Folding Mechanisms of Proteins with High Sequence Identity but Different Folds[†]

Kathryn A. Scott[‡] and Valerie Daggett*

Department of Medicinal Chemistry, University of Washington, Seattle, Washington 98195−7610

Received September 13, 2006; Revised Manuscript Received November 17, 2006

ABSTRACT: The problem of how a protein folds from a linear chain of amino acids to the three-dimensional structure necessary for function is often investigated using proteins with a low degree of sequence identity that adopt different folds. The design of pairs of proteins with a high degree of sequence identity but different folds offers the opportunity for a complementary study; in two highly similar sequences, which residues are the most important in directing folding to a particular structure? Here we use molecular dynamics simulations to characterize the folding−unfolding pathways of a pair of proteins designed by Bryan and co-workers [Alexander, P. A., et al. (2005) *Biochemistry 44*, 14045−14054; He, Y. N., et al. (2005) *Biochemistry 44*, 14055−14061]. Despite being 59% identical, the two protein sequences fold to two different structures. The first sequence folds to the $\alpha+\beta$ protein G structure and the second to the all-α-helical protein A structure. We show that the final protein structure is determined early along the folding pathway. In folding to the protein G structure, the single α-helix (α1) and the $\beta3-\beta4$ turn fold early. Formation of the hairpin turn essentially prevents folding to helical structure in this region of the protein. This early structure is then consolidated by formation of long-range hydrophobic interactions between α1 and the $\beta3-\beta4$ turn. The protein A sequence differs both in the residues that form the $\beta3-\beta4$ turn and also in many of the residues that form the early hydrophobic interactions in the protein G structure. Instead, in the protein A sequence, a more hierarchical mechanism is observed, with helices folding before many of the tertiary interactions are formed. We find that small, but critical, sequence differences determine the topology of the protein early along the folding pathway, which help to explain the process by which one fold can evolve into another.

The sequence of a protein typically contains all of the information for folding to the functional form (*1*). However, it is clear that not all amino acids are equally important in specifying which fold is adopted. For example, many amino acid sequences can form an α-helix or β-strand depending on the physical conditions or on their context within a protein (*2*−*5*). The question of which features specify a particular fold formed the origin of the Paracelsus challenge set forth by Rose and Creamer (*6*). The challenge was to convert one protein fold to another by changing no more than 50% of the original sequence. Three groups initially took up the Paracelsus challenge, using computational methods and rational design (*7*−*9*). Thornton and co-workers (*7*) chose the all-β antiviral protein BDS-I from *Anemonia sulcata* as a parent sequence (*10*) and the all-α B domain of staphylococcal protein A (protein A) as the target structure (*11*). They designed Paracelsin-43, a 43-residue sequence that is 53.5% identical in sequence to the parent (*7*). Although adopting predominantly random coil structure in water at room temperature, Paracelsin-43 was shown to form helix (though not a stable, globular fold) at low temperatures. Yuan and Clarke (*9*) chose the all-α phage 434 Cro sequence (*12*) as a parent sequence and the $\alpha+\beta$ B1 domain of streptococcal IgG protein G (protein G) as the target structure. The designed protein, Crotein-G, was 50% identical to 434 Cro and 62% identical to protein G; however, it did not fold cooperatively and aggregated at moderate concentrations (*9*). Regan and co-workers (*9*) were the first to meet all of the conditions of the Paracelsus challenge: they chose protein G and the all-α, homodimeric four-helix bundle Rop as their targets. The resulting protein, named Janus, remained 50% identical in sequence to the β-strand domain of protein G, but it folded into a nativelike, stable protein with the Rop fold (*8*). Later variants of Janus that were more identical to the parent sequence were also characterized (*13*).

Bryan, Orban, and co-workers recently applied a different methodology to the problem (*14*, *15*). Using primarily genetic selection, they developed a pair of sequences that were 59% identical, one of which folded to the protein G structure and the other to the protein A structure. They began with the sequences of IgG binding domains from streptococcal protein G (B3 domain) and staphylococcal protein A (a variant of the engineered Z domain). The first step was to introduce the protein A binding epitope into the protein G sequence. Directed evolution and phage display with an IgG binding assay were then used to select for variants of the protein G sequence that were more identical to protein A. Three “topological islands”, each consisting of fewer than 10 residues, were randomized to introduce protein A amino acids into the protein G sequence until no further stable variants

[†] Financial support was provided by the National Institutes of Health (Grant R01 GM 52789).

* To whom correspondence should be addressed. E-mail: daggett@u.washington.edu. Phone: (206) 685-7420. Fax: (206) 685-3252.

[‡] Current address: Structural Bioinformatics and Computational Biochemistry Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX13QU, U.K.

10.1021/bi061904l CCC: $37.00
Published on Web 01/18/2007

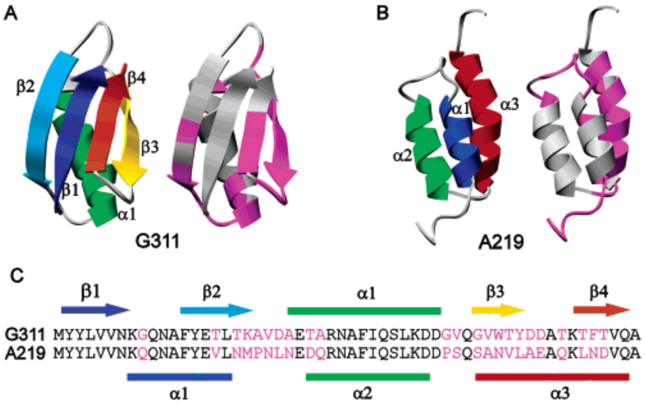
A
β2
β4
β3
β1
α1
G311
B
α1
α3
α2
A219
C
β1
β2
α1
β3
β4
G311 MYYLVVNKGQNAFYETLTKAVDAETARNAFIQSLKDDGVQGVWTYDDATKTFTVQA
A219 MYYLVVNKQQNAFYEVLNMPNLNEDQRNAFIQSLKDDPSQSANVLAEAQKLNDVQA
α1
α2
α3

FIGURE 1: (A) Two ribbon diagrams of G311. The left structure shows secondary structural elements colored from blue at the N-terminus to red at the C-terminus; *β*1 is colored blue, *β*2 light blue, α1 green, *β*3 yellow, and *β*4 red. The right structure shows residues that are identical in each protein colored gray and those that differ magenta. (B) Two ribbon diagrams of A219. The coloring is as for panel A, except α1 is colored blue, α2 green, and α3 red. (C) Sequence alignment for G311 and A219. Identical residues are colored black and those that differ magenta. Secondary structural elements are shown as arrows (*β*-strands) and boxes (α-helix). The protein sequences are 59% identical.

of protein G were formed. At this point, changes were made in the protein A sequence to increase the level of identity between the protein G and protein A sequences. The resulting proteins were shown to be monomeric in solution, to unfold reversibly in temperature denaturation experiments, and to have a stability of 2−3 kcal/mol. The authors designated the protein G variant G311 and the protein A variant A219. The structure of each protein was verified using NMR (Figure 1) (*15*).

The proteins developed by Bryan and co-workers make an attractive model system for a computational protein folding study. Molecular dynamics (MD)[1] simulations can give details of protein−protein and protein−water interactions at atomic resolution. However, despite recent increases in CPU speed, the protein folding time scale (from microseconds to seconds) is typically incompatible with the MD time scale (from nanoseconds to microseconds). One way to circumvent this problem is to study protein unfolding. The unfolding reaction is accelerated at high temperatures and becomes accessible to MD. This approach has been extensively validated, and it has been shown to yield good agreement with experimental protein folding transition states and intermediates for a variety of different proteins (*16*−*21*). Here we use high-temperature, all-atom MD simulations to investigate the unfolding pathways of G311 and A219. The transition state (TS) and denatured state ensemble is characterized for each protein. Considering the unfolding trajectories in reverse, we see significant differences early in folding. The denatured state of A219 exhibits very little residual structure. In contrast, in G311, a number of residues form highly dynamic, fragmented helical structure (or turns), adopting helical (ϕ and ψ) angles but showing rapid fluctuation in hydrogen bonding. In G311, we see early formation of the *β*3−*β*4 turn, followed by consolidation of this interaction with long-range hydrophobic interactions between the *β*3−*β*4 turn and α1. In A219, a much more hierarchical mechanism is seen, with helices forming early in folding and then docking to form the tertiary structure. In both cases, which fold will be adopted appears to be dictated early along the folding pathway.

METHODS

All MD simulations were performed using the program *in lucem* molecular mechanics, *il*mm (*22*). The potential energy function and the protocols for MD have been described previously (*23*−*25*). The first model from Protein Data Bank entries 1zxg and 1zxh was used as the starting structures for A219 and G311, respectively (*15*).

Eleven simulations of A219 were performed: 10 unfolding simulations at 498 K and a native state simulation at 298 K. For G311, 13 simulations were performed. Twelve of the simulations were under unfolding conditions, with two simulations at 498 K and 10 at 448 K. The native state simulation for G311 was carried out at 275 K, the temperature at which the NMR structure was determined. All simulations were carried out at neutral pH.

