Contribution of a Buried Hydrogen Bond to λ Repressor Folding Kinetics†

Jeffrey K. Myers and Terrence G. Oas*

Department of Biochemistry, Box 3711, Duke University Medical Center, Durham, North Carolina 27710

Received January 14, 1999; Revised Manuscript Received March 18, 1999

ABSTRACT: A hydrogen bond between the buried residues Asp 14 and Ser 77 in monomeric lambda repressor has been removed by mutation of these residues to alanine. Double mutant cycles show that the interaction stabilizes the native state of the protein by 1.5 kcal/mol. Removal of the interaction affects mainly the unfolding rates and not the folding rates, suggesting that this hydrogen bond is not substantially formed in the rate-limiting steps in the folding pathways of the protein. Mutations in two versions of λ_{6-85}, wild type and the faster folding G46A/G48A (WT*), show similar effects. Diffusion−collision correctly predicts the behavior of WT* but not of wild type. Our analysis suggests that folding of helix 3 is a crucial slow step along the various folding pathways and generally occurs before the formation of the 14−77 hydrogen bond. Experiments removing tertiary interactions, combined with experiments altering helical stability and diffusion−collision calculations, provide a strategy to unravel the folding mechanisms of small helical proteins.

It is generally accepted that proteins do not fold by a random search of all possible conformations. The discovery of fast, two-state protein folding has led to a "new view" which emphasizes the presence of multiple pathways to the native state rather than a single pathway (*1*−*3*). In this view, the direction of the folding process is due to a general bias toward the native state caused by energetically favorable native contacts, rather than a series of one or more intermediates defining a single pathway. Many intermolecular interactions are responsible for the stabilization of native protein structures (*4*, *5*), and the same interactions are likely to be important determinants of the folding kinetics of proteins. A monomeric version of the N-terminal domain of lambda repressor, λ_{6-85}, folds by a two-state mechanism on a microsecond time scale (*6*−*8*). We have chosen a specific hydrogen bond interaction between two residues and asked what role the interaction plays in the folding and unfolding kinetics. In the crystal structure of the N-terminal domain, the Oδ1 atom of Asp 14 forms a tertiary hydrogen bond with the Oγ of Ser 77 (*9*; see Figure 1). The side chains are inaccessible to solvent in the native structure. Using double mutant cycles, Marqusee and Sauer (*10*) showed that the hydrogen bond between Asp 14 and Ser 77 stabilizes the protein by about 1.5 kcal/mol. The 14−77 interaction is of interest in examining kinetic pathways because it represents the linkage of the two terminal regions of the protein. In a previous study, we proposed a modified diffusion−collision model to explain the folding of λ_{6-85} (*11*, *12*). The protein has been divided into five microdomains, which diffuse and collide to form eight native helix−helix interactions called "bonds".[1] As seen from the contact map in Figure 1, contacts between residues 14 and 77 contribute significantly to the "bond" between the first microdomain, consisting of helix 1, and the fifth microdomain, consisting of helix 5 and a loop structure.

Several proteins have been subjected to the "protein engineering method" (*13*−*17*) which truncates residues by mutation in order to study their roles in folding. In the present study, residues 14 and 77 have been mutated to alanine, singly and in combination. This has been done in both the wild-type sequence (WT) and WT*, a more stable variant of the protein where two glycines in helix 3 have been replaced with alanine (*7*). Previous studies showed that WT* and wild type (WT) exhibit quite different behavior, suggesting different folding mechanisms. It is of interest, then, to study the 14/77 mutations in both backgrounds.

Our diffusion−collision calculations indicate that the bond between microdomains 1 and 5 is formed relatively late in the folding of WT*. If bond formation occurs after the rate-limiting step, then the folding kinetics should not be perturbed by removal of interactions that contribute to the bond. The removal of the interacting atoms by single mutations Asp 14 to Ala and Ser 77 to Ala, and the double mutation, allows us to construct a double mutant cycle. Although each individual mutation removes many other contacts to other residues, the double mutant analysis allows determination of the energetic and kinetic consequences of removing just the 14−77 contacts (represented by the box at the intersection of the two lines in Figure 1). λ_{6-85} and λ_{6-85}* fold and unfold very quickly and follow a two-state folding mechanism where only native or denatured states are populated significantly (*7*). A technique useful for probing fast time scales in small proteins is NMR line-shape

† Supported by National Institutes of Health Grants GM45322 to T.G.O. and F32 GM18957 to J.K.M.

* To whom correspondence should be addressed (e-mail: oas@biochem.duke.edu).

[1] The term "bond" is used in diffusion−collision theory to refer to the sum of all native, atomic interactions between two microdomains. It is not to be confused with a specific bond between two atoms such as the 14−77 hydrogen bond. However, since all the contacts between two microdomains are assumed to form cooperatively, the 14−77 hydrogen bond acts as a probe of the entire 1−5 bond.

