Article No. mb982200 J. Mol. Biol. (1998) 284, 807–815

JMB

The Single Helix in Protein L is Largely Disrupted at the Rate-limiting Step in Folding

David E. Kim[1], Qian Yi[1], Sharon T. Gladwin[1], Jonathan M. Goldberg[2] and David Baker[1]*

[1]*Department of Biochemistry University of Washington Seattle, WA 98195, USA*

[2]*Department of Biochemistry Stanford University, Stanford CA 94305, USA*

**Corresponding author*

To investigate the role of helix formation in the folding of protein L, a 62 residue α/β protein, we studied the consequences of both single and multiple mutations in the helix on the kinetics of folding. A triple mutant with 11 additional carbon atoms in core residues in the amino-terminal portion of the helix folded substantially faster than wild type, suggesting that hydrophobic association with residues elsewhere in the protein occurs at the rate-limiting step in folding. However, helix-destabilizing mutations had little effect on the rate of folding; in particular, a triple glycine substitution on the solvent-exposed side of the helix increased the unfolding rate 56-fold while reducing the folding rate less than threefold. Thus, in contrast to the predictions of models of folding involving the coalescence of well-formed secondary structure elements, the single helix in protein L appears to be largely disrupted at the rate-limiting step in folding and unfolding.

Keywords: protein folding kinetics; protein L; α-helix formation; hydrophobic association; transition state

Introduction

The role of helix formation in the folding of small proteins has been the subject of considerable discussion in recent years (Burton *et al.*, 1998; Dill, 1990; Fersht, 1995; Karplus & Weaver, 1994; Kim & Baldwin, 1990; López-Hernández *et al.*, 1997; Sosnick *et al.*, 1996). Here, we probe the contribution of helix formation to the folding of a particularly simple model protein, the 62 residue IgG-binding domain of peptostreptococcal protein L. Protein L consists of a four-stranded β-sheet packed against a single α-helix (Figure 1, Wikström *et al.*, 1994) and is well suited for studying the role of helix formation in folding. The thermodynamics and kinetics of folding of a tryptophan-containing variant (referred to as protein L) are well characterized (Scalley *et al.*, 1997; Yi *et al.*, 1997), and a phage display selection method has been developed allowing the retrieval of functional folded variants from high-complexity libraries (Gu *et al.*, 1995). The selection strategy is based on the premise that, since the binding residues are not contiguous in sequence, the protein must be properly folded in order to bind IgG.

Here, we investigate the role of helix formation in the folding of protein L by studying the consequences of mutations distributed throughout the helix on the rate of folding and unfolding. We find that increasing the size of core residues in the helix significantly increases the folding rate, while helix-destabilizing mutations increase the unfolding rate and have little effect on the folding rate. Thus, the non-polar residues in the helix appear to engage in hydrophobic contacts in the rate-limiting step to folding without appreciable formation of the helix.

Results

In a previous study aimed at determining whether folding rates are extensively optimized by natural selection, functional folded variants of protein L were retrieved from libraries in which different segments of the protein were independently randomized (Kim *et al.*, 1998). In one of these libraries, the amino-terminal half of the helix was randomized (the carboxy terminus contributes to the binding site (Wikström *et al.*, 1995) and was not subjected to randomization). The frequency of the amino acid residues recovered at each position in the functional sequences relative to the expected

Abbreviation used: LYI, A29L/T30Y/A33I; GuHCl, guanidine hydrochloride.

