Mutational Analysis of the Folding Transition State of the C-Terminal Domain of Ribosomal Protein L9: A Protein with an Unusual β-Sheet Topology†

Ying Li,‡ Ruchi Gupta,‡ Jae-Hyun Cho,§ and Daniel P. Raleigh*,‡,§,||

Department of Chemistry, State University of New York, Stony Brook, New York 11794-3400, Graduate Program in Biochemistry and Structural Biology, State University of New York, Stony Brook, New York 11794, and Graduate Program in Biophysics, State University of New York, Stony Brook, New York 11794

Received July 27, 2006; Revised Manuscript Received October 23, 2006

ABSTRACT: The C-terminal domain of ribosomal protein L9 (CTL9) is a 92-residue $\alpha-\beta$ protein which contains an unusual three-stranded mixed parallel and antiparallel β-sheet. The protein folds in a two-state fashion, and the folding rate is slow. It is thought that the slow folding may be caused by the necessity of forming this unusual β-sheet architecture in the transition state for folding. This hypothesis makes CTL9 an interesting target for folding studies. The transition state for the folding of CTL9 was characterized by Φ-value analysis. The folding of a set of hydrophobic core mutants was analyzed together with a set of truncation mutants. The results revealed a few positions with high Φ-values (≥ 0.5), notably, V131, L133, H134, V137, and L141. All of these residues were found in the β-hairpin region, indicating that the formation of this structure is likely to be the rate-limiting step in the folding of CTL9. One face of the β-hairpin docks against the N-terminal helix. Analysis of truncation mutants of this helix confirmed its importance in folding. Mutations at other sites in the protein gave small Φ-values, despite the fact that some of them had major effects on stability. The analysis indicates that formation of the antiparallel hairpin is critical and its interactions with the first helix are also important. Thus, the slow folding is not a consequence of the need to fully form the unusual three-stranded β-sheet in the transition state. Analysis of the urea dependence of the folding rates indicates that mutations modulate the unfolded state. The folding of CTL9 is broadly consistent with the nucleation−condensation model of protein folding.

Most proteins need to fold into their unique native structures to fulfill their biological function (*1*, *2*). Thus, understanding how proteins fold is extremely important. Furthermore, a number of diseases are related to protein misfolding (*3*). It is now generally accepted that the information required for protein folding is stored in the primary sequence (*4*). However, our understanding of the rules which govern folding is far from complete. Many small single-domain globular proteins fold in a highly cooperative two-state process, and these have become popular model systems (*5*, *6*). Characterizing the transition state for folding is a critical step in describing the folding process. Somewhat surprisingly, the folding of a rather limited number of different topologies has been described.

One powerful approach to studying the structure and interactions formed in the transition state is Φ-value analysis (*7*, *8*). The Φ-value is defined as the change in the free energy of activation ($\Delta\Delta G^{\circ\ddagger}$) divided by the equilibrium free energy change ($\Delta\Delta G^{\circ}$).

$$\Phi = \Delta\Delta G^{\circ\ddagger}/\Delta\Delta G^{\circ} = (\Delta G^{\circ\ddagger}_{\mathrm{WT}} - \Delta G^{\circ\ddagger}_{\mathrm{mutant}})/(\Delta G^{\circ}_{\mathrm{WT}} - \Delta G^{\circ}_{\mathrm{mutant}}) \quad (1)$$

where $\Delta G^{\circ\ddagger}$ and ΔG° represent the activation free energy and folding free energy, respectively. $\Delta G^{\circ\ddagger}$ is directly related to the log of the folding rate (first-order rate constant k_f).

$$k_f = A e^{-\Delta G^{\circ\ddagger}/RT} \quad (2)$$

where A is the prefactor which can depend upon temperature.

Φ-Values have a useful quasi-structural interpretation provided the mutations do not alter the free energy of the unfolded state. If this condition is met, $\Delta\Delta G^{\circ\ddagger}$ is just equal to the change in the free energy levels of the transition state and $\Delta\Delta G^{\circ}$ is equal to the change in the native state free energies. In this case, high Φ-values are interpreted as indicating nativelike interactions in the transition state. Low Φ-values imply denatured-like interactions. Φ-Value analysis has been applied to a number of small, two-state folding proteins (*9*−*16*). In several cases, a few high-Φ-value sites have been found. This has been interpreted as evidence of the presence of a "folding nucleus", which is stabilized by weaker nativelike interactions that develop around the

† This work was supported by Grant GM 70941 from the National Institutes of Health.

* To whom correspondence should be addressed. Phone: (631) 632-9547. Fax: (631) 632-7960. E-mail: draleigh@notes.cc.sunysb.edu.

‡ Department of Chemistry.

§ Graduate Program in Biochemistry and Structural Biology.

|| Graduate Program in Biophysics.

10.1021/bi061516j CCC: $37.00
Published on Web 01/04/2007

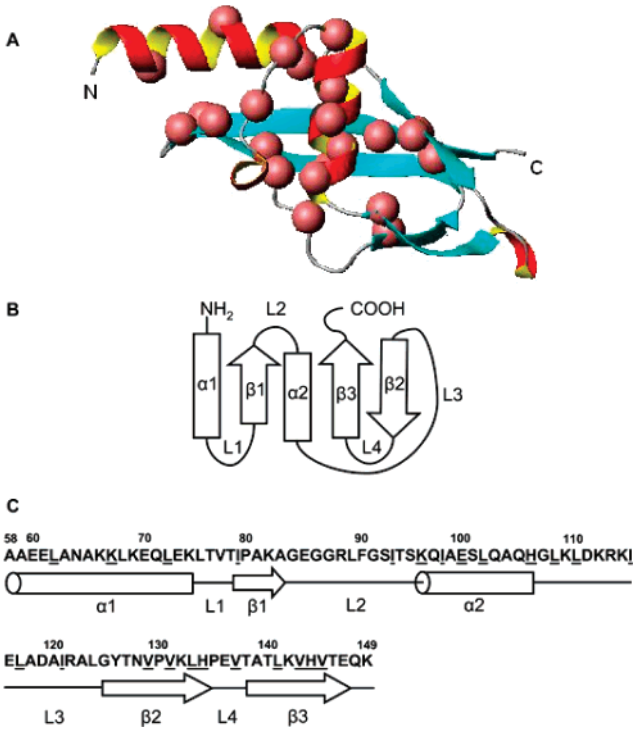

FIGURE 1: (A) Ribbon diagram of CTL9 generated using MOLMOL (*31*) and PDB entry 1DIV. The N- and C-termini are labeled. The sites of the point mutations are shown as spheres. (B) A diagram of the topology of CTL9 is shown together with (C) the primary sequence labeled with the corresponding secondary structures. Mutation sites are underlined.

nucleus. This model for the transition state is commonly called the "nucleation−condensation" model (*17*−*19*).