10.1021/bi990088x CCC: $18.00
Published on Web 05/07/1999

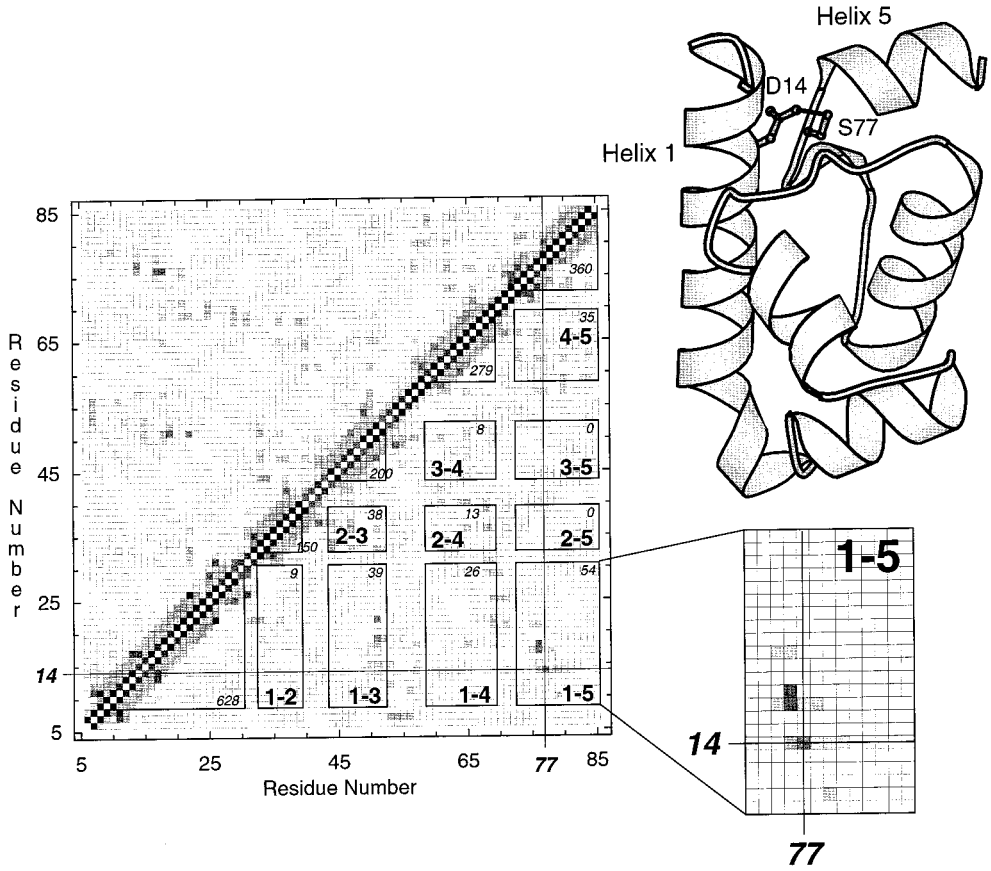

FIGURE 1: Contact map of λ_{6-85} based on the crystal structure. Contacts are defined as the close proximity of, or an overlap in, the van der Waals radii of two atoms. The darkness of each square is related to the number of contacts between two particular residues. Half-squares along the diagonal represent diffusion−collision microdomain boundaries. Off-diagonal boxes denote bonds between microdomains. The box representing the 1−5 bond is magnified at the right. Also to the right is a ribbon drawing of λ_{6-85} based on the crystal structure of the N-terminal domain of lambda repressor bound to DNA (9). Side chains of residues 14 and 77 are shown in ball-and-stick representation, and helices 1 and 5 are labeled. Ribbon diagram drawn using MOLSCRIPT (*36*).

analysis. This has been used to measure folding and unfolding kinetics of λ_{6-85} and the N-terminal domain of ribosomal protein L9 (*18*). NMR has been used to probe the kinetic role of a buried hydrogen bond in the folding and unfolding of λ_{6-85}.

MATERIALS AND METHODS

Mutagenesis, Protein Expression, and Purification. Single-stranded site-specific mutagenesis was performed on the plasmid containing the gene for wild type or WT* λ_{6-85} using standard methods (*19*). Mutations are labeled by the amino acid type being replaced, followed by the position of the mutation and the residue present in the mutant protein in single-letter amino acid code; so D14A means aspartate 14 replaced with alanine. Those in the G46A,G48A background are marked with an asterisk. Correct mutations were confirmed by automated DNA sequencing. Purification of the proteins followed previously published methods (*7*). The purity of the final product was confirmed by SDS−PAGE and analytical HPLC. The molecular weight was checked with electrospray mass spectrometry.

Equilibrium Denaturation Curves. Ten micromolar protein samples in a buffer of 20 mM potassium phosphate, pH 8.0, and 100 mM NaCl were titrated with urea in the same buffer in a Aviv circular dichroism spectrometer (model 202) equipped with an automatic titrator system. Samples were thermostated at 298 K with a built-in temperature controller. Concentrations of urea stock solutions were measured using the refractive index (*20*). A two-state unfolding model, with a linear dependence of ΔG on urea concentration, was fitted to the denaturation data (*20, 21*):

$$y = [(y_f + a_f[\text{urea}]) + (y_u + a_u[\text{urea}])\exp(m([\text{urea}] - [\text{urea}]_{1/2})/RT)]/[1 + \exp(m([\text{urea}] - [\text{urea}]_{1/2})/RT)] \quad (1)$$

where y is the observed CD signal, y_f and y_u are the intercepts of the pre- and post-transition baselines, a_f and a_u are the slopes of the baselines, $[\text{urea}]_{1/2}$ is the midpoint of the denaturation, and m is the linear dependence of ΔG on urea concentration. $[\text{urea}]_{1/2}$ times the m value gives ΔG^{H_2O}, the free energy of the two-state folding reaction in the absence of denaturant. $\Delta\Delta G_{N-D}$ is defined as the ΔG^{H_2O} of the reference (WT or WT*) minus the ΔG^{H_2O} of the mutant. The interaction free energy, $\Delta\Delta G_{int}$, is the sum of the $\Delta\Delta G$ of the single mutants minus the $\Delta\Delta G$ of the double mutant (*22*).

