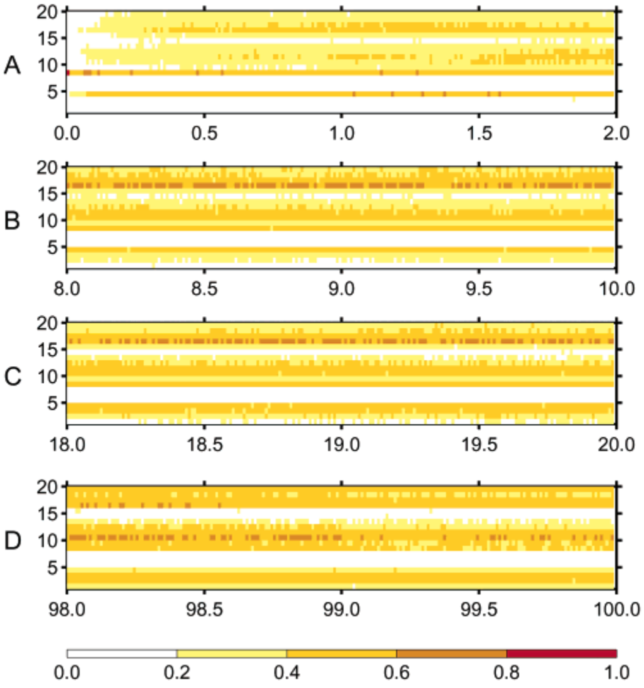

Figure 6. Fractional contacts of Trp_{25} side chain with other residues during the intervals (A) 0.0−2.0, (B) 8.0−10.0, (C) 18.0−20.0, and (D) 98.0−100 ns. Vertical axes represent residue indices (N_{20} is residue one), and horizontal axes represent time in ns.

N-terminal residues during which marginally stable helices also started to form. In the last 2 ns of the simulation, we observed that Trp_{25} side chain formed three major patches of contacts to Leu_{21}/Tyr_{22}, Gly_{29-30}, and Pro_{36-37}. About 50−60% of trajectories contained these contacts. All of these contacts were hydrophobic in nature, suggesting that the hydrophobic contacts bring the important residues together and hydrophobic forces are the driving force for the folding of this protein.

In the native NMR structure, the Trp_{25} side chain is completely buried inside a cage formed by Tyr_{22}, Pro_{31}, Pro_{37}, and Pro_{38}, constituting the main hydrophobic core. Thus, burial of Trp_{25} is intimately linked to the overall folding of the protein. Furthermore, the folding time of 4 μs was obtained from the fluorescence of the Trp_{25} during the thermal unfolding in temperature-jump experiments.[20] Because the fluorescence of Trp_{25} could be affected by both its burial and its close contact with Tyr_{22}, measurements of Trp_{25} burial and the distance between Trp_{25} and Tyr_{22} side chains can serve as a comparison with experiments.

The averaged fractional burial of Trp_{25} side chain over the simulations is shown in Figure 7. We considered Trp_{25} as 100% buried in the NMR structures[9] and 0% buried in the initial extended structure. Overall, the Trp_{25} side chain was ∼50% buried at the end of 100-ns simulations when averaged over 77 simulations. The Trp_{25} burial curve can be best fitted by a combination of three exponential functions with an RMS fitting error of 1.5%. The fastest relaxation was due to an extremely rapid collapse phase that had a time constant of ∼0.5 ns, and the second relaxation time was ∼28.0 ns. The slowest relaxation had a time constant of ∼3 μs. Similarly, the fitting of average center of mass (COM) distance between Tyr_{22} and Trp_{25} side chains also showed that the slowest relaxation had a time constant of ∼3 μs (data not shown). These relaxation time constants were consistent with other measurements discussed earlier (e.g., RMSD, tertiary contacts).

We then examined the Trp_{25} side-chain torsion distribution and compared it with that observed in the high-resolution X-ray structures. Figure 8 shows the distributions obtained from the simulations (red scatters in Figure 8A) and from X-ray structures (black scatters in Figure 8A). The latter was obtained from 6248 available Trp residues found in nonhomologous protein chains of less than 30% homology and better than 2.5-Å resolution in the PDB and the distributions were similar to those found by Dunbrack and co-workers.[33] Notably, χ_1 populated primarily in gauche$^+$ (g^+), trans (t), and gauche$^-$ (g^-) regions and χ_2 were in g^+ and g^- in simulations. These same regions are also well-populated by the high-resolution X-ray structures. The good overlap between the two data sets indicates that the Trp_{25} side-chain orientations observed in the simulations are conformationally allowed and well sampled in other proteins. For comparison, the Trp_{25} torsions in the NMR structures are also shown as blue dots; they are located in one of the six allowed regions, namely, (t, g^+). To a crude approximation, this contributes to the free energy barrier by about 1.0 kcal/mol due to the entropic cost.