In preparation for MD, the starting structure was first subjected to 1000 steps of steepest descent minimization in vacuo. Following minimization, the protein was solvated using water pre-equilibrated at the appropriate temperature such that the water box extended at least 10 Å from the protein for high-temperature simulations and 12 Å for 298 K simulations. The final target densities were 1.000 g/mL at 275 K, 0.997 g/mL at 298 K, 0.890 g/mL at 448 K, and 0.829 g/mL at 498 K (*26*, *27*). The water was subjected to 500 steps of steepest descent minimization, followed by 1000 steps of molecular dynamics. The water was then minimized again for 500 steps. Finally, the entire system was minimized for 500 steps. After these preparatory steps, the system was heated to the desired temperature with initial velocities assigned from a Maxwellian distribution, and the NVE microcanonical ensemble was used. The velocities of the atoms were adjusted until the system reached the desired temperature. Thereafter, atoms were allowed to move according to Newton's equations of motion with a 2 fs integration time period. For the simulation at 298 K, a 12 Å nonbonded interaction cutoff was used; for all other simulations, an 8 Å nonbonded interaction cutoff was used. In all cases, the nonbonded atom list was updated every 3 steps. Structures were saved for analysis every 1 ps.

Transition states were identified using conformational clustering based on the Cα root-mean-square deviation (Cα-rmsd) between structures (*28*). An all-by-all Cα-rmsd matrix was generated for the first 2 ns of each trajectory. Sammon's multidimensional scaling (*29*) was then used to reduce this matrix to three dimensions. The exits from the clusters were identified visually. Structures from the last 5 ps of the first cluster in each unfolding trajectory were used to characterize the TS.

RESULTS

Native State Simulations

In G311, the secondary structure is defined as follows: *β*1, residues 1−8; *β*2, residues 13−19; α1, residues 23−37; *β*3, residues 42−46; and *β*4, residues 51−56. In A129, α1 is residues 9−17, α2 residues 26−36, and α3 residues 40−

[1] Abbreviations: MD, molecular dynamics; TS, transition state; D, denatured state; TSE, transition state ensemble.

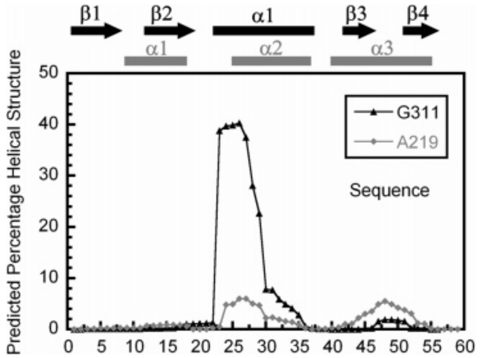

FIGURE 2: Predicted intrinsic helical propensities for the G311 (red) and A219 (black) sequences at 298 K. The values were predicted using AGADIR (*32*). The substitution of DG in A219 for TA in G311 is predicted to give rise to an ∼35% increase in intrinsic helical propensity in residues 23−28.

55. The native states of G311 and A219 were simulated for 21 ns at 275 and 298 K, respectively. In both cases, the simulation temperature matches that at which the NMR structures were determined. If the first nanosecond of the simulation for equilibration is excluded, the mean Cα-rmsd over the native trajectory is 2.2 Å for G311 and 2.8 Å for A219. The N- and C-termini of A219 are disordered in the NMR ensemble, and the Cα-rmsd excluding these regions is 2.3 Å.

Intrinsic Helical Propensity

Small changes in sequence can lead to significant changes in intrinsic secondary structural propensities. AGADIR (*30*, *31*) was used to assess the propensity for helix formation in the unfolded states of G311 and A219 (Figure 2). Despite having only helical structure in its native state, the sequence of A219 is predicted to have a low intrinsic helical propensity. Residues 23−28 in α2 and 46−52 in α3 are predicted to have a helical propensity of ∼5%. Where the sequences of α1 in G311 and α2 in A219 overlap, they are identical at all but two positions. The differences occur at positions 25 and 26, where DG in A219 is replaced with TA in G311. This small change in sequence is predicted to result in an ∼35% increase in helical propensity for residues 23−28 in G311 compared to A219. The change in the intrinsic propensity might be expected to lead to differences in the amount of residual helical structure present in the denatured state.

Unfolding Simulations

G311. Initially, two 21 ns, 498 K unfolding simulations, designated 498_1 and 498_2, were performed. Preliminary analysis showed that under these conditions the protein unfolded very rapidly such that the transition state (TS) was passed on the same time scale as heating. As a result, 10 further simulations were carried out at 448 K, two for 21 ns, 448_1 and 448_2, and eight for 2 ns, 448_3−448_10 (*18*). The four longer simulations all reached a Cα-rmsd of ∼12 Å from the starting structure by 5 ns and stayed at that level for the rest of the simulation. The denatured state (D) structures from the 448 and 498 K simulations are very similar, in terms of properties such as the number of contacts, the percentage and location of secondary structure, and the solvent accessible surface area (Table 1). As a result of the similarities between the properties of the denatured ensemble at 498 and 448 K, the conclusions about the folding pathway would be essentially the same regardless of which trajectories were analyzed in detail. Only 448 K trajectories are considered further here. Snapshots from simulation 448_2 are shown in Figure 3. The time points corresponding to the first TS were determined for the ten 448 K trajectories using Cα-rmsd clustering (see Methods). Transition states were identified at 387 ps in 448_1, 299 ps in 448_2, 157 ps in 448_3, 150 ps in 448_4, 225 ps in 448_5, 140 ps in 448_6, 270 ps in 448_7, 246 ps in 448_8, 150 ps in 448_9, and 235 ps in 448_10.

A219. Two 21 ns 498 K simulations, 498_1 and 498_2, and eight 2 ns 498 K simulations, 498_3−498_10, were performed. The two longer simulations reach a Cα-rmsd of 10 Å by ∼2 ns, followed by an increase to ∼20 Å after ∼10 ns. Snapshots of structures from trajectory 498_1 are shown (Figure 3). Transition states were identified as for G311. Two of the trajectories (498_3 and 498_10) did not cluster well enough for a transition state to be determined accurately. Transition states were identified in the other eight trajectories, at 235 ps in 498_1, 890 ps in 498_2, 120 ps in 498_4, 350 ps in 498_5, 260 ps in 498_6, 325 ps in 498_7, 400 ps in 498_8, and 450 ps in 498_9.

Transition States

G311. Representative structures from the TSE for each trajectory are shown in Figure 4 and average contact maps over the TSE in Figure 5. Mean properties over all members

Table 1: Mean Simulation Properties[a]

	Cα-rmsd (Å)	no. of contacts	no. of native contacts	no. of non-native contacts	no. of H-bonds	SASA (Å^2)	α(Φ,Ψ) (%)	*β*(Φ,Ψ) (%)
				G311				
N[b]	2.2 (0.1)	180 (2)	–	–	38 (3)	3920 (80)	33 (3)	41 (3)
TS[c]	4.2 (0.4)	146 (7)	121 (6)	25 (4)	32 (4)	4900 (200)	32 (4)	40 (4)
D[d] at 448 K	11.9 (0.5)	160 (8)	80 (7)	80 (8)	40 (4)	4600 (200)	45 (9)	16 (5)
D[e] at 448 K	11.9 (0.5)	162 (7)	86 (7)	76 (7)	42 (4)	4600 (200)	52 (5)	13 (3)
				A219				
N[b]	2.8 (0.1)	197 (3)	–	–	39 (3)	4340 (77)	62 (3)	10 (2)
TS[c]	5.9 (1.0)	152 (10)	121 (10)	31 (8)	33 (6)	5600 (260)	51 (8)	19 (5)
D[e]	16 (3)	120 (16)	84 (9)	36 (9)	27 (6)	6200 (500)	25 (9)	33 (6)

[a] The standard deviations are given in parentheses. [b] Native state properties are the mean over 1−21 ns of the 275 K simulation for G311 and the 298 K simulation for A219. [c] Transition state properties are the mean over all structures, six from each trajectory corresponding to a 5 ps window before the cluster exit. [d] G311 denatured state properties as the mean over 10−20 ns of trajectories 448_1 and 448_2. [e] Denatured state properties as the mean over 10−20 ns of trajectories 498_1 and 498_2.