NMR Line-Shape Analysis. Twenty samples were prepared containing 300 μM protein, 20 mM potassium phosphate, pD 8.0, 100 mM NaCl, 1 mM sodium azide, 1mM TMSP, in 99% D_2O, and various concentrations of deuterated urea. The urea concentration of each sample was determined as above. Concentrations of protein stock solutions were determined using UV spectrometry, using an extinction coefficient of 2910 M^{-1} cm^{-1} in 6 M GuHCl (*23*). Proton

Table 1: Equilibrium Stabilities from CD Urea Titration Curves in H_2O[a]

	$[urea]_{1/2}$ (M)	m [kcal/(mol·M)]	$\Delta G_{D-N}^{H_2O}$ (kcal/mol)	ΔG_{D-N}^{3M} (kcal/mol)	$\Delta\Delta G^{H_2O}$ (kcal/mol)	$\Delta\Delta G^{3M}$ (kcal/mol)
WT*	5.20	1.03	5.36	2.27	–	–
D14A*	3.96	0.96	3.79	0.92	1.57	1.35
S77A*	4.06	1.05	4.26	1.11	1.10	1.16
D14,S77A*	4.34	0.91	3.94	1.22	1.42	1.05
14−77* interaction[b]					1.25	1.46
WT	3.54	1.03	3.65	0.56		
D14A	2.19	0.88	1.93	−0.72	1.72	1.28
S77A	2.45	1.01	2.47	−0.56	1.18	1.12
D14,S77A	2.51	0.95	2.39	−0.47	1.26	1.03
14−77 interaction[b]					1.64	1.37

[a] Based on repetitive runs on WT. Errors are estimated to be ±0.1 M in $[urea]_{1/2}$ and ±0.05 kcal/(mol·M) in m for the CD curves. This leads to errors in ΔG_{D-N} of ±0.2 kcal/mol in water and ±0.1 kcal/mol in 3 M urea. [b] Interaction free energies based on double mutant cycles such as those shown in Figure 2: $\Delta\Delta G_{int} = \Delta\Delta G(D14A) + \Delta\Delta G(S77A) - \Delta\Delta G(D14,S77A)$.

NMR spectra were recorded at 298 K on a Varian Unity 500 MHz spectrometer with water and urea presaturation. The aromatic regions of the spectra, with chemical shifts relative to TMSP, were analyzed as described in ref *24* to give folding and unfolding rate constants at each urea concentration in the transition region. Uncertainties were calculated from errors in the global fitting of all spectra for each protein. The following model was fit to the plots of rate constants versus urea:

$$\ln k_f = \ln k_f^{H_2O} + m_f[\text{urea}] \quad (2)$$

$$\ln k_u = \ln k_u^{H_2O} + m_u[\text{urea}]$$

where k is the rate constant at a given urea concentration, k^{H_2O} is the rate constant in the absence of denaturant, and m is the slope. The fractional solvent accessibility of the transition state, α, is related to the slopes (*25*):

$$\alpha = m_u/(m_u - m_f) \quad (3)$$

The stability of the transition state relative to the parent protein is

$$\Delta\Delta G_{TS-D} = -RT \ln(k_u/k_u^*) \quad (4)$$

where k_u is the rate constant for the mutant and k_u^* is the rate constant for the parent. The Φ values are determined from the following:

$$\Phi_f = \Delta\Delta G_{TS-D}/\Delta\Delta G_{N-D} \quad (5)$$

$$\Phi_u = \Delta\Delta G_{TS-N}/\Delta\Delta G_{N-D}$$

The Φ values of a specific interaction are

$$\Phi_{int,f} = \Delta\Delta G_{TS-D,int}/\Delta\Delta G_{int} \quad (6)$$

$$\Phi_{int,u} = \Delta\Delta G_{TS-N,int}/\Delta\Delta G_{int}$$

where $\Delta\Delta G_{TS-D,int}$ is determined from a double mutant cycle of $\Delta\Delta G_{TS-D}$ and $\Delta\Delta G_{int}$ is determined from a double mutant cycle of $\Delta\Delta G_{N-D}$. In both cases the $\Delta\Delta G_{int}$ is the sum of the $\Delta\Delta G$ for each single mutation minus the $\Delta\Delta G$ of the double mutant (*22*).

RESULTS

Equilibrium Denaturation. The conformational stability of each protein was measured using urea denaturation, monitoring the state of the protein by circular dichroism. A two-state model, with a linear dependence of ΔG on urea concentration, was fitted to the data; the results are given in Table 1. In comparing the proteins, we will use both values extrapolated to 0 M denaturant and also values in 3 M urea. Values in 3 M urea are more accurate because they require little or no extrapolation from the region where experimental data are collected.

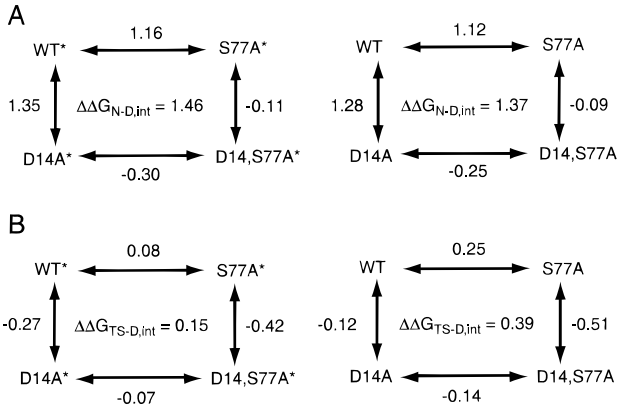

FIGURE 2: (A) Thermodynamic double mutant cycles for WT and WT* based on CD urea titrations (data in 3 M urea). Mutations are represented by the double arrow on each side of the box; $\Delta\Delta G_{N-D}$ (in kcal/mol) is shown for each mutation. The interaction free energy ($\Delta\Delta G_{N-D,int}$) is calculated from the differences across the box either horizontally or vertically. (B) Analogous cycles for changes in transition state free energy, $\Delta\Delta G_{TS-D}$.

The WT* protein is 1.7 kcal/mol more stable than wild type. The 14/77 mutants in the WT* background are more stable than the same mutants in the WT background by a similar amount, indicating the mutations at 14 and 77 are independent of the two alanine substitutions in helix 3.