E-mail address of the corresponding author: baker@ben.bchem.washington.edu

0022–2836/98/480807–09 $30.00/0

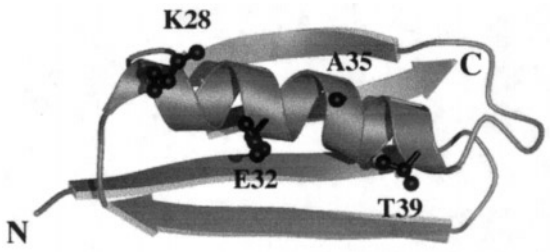
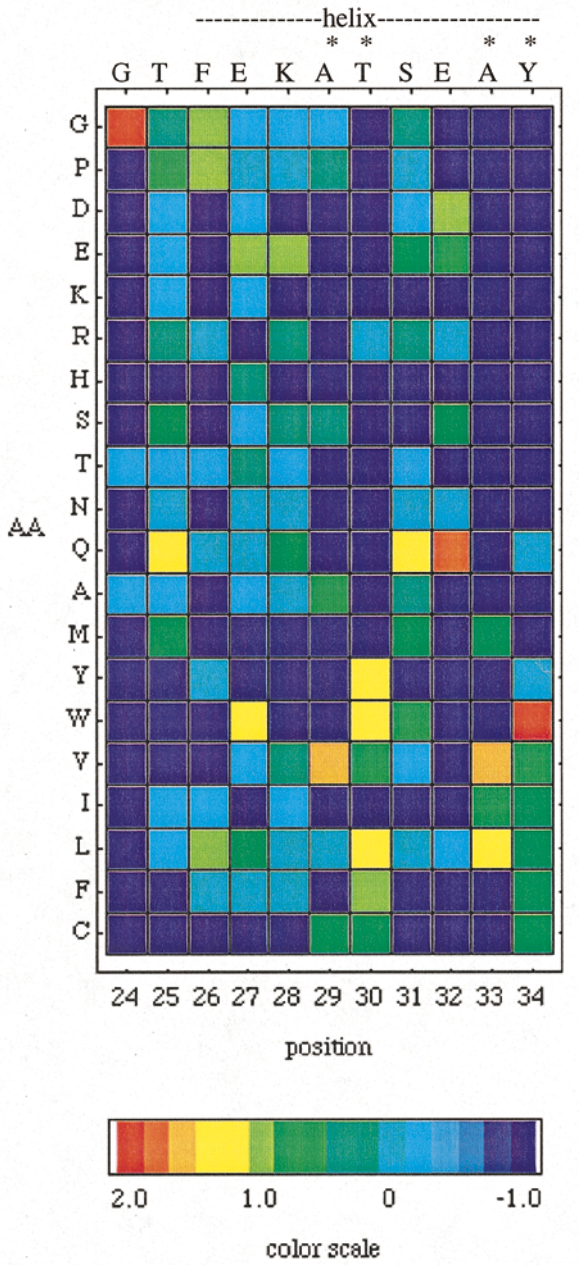


Figure 1. Backbone ribbon trace of protein L indicating sites of helix-destabilizing mutations. The Figure was prepared using MOLSCRIPT (Kraulis, 1991) and Raster3D (Bacon & Anderson, 1988; Merritt & Murphy, 1994).



frequency in the unselected population is displayed in Figure 2. Residues buried at the helix-sheet interface were replaced primarily with larger hydrophobic residues (asterisks in Figure 2), and the amino-terminus of the helix tolerated substantial sequence variation including proline residues. Biophysical characterization of two of the variants showed that although the multiple substitutions decreased stability, the folding rates were increased compared to wild-type (Table 1).

The sequence and kinetic data on the mutants retrieved from the combinatorial selection (Figure 2; Table 1) suggested that increases in the size of hydrophobic core residues may increase the folding rate. To investigate this possibility, we constructed and characterized a triple mutant, A29L/T30Y/A33I (LYI), based on the sequence of variant h-c (Table 1), with a substantial increase in the volume of the core residues. The circular dichroism spectrum of this mutant displayed a decrease in negative ellipticity at 222 nm (Figure 3A). To determine whether this change in ellipticity was due to a reduction in helical structure, the spectrum of LYI was subtracted from that of wild-type. The resulting difference spectrum (Figure 3B) exhibits minima at 205 and 222 nm, and closely resembles the spectrum of an isolated α-helix (Greenfield & Fasman, 1969). These results suggest a partial disruption of the helix in the triple mutant.

The stability of LYI was determined from GuHCl (guanidine hydrochloride) equilibrium denaturation experiments (see Materials and Methods). Despite the addition of 11 carbon atoms to the core, the triple mutant was only slightly destabilized compared to wild-type (Table 1).