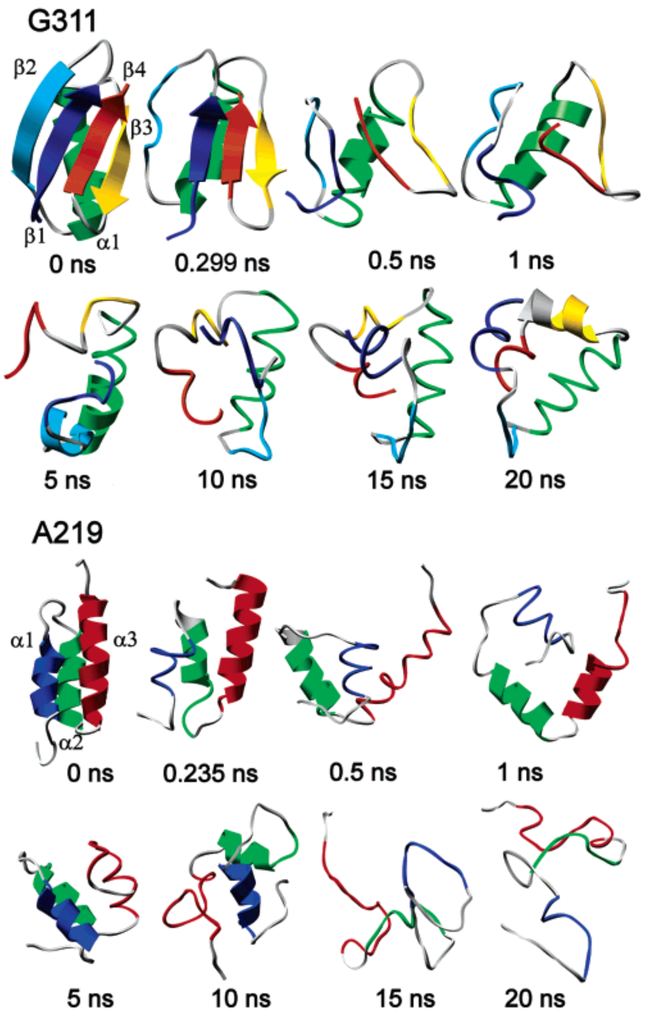

FIGURE 3: Snapshots from representative unfolding trajectories for G311 and A219. The proteins are colored as described in the legend of Figure 1. The second snapshot in each sequence represents the TS for that trajectory. In the top section (G311) are snapshots from trajectory 448_2. In all G311 trajectories, the core expands and the β-strands first become disordered while the helix remains well-folded. Late in the unfolding trajectories, residues in parts of the β1, β3, and β4 regions adopt helical ϕ and ψ angles. The β3−β4 turn remains in a nativelike conformation for the majority of the trajectory. In the bottom section (A219) are snapshots from trajectory 498_1. Early in the A219 trajectories, the core of the protein expands while the helices remain fairly well ordered, and after the TS, there are few persistent interhelical contacts. In both 21 ns 498 K simulations, there is some helical structure in α1−α3 for approximately the first 10 ns. After this point, the structures are highly disordered with few regular secondary structural elements.

of the ensemble are given in Table 1. The TSE is quite nativelike in terms of secondary structure but is expanded relative to the native state with a reduction in the number of contacts between α1 and the β-sheet. α1 is the most nativelike secondary structural element. In eight of the 10 trajectories, α1 shows a contiguous run of hydrogen bonds along the whole length for the majority of the TSE, while in the other two trajectories, α1 is kinked in the middle. In all simulations, the hydrogen bonding fluctuates between α- and 3_{10}-helix. For each native hydrogen bond, an $i \rightarrow i + 3$ interaction is present in ∼10−15% of the TSE and an $i \rightarrow i + 4$ interaction in ∼70−80%. In some structures, both hydrogen bonds are weakly formed at the same time. The β-structure exhibits greater heterogeneity across the TSE and can be described in terms of ϕ and ψ angles, hydrogen bonds, and contacts between strands. Residues 3−6 in β1 and 42−46 in β3 occupy beta ϕ and ψ space in ∼90% of the TSE, while residues 51−54 in β4 occupy beta ϕ and ψ space in ∼80% of the TSE. In contrast, β2 has no contiguous residues with high occupancy of beta ϕ and ψ angles. Residues 12, 14, 16, and 18 have a beta conformation in 70−90% of the TSEs and residues 13, 15, 17, and 19 in ∼20% of the TSE. Looking at hydrogen bonds and contacts gives a slightly different picture. There is hydrogen bonding between β1 and β2 in 15−30% of the TSE, with the hydrogen bond occupancy highest between residues N7 and Y14 (∼30%). Residues L4 and V6 in β1 form hydrogen bonds with F52 and V54 in β4 in ∼65% of the TSE, while the other three hydrogen bonds show ∼40% occupancy. In the β3−β4 hairpin, the first two hydrogen bonds between residues V42 and Q55 are formed in less than ∼20% of the TSE. The other four hydrogen bonds, between residues T44 and D46 in β3 and T51 and T53 in β4, are formed in ∼50% of structures. Side chain interactions between the strands are, however, sometimes maintained where hydrogen bonding is lost. The β1−β2 hairpin loses many side chain contacts, and only N7−Y14 and L4−L17 are present in greater than 50% of the TSE. The β1−β4 interface is better formed with four contacts present in >80% of the structures and seven contacts present in 60−80% of the structures. Similarly in the β3−β4 hairpin, three native contacts are present in >80% of the TSE and six contacts in 60−80% of the structures.

A219. Snapshots of a representative structure from the TSE from each trajectory are shown in Figure 4 and average contact maps over the TSE in Figure 5. Mean properties over all members of the ensemble are given in Table 1. The contact maps show that the majority of native interhelical side chain contacts are lost in the TSE. In terms of overall topology, the helices are no longer parallel now that the core has opened up. Some non-native interactions are formed; these appear to be nonspecific, with many different contacts formed in a low percentage of the ensemble. On average, any particular side chain contact is formed in <20% of the TSE. Specifically, the α1−α2 interface has three and the α1−α3 interface two contacts formed in 30−40% of the structures. The α2−α3 interface is slightly better formed with two contacts present in >50% of the structures and four in 30−50% of the structures. The protein is much more nativelike in terms of secondary structure. The characteristic helical $i \rightarrow i + 3$ and $i \rightarrow i + 4$ contacts are well-formed, especially toward the C-terminus of α2 and in α3 where they are present in at least 80% of the TSE. All helices show both α- and 3_{10}-helix hydrogen bonding patterns. At each position in α1, the $i \rightarrow i + 4$ hydrogen bond is present in ∼20−25% of the structures and the $i \rightarrow i + 3$ hydrogen bond in ∼20%. α2 frays at its termini; however, residues 28−36 form native $i \rightarrow i + 4$ hydrogen bonds in ∼70% of the TSE. α3 also formed well, showing native $i \rightarrow i + 4$ hydrogen bonds in ∼60% of the TSE and $i \rightarrow i + 3$ hydrogen bonds in ∼20%.

Denatured State

G311. Mean denatured state properties at both 448 and 498 K are given in Table 1. The properties and denatured state structures are very similar at 448 and 498 K, and only the 448 K trajectories are analyzed further. In this discussion,

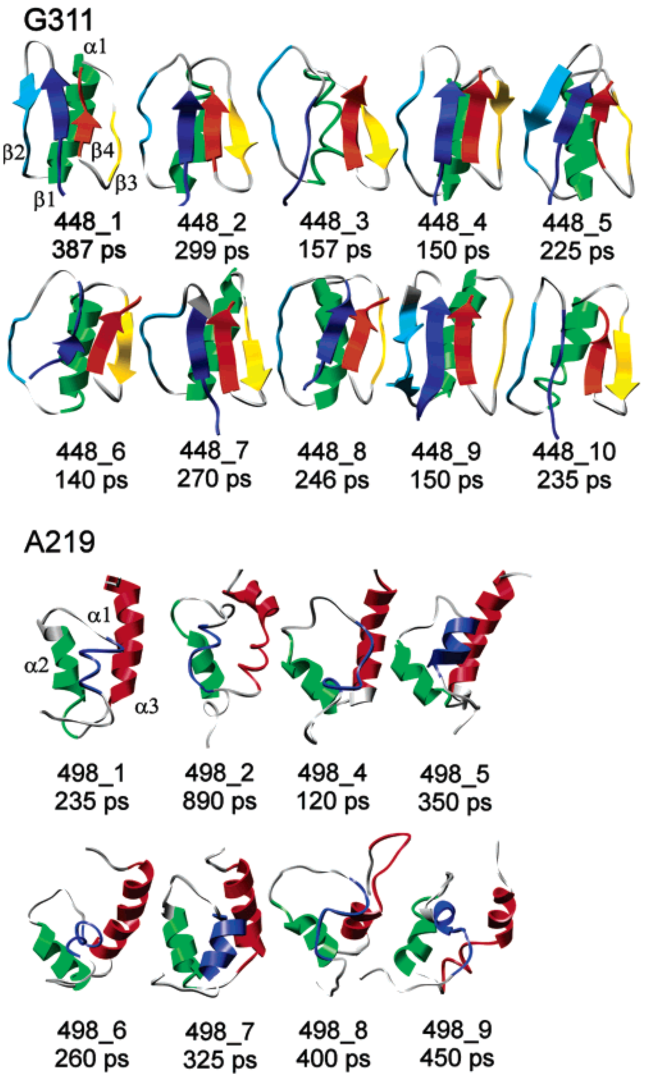

FIGURE 4: Snapshots for representative transition state structures from each of the unfolding trajectories (excluding the two A219 trajectories where the time point transition state could not be determined). In the top section (G311), the α-helix is well-formed and the core is expanded in all TS structures. The β-structure is more heterogeneous with native hydrogen bonds and ϕ and ψ angles in different strands in different simulations. In the majority of the trajectories, contacts between α1 and β2 are lost and those between α1 and the β1−β3−β4 sheet are consolidated early in unfolding. Interactions among β1, β3, and β4 are then well-formed in the transition state. The early unfolding events in trajectories 448_1, 448_5, and 448_9 are different; here contacts between α1 and β3 are lost early and those between α1 and the β1−β2−β4 sheet consolidated. In these transition state structures, there are interactions among β1, β2, and β4. In the bottom section (A219), in all transition state structures the core is much expanded, with <30% of the native contacts remaining. α2 and α3 are relatively well formed, and there are native contacts between the C-terminus of α2 and N-terminus of α3 in the majority of simulations. α1 is more heterogeneous, in terms of both secondary and tertiary contacts. The lower-contact order α1−α2 interface shows more contacts than the α1−α3 interface in most of the trajectories.