Single mutations of Asp 14 and Ser 77 to alanine destabilize the native state of the protein. The double mutant has a stability similar to that of either single mutant. Construction of a thermodynamic cycle (*22*) (shown in Figure 2A) gives a free energy of interaction between the two side chains ranging from 1.3 to 1.6 kcal/mol, indicating that the 14−77 interaction stabilizes the native state of the protein. This value is similar to that found by Marqusee and Sauer studying the same interaction in a longer version of the protein (*10*) and typical of the contribution of other hydrogen bonds to protein stability (*26*).

Kinetics by NMR Line-Shape Analysis. Aromatic ^1H NMR spectra at various urea concentrations were analyzed as

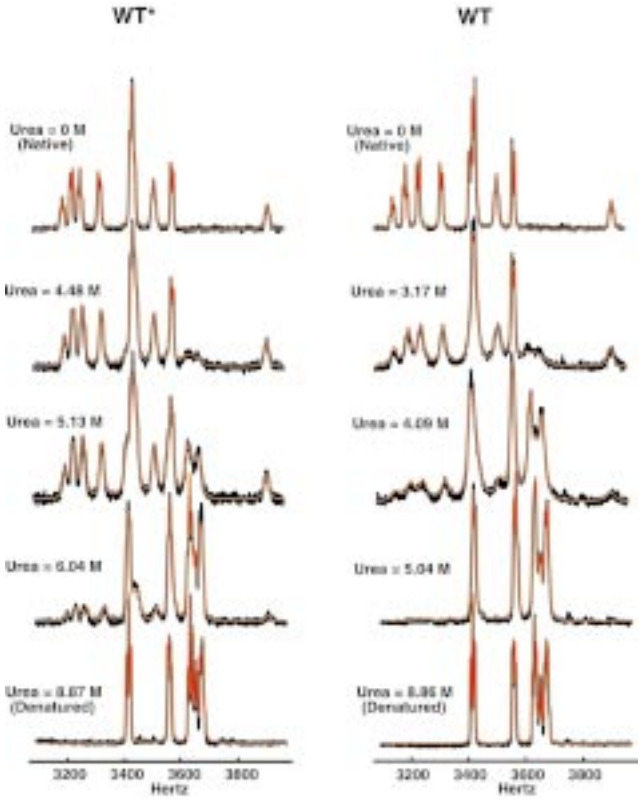

FIGURE 3: Raw data (black) and fitted (red) ^{1}H NMR spectra of WT* and WT at several urea concentrations. Top spectra are in 0 M urea representing the native state; bottom spectra in 8.9 M urea represent the denatured state. Three intermediate concentrations represent the exchanging (transition) region, where both native and denatured states are populated. Fits for other mutants are similar.

previously described (*24*). As an example, Figure 3 shows spectra for WT* and WT, shown in black, and the best-fit spectra, shown in red. A total of 20 spectra were collected for each protein, typically 5 each in the native and denatured baselines which are nonexchanging and 10 in the exchanging (transition) region where both the native and denatured states are at least 3% of the population and significant exchange broadening is observed. The kinetic parameters derived from these fits are given in Table 2. Figures 5 and 6 show the logarithm of the rate constants plotted versus urea concentration for mutants in the WT and WT* backgrounds, respectively. Plots of ln k_f and ln k_u versus urea are reasonably linear over the measured range of urea concentrations. Again, folding rates extrapolated to 0 M urea contain significant errors, so comparison will also be made in 3 M urea, which requires a smaller extrapolation. The results can be analyzed quantitatively by assuming a two-state kinetic mechanism where only the native (folded, N) and denatured (unfolded, D) states are present:

$$N \underset{k_u}{\overset{k_f}{\rightleftharpoons}} D \qquad (7)$$

There is a free energy barrier between N and D called the transition state (TS); the height of this barrier is directly proportional to the folding and unfolding rates (k_f and k_u). Initially we will refer to the transition state (TS) as a specific structure but later will argue it should be considered an ensemble of structures at the very least, and probably not be associated with structures at all but simply a collection of rate-limiting steps.

From the kinetic data the stability of the proteins under conditions of the NMR experiment (given in Table 3) can be compared with the stabilities measured by CD (Table 1). Proteins are typically stabilized by D_2O (*27*), and λ_{6-85} and its mutants are not exceptions. The urea midpoints increased by 0.3 to 0.6 M, and the *m* values increased slightly, causing ΔG_{D-N} to increase by ~1 kcal/mol for each mutant. The $\Delta\Delta G$ and $\Delta\Delta G_{int}$ values are very close to those calculated in H_2O from the CD curves.

WT and WT at 25 °C.* WT* folds at 25 °C with a k_f of around 50 000 s^{-1} (Table 2). This is slightly lower than the folding rate measured at 37 °C [88 000 s^{-1} (*7*)], but within the error of the measurement. Error becomes an important issue as the urea midpoint of WT* is high, and therefore a long extrapolation is required for comparison at 0 M denaturant. WT at 25 °C folds with a k_f around 11 000 s^{-1}, compared to 5 000 s^{-1} at 37 °C (*7*). The extrapolated values at the two temperatures for both WT and WT* are comparable, suggesting a shallow dependence of the folding rates on temperature. At room temperature, WT* folds about 5 times faster than WT and unfolds 9 times slower. This is in contrast to the situation at 37 °C where WT* folds 20 times faster but the unfolding rates are unchanged.