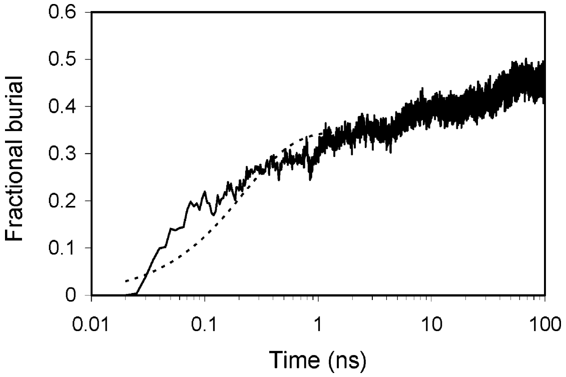

Figure 7. Fractional burial of the Trp side chain plotted on a logarithmic time scale. A 1.4-Å solvent probe radius was used. The multiphasic fitting is also shown. The rate of the slowest phase was ∼3 μs.

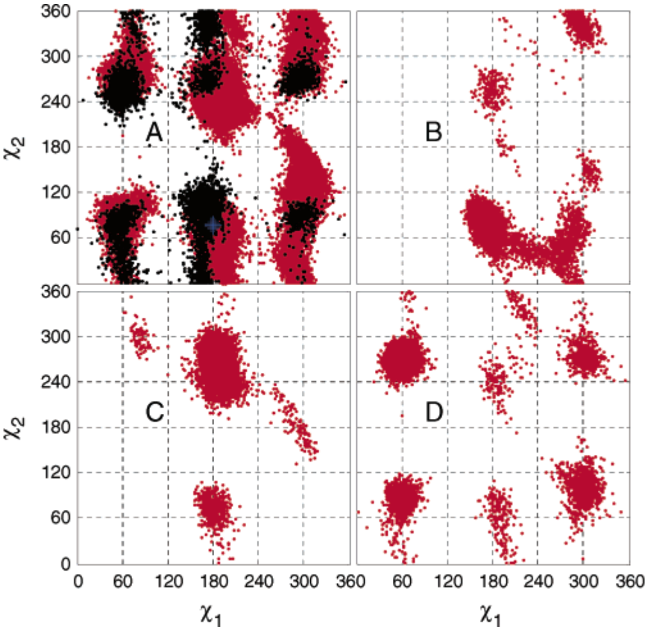

Figure 8. Scatter plots of the χ_1 and χ_2 torsion angles of the Trp_{25} side chain. (A) All saved snapshots in the 77 trajectories (red dots), the 38 NMR models (blue dots), and the 6248 Trp residues in PDB (black dots) chosen from the nonhomologous protein chains of better than 2.5-Å resolution. (B) Trajectory in which the native structure was produced. (C) Trajectory in which Trp side chain was flipped by 180° from the native orientation. (D) Trajectory that produced a random coil structure.

We further analyzed the Trp side-chain torsion distributions in the individual trajectories and found three typical distributions (Figure 8B−D), corresponding to three different scenarios. In the first scenario (Figure 8B), native structures were produced in which the Trp_{25} side chain was in the native (t, g^+) region. In this case, χ_1 and χ_2 fell quickly into the native region, although small populations in other nonnative regions occurred in the initial stage when native main-chain conformation was absent. In the second scenario (Figure 8C), the protein structures were nativelike, but the Trp_{25} side chain was flipped. In this case, most Trp_{25} conformations were distributed in (t, g^-) region, although other regions, including the native region, were also populated, albeit to a much decreased extent. Again, these poorly populated regions were sampled in the early part of the simulations. In the last category (Figure 8D), coiled structures were considered. In this case, the fluctuation was notably higher than in either of the two earlier cases. For example, in this particular trajectory, the majority of the structures accumulated in the (g^+, g^-) and (g^+, g^+) regions and, again, to a much lesser extent, in other regions, including the native region. Transitions in these regions showed that the Trp_{25} side chain is not simply locked randomly into one of the probable regions. Rather, the frequency of transition was linked to the main-chain conformations; larger fluctuations were observed in the random coil structures and partial formation of the native main-chain structures, and the tertiary contacts significantly reduced the level of fluctuation that effectively arrested the Trp_{25} side chain in one of the allowed regions. As a consequence, the protein had to overcome the kinetic barriers if the side chain were to flip to the native conformation, which requires partial unfolding to allow the side chain to change its conformation. This can be an unfavorable step because the compact structure is partially stabilized by the Trp_{25} side chain. Hence, incorrect burial of the Trp_{25} side chain significantly impairs its folding speed, and unfolding of this state is likely to be a major contributor to the rate-limiting step. This is one of the reasons for the relatively low folding rate observed in our GB simulations, despite the lack of solvent viscosity, which otherwise shows a very high folding rate with respect to explicit solvent simulation. Because of the additional entropic cost to bring the side chain to its native conformation, the step leading to the correct packing is even less favorable. Therefore, our simulations suggest that the correct packing of Trp_{25} side chain was the rate-limiting step.