Here we characterize the folding transition state of a protein with an interesting and unusual topology, the C-terminal domain of ribosomal protein L9 (CTL9).[1] Intact CTL9 binds the 23S ribosomal RNA and is believed to play a structural role in the ribosome. CTL9 is a 92-residue $\alpha-\beta$ protein which contains two α-helices and a three-stranded β-sheet. The three-stranded β-sheet has a very unusual mixed parallel and antiparallel architecture. Such a mixed packing is rare, but what is even rarer is the $\beta1-\beta3-\beta2$ topology with strands 1 and 3 parallel (Figure 1). In a recent survey of β-sheet topology, only 20% of the 225 proteins which contain three-stranded sheets adopted a mixed parallel, antiparallel sheet. Furthermore, the topology with strands 1 and 3 running parallel was found in only six cases, while the mixed topology with strands 1 and 2 parallel was found 19 times (*20*). There have been no reported studies of the folding of this topology, and in fact, there have been very few studies of the folding of single-layer sheets. Interestingly, the folding rate of CTL9 is noticeably lower than that predicted by its contact order (*21*). It is natural to ask if the folding is slow because the formation of this topology is inherently difficult or, stated differently, if the slow folding is a consequence of the need to form the unusual β-sheet topology in the transition state for folding. We conducted a Φ-value analysis of the folding of CTL9 to address this question.

[1] Abbreviations: CD, circular dichroism; CTL9, C-terminal domain of ribosomal protein L9 from *Bacillus stearothermophilus*; TS, transition state; UV, ultraviolet; WT, wild-type protein.

MATERIALS AND METHODS

Mutagenesis, Protein Expression, and Purification. Primers were purchased from Operon, and plasmids containing those primers were generated and amplified by PCR (polymerase chain reaction). Mutations were confirmed by DNA sequencing. For helix truncation mutants, primers were designed to be complementary DNA fragments of the wild-type CTL9 plasmid without the truncation region. Proteins were overexpressed and purified as previously described (*22*). Purity was confirmed by reverse phase HPLC (high-performance liquid chromatography). Identities were confirmed by MALDI-TOF-MS (matrix-assisted laser desorption ionization time-of-flight mass spectrometry) or ESI (electrospray mass spectroscopy). The yield was from 60 to 120 mg/L of Luria broth medium.

Circular Dichroism Spectroscopy. Experiments were performed using Aviv model 62A DS and 202SF spectrometers at 25 °C. ^{1}H NMR and far-UV wavelength scans were taken for all mutants and compared with those of wild-type CTL9 to confirm the native fold. Urea-induced denaturation experiments were carried out using a titrator unit connected to the CD spectrometer. The protein concentration was 8−12 μM in 20 mM sodium phosphate and 100 mM sodium chloride buffer at pH 8.0. Signals were recorded at 222 nm. Urea denaturation curves were fit using standard methods (*23*). A two-state model was used for all mutants.

$$\Delta G_U^{o} = \Delta G_U^{o}(H_2O) - m[\text{denaturant}] \qquad (3)$$

where ΔG_U^{o} is the apparent free energy for the N to D transition. The folded and unfolded baselines are assumed to be linear functions of denaturant.

Stopped-Flow Fluorescence. Stopped-flow fluorescence experiments were performed using an Applied Photophysics SX.18MV stopped-flow reaction analyzer equipped for asymmetric mixing at a ratio of 10:1 (v:v). The only tyrosine residue in the primary sequence of CTL9 was used as the fluorescent probe, with excitation at 276 nm and emission monitored at 305 nm. The folding measurements were initiated with an 11-fold dilution of the prepared completely denatured protein solution into lower concentrations of urea. The unfolding measurements were initiated with an 11-fold dilution into higher concentrations of urea. Final protein concentrations were approximately 50−100 μM. The resulting curves at given urea concentrations were fit using a double-exponential equation to obtain the first-order rate constants for each phase. The major phase was always used for analyses. The minor slow phase is believed to be due to proline isomerization (*22*). Each curve was an average of three to five individual shots. The chevron plots of ln k_{obs} versus urea concentration were fit to the following equation:

$$\ln k_{obs} = \ln\{k_f(H_2O)\exp[(m_f[\text{denaturant}])/(RT)] + k_u(H_2O)\exp[(m_u[\text{denaturant}])/(RT)]\} \qquad (4)$$

where $k_f(H_2O)$ and $k_u(H_2O)$ are the folding and unfolding rate constants, respectively, in the absence of denaturant and m_f and m_u are constants that describe how ln k_f and ln k_u, respectively, vary as a function of the denaturant concentration. The urea concentration was determined by refractom-

Table 1: Mutants Designed To Probe the Folding Mechanism of CTL9

secondary structure of CTL9	mutations
helix α1 from residue 58 to 74	L72A, HT62,[a] HT67,[a] HT72[a]
loop 1 from residue 74 to 77	
strand β1 from residue 77 to 81	I79A
loop 2 from residue 81 to 95	I93A
helix α2 from residue 95 to 106	I98A, L102A, H106Q, K96G, K96A, E100G, E100A
loop 3 from residue 106 to 125	L108A, L110A, I115A, L117A, I121A
strand β2 from residue 125 to 134	V129A, V131A, L133A
loop 4 from residue 134 to 137	H134Q, V137A
strand β3 from residue 137 to 147	L141A, V143A, H144Q, V145A

[a] HT62, HT67, and HT72 are helix truncation mutants whose N-terminal residues are L62, K67, and L72, respectively.

etry. All experiments were conducted at 25 °C in H_2O buffer containing 20 mM sodium phosphate and 100 mM NaCl at pH 8.0.

RESULTS AND DISCUSSION

Design of Mutants. Alanine mutations were used to probe the role of hydrophobic core residues. Mutations were selected to test the importance of strand−strand packing and helix−strand packing. Mutations were introduced into all secondary structure elements except a very short loop, L1. A complete list of mutations is given in Table 1, and the location of the mutations is depicted in Figure 1. The fractional solvent accessibility of each mutated site is listed in the Supporting Information. All of the mutants adopted the native fold, as judged by CD.

The C-terminal domain and the N-terminal domain of L9 fold independently, and the long α-helix connecting these two domains has been shown to be stable in isolation (*21*, *24*). A series of helix truncation mutants of different lengths were designed to probe the role of the long N-terminal helix, α1, in folding. They are called HTs (helix truncations) with a residue number, corresponding to the new N-terminus. For example, HT62 stands for the helix truncation mutant without the first four residues (58−61), but retaining residue 62 as the N-terminus.

The development of structure in the second helix, α2, which connects strands β1 and β2 was studied using a strategy proposed by Fersht and co-workers (*25*). The idea is to select solvent-exposed sites and mutate them to a residue with high helical propensity, usually alanine, and separately to a residue with low helical propensity, usually glycine. Comparison of the effects of the two mutants provides information about the development of helical structure in the folding transition state. One complicating factor is that mutations to glycine will change the entropy of the unfolded state and thus its free energy. If helix formation is critical, the mutations should have significant effects. In contrast, if helix formation is not important, then the mutations should have only minor effects upon the folding rates. Two surface residues were chosen, K96 and E100.