Equilibrium *m* values are proportional to the change in solvent-accessible surface area from native to denatured states (*28*). Analogously, the *k* vs [urea] slopes should be proportional to the change in surface area between the N and D states and the transition state (*29*). This is represented by the α value (*25*). At 37 °C the α value is dramatically different for WT* and WT [0.6 for WT and 0.2 for WT*, (*7*)]. This was interpreted as WT having a much more solvent-exposed transition state than WT* and suggested a major change in mechanism as a result of the Gly to Ala mutations in helix 3. At 25 °C the α values are identical for the two proteins (Table 2), suggesting that the solvent

Table 2: Kinetic Parameters from NMR Line Shape Analysis

	$k_f \times 10^3$ [a] (s^{-1})	k_f^{3M} [b] (s^{-1})	$-m_f$ (M^{-1})	k_u [a] (s^{-1})	k_u^{3M} [b] (s^{-1})	m_u (M^{-1})	α[c]
WT*	50.7 ± 36	751 ± 240	1.42 ± 0.13	1.32 ± 0.86	5.78 ± 1.7	0.49 ± 0.12	0.26 ± 0.08
D14A*	100. ± 12	1130 ± 57	1.50 ± 0.03	40.0 ± 5.4	109 ± 5.3	0.34 ± 0.03	0.18 ± 0.02
S77A*	62.7 ± 28	629 ± 110	1.53 ± 0.10	7.43 ± 2.9	29.1 ± 3.7	0.46 ± 0.09	0.23 ± 0.05
D14,S77A*	93.7 ± 37	1280 ± 210	1.43 ± 0.08	24.3 ± 10	77.7 ± 13	0.39 ± 0.10	0.21 ± 0.06
WT	11.1 ± 3	190 ± 18	1.36 ± 0.06	9.24 ± 2.8	40.5 ± 3.6	0.49 ± 0.07	0.26 ± 0.04
D14A	11.5 ± 3	233 ± 15	1.30 ± 0.07	152 ± 34	508 ± 36	0.40 ± 0.08	0.24 ± 0.05
S77A	10.2 ± 1	124 ± 2.2	1.47 ± 0.02	43.5 ± 3.2	234 ± 5.1	0.56 ± 0.02	0.28 ± 0.02
D14,S77A	18.1 ± 4	283 ± 19	1.39 ± 0.07	96.2 ± 24	283 ± 19	0.44 ± 0.08	0.24 ± 0.05

[a] Rates extrapolated to 0 M urea. [b] Rates extrapolated or interpolated to 3 M urea. [c] $m_u/(m_u - m_f)$, related to solvent accessibility of apparent transition state relative to native and denatured states.

Table 3: Equilibrium Stabilities from NMR Line Shape Analysis in D_2O[a]

	$[urea]_{1/2}$ (M)	m [kcal/(mol·M)]	$\Delta G_{D-N}^{D_2O}$ (kcal/mol)	ΔG_{D-N}^{3M} (kcal/mol)	$\Delta\Delta G^{D_2O}$ (kcal/mol)	$\Delta\Delta G^{3M}$ (kcal/mol)
WT*	5.53	1.13	6.25	2.86	–	–
D14A*	4.25	1.08	4.59	1.35	1.66	1.51
S77A*	4.55	1.18	5.37	1.83	0.88	1.03
D14,S77A*	4.54	1.08	4.90	1.66	1.35	1.20
14−77* interaction[b]					1.19	1.34
WT	3.82	1.10	4.20	0.90		
D14A	2.54	1.01	2.57	−0.46	1.63	1.36
S77A	2.67	1.20	3.20	−0.40	1.00	1.30
D14,S77A	2.86	1.08	3.09	−0.15	1.11	1.05
14−77 interaction[b]					1.52	1.61

[a] Errors from the line-shape fitting give an error in ΔG of ±0.3 kcal/mol in D_2O and ±0.1 kcal/mol in 3 M urea. [b] Interaction free energies based on double mutant cycles such as those shown in Figure 2: $\Delta\Delta G_{int} = \Delta\Delta G(D14A) + \Delta\Delta G(S77A) - \Delta\Delta G(D14,S77A)$.

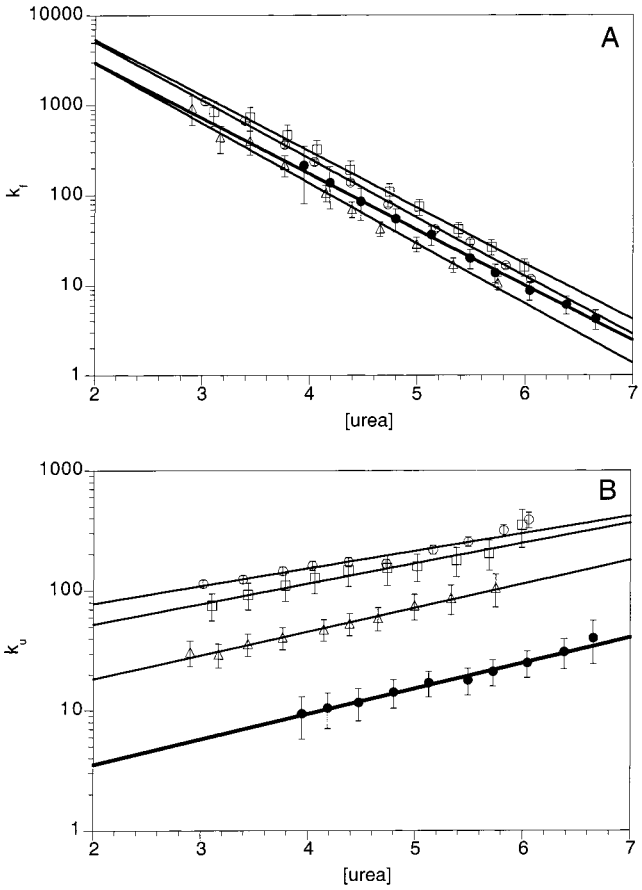

FIGURE 4: Plots of kinetic rate constants of folding (A) and unfolding (B) versus urea concentration for WT* and mutants in the WT* background: WT* (closed circles), D14A* (open circles), S77A* (open triangles), and D14,S77A* (open squares). Lines are fits of the two-state kinetic model (see Materials and Methods).