W25F Mutant. The phenylalanine side chain has only one six-membered aromatic ring and is considerably smaller than the tryptophan side chain and is the closest substitute to tryptophan among all 20 naturally occurring amino acids. Although the NMR and CD studies of the W25F mutant showed that its structure is less stable than the wild-type structure,[34] we chose W25F to investigate the effects of a less bulky aromatic side chain that still has the potential to form key contacts including stacking with Pro_{31} (important in stabilizing the 3_{10}-helix and the overall structure). Furthermore, as we have discussed earlier, because of the entropic cost of the asymmetric Trp side chain increases the free energy barrier of native packing by about 1.0 kcal/mol, substituting it with a symmetric aromatic residue allowed us to examine the folding process without this asymmetric entropy cost.

Of the 14 W25F trajectories (100 ns each), 8 were within a 2-Å RMSD of the wild-type native, 7 of which were within less than 1 Å. This was remarkable since considerably fewer wild-type trajectories could fold to the native structure within the same time frame. It clearly suggested that the mutant could easily fold to nativelike main-chain structure. One can extrapolate a folding rate of about 0.2 μs based on simple two-state kinetics, again without considering the reverse process. Caution must be taken, however, because solvent viscosity was completely absent in our model. Thus, the folding rate, when solvent viscosity is considered, should be considerably lower. Nevertheless, our result strongly suggests that the W25F mutant folded much more rapidly than the wild type, despite its poor stability. This implies that the ruggedness of the folding funnel plays an important role and, because W25F mutant is considerably less stable than the wild type (discussed below), stability appears to play a less important role in determining the folding rates.

Interestingly, even though the main-chain RMSD value fell within 1 Å, it did not stay within that region; instead, the RMSD fluctuated considerably, suggesting that the W25F mutant had a lower thermodynamic stability and a reduced kinetic barrier separating the native and nonnative states. At around 50 ns, the population within 2.5-Å of the main-chain RMSD reached 42% but then quickly diminished to 20% by 60 ns. This trend continued until the population decreased to about 10% near 80 ns. Thus, the W25F mutant was much less stable than the wild type.

There were 27 clusters with more than 1000 snapshots. They were classified into five classes by visual inspection (Figure 9A−E). These included (i) structures that highly resemble the native wild-type structure, (ii) nativelike structures, (iii) structures with only α-helical region, (iv) hairpins, and (v) random coils. In the native clusters (i), the side-chain packing closely resembled the native structure (Figure 9A), and the Phe_{25} residue was completely buried, as shown by the solvent accessible surface area (data not shown). Approximately 12.6% of all snapshots fell within this category, which roughly translated to a folding rate of 0.7 μs, much faster than the wild-type rate. Consistently, the 0.7-μs value was close to the formation rates of the α-helix and tertiary contacts in the mutant; both α-helix and tertiary contacts formed at a rate of about 1.0 μs as measured by the slowest phase among three phases (data not shown). Again, a lower rate is expected in reality because of solvent viscosity. Nevertheless, this was much faster than the folding rate of the wild type (2.0−6.0 μs).