Figure 2. Sequence variation in functional protein L mutants recovered from a phage library displaying protein L variants with the amino-terminal half of the helix randomized. The frequency of occurrence of each of the amino acids was roughly equal prior to selection (Kim *et al.*, 1998). The wild-type sequence and residue numbers are represented above and below the plot, respectively. The helix and buried residues (*) are indicated above the wild-type sequence. Each box represents the logarithm of the ratio of the frequency of occurrence in the selected population to the frequency in the unselected population according to the color scale at the bottom of the Figure; pseudo-counts were added to avoid taking the logarithm of zero. Red and blue indicate a strong selection for and against residues at a particular position, respectively, and green indicates a frequency similar to the unselected population. The amino acids are arranged from polar to non-polar in rows 3-19, with glycine, proline and cysteine in rows 1, 2 and 20, respectively. Features to note are the recovery of proline residues in the amino-terminal region of the helix, the high frequency of valine and leucine, and low frequency of alanine at position 33, the strong bias towards hydrogen bonding residues at position 32, and the strong conservation of glycine in the turn preceding the helix at position 24.

Table 1. Thermodynamic and kinetic parameters

	Thermodynamic		Kinetic				
	m[a] (kcal mol^{-1} M^{-1})	$\Delta\Delta G^{2M}_{U-F}$ (kcal mol^{-1})	$-m_f$ (kcal mol^{-1} M^{-1})	m_u	$k_f^{H_2O}$ (s^{-1})	k_u^{2M}[e]	$\Phi_F^{H_2O}$
Wild type	1.9		1.5	0.5	61	0.05	
A. *Combinatorial mutants*[b]							
h-c	2.2	1.6	1.5	0.7	158	1.6	
h-d	3.1	4.4	2.2	0.3	90	12	
B. *Triple mutants*							
A29L, T30Y, A33I	2.2	0.7	1.3	0.7	205	0.65	−0.90[d]
E32G, A35G, T39G	2.1	2.9	1.8	0.7	25	2.8	0.19
C. *Helix destabilized mutants*							
K28G	1.7	−0.1	1.3	0.5	47	0.05	[c]
E32G	1.9	0.7	1.5	0.6	48	0.20	0.12
E32I	2.0	0.9	1.5	0.6	55	0.23	0.06
A35G	2.0	1.3	1.5	0.5	34	0.24	0.29
T39G	1.9	0.2	1.5	0.6	109	0.10	−1.7[d]

Standard errors obtained from least squares fitting ranged from 2–20%, and are likely to be underestimates of the true values.
[a] Dependence of the free energy of unfolding on GuHCl concentration.
[b] The sequences of h-c, h-d and wild-type from position 24 to 34 are GEYHTLYQEIW, GSGLEALGSVL, and GTFEKATSEAY, respectively (Kim *et al.*, 1998).
[c] $\Phi_{F_2}^{H_2O}$ could not be determined accurately because K28G was not destabilized.
[d] Negative $\Phi_{F_2}^{H_2O}$ values were due to increased folding rates compared to wild-type.
[e] Unfolding rate constant in 2 M GuHCl. All remaining parameters are described in the text.

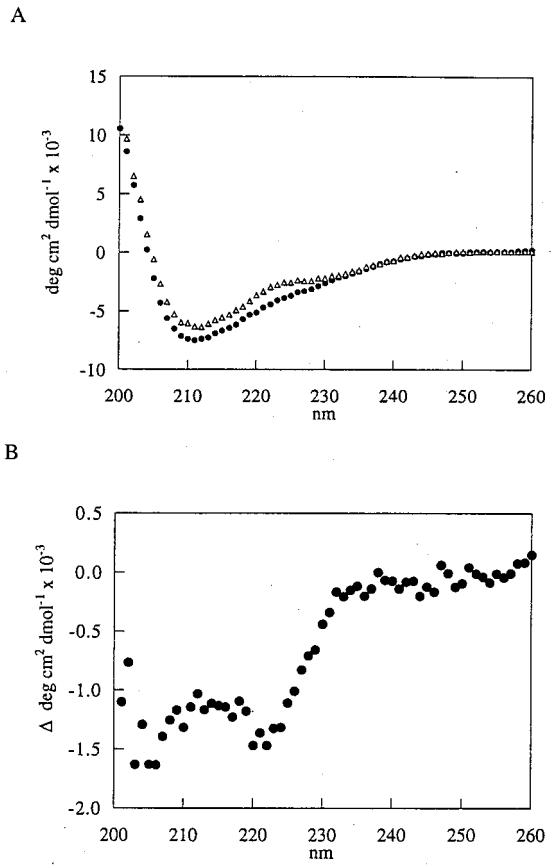


Figure 3. The helix is partially disrupted in the LYI mutant. A, Circular dichroism spectra of LYI (A29L, T30Y, and A33I; open triangles) and wild-type protein L (closed circles). B, Difference spectrum of LYI relative to wild-type protein L.