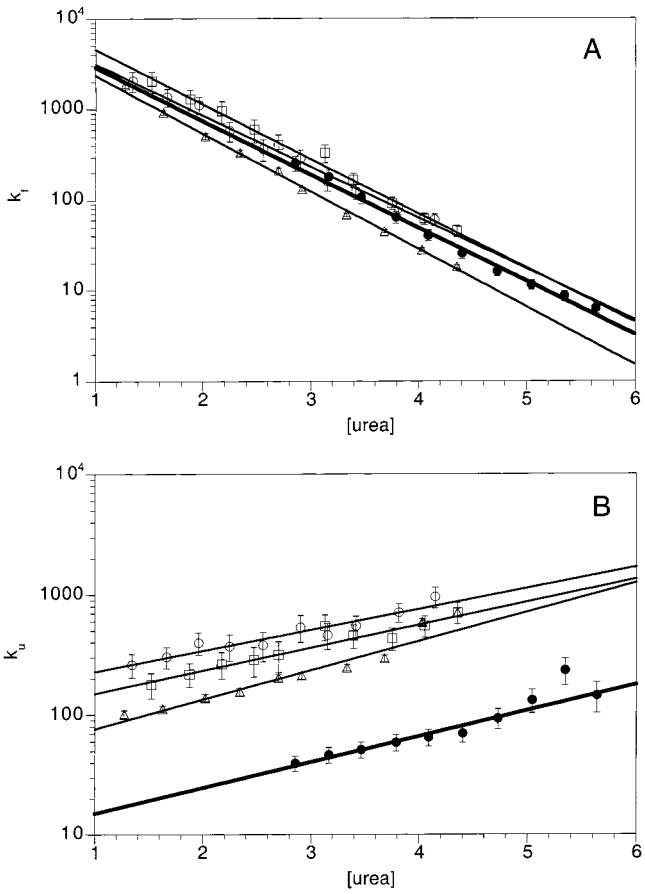

FIGURE 5: Plots of kinetic rate constants of folding (A) and unfolding (B) versus urea concentration for WT and mutants in the WT background: WT (closed circles), D14A (open circles), S77A (open triangles), and D14,S77A (open squares). Lines are fits of the two-state kinetic model (see Materials and Methods).

exposure of the two transition states is similar. This raises the possibility that the kinetic mechanisms of WT and WT* are highly temperature dependent, even though the folding rates themselves are not.

The 14/77 Variants. The data in Figures 4 and 5 show that the mutations did not dramatically change the folding rate in either background but did change the unfolding rates in a similar manner in both WT and WT* backgrounds. From the 0 M urea data in Table 2, it can be seen that Asp 14 to Ala increased k_u by 30-fold in D14A* and 16-fold in D14A. Ser 77 to Ala increased k_u by a factor of 6 in S77A* and 5 in S77A. The double mutants D14,S77A and D14,S77A* show increases intermediate of the two single mutants. At 3 M urea similar trends are evident (Table 2), although the magnitudes are different because of the slightly different values for the slopes. However, the α values for the mutants do not show substantial changes from their respective backgrounds.

From the changes in folding and unfolding kinetics, $\Delta\Delta G_{TS-D}$ and $\Delta\Delta G_{TS-N}$ can be calculated, which represent the change in the free energy of the transition state relative to the denatured and native states, respectively. This change relative to the difference in free energies between the N and D states is termed a Φ value (*14*, *30*, *31*). Φ_f values of 0 indicate that the mutation does not affect the folding kinetics and is therefore not important in the transition state. A Φ_f

Table 4: Φ Values[a]

	$\Delta\Delta G_{TS-D}$ (kcal/mol)	$\Delta\Delta G_{TS-N}$ (kcal/mol)	Φ_f	Φ_u
D14A*	−0.27 ± 0.20	−1.74 ± 0.18	−0.18 ± 0.12	1.15 ± 0.17
S77A*	0.08 ± 0.23	−0.96 ± 0.19	0.08 ± 0.23	0.93 ± 0.21
D14,S77A*	−0.34 ± 0.23	−1.54 ± 0.20	−0.28 ± 0.19	1.28 ± 0.21
14−77* interaction	0.15 ± 0.38	1.16 ± 0.33	0.11 ± 0.28	0.87 ± 0.26
D14A	−0.12 ± 0.06	1.50 ± 0.07	−0.08 ± 0.04	1.10 ± 0.07
S77A	0.25 ± 0.06	1.04 ± 0.05	0.19 ± 0.05	0.80 ± 0.06
D14,S77A	−0.26 ± 0.06	1.29 ± 0.07	−0.25 ± 0.06	1.23 ± 0.09
14−77 interaction	0.39 ± 0.10	1.25 ± 0.11	0.24 ± 0.06	0.78 ± 0.09

[a] Calculated using rates and stabilities in 3 M urea.

of 1 indicates that the free energy perturbation affecting the native conformation equally affects the transition state so the residue is folded in the transition state. The opposite is true for Φ_u values. The Φ values for the single mutants are all close to 0 for Φ_f or 1 for Φ_u[2] (Table 4; Φ values are given in 3 M urea as the errors are quite high for values extrapolated to 0 M). This alone indicates that the hydrogen bond is probably not very important in the transition state. However, we can even more specifically target the contribution of the hydrogen bond. A thermodynamic cycle can be constructed from the $\Delta\Delta G_{TS-D}$ and $\Delta\Delta G_{TS-N}$ values to estimate the contribution of the 14−77 hydrogen bond to the average transition state free energy (*14*). The values found in this manner for the hydrogen bond interaction are given in Table 4. For WT* $\Phi_{f,int}$ is −0.14 at 0 M urea and 0.11 in 3 M urea, and the $\Phi_{u,int}$ is 1.11 at 0 M urea and 0.87 in 3 M urea, suggesting that most pathways lack the hydrogen bond between 14 and 77 at the transition state and that the interaction is formed late in the folding process, in agreement with our diffusion−collision calculation (see below). The Φ values for WT are similar: $\Phi_{f,int}$ is 0.21 in water and 0.24 in 3 M urea, and $\Phi_{u,int}$ is 0.78 in both water and 3 M urea.