In category ii, although the main-chain conformation appeared similar to the native structure including all secondary structures, the side-chain packing was incorrect (Figure 9B). About 23% of all saved structures belonged to this category. Categories iii−v were, respectively, well-formed α-helix with wrong overall conformation (Figure 9C, 24%), nonnative hairpin structures (Figure 9D, 8%), and random coils (Figure 9E). Interestingly, all of these detectable intermediate states of the W25F mutant closely resemble those of the wild-type Trp cage. This suggests that folding of the Trp cage and its W25F mutant most likely proceed through similar pathways and sample a similar set of states.

In summary, the W25F simulations suggested that the mutant can fold much more rapidly than the wild-type protein.

Discussion

We investigated the possibility of fitting the kinetic data with single exponential functions. In all cases, the obtained time constants ranged from 0.1 μs to 0.4 μs, depending on the time intervals used in the fitting. A general trend was that a higher rate could be obtained when fitting was done against shorter simulation data, in agreement with earlier study by Caflisch and co-workers.[37] The faster end of our single-exponential fitting data, which was about 0.1 μs, was close to those found by Zagrovic and Pande in the ultralow-viscosity regime.[38] Fur-

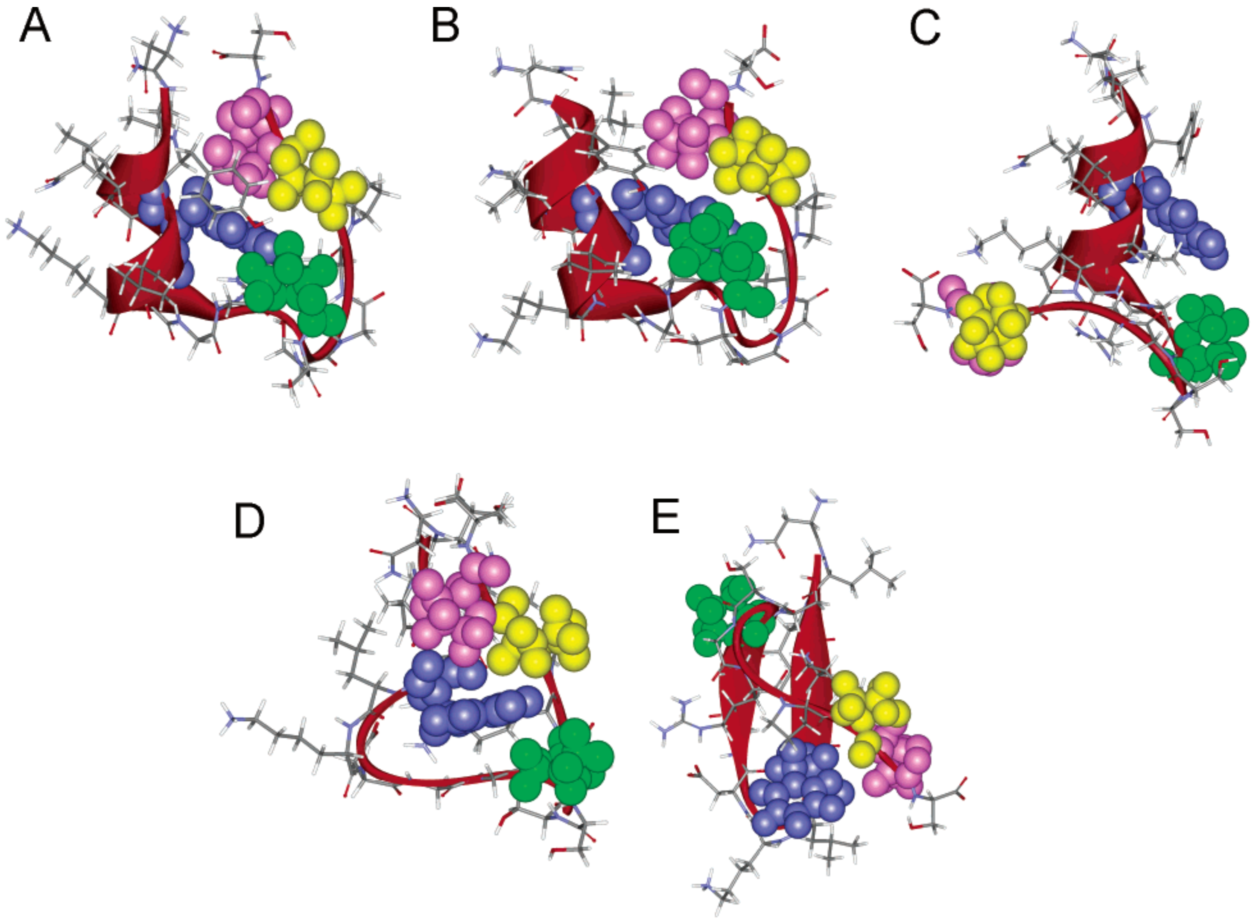

Figure 9. Representative structures of the W25F mutant: (A) native structure, (B) structure in the vicinity of the native structural basin, (C) cluster with only a helical region, (D) clusters with hairpins, (E) random coil structure. Color and side chain models are identical to Figure 3.