Stopped-flow kinetic studies showed that the folding rate of LYI was more than three times greater than that of the wild-type protein (Table 1), despite the partial disruption of the helix suggested by the CD spectrum. Both the folding and unfolding kinetics were similar to those of variant h-c over a broad range of GuHCl concentrations (Figure 4),

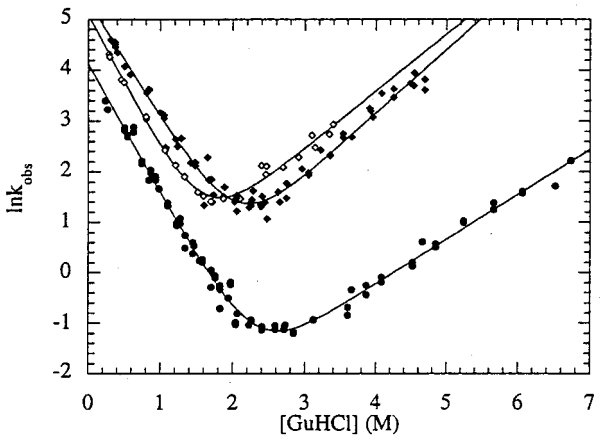


Figure 4. Observed folding and unfolding rates of LYI, h-c, and wild-type. At low GuHCl concentrations (left branch of curves), the relaxation rate is dominated by the folding rate; data points were collected by rapidly diluting denatured protein with buffer containing variable amounts of denaturant using a stopped-flow instrument. At high GuHCl concentrations (right branch of curves), the relaxation rate is dominated by the unfolding rate; data points were collected by rapidly mixing folded protein with GuHCl. Wild-type (filled circles); LYI (filled diamonds); and h-c (open diamonds).

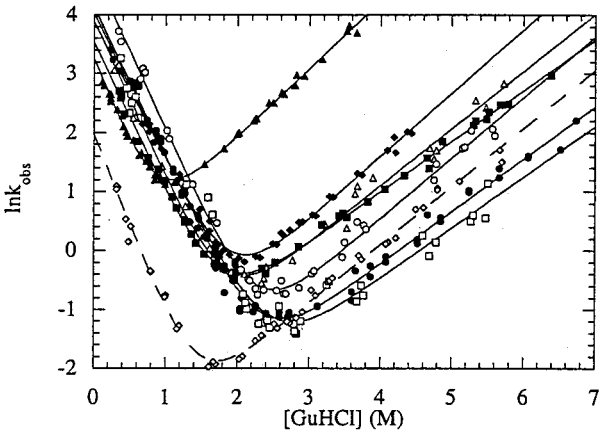


Figure 5. Folding and unfolding rates of helix-destabilizing mutants. Wild-type (filled circles), K28G (open squares) E32I (filled diamonds), E32G (open triangles), A35G (filled squares), T39G (open circles), and triple glycine mutant, E32G, A35G and T39G (filled triangles). The helix-destabilizing mutants increase the rate of unfolding but have relatively little effect on the folding rate. The denaturant dependence of the folding and unfolding rates of the G15A substitution in the first beta turn (Gu *et al.*, 1997), which substantially decreases the folding rate, is also shown for comparison (open diamonds).

suggesting that the increase in folding rates observed for variants h-c and h-d (Table 1) were likely to result from the increases in the size of hydrophobic core residues in the helix. Since LYI is not significantly destabilized, the increase in folding and unfolding rates suggests that the increase in size of the core residues stabilizes the folding transition state relative to the folded and unfolded states. Taken together, the CD data and kinetic data suggest that the rate-limiting step in folding of protein L does not require complete formation of the helix, but does involve hydrophobic association. A more complete interpretation of these results will require higher-resolution structural characterization of the LYI mutant.

To more directly probe the extent of helix formation in the folding transition state, we studied the effects of helix-destabilizing substitutions at solvent-exposed positions (K28G, E32I, E32G, A35G and T39G) on the kinetics of folding and unfolding (Figure 1). Since these mutations affect only local interactions, they probe the consequences of reductions in the population of molecules with fully formed helix on the rate of folding without potentially complicating changes in tertiary interactions. A similar approach was used to study the role of helix formation in the fully helical coiled-coil region of GCN4 (Sosnick *et al.*, 1996).