the denatured state properties are an average over 10−20 ns in trajectories 448_1 and 448_2. Contact maps for the denatured state ensemble are shown in Figure 5. The denatured state structures are somewhat surprising, showing adoption of helical ϕ and ψ angles in four different regions, three of which do not form helical structure in the denatured state. Residues forming α1 in the native state adopt helical ϕ and ψ angles throughout simulation 448_2, while in simulation 448_1, residues 23−28 in α1 become disordered after 10 ns. In addition, residues 1−9 adopt helical ϕ and ψ angles for much of the denatured state in simulations 448_1 and 448_2, while residues 42−46 and 52−56 adopt helical ϕ and ψ angles in simulation 448_2 only. The protein also shows some $i \rightarrow i + 3$ and $i \rightarrow i + 4$ contacts, particularly for residues 1−9 where they are present in >60% of the denatured ensemble (Figure 5). It is important to note that, although helical ϕ and ψ angles and $i \rightarrow i + 3$ or 4 contacts are seen for these residues, the structures are highly dynamic. Thus, while helical ϕ and ψ angles are adopted on the nanosecond time scale, the hydrogen bonding patterns switch among 3_{10}-, α-, or π-helical and no hydrogen bonding patterns on a much faster time scale. Further, the contacts described above are on a per-residue basis, with atom−atom contacts changing on a faster time scale. The dynamic nature of the helical structure is illustrated by comparison of the pattern of $i \rightarrow i + 4$ α-helical hydrogen bonds in the native state and in the denatured state (Figure 6). In this analysis, we require at least six contiguous hydrogen bonds for a segment to be considered helical. In the native state, 94% of the time points have at least one contiguous segment with six or more hydrogen bonds. In contrast, in the denatured state, ∼25% of time points have at least one contiguous segment with six or more hydrogen bonds. The distribution of lengths of hydrogen-bonded segments also differs between the two states, with the native state showing much longer segments of α-helical $i \rightarrow i + 4$ hydrogen bonds. Thus, local, unstable helical structure is adopted, not well-ordered, long-lived helices.

There are very few long-lived contacts between side chains in the denatured state. Typical contacts are present in only 10−20% of the denatured ensemble. A very small number of short-range side chain contacts (greater than $i \rightarrow i + 1$ but less than $i \rightarrow i + 5$) are present in more than 80% of the denatured ensemble. Persistent native contacts are the D46−T49 contact in β3 and the β3−β4 turn (94%) and the T49−T51 contact in the β3−β4 turn and β4 (99%), while the non-native contacts are the D46−F52 and D46−T53 contacts between β3 and β4. A small number of long-range interactions are present in more than 50% of the denatured ensemble. These include the native F30−F53 interaction between α1 and β4 and the non-native Y3−T53 interaction between β1 and β4 and the non-native E15−R27 interaction between β2 and α1.

The β1−β2 and β3−β4 hairpins behave differently in the denatured state. The β1−β2 turn is comprised of G9, Q10, N11, and A12 and the β3−β4 turn D47, A48, T49, and K50. In the native state, both turns have a stable conformation. The turns also adopt a nativelike conformation in the TSE. However, a few 100 ps after the TS, the β1−β2 turn becomes disordered. In contrast, the β3−β4 turn is much less mobile in the high-temperature simulations, with residues 48−50 in particular adopting ϕ and ψ angles close to those seen in the native structure.

A219. Denatured state properties are an average over 10−20 ns in trajectories 498_1 and 498_2. Mean properties are given in Table 1 and contact maps in Figure 5. Unlike that of G311, there is little residual structure in the denatured ensemble of A219. After the transition state, α2 and α3 remain relatively well formed for ∼10 ns in both trajectories,

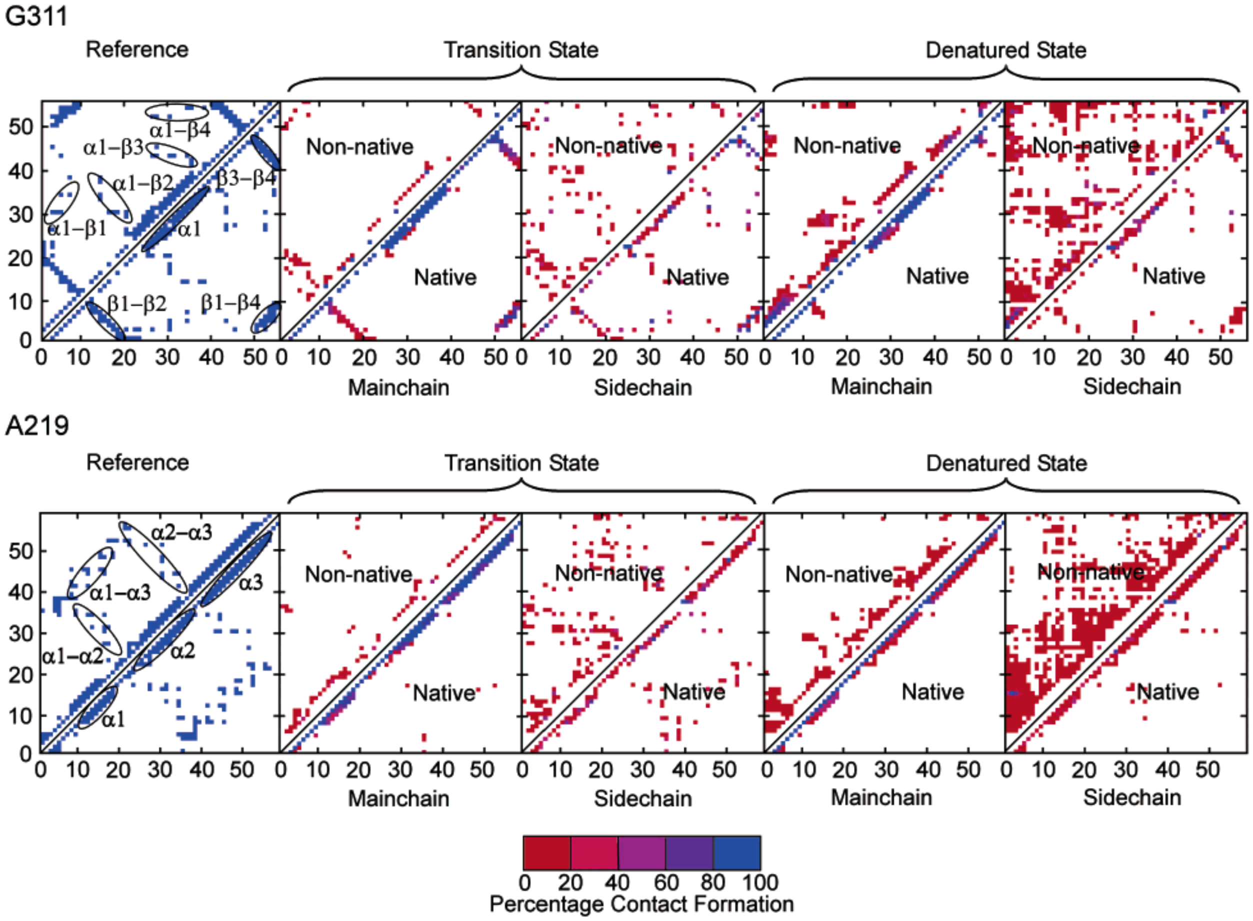

FIGURE 5: Contact maps for G311 and A219. Contacts are colored from red, if present in at least 0.05% of the ensemble, to blue, if present in 100% of the ensemble. The contact map of the starting structure is shown as a reference. Transition state contact maps show an average over all structures in the transition state ensemble, six structures from each trajectory where a transition state could be identified. Denatured state contact maps are an average over 10−20 ns in 448_1 and 448_2 for G311 and over 10−20 ns in 498_1 and 498_2 for A219.

while α1 is only present in 498_1. After this time point, all helices become unfolded in both trajectories. In the denatured ensemble, one to two turns of helical structure are formed at various times, on a time scale of hundreds of picoseconds, in particular over residues 11−17 in α1 and 29−35 in α2. There are no persistent, long-range native interactions, and the only high-occupancy long-range non-native interactions are salt bridges, in particular between Y2 and E15 and between Y3 and E15.

DISCUSSION

G311 Unfolding Pathway. The G311 unfolding trajectories can be placed into two categories on the basis of early events and the TS structure. The more common pathway (seven of 10 trajectories) exemplified by 448_2 is described in detail here (Figures 7 and 8). The first event in unfolding is the opening and repacking of the hydrophobic core. The *β*1−*β*2 hairpin dissociates, and at the same time *β*2 loses contact with α1 except near the *β*2−*β*2 loop. The *β*3−*β*4 turn end of the *β*3−*β*4 hairpin then moves away from α1 as the *β*1−*β*3−*β*4 sheet and α1 rotate relative to each other so that they are almost parallel and the core rearranges. Y3 and V5 in *β*1, F30 and I34 in α1, W43 and Y45 in *β*3, and F52 and V54 in *β*4 are all buried in the core of the native state (<10% SASA). These residues retain long-range interactions following rearrangement, though the specific atom−atom interactions are non-native. Additional interactions with I31 are also seen. A subset of the interactions seen after rearrangement is present in the TS (different interactions in different trajectories). In 448_2, V5 (*β*1) interacts with F30 (α1), F52 (*β*4), and V54 (*β*4), W43 (*β*3) interacts with I31 (α1), F52 (*β*4), and V54 (*β*4), and L34 (α1) interacts with V54 (*β*4). Following the TS, there is further disruption of the core interactions and the *β*-strands rotate ~45° relative to α1. A small number of α1−*β*-strand interactions are still retained; in particular, F30 (α1)−V4 (*β*1) and −V54 (*β*4) and I31−W43 (*β*3) are retained until 2.5 ns. At the same time, as the strands change orientation they begin to dissociate, with hydrogen bonds being lost before side chain interactions.