thermore, judging from the obvious multiphase behavior, it is not surprising that reliable fitting could not be accomplished by single exponential functions. It is interesting to note that a very recent folding study of Trp-cage protein with low solvent viscosity by Hagen and co-workers[39] demonstrated that internal friction of Trp cage contributed to the folding rate of the protein. Thus, the observed relatively lower folding rate of Trp cage (in comparison to other GB simulations of small peptides/proteins) is not surprising.

Direct application of all-atom MD simulations to study folding processes of proteins began only recently.[16,17] Because of the limitations in computer power, the focus has remained on small proteins, including BBA5[4] and Trp cage.[18,21−23] Recent kinetic (unfolding) experiments on Trp cage suggested a ∼4-μs folding time,[20] making it one of the fastest-folding mini-protein known to date. Because of its small size and high folding rate, the Trp cage presented an interesting case for computer modeling. To date, three different force fields, including two variants of the same force field,[21,23] have been applied to study Trp cage. Several groups, including ours, have reported folding of Trp cage into its native structural basin.[18,21−23] Although detailed information on the folding kinetics and events is yet to emerge, results from these simulations already showed a strong consensus and demonstrated that the modeling methodologies are rapidly converging on the level of accuracy and maturity. With further development of these models, such as the new force field used in this study,[19] we are optimistic that simulations of more topologically challenging proteins will soon be possible.

Among the reported modeling study of the Trp cage, Simmerling et al.[21] simulated this protein prior to the release of the NMR structure. A variant of Cornell et al. force field[41] was used in the study in which the main-chain dihedral parameters were obtained by optimizing against a set of peptide decoys. They observed that the mini-protein fell inside the native structural basin within 10 ns. This was close to the value reported in our earlier study[18] and is in qualitative agreement with the fast-folding rates (30−40 ns) shown in this study. Obviously, it was not possible to observe the slow pathway from a single (short) simulation.

Snow et al.[22] simulated Trp cage using the Folding@Home distributed computing project. Their calculated folding time (based on an RMSD cutoff of 2.5−3.0 Å) was 8.7−1.5 μs and agreed qualitatively with experiments. The detail of the folding process was not reported in this study. Instead, the study focused mainly on the unfolded-state ensemble. Although our analyses focused on the sequence of events leading to the native state (from the fully extended structure), as we noted before, the simulations are expected to produce an ensemble mimicking the unfolded state. Thus, our earlier descriptions on the clusters and snapshots are applicable to the unfolded ensemble. These include the substantial secondary structures, diverse conformations of Trp_{25} side chain, and partial formation of tertiary contacts, consistent with the observations made by Snow et al.[22] In particular, the PPII helix is expected to be stable, and the α-helix and, to a lesser extent, the 3_{10}-helix are expected to be partially formed in the unfolded state.

Pitera et al.[23] studied the folding thermodynamics of Trp-cage motif using the replica-exchange method and the Cornell et al. force field[41] with an updated set of torsion parameters.[42] They showed that the protein folded cooperatively with the all-atom model. Although the calculated melting temperature of

~400 K was notably higher than the experimental value of 310 K, the ability to show cooperative folding using an all-atom model was encouraging. This has significant implications because the smoothness of the folding funnel has been linked to the cooperativity of protein folding.[43] The ability to fold a small protein cooperatively using this type of model suggests that the inherent approximations appear to be acceptable.

Nikiforovich et al.[44] applied a "chain-grow" algorithm and concluded that Trp cage folded through pathways driven by local sequence propensity. This conclusion is consistent with our simulations in which secondary structures and short-range contacts tend to form early.