In addition, point mutants of each of three histidines in CTL9 have previously been characterized (*30*). In that study, His to Gln mutants were studied and the choice of Gln was dictated by the desire to replace His with a polar but uncharged residue that had a similar volume. The results of that study are included in this analysis for completeness.

Table 2: Thermodynamic Parameters and Φ-Values for CTL9 Mutants[a]

	m_{eq} (kcal mol^{-1} M^{-1})	$\Delta G°$(equilibrium) (kcal/mol)	θ_m(equilibrium)	Φ[b]
WT	1.05 ± 0.01	6.49 ± 0.08	0.70	−
L72A	1.18 ± 0.01	2.72 ± 0.04	0.75	0.29
I79A	1.17 ± 0.01	3.00 ± 0.02	0.70	0.13
I93A	1.16 ± 0.01	3.30 ± 0.03	0.61	0.08
I98A	1.20 ± 0.01	2.21 ± 0.04	0.78	0.09
L102A	1.14 ± 0.01	4.12 ± 0.02	0.67	0.19
H106Q[c]	1.10 ± 0.01	4.19 ± 0.03	0.75	0.04
L108A	1.15 ± 0.01	4.16 ± 0.04	0.70	0.02
L110A	1.07 ± 0.02	3.28 ± 0.06	0.78	0.05
I115A	1.13 ± 0.02	2.97 ± 0.06	0.72	0.03
L117A	1.08 ± 0.01	5.67 ± 0.04	0.71	0.14
I121A	1.15 ± 0.01	3.61 ± 0.03	0.65	0.07
V129A	1.13 ± 0.01	4.68 ± 0.04	0.78	0.23
V131A	1.16 ± 0.01	4.56 ± 0.06	0.71	0.59
L133A	1.14 ± 0.01	3.55 ± 0.03	0.73	0.63
H134Q[c]	1.12 ± 0.01	4.41 ± 0.04	0.68	0.55
V137A	1.12 ± 0.01	4.01 ± 0.04	0.69	0.46
L141A	1.20 ± 0.01	2.93 ± 0.04	0.78	0.45
V143A	1.18 ± 0.01	3.24 ± 0.04	0.76	0.22
H144Q[c]	0.99 ± 0.01	5.42 ± 0.04	0.69	−0.07
V145A	1.12 ± 0.01	4.57 ± 0.02	0.71	0.19
HT62	1.10 ± 0.01	5.24 ± 0.05	0.72	0.41
HT67	1.05 ± 0.01	2.10 ± 0.04	0.77	0.23
HT72[d]	−	−	−	−
K96G	1.05 ± 0.01	6.29 ± 0.05	0.70	−[e]
K96A	0.91 ± 0.01	7.10[f]	0.77	0.20[g]
E100G	1.13 ± 0.01	4.97 ± 0.04	0.61	0.23
E100A	1.01 ± 0.02	5.51 ± 0.10	0.73	−0.02

[a] All experiments were performed at 25 °C in 20 mM sodium phosphate, 100 mM NaCl buffer at pH 8.0. Standard errors to the fits are given; however, they are likely to underestimate the experimental uncertainty. The standard deviation for $\Delta G°$ is estimated to be ±0.10 kcal/mol on the basis of repeated measurements on the wild type. [b] Calculated using k_f and equilibrium $\Delta G°$ values. [c] Data from Horng et al. (*30*). [d] Data not available, since this protein did not fold. [e] The Φ-value could not be accurately determined due to the very small $\Delta\Delta G°$. [f] A precise value of $\Delta G°$ is not available because of the poorly defined post-transition region in the urea denaturation curve. The value was estimated using the C_m value, the midpoint urea concentration, and the wild-type *m* value. [g] Because of the small value of $\Delta\Delta G°$, this Φ-value is thought to be less reliable.

Equilibrium Stability of the Mutants. The stability of CTL9 is strongly pH dependent, up to pH 7, because of the three histidine residues at positions 106, 134, and 144 (*22*). We conducted all measurements at pH 8.0 and 25 °C. Under these conditions, the mutants were more stable. In addition, minor variations in pH have an only small effect upon folding or stability at this pH. Thermodynamic parameters for all of the mutants derived from urea denaturation studies are listed in Table 2. All of the denaturation curves were well fit using the standard two-state model (*23*) (data not shown). All hydrophobic core and helix truncation mutants were destabilizing. The average difference in stability from the wild type was ∼3 kcal/mol, which is significantly larger than the experimental error. This allows an accurate calculation of the Φ-value. All of the mutants had *m* values, which are related to the change in solvent accessible surface area between the folded and unfolded states, close to the value for the wild type. The most destabilizing core mutant was I98A. This site is located in helix α2 and is part of the hydrophobic interface with the three-stranded β-sheet. Mutation of the charged surface residues, K96 and E100, does not significantly change the stability of the protein or its folding rate. The K96A mutant was slightly stabilizing, and it was difficult to obtain a reliable $\Delta G°$ and *m* value because