Interestingly, once the *β*1−*β*4 interactions are no longer present, we see some re-formation of the *β*3−*β*4 hairpin in the correct register, and at ~1.8 and 2.5−3 ns, native hydrogen bonds form between the strands (Figure 9). The *β*3−*β*4 turn retains nativelike ϕ and ψ angles throughout the trajectory (most likely because D47 and K50 interact either with each other or with backbone carbonyls). Before ~10 ns, the conformation of the turn would appear to promote contacts between *β*3 and *β*4. Later in unfolding, the C-terminus and regions of *β*1 and *β*3 adopt helical ϕ and ψ angles, forming highly dynamic fragmented helical structure with rapid interconversion of 3_{10}-, α-, or π-helical and no sustained hydrogen bonding patterns.

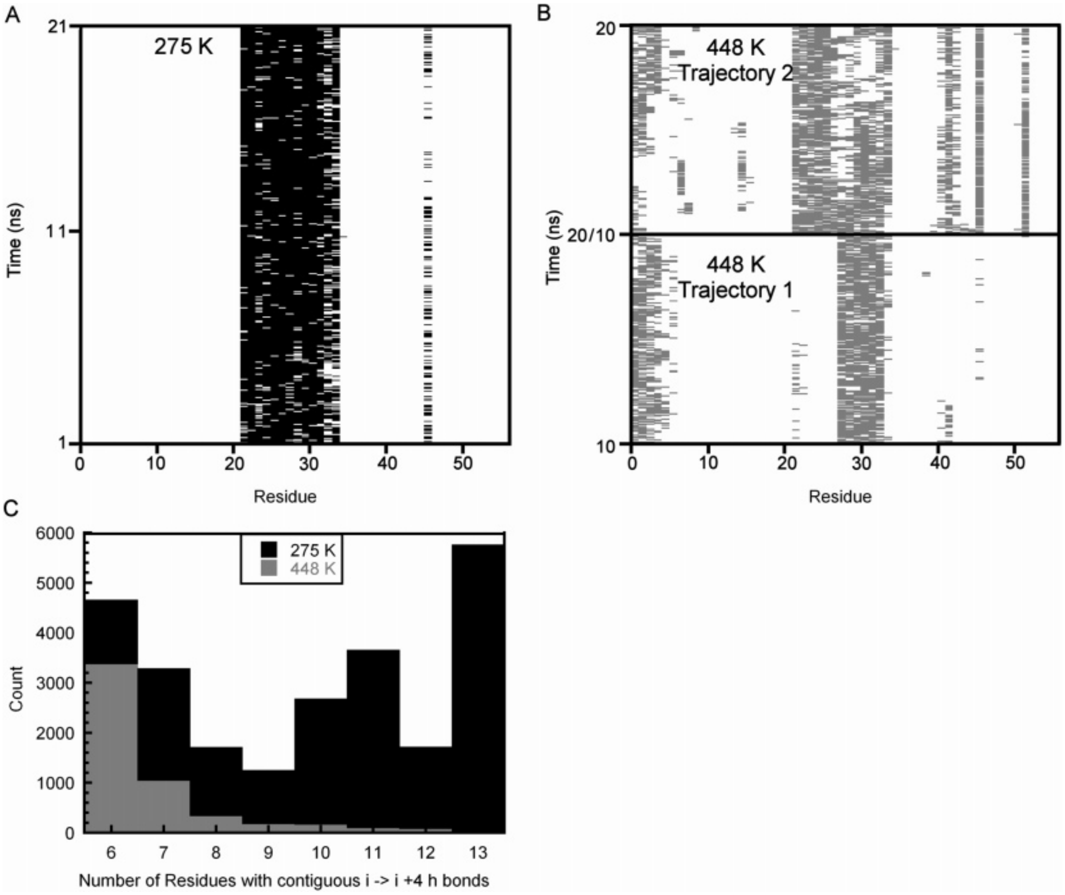

FIGURE 6: Hydrogen bonding in G311. (A and B) Plots showing how the presence of $i \rightarrow i + 4$ hydrogen bonds varies with time. A colored bar is shown for the *i*th residue at each point when a hydrogen bond is present. Panel A shows the native simulation at 275 K and panel B the denatured state defined as 10−20 ns of simulations 448_1 and 448_2. (C) Histogram showing the lengths of contiguous $i \rightarrow i + 4$ hydrogen-bonded segments for the native and denatured states of G311. Only segments having at least six contiguous hydrogen bonds are considered. In each case, the maximum length of the hydrogen bond is 2.6 Å and the maximum angular variation is ±35° from linearity.

In three of the trajectories, the β3−β4 hairpin loses hydrogen bonds and some side chain interactions early in unfolding. The remaining β1−β2−β4 sheet moves parallel to α1, and the core rearranges. In this case, the rearrangement is less marked and fewer contacts are lost between α1 and β3 than between α1 and β2 in the other trajectories. The network of interactions includes residues Y14, T16, and T18 from β2 to a greater extent than the more common pathway. Following the TS, the events are then very similar to those described above. The core expands further, and the strands rotate relative to α1, with F52 (β4) interacting with F30 (α1). After the β1−β4 interaction is lost, the β3−β4 hairpin reforms, much as in trajectory 448_2.

The circular dichroism (CD) signal of G311 is observed experimentally to decrease with an increase in temperature as the protein unfolds. We do not believe this observation to be inconsistent with the results of the simulations. The helical structure seen in our simulations is irregular and highly dynamic in nature, with few long contiguous hydrogen-bonded segments. It has been demonstrated that, although having a larger ellipticity per helical unit than originally believed, short helical segments have a lower ellipticity per helical unit than longer helices and show a shift in the negative maximum at 222 toward shorter wavelengths (*32*, *33*). CD signals appear to be very sensitive to helix length and geometry, needing many successive helical residues and well-formed hydrogen bonds for a good signal (*32*).

Comparison with experimental studies on the full-length wild-type protein G and on the β3−4 hairpin in isolation suggests that the behavior of G311 is similar to that of the wild-type protein (*34*−*39*). Orban and co-workers characterized the acid-unfolded state of the protein G F30H mutant. They showed that the acid-denatured state populates native residual structure, defined as any deviation from random coil ϕ and ψ space and likely to be weakly populated, in regions corresponding to the β1−β2 turn, α1, and the β3−β4 hairpin. In particular, the residues comprising the β3−β4 hairpin show significant line broadening in the HSQC spectrum, showing that this region of the protein is in slow exchange with other conformers in the denatured ensemble. It is also interesting to note that a number of non-native turnlike interactions are also present in regions corresponding to β1, β2, and the α1−β3 loop. That the β3−β4 hairpin is structured early in folding is also consistent with the observation that the 16-residue β3−β4 hairpin of protein G is stable in isolation. The sequence of β3−β4 was used in the first investigation of the structure, thermodynamics, and kinetics of β-hairpin folding. It has been shown to exhibit two-state behavior in kinetic and equilibrium studies and to fold on the microsecond time scale (*35*−*37*).

Baker and co-workers have characterized the transition states of wild-type protein G and the structurally similar IgG binding domain of peptostreptococcal protein L (*38*, *39*). Protein engineering Φ value analysis of protein G shows Φ

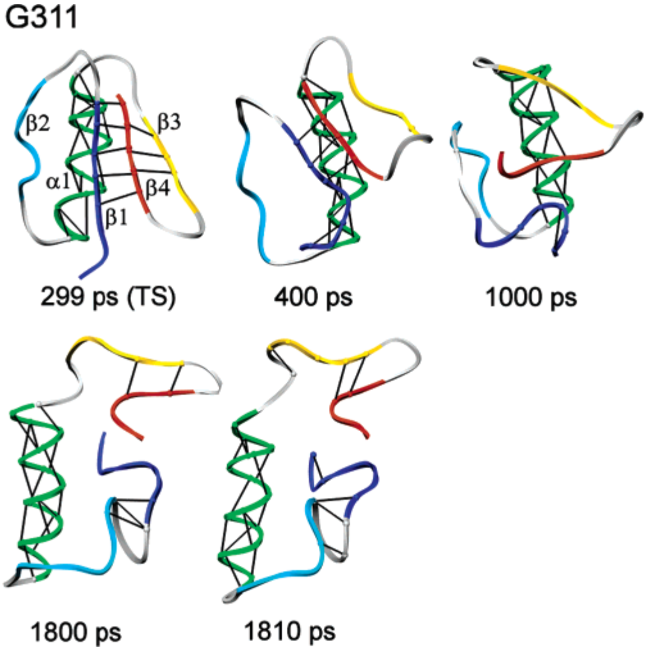

FIGURE 7: Snapshots illustrating hydrogen bond formation and the relative orientation of α1 and the β-strands in trajectory 448_2 of G311. In the TS, this trajectory shows hydrogen bonding between β1 and β4 and between β3 and β4, and α1 is roughly parallel to the β1−β4−β3 sheet. Within several hundred picoseconds of the TS, the core of the protein loosens, β3 dissociates, and the strands rotate relative to α1 (400 ps). β1 and β4 then dissociate (1000 ps). The β3−β4 hairpin re-forms, though the structure is highly dynamic.

values close to 1 in the β3−β4 turn, moderate Φ values in strands β1, β3, and β4, and low Φ values in β2 and α1. This suggests a mechanism in which the β3−β4 turn is largely formed and the β1−β3−β4 sheet partially formed in the transition state, consolidation of the β1−β2 hairpin structure, and much of the α-helix formation then occurs after the rate-limiting step. In seven of the 10 G311 simulations, the structure of the β-strands in the transition state is consistent with the experimental data for the wild-type protein. In the other three simulations, the structure is more like that seen for protein L, where the first β-hairpin is well-formed in the transition state. That β3−β4 hairpin should exhibit similarity with the wild type is not unexpected, given than 14 of the 16 residues are identical in the two proteins. The most significant difference between the simulated transition states of G311 and the experimentally observed behavior of the wild-type protein occurs in α1. In the simulation, this helix becomes structured early in folding, while the experimental Φ values suggest that much of the helix formation occurs after the rate-determining step. Again, this behavior might be expected on the basis of sequence comparison between G311 and the wild-type protein; α1 differs at a number of different residues, and the sequence of α1 in G311 is predicted, using AGADIR, to have approximately double the intrinsic helical propensity of the wild-type protein (*30, 31*).