Packing of the Trp25 Side Chain Is the Rate-Limiting Step. Although the tendency to form short-range contacts might be a general feature of protein folding, as demonstrated by the successes of contact order theory, this, as we suggested earlier,[18] does not imply that the rate-limiting step is to form long-range contacts in a particular protein. Instead, our simulations suggested a quite different picture. Here, the native contacts, as defined by C_β distances, which mainly take into account the chain entropy, can form quite early; the rate-limiting step for the folding of the Trp-cage mini-protein is the packing of the bulky Trp side chain; and formation of the nonnative hydrophobic contacts significantly reduces the overall folding rate. This is understandable as the free energy barriers comprise both entropic and enthalpic contributions. In the case of the Trp cage, the dominant contribution is not the chain entropy, perhaps because of its small size. As we discussed in our previous work,[18] a common factor that is shared by all proteins is the entropic cost of bringing distant residues together. Thus, when a diverse set of proteins is analyzed, it is expected that one should find a correlation between chain entropy and the folding rates. However, for individual proteins, the role of specific interactions can be important and sometimes can even be dominant. This view is complementary to the chain-entropy view and is consistent with many of the recent studies in which diverse folding rates have been observed in proteins and their mutants.[46,47]

Fast and Slow Pathways and Folding Free Energy Barrier. Among the two folding rates obtained from our simulations, the faster rate, being between ~30 and ~50 ns, was about 2 orders of magnitude faster than the slower rate obtained in our simulations. If we take the rates obtained using a 2.0-Å RMSD cutoff, the difference between the two rates was 220 times (Table 1). In the fast pathway, the Trp_{25} side chain was close to its native conformation after the initial collapse, and partial formation of the rest of the protein, including secondary structures and main-chain tertiary contacts, was also completed during the initial collapse. Thus, during the folding, Trp_{25} side chain smoothly adopts its native conformation inside the cage. Hence, this pathway could be viewed as the one without the significant delay caused by the free energy barrier of reorienting the Trp_{25} side chain to its native conformation. In contrast, the majority of the protein has to go through the slow pathway and overcome the free energy barrier to unfold the nonnative Trp cage and to reorient Trp_{25} side chain. Thus, the ratio of these two rates allows us to estimate the folding free energy barrier, which is approximately 3.2 kcal/mol. Among which, about 1.0 kcal/mol is the contribution due to the multiple Trp_{25} side chain conformations. Therefore, the free energy associated with the unfolding of the nativelike nonnative states is about 2.2 kcal/mol. Their unfolding rates are on the order of 0.6~1.0 μs (based on 4.0- and 6.0-μs folding rates, respectively). Given the multiple Trp_{25} conformations, these states could have different stabilities, and the most stable nonnative structure would determine the folding rate of the protein. Given the existence of a well-populated cluster of highly nativelike structure in which the Trp_{25} was in the cage but was flipped by 180° relative to the native conformation, this could be the candidate of the nonnative structure that determined the folding rate of the protein.

In the studies of protein folding mechanisms, it is imperative that one can apply the understanding of the folding process to predict experimental observables. This serves both to validate the theoretical work and to facilitate new studies. In the case of the Trp-cage mini-protein, upon mutation of Trp to a smaller residue, a significant enhancement of the folding rate was observed. Thus, a direct scrutiny to our results is to measure the folding rate of the mutant. One caveat is that this mutant can be unstable. Alternatively, one can consider using a nonnatural amino acid whose side chain is symmetric and whose shape and size are comparable those of Trp. This mutation can reduce the entropic cost to bring the side chain to its native conformation. The third choice is to use a nonnatural amino acid based on the Phe residue by substituting the H_ζ with a small hydrophobic group. Our simulations suggest that all of these mutations can enhance the folding rate and can potentially push the expected 1.0-μs[48] folding rate limit.

Concluding Remarks

Detailed analysis of extensive long-time folding simulations on the Trp-cage mini-protein and its W25F mutant allowed us to explore the folding pathways. The calculated folding time agrees qualitatively with experimental observations. The sequential events of secondary structure formation were observed. The robust PPII helix forms extremely rapidly within the first 1 ns of the simulation and does not depend on the Trp residue. Although the short α-helix starts forming early in the simulation, its formation is only completed by ~1 μs. Both the growth rate of α-helix and the stability of 3_{10}-helix are dependent on the residue at the 25th position. The formation of the native main-chain topology is initiated early in the folding process, which is accelerated by the initial hydrophobic contacts between Trp_{25} and C-terminal prolines. However, packing of the Trp_{25} side chain into the native conformation takes longer. This is the rate-limiting step of folding and is amplified by the presence of the nonnative hydrophobic contacts. Because of its large surface area, Trp also stabilizes the 3_{10}-helix and the overall architecture of the protein by stacking with Pro_{31}. However, because of the asymmetric nature of the indole ring, the Trp side chain conformation can be different. This imposes an entropic cost and delays the formation of the native structure of the protein. The simulations on W25F mutant have confirmed these observations in which the native structures form markedly faster than in the wild type.