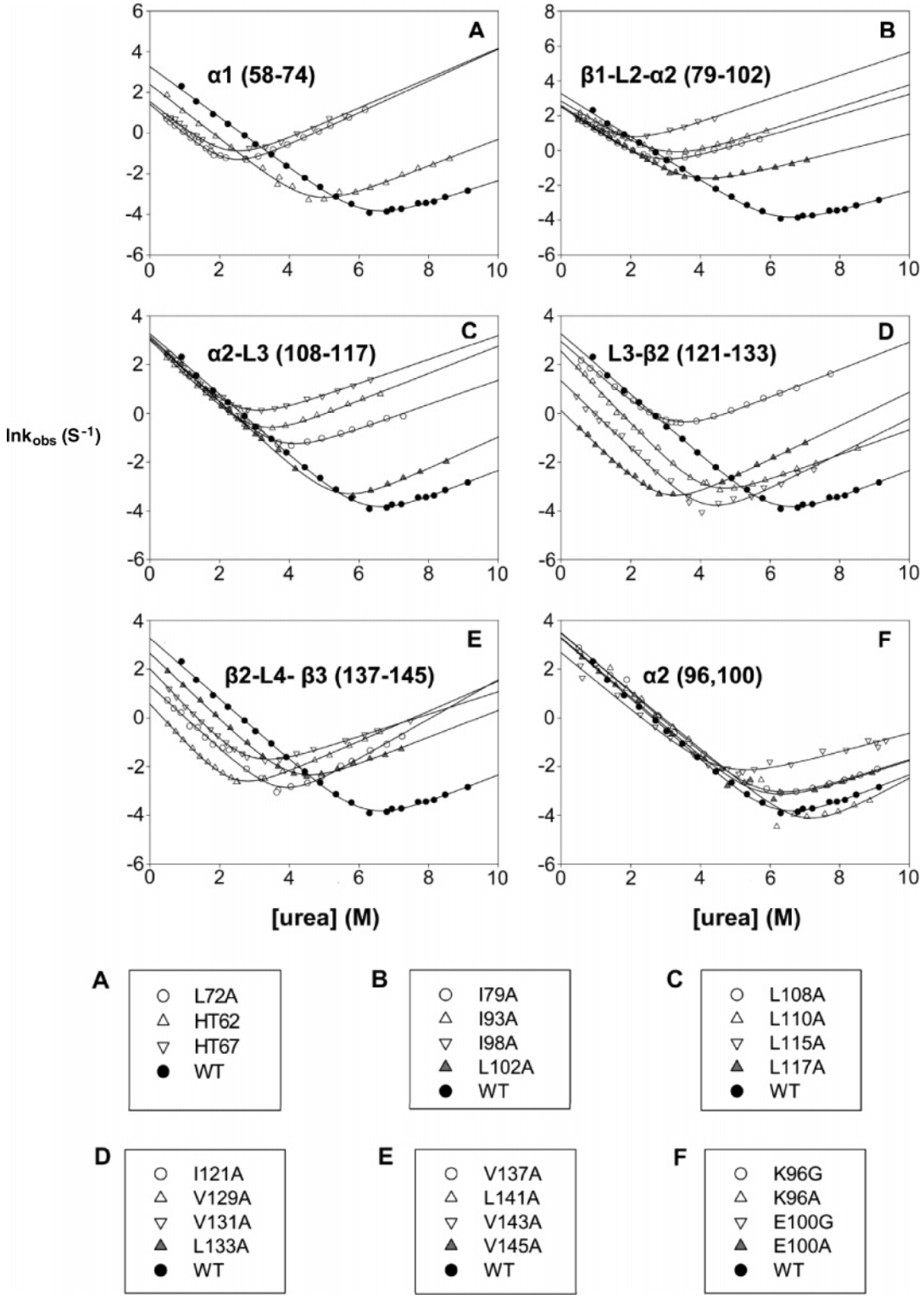

FIGURE 2: Summary of chevron plots for CTL9 mutants. Data for the wild type are included in all plots. The solid line represents the best fit to the standard equation for two-state folding.

of the poorly defined post-transition region in the urea denaturation curve. Its stability was estimated using the C_m value, the midpoint urea concentration, which can be determined accurately, and the wild-type *m* value.

Φ-*Value Analysis Reveals that Most Mutations Have a Minor Effect upon the Refolding Rate.* The natural logarithms of the sum of the folding and unfolding rate constants, k_{obs}, were plotted versus the urea concentration (Figure 2). These plots, so-called "chevron plots", are all v-shaped curves without rollover at low urea concentrations, indicating that all mutants folded in a two-state fashion. Kinetic parameters were derived by fitting the chevron plot as described in Materials and Methods. $\Delta G°$ and *m* values were also calculated from kinetic data, most of which exhibited excellent agreement with the equilibrium data. Table 3 summarizes all of the kinetic parameters. As expected, most of the mutants folded slower than the wild type. The only exceptions were the surface-charged mutations in helix α2,

Table 3: Kinetic Parameters and Φ-Values for CTL9 Mutants[a]

	k_f (s^{-1})	k_u (s^{-1})	m_f (kcal mol^{-1} M^{-1})	m_u (kcal mol^{-1} M^{-1})	ΔG°(kinetic) (kcal/mol)	m(kinetic) (kcal mol^{-1} M^{-1})	θ_m[b]	Φ[b]
WT	26.3 (±1.5)	3.9×10^{-4} ($\pm 1.0 \times 10^{-4}$)	−0.74 (±0.01)	0.33 (±0.02)	6.59 (±0.18)	1.07 (±0.03)	0.69	–
L72A	4.12 (±0.13)	0.03 ($\pm 1.0 \times 10^{-3}$)	−0.88 (±0.02)	0.46 ($\pm 7.4 \times 10^{-3}$)	3.03 (±0.09)	1.34 (±0.03)	0.66	0.31
I79A	12.5 (±0.5)	0.07 (±0.01)	−0.82 (±0.02)	0.35 (±0.02)	3.05 (±0.12)	1.17 (±0.04)	0.70	0.13
I93A	17.0 (±0.6)	0.07 (±0.01)	−0.71 (±0.02)	0.38 (±0.02)	3.25 (±0.10)	1.09 (±0.04)	0.65	0.08
I98A	14.0 (±1.4)	0.60 (±0.10)	−0.93 (±0.08)	0.31 (±0.02)	1.87 (±0.15)	1.24 (±0.10)	0.75	0.08
L102A	12.3 (±0.6)	0.02 ($\pm 3.6 \times 10^{-3}$)	−0.76 (±0.02)	0.29 (±0.02)	3.86 (±0.08)	1.05 (±0.04)	0.72	0.17
H106Q[c]	22.7 (±2.9)	0.02 ($\pm 6.0 \times 10^{-3}$)	−0.77 (±0.03)	0.25 (±0.03)	4.19 (±0.24)	1.10 (±0.06)	0.75	0.04
L108A	23.9 (±1.2)	0.03 ($\pm 4.9 \times 10^{-3}$)	−0.81 (±0.02)	0.30 (±0.02)	4.04 (±0.04)	1.11 (±0.04)	0.73	0.02
L110A	20.3 (±1.1)	0.06 (±0.01)	−0.83 (±0.02)	0.33 (±0.02)	3.47 (±0.10)	1.16 (±0.04)	0.72	0.05
I115A	21.8 (±1.2)	0.18 (±0.02)	−0.81 (±0.03)	0.29 (±0.01)	2.84 (±0.09)	1.10 (±0.04)	0.74	0.03
L117A	21.6 (±0.9)	6.0×10^{-4} ($\pm 2.0 \times 10^{-4}$)	−0.77 (±0.01)	0.38 (±0.02)	6.21 (±0.24)	1.15 (±0.03)	0.67	0.31
I121A	19.1 (±0.8)	0.07 ($\pm 6.0 \times 10^{-3}$)	−0.75 (±0.02)	0.33 (±0.01)	3.32 (±0.07)	1.08 (±0.03)	0.69	0.06
V129A	12.9 (±0.6)	3.1×10^{-3} ($\pm 5.0 \times 10^{-4}$)	−0.88 (±0.02)	0.30 (±0.02)	4.94 (±0.10)	1.18 (±0.04)	0.75	0.26
V131A	3.84 (±0.36)	6.0×10^{-4} ($\pm 3.0 \times 10^{-4}$)	−0.82 (±0.03)	0.42 (±0.04)	5.19 (±0.44)	1.25 (±0.07)	0.66	0.82
L133A	1.12 (±0.05)	2.4×10^{-3} ($\pm 3.0 \times 10^{-4}$)	−0.83 (±0.02)	0.41 (±0.01)	3.64 (±0.09)	1.23 (±0.03)	0.67	0.63
H134Q[c]	3.1 (±0.4)	2.2×10^{-3} ($\pm 4.0 \times 10^{-4}$)	−0.83 (±0.03)	0.39 (±0.02)	4.41 (±0.06)	1.12 (±0.05)	0.68	0.55
V137A	3.78 (±0.34)	1.4×10^{-3} ($\pm 5.0 \times 10^{-4}$)	−0.77 (±0.03)	0.48 (±0.03)	4.68 (±0.29)	1.25 (±0.06)	0.62	0.60
L141A	1.76 (±0.09)	9.4×10^{-3} ($\pm 8.0 \times 10^{-4}$)	−0.93 (±0.02)	0.37 (±0.01)	3.10 (±0.07)	1.30 (±0.03)	0.72	0.46
V143A	7.51 (±0.33)	0.03 ($\pm 2.5 \times 10^{-3}$)	−0.90 (±0.02)	0.27 (±0.01)	3.27 (±0.08)	1.17 (±0.03)	0.77	0.22
H144Q[c]	29.4 (±2.0)	1.5×10^{-3} ($\pm 4.0 \times 10^{-4}$)	−0.75 (±0.01)	0.33 (±0.02)	5.42 (±0.59)	0.99 (±0.03)	0.69	−0.07
V145A	14.2 (±0.6)	5.4×10^{-3} ($\pm 8.0 \times 10^{-4}$)	−0.80 (±0.01)	0.33 (±0.02)	4.66 (±0.10)	1.13 (±0.03)	0.71	0.19
HT62	11.0 (±1.1)	1.1×10^{-3} ($\pm 3.0 \times 10^{-4}$)	−0.79 (±0.02)	0.39 (±0.03)	5.46 (±0.22)	1.18 (±0.05)	0.67	0.46
HT67	4.67 (±0.40)	0.04 (±0.01)	−0.81 (±0.05)	0.43 (±0.03)	2.79 (±0.24)	1.25 (±0.08)	0.65	0.27
HT72[d]	–	–	–	–	–	–	–	–
K96G	33.2 (±2.8)	2.0×10^{-3} ($\pm 9.0 \times 10^{-4}$)	−0.73 (±0.02)	0.27 (±0.03)	5.76 (±0.43)	1.00 (±0.05)	0.73	–[e]
K96A	32.1 (±6.6)	4.4×10^{-5} ($\pm 1.0 \times 10^{-4}$)	−0.70 (±0.04)	0.45 (±0.20)	8.00 (±0.85)	1.15 (±0.24)	0.61	0.08
E100G	14.5 (±1.8)	0.012 ($\pm 3.7 \times 10^{-3}$)	−0.69 (±0.03)	0.23 (±0.02)	4.21 (±0.30)	0.92 (±0.05)	0.75	0.15
E100A	27.1 (±3.3)	1.8×10^{-3} ($\pm 1.3 \times 10^{-3}$)	−0.73 (±0.03)	0.27 (±0.05)	5.70 (±0.80)	1.00 (±0.08)	0.73	−0.02