A219 Unfolding Pathway. The A219 unfolding pathway is quite similar in the different trajectories; herein, we describe the unfolding pathway of trajectory 498_1 (Figure 8). There is expansion of the core early in unfolding. In the majority of the trajectories, including 498_1, this occurs simultaneously with fraying at the C-terminus of α2. α2 and α3 slide relative to each other such that F30, which is intercalated with V44 and A48 in the native structure, now lies below V44. α1 remains in native register with α2 and is therefore out of register with α3. Prior to the TS, residues F13 and L17 in α1, F30 and L34 in α2, and V44 and V45 in α3 form long-range interactions between helices. In the TS, a small subset of these contacts formed, with different simulations typically showing different contacts. In trajectory 498_1, α2 and α3 are well-formed in the TS and interact around the α2−α3 loop, where L45 interacts with S30 and S41 with S33 and L34. In the TS, α1 and α2 are rotated relative to each other when compared with the native state, and the primary interactions are F13−F30 and Q10−F30 contacts. After the TS, the core opens further as the helices move away from each other. By 500 ps, the only side chain contacts with order of >5 (5 or more intervening residues) occur near the α1−α2 (F13−L22) and α2−α3 regions (L34−S41 and S33−S41). At 2.5 ns, α1 breaks all contact with α2, and at 3 ns, the C-terminal turn of α3 unfolds. The small cluster of interactions around the α2−α3 loop persists until 7.5 ns into the simulation, where the S33−S41 contact is broken. Helices α2 and α3 fully unfold at ∼10 ns, and α1 unfolds at 12.5 ns. Small regions of helical structure form throughout the denatured state, particularly in residues 30−40. The unfolding pathway of A219 differs somewhat from the protein A, B, and E domain unfolding pathways described previously in this laboratory (*33*). In the unfolding simulations of the B and E domains, H3 was the most persistent helix under denaturing conditions with H1 and H2 losing structure earlier in the simulation. The relative stability of H1 and H2 differed between domains; in the E domain, structure is lost first in H2 and then H1, while in the B domain, H1 lost helicity at a time similar to that of H2. In contrast, the helices in A219 are more similar in stability, leading to a more diffusion−collision-like folding mechanism.

Comparison of A219 and G311. We are interested in the features of the G311 and A219 sequences that direct folding to one structure or the other. For the purpose of this discussion, we consider the unfolding trajectories in reverse as a description of the folding pathway. The key events along the folding pathway of G311 begin with α1 in a helical conformation and with the adoption of native ϕ and ψ angles in the residues of the β3−β4 turn in the denatured state. Contacts are formed in the β3−β4 hairpin which then makes a small number of hydrophobic interactions with α1. β1 makes hydrophobic interactions with β4, and some of the β3−β4 contacts are lost as the β1−β4 interface is formed. The β1, β3, and β4 strands form further interactions with α1 in the approach to the TS. Following the TS, the core rearranges and the β1−β2 hairpin becomes fully structured. For A219, folding begins with a denatured state with little residual helical structure and many different short-lived side chain contacts. Helical structure is formed in α1−α3 before long-lived, long-range side chain interactions are made. Helices α2 and α3 first interact near the α2−α3 loop before making a small number of contacts with α1. At the TS, helices are better formed than the contacts between them so that most of the interhelical contacts are formed in the rearrangement following the TS.

The first thing to consider in the comparison of G311 and A219 is that, although the sequences are 59% identical,

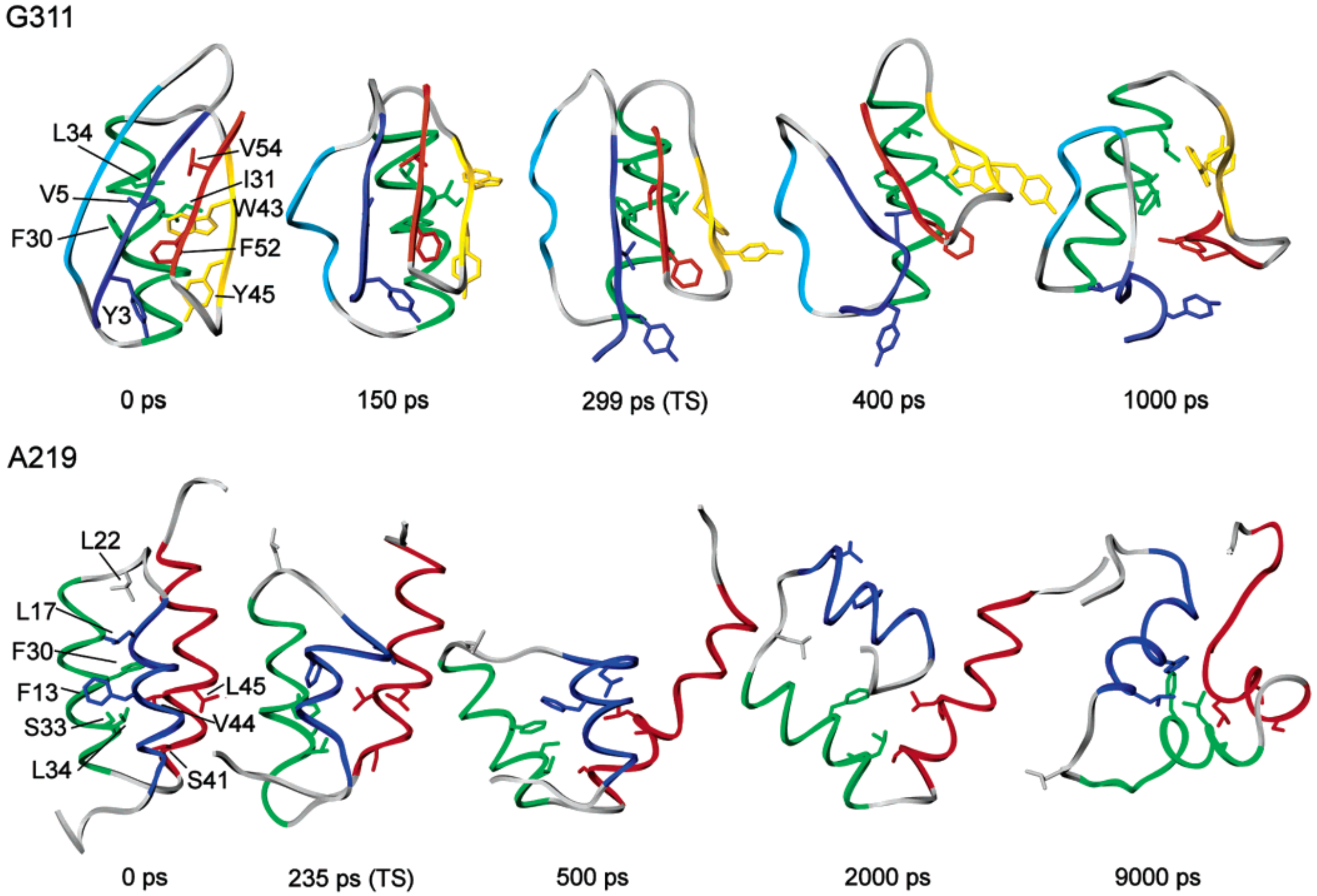

FIGURE 8: Snapshots of early unfolding events in G311 448_2 and A219 498_1. In the top panel (G311), side chains involved in long-range contacts in >50% of the TSE are displayed and colored according to which element of secondary structure they form. Early in unfolding, the core of the protein opens with a hinge motion about the $\beta 1-\beta 2$ and $\alpha 1-\beta 3$ turns. Contacts between $\alpha 1$ and $\beta 2$ are lost, and the $\alpha 1-\beta 1-\beta 3-\beta 4$ sheet interface rearranges (150 ps). The TS structure is similar to that at 150 ps, but with fewer long-range contacts. After the TS, the β-strands dissociate and the core contacts are further disrupted (400 and 1000 ps). In the bottom panel (A219), side chains involved in long-range contacts in >35% of the TSE are displayed and colored according to which element of secondary structure they form. In the TS, all three helices are relatively well ordered but there are few core contacts. The specific contacts differ between trajectories, so few contacts are present in >35% of the TSE. After the TS, the core opens further and there are few long-lived long-range interhelical interactions. The most persistent interactions are those between $\alpha 2$ and $\alpha 3$ near the $\alpha 2-\beta 3$ loop; these interactions consolidate after the transition state and are present for much of the first 7 ns of the trajectory.