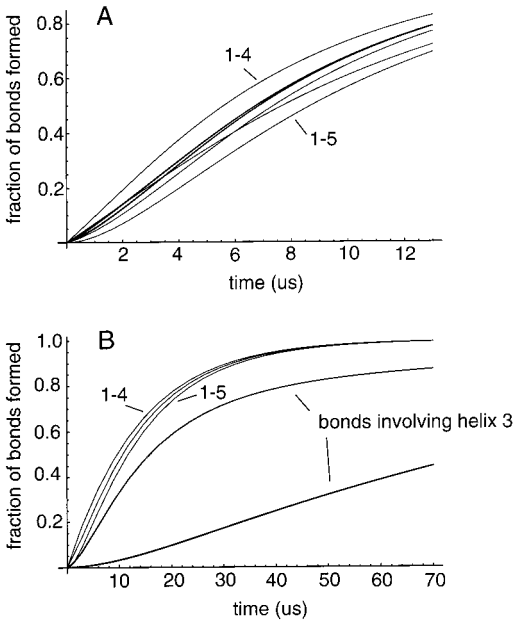

FIGURE 6: Predictions of rates of formation of individual bonds in WT* (A) and WT (B) based on the diffusion−collision model presented in ref 11.

DISCUSSION

Two types of mutation are presently being used to study λ_{6-85}. First are mutations on solvent-exposed faces of helices which affect helix stability, the simplest being alanine to glycine interchanges (*8*). The second type involves removing tertiary interactions to study the interactions between helices. Since diffusion−collision theory predicts rate changes on the basis of helix stability and can predict the relative rates of formation of helix to helix bonds, the kinetic results of these two types of mutations are directly interpretable with the model. The WT to WT* mutations are the first type, and they caused dramatic changes in the properties of the transition state at 37 °C. The model interpretation was that collisions involving helix 3 are highly unproductive in WT and that most of the protein folds before helix 3 does (*11*). As a result, the folding of the protein goes through a bottleneck waiting for helix 3. The alanine substitutions in helix 3 in WT* stabilize the helix so that now there are a variety of pathways open to the protein. However, at 25 °C the transition states become more alike, based on the identical WT* and WT α values and the similar Φ values for the 14/77 mutants in the two backgrounds. Additionally, WT data collected at 30 °C show an α value and rates similar to those at 25 °C (R. Burton, unpublished results) so that a 7° change in temperature is apparently responsible for a dramatic change in the transition state. This plasticity of the apparent transition state might be a characteristic of proteins which fold by multiple parallel pathways. If a mutation slows down the folding through one set of pathways, the protein tends to take others, and the apparent properties of the transition state change.

It is not possible to ask our model "what does the transition state look like?". The diffusion−collision model does not feature a single transition state; rather it has a number of pathways with discrete steps, each with characteristic forward and reverse rates. However, the model does predict rates of formation of the eight microdomain−microdomain bonds in lambda repressor. From the parameters and calculations given in Burton et al. (*11*), Figure 6 was generated, showing the fractional occupancy of each bond as a function of time, starting from a state at time 0 with no bonds formed. On the average, bond 1−4 forms first, and bond 1−5 forms last in the WT* protein (Figure 6A). Therefore, it is likely that the 1−5 bond is formed after the transition state in most pathways, and the results of the experiments presented above are consistent with our theoretical calculation for WT*. Formally, fractional Φ values can arise from a fractional amount of interaction free energy at the TS or by parallel pathways where the interaction is present in the TS of some

[2] Negative Φ_f values and Φ_u values greater than 1 indicate that while destabilizing the native state, the mutations stabilize the transition state.

pathways but not others. Our diffusion−collision model is more directly related to the second interpretation. However, the Φ values for our mutations are close enough to 0 and 1 to allow reasonably unambiguous interpretation. The prediction for the WT protein is that 1−5 forms before bonds involving helix 3 (Figure 6B). Our experimental results suggest that this is incorrect. If the rest of the protein is folded and waiting for helix 3 to fold and form native interactions, then the 14−77 hydrogen bond should show a $\Phi_f \approx 1$, rather than $\Phi_f \approx 0$ as we see experimentally. Helix 3 may fold in most pathways before certain other parts of the molecule can fold, like the 1−5 bond for example.

Since diffusion−collision is a multiple-pathway model, we do not consider the structure of the transition state as a single structure. The average properties of the apparent transition state are the weighted average of the various rate-limiting steps. For WT*, the model predicts (and our experiments suggest) that most molecules follow paths which form the 1−5 bond late in the process, after the rate-limiting steps. The two important considerations for diffusion−collision are proximity in sequence and intrinsic helical stability. Helices 1 and 4 of lambda repressor are predicted to have high intrinsic helix stabilities, while helices 2, 3, and 5 are quite low (*11*). The distance between microdomains 1 and 5 and the low helix propensity of helix 5 hinder their early coalescence, and they come together only when other parts of the protein have already formed. From our results, combined with earlier results from alanine to glycine mutations in WT* (*8*) and the diffusion−collision model (*11*), a picture of the folding of WT* is beginning to emerge. Helices 1 and 4 are formed early. This could be due to bonding to each other or neighboring helices. The rate-limiting steps are the formation and bonding of helices 2 and/or 3, which have lower intrinsic helical stabilities. The bond between the two most distant helices, 1 and 5, forms late, after the rate-limiting steps. Similar studies can be carried out using double mutant cycles with residues representing contacts between other microdomains of the protein, using the contact map shown in Figure 1 as a guide. For example, Figure 4A suggests that a $\Phi_f \approx 1$ would be found for interactions in the 1−4 bond. This type of experiment can also test the assumption that all contacts in a particular bond form simultaneously by examining interactions between different residues comprising the bond.