[a] All experiments were performed at 25 °C in 20 mM sodium phosphate, 100 mM NaCl buffer at pH 8.0. Standard errors to the fits are given; however, they may underestimate the experimental uncertainty. The standard deviation for k_f for the wild type is estimated to be ± 2 s^{-1} on the basis of repeated measurements. [b] Calculated using k_f and k_u values. [c] Data from Horng et al. (*30*). [d] Data not available, since this protein did not fold. [e] The Φ-value could not be accurately determined due to the very small $\Delta\Delta G^\circ$.

K96A, K96G, and E100A, which folded slightly faster, although the increase in the folding rate for these mutants was not large, especially considering the experimental uncertainty. The Φ-values and θ_m, which is defined as the ratio of m_f to m, were calculated both from kinetic data and from combining kinetic and equilibrium data as presented in Tables 2 and 3. In general, θ_m reflects the position of the transition state along the folding reaction coordinate in terms of solvent accessibility. m_f and θ_m for all mutants were very similar to the respective values for the wild type, suggesting that the mutations do not perturb the folding mechanism. There is more variation in m_u. The unfolding branch of some of the chevron plots for the mutants in α2 is not defined as well as the folding branch because of the relatively high C_m values. Thus, k_u is not always defined with high precision at those sites. So we avoid making any conclusions on the basis of any apparent variations in k_u.

The Φ-values calculated from the kinetic data alone were consistent with the ones calculated from k_f and equilibrium $\Delta\Delta G^\circ$ values. Importantly, none of the small variations affect the interpretation of the data, and the conclusions are independent of the set of Φ-values which was used. K96G and K96A were not suitable for Φ-value analysis, because the $\Delta\Delta G^\circ$ values were too close to zero. Given the fact that their folding rates were very similar to the that of the wild type, it is safe to say that these mutations do not significantly perturb the folding of CTL9.

A histogram of the Φ-values versus residue number is shown in Figure 3A. A few mutation sites with high Φ-values clearly stand out. They are V131, L133, H134, V137, and L141 and are all located in the $\beta 2-\beta 3$ hairpin. Figure 4A shows a ribbon diagram of CTL9 with positions color-coded with respect to Φ-values. A 90°-rotated view is shown in panel B. Blue spheres represent mutation sites with small Φ-values (<0.5), and red spheres represent the sites with higher Φ-values (≥0.5). One possible misinterpretation in Φ-value analysis is the assumption that mutants with small to medium Φ-values always have smaller effects upon folding than mutants with large Φ-values. This does not have to be true, since Φ-values involve normalization by $\Delta\Delta G^\circ$. Thus, a site with a high Φ-value but small $\Delta\Delta G^\circ$ could have a weaker effect on the folding rate than a site with a smaller Φ-value but larger $\Delta\Delta G^\circ$. For example, the Φ-value for the L72A mutant is ∼0.3, but this mutant destabilized the protein by more than 3.5 kcal/mol. The folding rate of this mutant was more than 6 times slower than that of the wild type and fell in the range of folding rates that were observed for the high-Φ-value mutants. Thus, residue 72 is as important energetically as some of the sites with higher Φ-values. A histogram of $RT \ln(k_f^{WT}/k_f^{mutant})$ versus residue number is shown in Figure 3B, and a color-coded ribbon diagram of CTL9 is shown in panels C and D of Figure 4. In general, the plots of Φ and $RT \ln(k_f^{WT}/k_f^{mutant})$ appear very similar. Residue 72 appears to be more important when $RT \ln(k_f^{WT}/k_f^{mutant})$ is considered while residue 117 appears to be less important, but the overall trend is very similar. H144Q and E100A have small negative Φ-values. Negative Φ-values are often interpreted as an indication of non-native interactions in the transition state, but it is extremely unlikely that this is the case here because both Φ-values are very small (−0.07 and −0.02) and are within experimental uncertainty of zero.