Acknowledgment. Computer time was provided by Pittsburgh Supercomputer Center. This work was supported by NIH Research Resource (RR-15588, PI Lenhoff). Financial support was also provided to Y.D. by the state of Delaware, University of Delaware Research Fund, and NIH NIGMS (GM64458 and GM67168). Usage of the VMD, UCSF Midas, and WebLab viewer graphics packages is gratefully acknowledged.

References and Notes

(1) Chan, C. K.; Hofrichter, J.; Eaton, W. A. *Science* **1996**, *274*, 628.

(2) Ballew, R. M.; Sabelko, J.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5759.

(3) Callender, R. H.; Dyer, R. B.; Gilmanshin, R.; Woodruff, W. H. *Annu. Rev. Phys. Chem*. **1998**, *49*, 173.

(4) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102.

(5) McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. *Nat. Struct. Biol.* **1997**, *4*, 180.

(6) Hill, R. B.; Raleigh, D. P.; Lombardi, A.; Degrado, N. F. *Acc. Chem. Res.* **2000**, *33*, 745.

(7) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82.

(8) Struthers, M. D.; Cheng, R. P.; Imperiali, B. *Science* **1996**, *271*, 342.

(9) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425.

(10) Leopold, P. E.; Montal, M.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8721.

(11) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167.

(12) Boczko, E. M.; Brooks, C. L., III. *Science* **1995**, *269*, 393.

(13) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928.

(14) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985.

(15) Weikl, T. R.; Dill, K. A. *J. Mol. Biol.* **2003**, *329*, 585.

(16) Duan, Y.; Wang, L.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9897.

(17) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740.

(18) Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. *J. Mol. Biol.* **2003**, *327*, 711.

(19) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem*. **2003**, *24*, 1999.

(20) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952.

(21) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258.

(22) Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548.

(23) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587.

(24) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; III., T. E. C.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 7*; University of California, San Francisco: San Francisco, CA, 2002.

(25) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489.

(26) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(27) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.

(28) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Comput. Phys.* **1984**, *81*, 3684.

(29) Chowdhury, S.; Zhang, W.; Wu, C.; Xiong, G.; Duan, Y. *Biopolymers* **2003**, *68*, 63.

(30) Bashford, D.; Weaver, D. L.; Karplus, M. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1243.

(31) Karplus, M.; Weaver, D. L. *Protein Sci.* **1994**, *3*, 650.

(32) Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190.

(33) Dunbrack, R. L.; Karplus, M. *Nat. Struct. Biol.* **1994**, *1*, 334.

(34) Barua, B.; Andersen, N. H. *Lett. Pept. Sci.* **2002**, *8*, 221.

(35) Rohl, C. A.; Doig, A. J. *Protein Sci.* **1996**, *5*, 1687.

(36) Thornton, J. M.; Jones, D. T.; Macarthur, M. W.; Orengo, C. M.; Swindells, M. B. *Philos. Trans. R. Soc. London B: Biol.ogical Sci.ences* **1995**, *348*, 71.

(37) Rao, F.; Caflisch, A. *J. Chem. Phys.* **2003**, *119*, 4035.

(38) Zagrovic, B.; Pande, V. S. *J. Comput. Chem.* **2003**, *24*, 1432.

(39) Qiu, L.; Hagen, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 3398.

(40) Berendsen, H. J. C. *Science* **1998**, *282*, 642.

(41) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(42) Wang, J. M.; Kollman, P. A. *J. Comput. Chem.* **2001**, *22*, 1219.

(43) Zhou, Y.; Karplus, M. *Nature* **1999**, *401*, 400.

(44) Nikiforovich, G. V.; Andersen, N. H.; Fesinmeyer, R. M.; Frieden, C. *Proteins* **2003**, *52*, 292.

(45) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1990**, *59*, 631.

(46) Burton, R. E.; Huang, G. S.; Daugherty, M. A.; Calderone, T. L.; Oas, T. G. *Nat. Struct. Biol.* **1997**, *4*, 305.

(47) Myers, J. K.; Oas, T. G. *Nat. Struct. Biol.* **2001**, *8*, 552.

(48) Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11615.