residues that differ are not evenly spread along the sequence. Instead, residues 1−20, which form the $\beta 1-\beta 2$ hairpin in G311 and the N-terminal tail and $\alpha 1$ in A219, are 75% identical. Residues 24−38, which make up approximately $\alpha 1$ in G311 and $\alpha 2$ in A219, are 80% identical. Finally, residues forming the $\beta 3-\beta 4$ hairpin in G311 and $\alpha 3$ in A219 are 35% identical. Second, as illustrated in Figure 9, the packing of the deeply buried (less than 10% solvent accessible surface area) residues is very different in the two native structures. Both proteins can be considered as having three local structural elements, $\alpha 1-\alpha 3$ in A219 and $\beta 1-\beta 2$, $\alpha 1$, and $\beta 3-\beta 4$ in G311. $\alpha 1$ in G311 and $\alpha 2$ in A219 both present essentially the same set of residues to the rest of the protein. The different elements of structure interact with different faces of this helix in the two folds (Figure 9, right panel). It is also interesting to note that $\beta 2$ in G311 has no deeply buried residues.

Our simulations give unexpected results for the two different proteins. From the denatured state structures, it would be easy to incorrectly identify the proteins: the denatured state of G311 has more residues with helical ϕ and ψ angles than A219 which is much more random coil-like. Given this observation, why does G311 not fold to the protein A structure? In the protein A structure, residues forming the G311 $\beta 3-\beta 4$ hairpin adopt a helical conformation. In much of the denatured state of G311 simulation 448_2, many of the residues that form the $\beta 3-\beta 4$ hairpin actually have helical (α_R) ϕ and ψ angles. One important exception is K50 in the $\beta 3-\beta 4$ turn which adopts an α_L conformation throughout the denatured state in both simulations. The α_L conformation of K50 would prevent a single contiguous helix analogous to $\alpha 3$ in A219 from forming. Rather than forming a single helix, the aromatic residues in $\beta 3$ and $\beta 4$ form interactions with F30 in $\alpha 1$ early in folding (the F52−F30 interaction is present in 50% of the denatured ensemble). These interactions direct the face of the helix with which the $\beta 3-\beta 4$ turn will interact early in folding. Some of these interactions are sacrificed later in folding, when the hydrophobic residues of $\beta 4$ interact with $\beta 1$ and form the $\beta 1-\beta 4$ interfaces, but by this point, the topology is fixed.

Why does A219 not fold to the protein G structure? In this case, there is a lower intrinsic α-helical propensity in the central helix, and all three helices begin to form at a similar time point. To adopt the protein G structure, $\alpha 3$ would

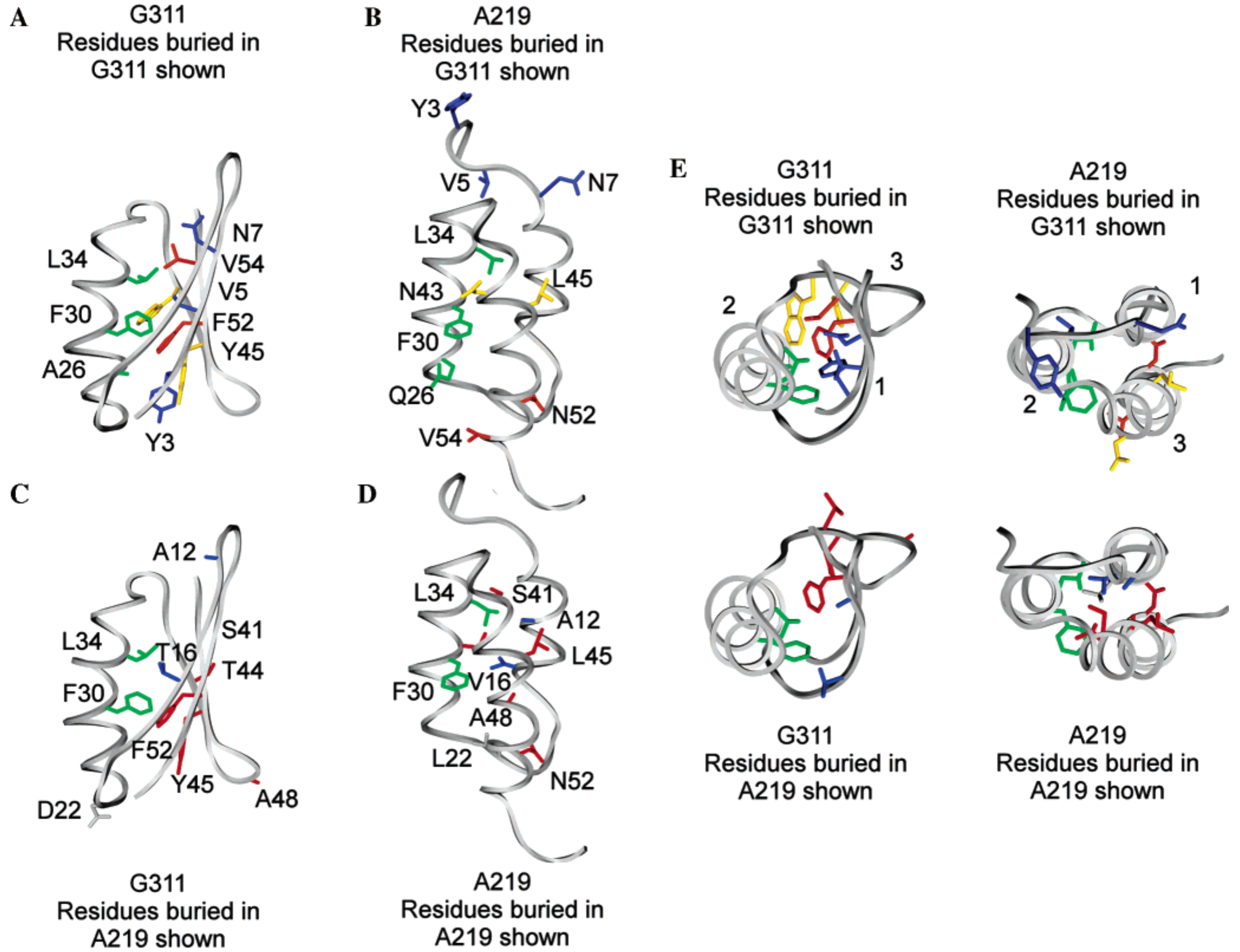

FIGURE 9: Comparison of core (buried, SASA < 10%) residues in G311 and A219. (A) G311 with residues buried in G311 shown. Residues are colored according to which secondary structural element they form. (B) A219 with residues buried in G311 shown. Residues are colored according to which secondary structural element they form in G311. (C) G311 with residues buried in A219 shown. Residues are colored according to which secondary structural element they form in A219. (D) A219 with residues buried in A219 shown. Residues are colored according to which secondary structural element they form. Only residues at positions 30, 34, 45, and 52 appear in the core of both proteins. Of these four residues, only two have the same residue type, F30 and L34. (E) The four structures on the far right are rotated up by 90° relative to their corresponding structure on the left.

have to form a β-hairpin. In G311, the $\beta3-\beta4$ turn adopts native ϕ and ψ angles early in folding. The equivalent residues in A219 are much more disordered in the denatured state, with all residues adopting a number of different ϕ and ψ angles before finally folding with the N-terminal region of α3. Further, in G311, the $\beta3-\beta4$ hairpin has four large hydrophobic residues which form part of the core in the folded structure: W43 and Y45 in β3 and F52 and V54 in β4. In the A219 sequence, the residues at these positions are N43, L45, N52, and V54. These residues would not be suitable for packing in the hydrophobic core of G311. Instead, in A219 helix is formed, and then local interactions in the α2−α3 hinge region fix the relative orientation of the α2−α3 hinge and all three helices come together late in folding. Another consideration is that in protein G there is a glycine residue at position 38, the first position in the loop between α1 and the $\beta3-\beta4$ turn. In the native state, this loop exhibits a higher conformational flexibility than, for example, the $\beta3-\beta4$ and $\beta1-\beta2$ turns; however, glycine 38 adopts positive ϕ and ψ angles for the majority of the simulation. In the A219 sequence, residue 38 is proline. Proline is unable to adopt positive ϕ and ψ angles and thus may disfavor the formation of the protein G structure.