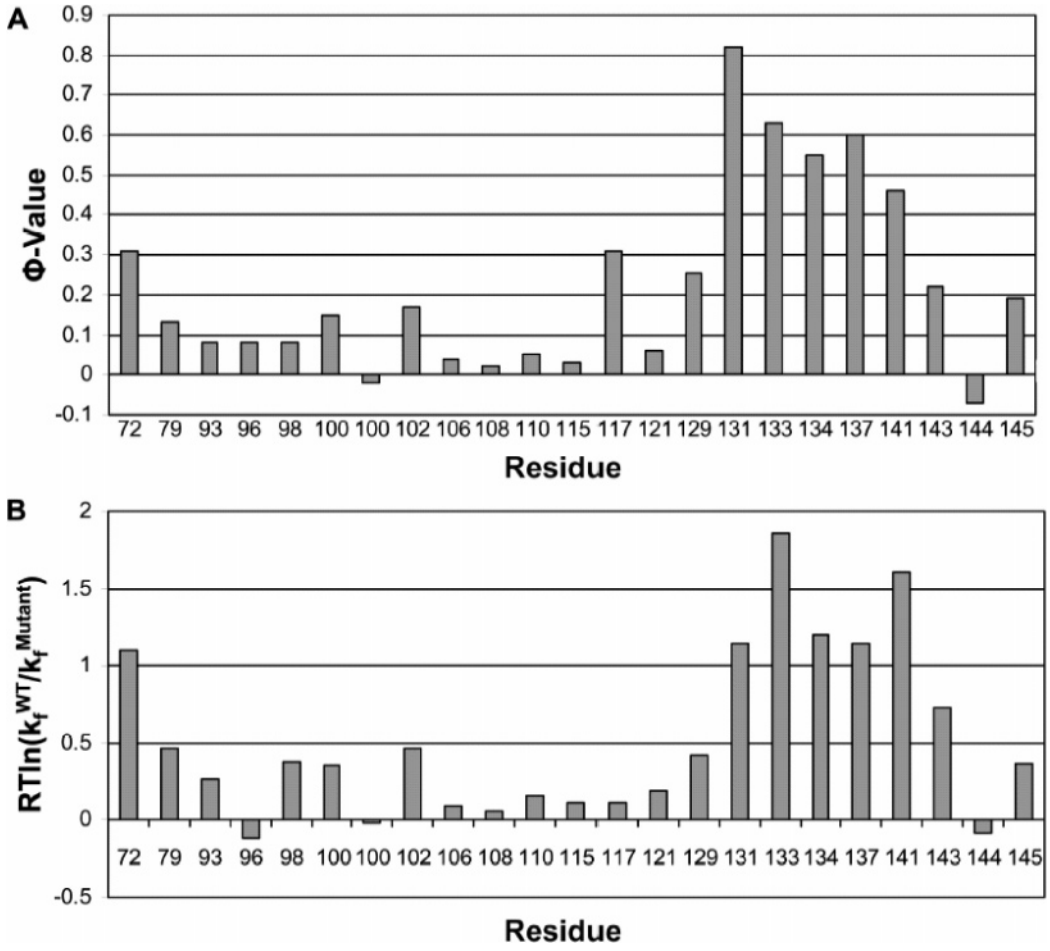

FIGURE 3: (A) Histogram of Φ-values vs residue number. (B) Histogram of $\Delta\Delta G^{\ddagger}$ [RT ln(k_f^{WT}/k_f^{mutant})] vs residue number.

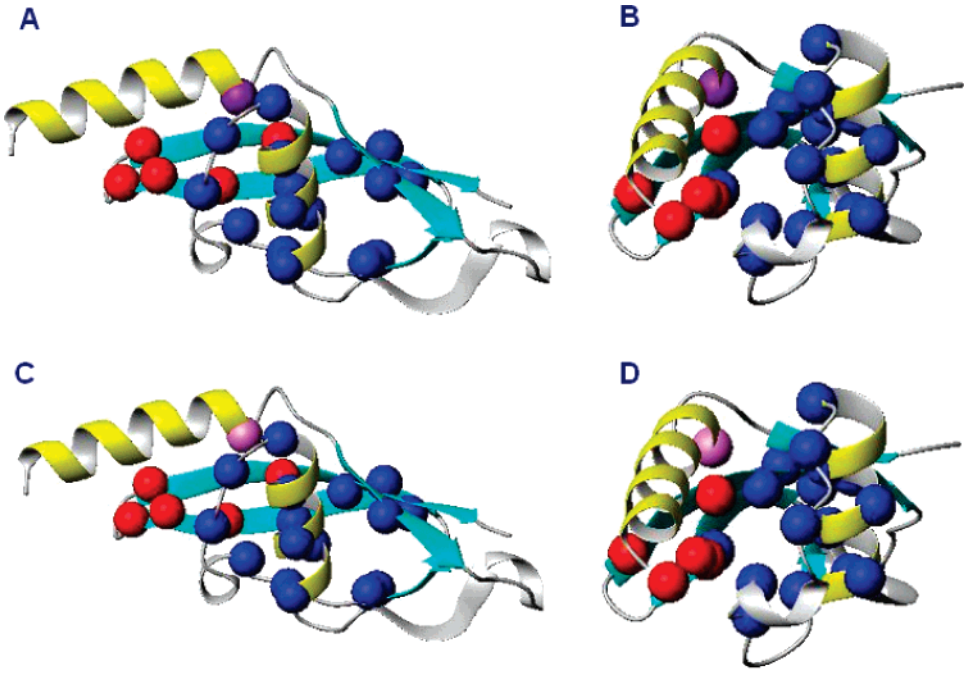

FIGURE 4: (A) Ribbon diagram of CTL9 with positions color-coded with respect to Φ-values. (B) Panel A rotated 90°. (C) Ribbon diagram of CTL9 with positions color-coded with respect to the values of RT ln(k_f^{WT}/k_f^{mutant}). (D) Panel C rotated 90°. Blue spheres represent mutation sites with low Φ-values (<0.5) or low values of RT ln(k_f^{WT}/k_f^{mutant}) (<1), and red spheres represent sites with high Φ-values (≥0.5) or high values of RT ln(k_f^{WT}/k_f^{mutant}) (>1). Plots were generated using MOLMOL (*31*).

Formation of the β-Hairpin (β2−L4−β3) Is Involved in the Rate-Limiting Step in Folding. Figure 4 clearly shows that all high-Φ-value mutation sites are in the β2−L4−β3 hairpin region. The residues with large values of RT ln(k_f^{WT}/k_f^{mutant}) are also found in this region. One residue, L72, which has a large value of RT ln(k_f^{WT}/k_f^{mutant}) is not part of this

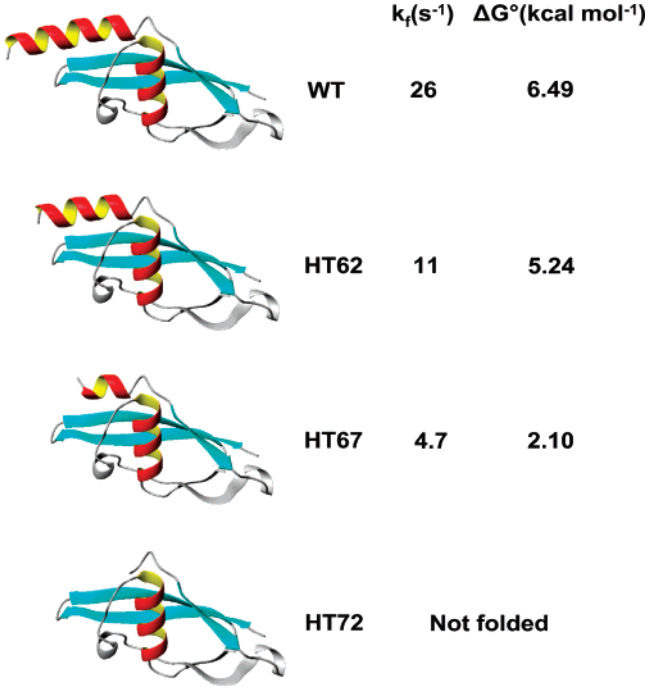

FIGURE 5: Helix truncation mutants mapped onto the native structure. The folding rate and the equilibrium $\Delta G°$ in the absence of denaturant are listed. HT72 did not fold. This figure was generated using MOLMOL (*31*).

sequence; however, it packs directly against this region. The analysis indicates the importance of formation of structure in this region in the folding process.