The stabilities of the mutants were determined using equilibrium GuHCl denaturation experiments. K28G and T39G, at the amino and carboxy termini of the helix (Figure 1), respectively, do not significantly reduce the stability of the protein (Table 1). This may be because partial fraying at the ends of the helix is compensated by the additional degrees of freedom in the adjoining turns (mutations at helix termini in isolated helical peptides have relatively little effect on helix content (Chakrabartty *et al.*, 1991)). The small change in stability could also reflect the elimination of an unfavorable charge-helix dipole interaction (K28G) and the low helix propensity of threonine. However, mutations at solvent-exposed positions in the middle of the helix (E32G, E32I, and A35G) destabilize the native state by 0.7-1.3 kcal/mol. The E32I mutant was constructed, in part, because very few hydrophobic residues were recovered at this position in the phage selection (Figure 2). Since the point mutation is no more destabilized than many of the variants recovered in the selection (Table 1), the conservation of residues capable of hydrogen bonding probably reflects a role for the glutamate in IgG binding (the IgG-binding site defined by chemical shift perturbation studies does not include this residue, but does include nearby residues in the carboxy-terminal region of the helix; Wikström *et al.*, 1995)).

The characterization of the kinetics of folding and unfolding revealed that the helix-destabilizing mutations increase the rate of unfolding significantly more than they decrease the rate of folding. The modest destabilization of the T39G mutant and the larger destabilization of the E32G, E32I and A35G mutants are mostly accounted for by increases in the unfolding rates (Table 1 and Figure 5), suggesting that the mutations destabilize the native state considerably more than the transition state. Interestingly, the T39G mutation increased the rate of folding, suggesting that the transition state ensemble includes conformations with chain reversals or bends at the C terminus of the helix. The simplest explanation for the much larger effect of the mutations on the native state than on the transition state is that the helix is not completely formed in the transition state.

A possible requirement for formation of only a portion of the helix in the folding transition state could potentially have been missed in the single mutant studies, since only a portion of the helix was destabilized in each experiment. To destabilize helix formation along the entire length of the helix, a triple mutant was designed in which the solvent-exposed residues E32, A35 and T39 (Figure 1) were simultaneously changed to glycine residues. The addition of three glycine residues in the helix dramatically decreases the helix propensity (Chakrabartty *et al.*, 1994) and thus should decrease the population of helix in the unfolded state. The mutant is significantly destabilized ($\Delta\Delta G^{2M}_{U-F} = 2.9$ kcal/mol), and the loss in stability is almost entirely due to an increase in the unfolding rate; the unfolding rate increased 56-fold while the folding rate decreased less than threefold (Table 1, Figure 5). These results, taken together with the helix-destabilizing point mutations

described above, strongly suggest that the helix is disrupted at the rate-limiting step in folding.

The kinetic results described above may be conveniently summarized using the Φ_F value parameter ($\Delta\Delta G_{\ddagger-U}/\Delta\Delta G_{F-U}$), which reflects the extent to which interactions removed by mutation are present in the transition state in a simple two-state model; a value of 1 indicates that the site of mutation is as structured in the transition state as in the native state, whereas a value of zero reflects the absence of such structure in the transition state (Matouschek *et al.*, 1989). An important requirement for this analysis is that the nature of the transition state, inferred by the estimated amount of hydrophobic surface exposed in the transition state relative to the folded state (m_u/m), is not perturbed upon mutation. The m_u and m values for all of the helix-destabilizing mutants are similar to those of wild-type (Table 1). The Φ_F values for the helix-destabilizing mutants are all less than 0.30 (Table 1). In particular, the Φ_F values for the two glycine substitutions in the center of the helix were 0.12 and 0.29, and the triple glycine mutant had a low Φ_F value of 0.19, suggesting that the helix is not appreciably structured along its entire length in the folding transition state ensemble.