CONCLUSIONS

The design of pairs of proteins with highly identical sequences but different folds offers the opportunity to investigate which residues are the most important in directing folding to a given structure. In particular, at what stage in folding is the final topology determined? Bryan, Orban, and co-workers recently developed a pair of proteins with highly identical sequences but different folds that are an ideal model system for computational studies. Despite the sequences having 59% identical sequences, one sequence folds to the α+β protein G structure and the second to the all-α-helical protein A structure. We have used MD simulations to characterize the folding−unfolding pathway of this pair of proteins at atomic resolution. By considering the unfolding simulations in reverse, the final protein structure is determined early along the folding pathway. Important differences are seen in the early folding behavior of residues forming α1 in G311 and α2 in A219 (80% identical) and in the residues that form the $\beta3-\beta4$ hairpin in G311 and α3 in A219 (35% identical). Despite the highly identical sequences of residues forming α1 in G311 and α2 in A219, there is a significant difference in the predicted intrinsic helical propensity of the two sequences. In the simulations, we see that α1 in G311 (which has the higher helical propensity) adopts helical ϕ and ψ angles in much of the denatured ensemble; in contrast, the majority of residues forming α2 in A219 adopt a random coil-like conformation in the denatured ensemble. The sequence of the $\beta3-\beta4$ hairpin in G311 is most different from the corresponding region in A219 (α3). The $\beta3-\beta4$ turn becomes structured early in folding, with the α_L conformation of K50 essentially blocking the adoption of helical structure. Long-range interactions between the $\beta3-\beta4$ turn and α1 fix the protein G topology early in folding. In contrast, for the protein A sequence, a more hierarchical folding mechanism is observed. All three helices become structured early in folding, and then docking occurs first between α2 and α3. The most similar sequences, forming the $\beta1-\beta2$ hairpin in G311 and α1 in A219, become fully structured later in folding. Small sequence differences lead to different key events early in folding, which in turn determine the folded topology.

ACKNOWLEDGMENT

We are grateful to Dr. Jane Clarke for helpful discussions and for comments on the manuscript.

REFERENCES

1. Anfinsen, C. B. (1973) The rules that govern the folding of protein chains, *Science 181*, 223−230.
2. Kabsch, W., and Sander, C. (1984) The use of sequence homologies to predict protein structure: Identical pentapeptides can have completely different conformations, *Proc. Natl. Acad. Sci. U.S.A. 81*, 1075−1078.
3. Cerpa, R., Cohen, F. E., and Kuntz, I. D. (1996) Conformational switching in designed peptides: The helix/sheet transition, *Folding Des. 1*, 91−101.
4. Minor, D. L., Jr., and Kim, P. S. (1996) Context-dependent secondary structure formation of a designed protein sequence, *Nature 380*, 730−734.
5. Cregut, D., Civera, C., Macias, M. J., Wallon, G., and Serrano, L. (1999) A tale of two secondary structure elements: When a β-hairpin becomes an α-helix, *J. Mol. Biol. 292*, 389−401.
6. Rose, G. D., and Creamer, T. P. (1994) Protein-Folding: Predicting Predicting, *Proteins: Struct., Funct., Genet. 19*, 1−3.
7. Jones, D. T., Moody, C. M., Uppenbrink, J., Viles, J. H., Doyle, P. M., Harris, C. J., Pearl, L. H., Sadler, P. J., and Thornton, J. M. (1996) Towards meeting the Paracelsus Challenge: The design, synthesis, and characterization of Paracelsin-43, an α-helical protein with over 50% sequence identity to an all-β protein, *Proteins: Struct., Funct., Genet. 24*, 502−513.
8. Dalal, S., Balasubramanian, S., and Regan, L. (1997) Protein alchemy: Changing β-sheet into α-helix, *Nat. Struct. Biol. 4*, 548−552.
9. Yuan, S. M., and Clarke, N. D. (1998) A hybrid sequence approach to the Paracelsus Challenge, *Proteins: Struct., Funct., Genet. 30*, 136−143.
10. Driscoll, P. C., Gronenborn, A. M., Beress, L., and Clore, G. M. (1989) Determination of the 3-Dimensional Solution Structure of the Antihypertensive and Antiviral Protein Bds-I from the Sea-Anemone *Anemonia sulcata*: A Study Using Nuclear Magnetic-Resonance and Hybrid Distance Geometry-Dynamical Simulated Annealing, *Biochemistry 28*, 2188−2198.
11. Gouda, H., Torigoe, H., Saito, A., Sato, M., Arata, Y., and Shimada, I. (1992) 3-Dimensional Solution Structure of the B-Domain of Staphylococcal Protein-a: Comparisons of the Solution and Crystal-Structures, *Biochemistry 31*, 9665−9672.
12. Wolberger, C., Dong, Y., Ptashne, M., and Harrison, S. C. (1988) Structure of a Phage 434 Cro/DNA Complex, *Nature 335*, 789−795.
13. Dalal, S., and Regan, L. (2000) Understanding the sequence determinants of conformational switching using protein design, *Protein Sci. 9*, 1651−1659.
14. Alexander, P. A., Rozak, D. A., Orban, J., and Bryan, P. N. (2005) Directed evolution of highly homologous proteins with different folds by phage display: Implications for the protein folding code, *Biochemistry 44*, 14045−14054.
15. He, Y. N., Yeh, D. C., Alexander, P., Bryan, P. N., and Orban, J. (2005) Solution NMR structures of IgG binding domains with artificially evolved high levels of sequence identity but different folds, *Biochemistry 44*, 14055−14061.
16. White, G. W. N., Gianni, S., Grossmann, J. G., Jemth, P., Fersht, A. R., and Daggett, V. (2005) Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding, *J. Mol. Biol. 350*, 757−775.
17. Jemth, P., Day, R., Gianni, S., Khan, F., Allen, M., Daggett, V., and Fersht, A. R. (2005) The structure of the major transition state for folding of an FF domain from experiment and simulation, *J. Mol. Biol. 350*, 363−378.
18. Day, R., and Daggett, V. (2005) Ensemble versus single-molecule protein unfolding, *Proc. Natl. Acad. Sci. U.S.A. 102*, 13445−13450.
19. Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M. V., Alonso, D. O. V., Daggett, V., and Fersht, A. R. (2003) The complete folding pathway of a protein from nanoseconds to microseconds, *Nature 421*, 863−867.
20. Day, R., Bennion, B. J., Ham, S., and Daggett, V. (2002) Increasing temperature accelerates protein unfolding without changing the pathway of unfolding, *J. Mol. Biol. 322*, 189−203.
21. Daggett, V. (2001) Validation of protein-unfolding transition states identified in molecular dynamics simulations, in *From Protein Folding to New Enzymes*, pp 83−93, Portland Press, London.
22. Beck, D. A. C., Alonso, D. O. V., and Daggett, V. (2006) In lucem molecular mechanics, Computer Program, University of Washington.
23. Beck, D. A. C., and Daggett, V. (2004) Methods for molecular dynamics simulation of protein folding/unfolding in solution, *Methods 34*, 112−120.
24. Levitt, M., Hirshberg, M., Sharon, R., Laidig, K. E., and Daggett, V. (1997) Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution, *J. Phys. Chem. B 101*, 5051−5061.
25. Levitt, M., Hirshberg, M., Sharon, R., and Daggett, V. (1995) Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution, *Comput. Phys. Commun. 91*, 215−231.
26. Kell, G. S. (1967) Precise representation of volume properties of water at one atmosphere, *J. Chem. Eng. Data 12*, 66−69.
27. Haar, L., Gallagher, J. S., and Kell, G. S. (1984) *NBS/NRC Stram Tables: Thermodynamic and Transport Properties and Computer Programs for Vapor and Liquid States of Water in SI Units*, Hemisphere Publishing Corp., Washington, DC.
28. Li, A., and Daggett, V. (1994) Characterization of the transition state of protein unfolding by use of molecular dynamics, *Proc. Natl. Acad. Sci. U.S.A. 91*, 10430−10434.
29. Sammon, J. W. (1969) A Nonlinear Mapping for Data Structure Analysis, *IEEE Trans. Comput. C 18*, 401−409.
30. Lacroix, E., Viguera, A. R., and Serrano, L. (1998) Elucidating the folding problem of α-helices: Local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters, *J. Mol. Biol. 284*, 173−191.

31. Munoz, V., and Serrano, L. (1997) Development of the multiple sequence approximation within the agadir model of α-helix formation. Comparison with Zimm-Bragg and Lifson-Roig formalisms, *Biopolymers 41*, 495−509.
32. Manning, M. C., and Woody, R. W. (1991) Theoretical CD studies of polypeptide helices: Examination of important electronic and geometric factors, *Biopolymers 31*, 569−586.
33. Chin, D. H., Woody, R. W., Rohl, C. A., and Baldwin, R. L. (2002) Circular dichroism spectra of short, fixed nucleus alanine helices, *Proc. Natl. Acad. Sci. U.S.A. 99*, 15416−15421.
34. Sari, N., Alexander, P., Bryan, P. N., and Orban, J. (2000) Structure and dynamics of an acid-denatured protein G mutant, *Biochemistry 39*, 965−977.
35. Blanco, F. J., Rivas, G., and Serrano, L. (1994) A short linear peptide that folds into a native stable β-hairpin in aqueous solution, *Nat. Struct. Biol. 1*, 584−590.
36. Blanco, F. J., and Serrano, L. (1995) Folding of protein-G B1 domain studied by the conformational characterization of fragments comprising its secondary structure elements, *Eur. J. Biochem. 230*, 634−649.
37. Munoz, V., Thompson, P. A., Hofrichter, J., and Eaton, W. A. (1997) Folding dynamics and mechanism of β-hairpin formation, *Nature 390*, 196−199.
38. McCallister, E. L., Alm, E., and Baker, D. (2000) Critical role of β-hairpin formation in protein G folding, *Nat. Struct. Biol. 7*, 669−673.
39. Kim, D. E., Fisher, C., and Baker, D. (2000) A breakdown of symmetry in the folding transition state of protein L, *J. Mol. Biol. 298*, 971−984.

BI061904L