Role of N-Terminal Helix α1 (residues 58−74). The N-terminal helix projects away from the globular core of the domains. In the crystal structure of intact L9 (*26*), residues 58−61 make no contact with the globular portion of the protein. Subsequent residues in the helix pack against and shield one face of the $\beta2-\beta3$ hairpin. We prepared and analyzed a set of truncation mutants to probe this region of the structure. The truncation of the N-terminal helix, α1, affected both the folding rate and stability significantly (Table 3). A ribbon diagram of the helix truncation mutants with the corresponding folding rates and equilibrium free energies is shown in Figure 5. The more drastic the helix truncation, the slower the folding rate and the lower the stability. HT62 removes the first turn of the helix and leads to a $\Delta\Delta G°$ of 1.10−1.25 kcal/mol. The folding rate is decreased by a factor of more than 2, and the Φ-value is 0.41−0.46. Truncation of the helical region is expected to decrease the propensity of the helix to form, and the decrease in folding rate demonstrated that this effect is coupled to folding. Further truncation leads to HT67 in which the next turn (plus an additional residue) is deleted. This mutant is even less stable with a $\Delta\Delta G°$ relative to the wild type between 3.80 kcal/mol (from kinetic data) and 4.39 kcal/mol (from equilibrium data) and has a folding rate which is 5-fold slower. This truncation eliminates one hydrophobic residue, A63, which forms hydrophobic contacts with residues in the $\beta2-\beta3$ hairpin. The Φ-value is 0.23−0.27. HT72, which fully eliminated the helix, was not produced in the soluble fraction during expression. This might be either because the unstructured polypeptide chain was degraded by cellular proteases or because the protein misfolded and formed inclusion bodies. The truncation mutants clearly affect the folding of CTL9, arguing for a role for this part of the structure in the folding transition state.

Role of Helix α2 (residues 95−106). All mutations in this region, I98A, L102A, H106Q, K96A, K96G, E100A, and E100G, have small Φ-values and small effects upon the folding rate even though some significantly destabilized the protein. This indicates that helix α2 contributes to the stability of the folded state but not to transition state stabilization. Φ-Values close to zero like those observed here indicated that the interactions being probed are no more developed in the transition state than they are in the unfolded state. This need not imply that the region is unstructured in the transition state; it simply indicates that the structure is no more developed than in the unfolded state. The major contribution to the equilibrium stability can be rationalized from the three-dimensional structure. The β-hairpin region is sandwiched between the two α-helices, and helix α2 plays an important role by protecting the hydrophobic core against solvent exposure.

Formation of Strand β1 (residues 77−81), L2 (residues 81−95), and L3 (residues 106−125) Is Not Critical for the Formation of the Transition State. The first strand is short, consisting of five residues. Unlike mutations in the other two strands, the only mutation site in strand $\beta1$, I79, exhibited a small Φ-value, 0.13, and a small $RT \ln(k_f^{WT}/k_f^{mutant})$ value, indicating that $\beta1$ was not well structured in the transition state. This is understandable, because this strand is short and relatively far from the folding nucleus. Folding of CTL9 does not depend upon full formation of the mixed parallel and antiparallel β-sheet structure. All other mutants in loop 2 and loop 3 exhibited very small Φ-values and small $RT \ln(k_f^{WT}/k_f^{mutant})$ values.

Do Mutations Affect the Unfolded State? The unfolded state of CTL9 in the absence of denaturant has been shown to be relatively compact compared to the fully unfolded state populated in high concentrations of denaturants (*27*). pH-dependent hydrophobic cluster formation has been proposed to contribute to the compaction. Thus, it is natural to inquire if the mutations alter the properties of the denatured state ensembles. Kiefhaber and colleagues have pointed out that plots of m_f versus m_{eq} offer a sensitive probe of potential movement in the relative position of the denatured state and the transition state (*28*, *29*). m_f values range from −0.69 to −0.93 for the mutations studied here, while the equilibrium m values range from 0.91 to 1.20. If only hydrophobic core mutations are considered, then m_f ranges from −0.79 to −0.93 and m_{eq} from 1.05 to 1.20. There is a clear correlation between m_f and m_{eq} (Figure 6), although there is obvious scatter in the data. This is consistent with the notion that the mutations have some effect on the unfolded state. The magnitude of m_f increases for almost all of the hydrophobic core mutants, although the effects are relatively modest; the largest change is 0.18 or approximately 25%, while for 12 of the 17 hydrophobic mutants, the change is on the order of ≤10%. This is consistent with some expansion of the unfolded state upon mutation. A critical assumption in the structural interpretation of Φ-values is that mutations do not induce significant energetic effects upon the denatured state. In general, this is almost impossible to determine experimentally. The hydrophobic core mutations studied here have relatively small but correlated effects upon m_f and m_{eq}. The relatively small changes suggest, but of course do not prove,

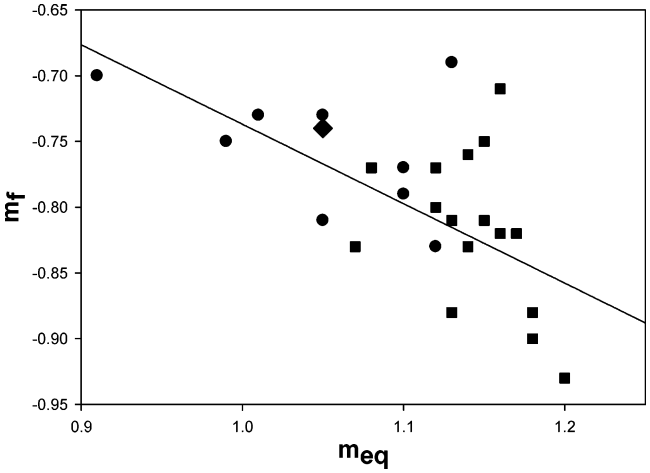

FIGURE 6: Plot of m_f values vs m_{eq} values for all of the mutants, including the helix truncation mutants. The solid line is the linear fit ($r = 0.66$). The point for the wild type is shown as a diamond. The points for the hydrophobic core mutants are shown as squares. The points for the truncation mutants, K96, E100 mutants and His to Gln mutants, are shown as circles. The *r* value decreases to 0.571 if only hydrophobic core mutations are considered.

that the mutational effects upon the unfolded state are relatively small. For unfolded state effects to significantly change our interpretation of the transition state, there would have to be differential effects; i.e., the changes caused by mutations in the β-hairpin region would have to have different effects on the unfolded state than mutations of other hydrophobic residues.