Discussion

The role of the helix in folding

The results described here suggest that the helix in protein L is largely disrupted at the rate-limiting step in folding and unfolding. The helix-destabilizing mutations have much larger effects on the rate of unfolding than on the rate of folding; the triple glycine substitution in particular increases the unfolding rate 56-fold, while decreasing the folding rate less than three-fold. The increase in folding rate accompanying the introduction of 11 additional carbon atoms to the core of the LYI mutant suggests that hydrophobic interactions involving residues in the helix partially stabilize the transition state ensemble despite the absence of well-formed helix. In the reverse direction, the large effects of the helix-destabilizing mutations on the unfolding rate suggests that disruption of the helix is a key step in disassembly of the hydrophobic core in unfolding.

The Φ values for the helix-destabilizing mutations are greater than zero (0.06–0.29). How are these partial Φ values best interpreted? For hydrophobic core mutations, partial Φ values would result if a residue made some but not all of its interactions in the transition state (Fersht *et al.*, 1992). For solvent-exposed residues, this interpretation seems less plausible, since such residues make relatively few interactions even in the native state. Perhaps the simplest interpretation is that the local structural element containing the solvent-exposed residue is partially ordered in the transition state ensemble (Itzhaki *et al.*, 1995a; Sosnick *et al.*, 1996). For glycine substitutions on the solvent-exposed face of an α-helix, Φ values may report on the fraction of the Ramachandran plot sampled in the transition state ensemble relative to the denatured and native states (Sosnick *et al.*, 1996). Alanine to glycine substitutions can increase chain entropy significantly because of the larger range of backbone torsion angles accessible to glycine residues; the entropy of the denatured state is expected to increase considerably more than that of the native state because the extensive interactions in the native state effectively confine the values of the torsion angles. Partial Φ values for glycine substitutions would result if the surrounding chain segment in the transition state was less constrained than in the native state but more constrained than in the denatured state. For example, if a glycine substitution produced three- and four-fold increases in the number of states accessible in the transition state and denatured state ensembles relative to the native state, respectively, the resulting Φ value due to chain entropy changes alone would be $(\ln 4 - \ln 3)/(\ln 4 - \ln 1) = 0.21$. While this ''partial ordering'' interpretation seems plausible, other interpretations cannot be ruled out (see the Appendix).

In contrast to our results with protein L, the transition states of previously studied monomeric helix-containing proteins all appear to have well-formed helices: Φ values of greater than 0.5 were observed for helix-destabilizing mutations in barnase (Serrano *et al.*, 1992), CI2 (Itzhaki *et al.*, 1995a) and two of the helices in λ repressor (Burton *et al.*, 1997). The protein L results are more similar to those obtained for the dimeric coiled-coil region of GCN4: alanine to glycine mutations on the solvent-exposed surface of GCN4 increased the unfolding rate but had little effect on the folding rate (Sosnick *et al.*, 1996). The differences between protein L and the other monomeric helix-containing proteins may be related to the differences in the topology of the proteins and to our previous finding that the first β-hairpin is formed at the rate-limiting step in folding of protein L (Gu *et al.*, 1997; the folding kinetics of a glycine to alanine substitution in this turn are shown in Figure 5 for comparison). The folding nucleus of monomeric proteins may consist of a relatively well-formed local structural element making loose hydrophobic interactions with residues elsewhere in the chain; for protein L, this local structural element may be the β-hairpin rather than a part of the helix. The important role of the first hairpin may also explain the apparent lack of a role for either the amino or carboxy terminus of the helix in the early formation of structure in protein L; in CI2, a helix N-cap was one of the most ordered regions of the protein in the transition state (Itzhaki *et al.*, 1995a), while in barnase a helix C-cap was found to be structured in the transition state (Serrano *et al.*, 1992). The observation of rapid folding in the protein L variants obtained in the phage selection

is perhaps the most dramatic illustration of the lack of importance of regular structure at the helix amino terminus; the sequences often contain proline and generally have low helix propensity (Chakrabartty *et al.*, 1994; Figure 2), and the CD spectra of the variants are consistent with partial disruption of the helix (data not shown).