At present, diffusion−collision is probably only applicable to small, helical proteins that have reasonable intrinsic helical stability and do not fall into non-native kinetic traps, since there is no provision in the model for non-native structure. If the intrinsic stabilities of the secondary structure are too low, then the protein might initiate folding by random hydrophobic collapse and then proceed to organize to the native state while collapsed. Diffusion−collision is probably not appropriate for reactions which are not diffusion limited. The model should be useful to describe the folding of small helical proteins that contain relatively polar sequences and reasonably high intrinsic helical stabilities. For example, myoglobin is another system for which diffusion−collision calculations can be compared with experiment (*32*, *33*). The model is consistent with the hierarchic view of protein folding, in which local factors dominate early folding events, and the native structure is built-up progressively from smaller structural elements (*34*, *35*).

Diffusion−collision holds the promise of de novo prediction of folding and unfolding kinetics based solely on the native structure. With ever increasing ability to predict native structures from sequence and continued understanding of the factors contributing to protein stability, the structure and the folding thermodynamics and kinetics of a protein will one day be accurately predictable directly from its sequence. Considering the enormous number of protein sequences being found by DNA sequencing of the genomes of humans and other organisms, an advanced predictive ability will be necessary to take full advantage of the sequence data.

ACKNOWLEDGMENT

We thank Randy Burton for help with diffusion−collision calculations and NMR line-shape analysis and Tiffany Calderone for help with mutagenesis and protein expression. We also thank Gordon Hammes and the members of the Oas laboratory group for helpful discussions.

REFERENCES

1. Baldwin, R. L. (1995) *J. Biomol. NMR 5*, 103−109.
2. Dill, K. A., and Chan, H. S. (1997) *Nat. Struct. Biol. 4*, 10−19.
3. Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z., and Wolynes, P. G. (1996) *Folding Des. 1*, 441−450.
4. Matthews, B. W. (1995) *Adv. Protein Chem. 46*, 240−311.
5. Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) *FASEB J. 10*, 75−83.
6. Huang, G. S., and Oas, T. G. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6878−6882.
7. Burton, R. E., Huang, G. S., Daugherty, M. A., Fullbright, P. W., and Oas, T. G. (1996) *J. Mol. Biol. 263*, 311−322.
8. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L., and Oas, T. G. (1997) *Nat. Struct. Biol. 4*, 305−310.
9. Beamer, L. J., and Pabo, C. O. (1992) *J. Mol. Biol. 227*, 177−196.
10. Marqusee, S., and Sauer, R. T. (1994) *Protein Sci. 3*, 2217−2225.
11. Burton, R. E., Myers, J. K., and Oas, T. G. (1998) *Biochemistry 37*, 5337−5343.
12. Karplus, M., and Weaver, D. L. (1994) *Protein Sci. 3*, 650−668.
13. Fersht, A. R. (1993) *FEBS Lett. 325*, 5−16.
14. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) *J. Mol. Biol. 254*, 260−288.
15. Milla, M. E., Brown, B. M., Waldburger, C. D., and Sauer, R. T. (1995) *Biochemistry 34*, 13194−13919.
16. Martinez, J. C., Pisabarro, M. T., and Serrano, L. (1998) *Nat. Struct. Biol. 5*, 721−729.
17. Grantcharova, V. P., Riddle, D. S., Santiago, J. V., and Baker, D. (1998) *Nat. Struct. Biol. 5*, 714−720.
18. Kuhlman, B., Boice, J. A., Fairman, R., and Raleigh, D. P. (1998) *Biochemistry 37*, 1025−32.
19. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol. 154*, 367−382.
20. Pace, C. N. (1986) *Methods Enzymol. 131*, 266−280.
21. Nicholson, E. M., and Scholtz, J. M. (1996) *Biochemistry 35*, 11369−11378.
22. Horovitz, A., and Fersht, A. R. (1990) *J. Mol. Biol. 214*, 613−617.
23. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) *Protein Sci. 4*, 2411−2423.
24. Burton, R. E., Busby, R. S., and Oas, T. G. (1998) *J. Biomol. NMR 11*, 355−360.
25. Chen, B.-L., Baase, W. A., Nicholson, H., and Schellman, J. A. (1992) *Biochemistry 31*, 1464−1476.
26. Myers, J. K., and Pace, C. N. (1996) *Biophys. J. 71*, 2033−2039.

27. Hvidt, A., and Nielsen, S. O. (1966) *Adv. Protein Chem. 21*, 287−386.
28. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) *Protein Sci. 4*, 2138−2148.
29. Tanford, C. (1970) *Adv. Protein Chem. 24*, 1−95.
30. Matouschek, A., Serrano, L., and Fersht, A. R. (1994) in *Mechanisms of protein folding* (Pain, R. H., Ed.) pp 137−159, IRL Press, Oxford.
31. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) *J. Mol. Biol. 224*, 771−782.
32. Pappu, R. V., and Weaver, D. L. (1998) *Protein Sci. 7*, 480−490.
33. Cavagnero, S., Dyson, H. J., and Wright, P. E. (1999) *J. Mol. Biol. 285*, 269−282.
34. Baldwin, R. L., and Rose, G. D. (1999) *Trends Biochem. Sci. 24*, 26−33.
35. Baldwin, R. L., and Rose, G. D. (1999) *Trends Biochem. Sci. 24*, 77−83.
36. Kraulis, P. K. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI990088X