CONCLUSIONS

In this study, a series of point mutants and helix truncation mutants were designed to probe the folding mechanism of CTL9. Φ-Value analysis and analysis of the effect upon folding rates showed that formation of the β-hairpin (β2−L4−β3) is involved in the rate-determining step. Hydrophobic residues in this region pack against N-terminal α-helix α1, and truncation mutants of this helix affect the folding rate. Hence, packing of this region against the β-hairpin is also involved in the rate-determining step. In contrast, the other helix, α2, is important for stability but not for formation of transition state structure. The third strand is also not critical for the formation of the transition state. The core residues with larger Φ-values or $RT \ln(k_f^{WT}/k_f^{mutant})$ values cluster together in the native structure. This is broadly consistent with the classic nucleation−condensation model of folding (*17*−*19*). The data clearly show that slow folding is not due to a requirement to fully form the unusual three-stranded sheet in the transition state.

SUPPORTING INFORMATION AVAILABLE

Tables of the relative solvent accessibility for all point mutants and helix truncation mutants examined in this study. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Agard, D. A. (1993) To fold or not to fold, *Science 260*, 1903−1904.
2. Watson, J. D. (1963) Involvement of RNA in the synthesis of proteins, *Science 140*, 17−26.
3. Chaudhuri, T. K., and Paul, S. (2006) Protein-misfolding diseases and chaperone-based therapeutic approaches, *FEBS Lett. 273*, 1331−1349.
4. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science 181*, 223−230.
5. Jackson, S. E., and Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition, *Biochemistry 30*, 10428−10435.
6. Jackson, S. E. (1998) How do small single-domain proteins fold? *Folding Des. 3*, R81−R91.
7. Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E., and Fersht, A. R. (1995) Movement of the position of the transition state in protein folding, *Biochemistry 34*, 13656−13662.
8. Li, A., and Daggett, V. (1994) Characterization of the transition state of protein unfolding by use of molecular dynamics: Chymotrypsin inhibitor 2, *Proc. Natl. Acad. Sci. U.S.A. 91*, 10430−10434.
9. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M., and Dobson, C. M. (1999) Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding, *Nat. Struct. Biol. 6*, 1005−1009.
10. Fowler, S. B., and Clarke, J. (2001) Mapping the folding pathway of an immunoglobulin domain: Structural detail from Φ value analysis and movement of the transition state, *Structure 9*, 355−366.
11. Fulton, K. F., Devlin, G. L., Jodun, R. A., Silvestri, L., Bottomley, S. P., Fersht, A. R., and Buckle, A. M. (2005) PFD: A database for the investigation of protein folding kinetics and stability, *Nucleic Acids Res. 33*, D279−D283.
12. Hamill, S. J., Steward, A., and Clarke, J. (2000) The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology, *J. Mol. Biol. 297*, 165−178.
13. Kragelund, B. B., Robinson, C. V., Knudsen, J., Dobson, C. M., and Poulsen, F. M. (1995) Folding of a four-helix bundle: Studies of acyl-coenzyme A binding protein, *Biochemistry 34*, 7217−7224.
14. Martinez, J. C., and Serrano, L. (1999) The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved, *Nat. Struct. Biol. 6*, 1010−1016.
15. Oliveberg, M. (2001) Characterisation of the transition states for protein folding: Towards a new level of mechanistic detail in protein engineering analysis, *Curr. Opin. Struct. Biol. 11*, 94−100.
16. Villegas, V., Zurdo, J., Filimonov, V. V., Aviles, F. X., Dobson, C. M., and Serrano, L. (2000) Protein engineering as a strategy to avoid formation of amyloid fibrils, *Protein Sci. 9*, 1700−1708.
17. Tan, Y. J., Oliveberg, M., and Fersht, A. R. (1996) Titration properties and thermodynamics of the transition state for folding: Comparison of two-state and multi-state folding pathways, *J. Mol. Biol. 264*, 377−389.
18. Fersht, A. R. (1995) Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications, *Proc. Natl. Acad. Sci. U.S.A. 92*, 10869−10873.
19. Fersht, A. R. (1997) Nucleation mechanisms in protein folding, *Curr. Opin. Struct. Biol. 7*, 3−9.
20. Zhang, C., and Kim, S. H. (2000) The anatomy of protein β-sheet topology, *J. Mol. Biol. 299*, 1075−1089.
21. Sato, S., Luisi, D. L., and Raleigh, D. P. (2000) pH jump studies of the folding of the multidomain ribosomal protein L9: The structural organization of the N-terminal domain does not affect the anomalously slow folding of the C-terminal domain, *Biochemistry 39*, 4955−4962.
22. Sato, S., and Raleigh, D. P. (2002) pH-dependent stability and folding kinetics of a protein with an unusual α-β topology: The C-terminal domain of the ribosomal protein L9, *J. Mol. Biol. 318*, 571−582.
23. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves, *Methods Enzymol. 131*, 266−280.
24. Sato, S., Kuhlman, B., Wu, W. J., and Raleigh, D. P. (1999) Folding of the multidomain ribosomal protein L9: The two domains fold independently with remarkably different rates, *Biochemistry 38*, 5643−5650.
25. Serrano, L., Sancho, J., Hirshberg, M., and Fersht, A. R. (1992) α-Helix stability in proteins. I. Empirical correlations concerning substitution of side-chains at the N and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces, *J. Mol. Biol. 227*, 544−559.

26. Hoffman, D. W., Davies, C., Gerchman, S. E., Kycia, J. H., Porter, S. J., White, S. W., and Ramakrishnan, V. (1994) Crystal structure of prokaryotic ribosomal protein L9: A bi-lobed RNA-binding protein, *EMBO J. 13*, 205−212.
27. Li, Y., Picart, F., and Raleigh, D. P. (2005) Direct characterization of the folded, unfolded and urea-denatured states of the C-terminal domain of the ribosomal protein L9, *J. Mol. Biol. 349*, 839−846.
28. Bodenreider, C., and Kiefhaber, T. (2005) Interpretation of protein folding ϕ values, *J. Mol. Biol. 351*, 393−401.
29. Sanchez, I. E., and Kiefhaber, T. (2003) Hammond behavior versus ground state effects in protein folding: Evidence for narrow free energy barriers and residual structure in unfolded states, *J. Mol. Biol. 327*, 867−884.
30. Horng, J. C., Cho, J. H., and Raleigh, D. P. (2005) Analysis of the pH-dependent folding and stability of histidine point mutants allows characterization of the denatured state and transition state for protein folding, *J. Mol. Biol. 345*, 163−173.
31. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51−55.

BI061516J