Perhaps the two most characteristic features of globular proteins are the presence of a hydrophobic core and the large amounts of regular secondary structure (sheets and helices) and thus there has been considerable debate over the past decade over the relative timing of hydrophobic collapse and secondary structure formation in folding (Dill, 1990; Fersht, 1995; Karplus & Weaver, 1994; Kim & Baldwin, 1990). Our results indicate that for protein L, hydrophobic interactions of non-polar residues in the helix with residues elsewhere in the protein are more critical to the rate-limiting step in folding than is the formation of the helix. It is interesting to note that almost all protein tertiary structure prediction methods (Bowie & Eisenberg, 1994; Monge *et al.*, 1994; Simons *et al.*, 1997; Skolnick *et al.*, 1997; Srinivasan & Rose, 1995) assemble tertiary structures hierarchically from predicted local structural elements, typically helices and strands; the results reported here suggest that this may not be the way many proteins fold.

Materials and Methods

All reagents, solutions, and enzymes used for molecular biology procedures were as described by Gu *et al.* (1995) and Scalley *et al.* (1997). Point mutations were made with the QuickChange site-directed mutagenesis kit (Strategene). Expression and purification were carried out as described by Gu *et al.* (1995). All mutants were verified by DNA sequencing and mass spectrometry. Equilibrium CD and fluorescence measurements, stopped-flow experiments, and data analysis were carried out as described by Gu *et al.* (1997) and Scalley *et al.* (1997).

The free energy of folding of protein L appears to be a non-linear function of denaturant and, as a result, standard linear extrapolation estimates of the free energy of unfolding in the absence of denaturant ($\Delta G^{H_2O}_{U-F}$) may involve considerable errors (Yi *et al.*, 1997). To avoid the extrapolation, the change in free energy of unfolding between mutant and wild-type ($\Delta\Delta G^{2M}_{U-F}$) was calculated at 2 M GuHCl.

The refolding and unfolding data were fit by a two-state model using the following equation:

$$\ln k_{obs}[\mathrm{GuHCl}] = \ln[k_f^{H_2O}\exp(-m_f[\mathrm{GuHCl}]) + k_u^{H_2O}\exp(m_u[\mathrm{GuHCl}])] \quad (1)$$

where k_{obs} [GuHCl] is the observed relaxation rate at a given concentration of GuHCl, $k_f^{H_2O}$ and $k_u^{H_2O}$ are the refolding and unfolding rate constants in the absence of denaturant, respectively, and m_f and m_u describe the dependence of the logarithm of the rate constant on GuHCl concentration. $\Phi_F^{H_2O}$ (Itzhaki *et al.*, 1995a) was determined for each point mutant using the following equations:

$$\Phi_F^{H_2O} = \Delta\Delta G^{H_2O}_{\ddagger-U}/\Delta\Delta G_{F-U} \quad (2)$$

$$\Delta\Delta G^{H_2O}_{\ddagger-U} = -RT\ln(k^{H_2O}_{f,wt}/k^{H_2O}_{f,mut}) \quad (3)$$

$$\Delta\Delta G_{F-U} = \langle m\rangle([\mathrm{GuHCl}]_{50\%mut} - [\mathrm{GuHCl}]_{50\%,wt}) \quad (4)$$

where $k^{H_2O}_{f,wt}$ and $k^{H_2O}_{f,mut}$ are the folding rate constants in the absence of denaturant for wild-type and mutant, respectively, $\Delta\Delta G_{F-U}$ is the change in free energy of folding between wild-type and mutant in the absence of denaturant, $\langle m\rangle$ is the average m value calculated from the mutants and wild-type ($\langle m\rangle = 2.0$ kcal mol^{-1} M^{-1}), and $[\mathrm{GuHCl}]_{50\%,mut}$ and $[\mathrm{GuHCl}]_{50\%,wt}$ are the concentrations of GuHCl where 50% of the protein is folded for mutant and wild-type, respectively.

Acknowledgments

We thank Carol Rohl and members of the Baker laboratory for helpful comments on the manuscript, and Tobin Sosnick for helpful discussions. This work was supported by a grant from the NIH (GM5188) and by young investigator awards to D.B. from the NSF and the Packard Foundation.

References

Bacon, D. J. & Anderson, W. F. (1988). A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graph.* **6**, 219–220.

Bowie, J. U. & Eisenberg, D. (1994). An evolutionary approach to folding small α-helical proteins that uses sequence information and an empirical guiding fitness function. *Proc. Natl Acad. Sci. USA*, **91**, 4436–4440.

Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L. & Oas, T. G. (1997). The energy landscape of a fast-folding protein mapped by Ala → Gly substitutions. *Nature Struct. Biol.* **4**, 305–310.

Burton, R. E., Myers, J. K. & Oas, T. G. (1998). Protein folding dynamics: quantitative comparison between theory and experiment. *Biochemistry*, **37**, 5337–5343.

Chakrabartty, A., Schellman, J. A. & Baldwin, R. L. (1991). Large differences in the helix propensities of alanine and glycine. *Nature*, **351**, 586–588.

Chakrabartty, A., Kortemme, T. & Baldwin, R. L. (1994). Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* **3**, 843–852.

Cleland, W. W. (1975). Partition analysis and the concept of net rate constants as tools in enzyme kinetics. *Biochemistry*, **14**, 3220–3224.

Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, **29**, 7133–7155.

Fersht, A. R. (1995). Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl Acad. Sci. USA*, **92**, 10869–10873.

Fersht, A. R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782.

Fersht, A. R., Itzhaki, L. S., El Masry, N. F., Matthews, J. M. & Otzen, D. E. (1994). Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc. Natl Acad. Sci. USA*, **91**, 10426–10429.
Gilmanshin, R., Williams, S., Callender, R. H., Woodruff, W. H. & Dyer, R. B. (1997). Fast events in protein folding: relaxation dynamics of secondary and tertiary structure in native apomyoglobin. *Proc. Natl Acad. Sci. USA*, **94**, 3709–3713.
Greenfield, N. & Fasman, G. D. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*, **8**, 4108–4116.
Gu, H., Yi, Q., Bray, S. T., Riddle, D. S., Shiau, A. K. & Baker, D. (1995). A phage display system for studying the sequence determinants of protein folding. *Protein Sci.* **4**, 1108–1117.
Gu, H., Kim, D. & Baker, D. (1997). Contrasting roles for symmetrically disposed β-turns in the folding of a small protein. *J. Mol. Biol.* **274**, 588–596.
Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995a). The structure of the transition state for folding of chymotrypsin inhibitor 2 analyzed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* **254**, 260–288.
Itzhaki, L. S., Neira, J. L., Ruiz-Sanz, J., de Prat, Gay G. & Fersht, A. R. (1995b). Search for nucleation sites in smaller fragments of chymotrypsin inhibitor 2. *J. Mol. Biol.* **254**, 289–304.
Karplus, M. & Weaver, D. L. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci.* **3**, 650–668.
Kim, D. E., Gu, H. & Baker, D. (1998). The sequences of small proteins are not extensively optimized for rapid folding by natural selection. *Proc. Natl Acad. Sci. USA*, **95**, 4982–4986.
Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59**, 631–660.
Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
López-Hernández, , Cronet, P., Serrano, L. & Muñoz, V. (1997). Folding kinetics of Che Y mutants with enhanced native α-helix propensities. *J. Mol. Biol.* **266**, 610–620.
Matouschek, A., Kellis, J. T., Serrano, L. & Fersht, A. R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature*, **340**, 122–126.
Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta. Crystallog. sect. D*, **50**, 869–873.
Monge, A., Friesner, R. A. & Honig, B. (1994). An algorithm to generate low-resolution protein tertiary structures form knowledge of secondary structure. *Proc. Natl Acad. Sci. USA*, **91**, 5027–5029.
Scalley, M. L., Yi, Q., Gu, H., McCormack, A., Yates, J. R., III & Baker, D. (1997). Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry*, **36**, 3373–3382.
Serrano, L., Matouschek, A. & Fersht, A. R. (1992). The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analyzed by a protein engineering procedure. *J. Mol. Biol.* **224**, 805–818.
Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225.
Skolnick, J., Kolinski, A. & Ortiz, A. R. (1997). MONSSTER: a method for folding globular proteins with a small number of distance restraints. *J. Mol. Biol.* **265**, 217–241.
Sosnick, T. R., Jackson, S., Wilk, R. R., Englander, S. W. & DeGrado, W. F. (1996). The role of helix formation in the folding of a fully α-helical coiled coil. *Proteins: Struct. Funct. Genet.* **24**, 427–432.
Srinivasan, R. & Rose, G. D. (1995). LINUS: A hierarchic procedure to predict the fold of a protein. *Proteins: Struct. Funct. Genet.* **22**, 81–99.
Thompson, P. A., Eaton, W. A. & Hofrichter, J. (1997). Laser temperature jump study of the helix ↔ coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model. *Biochemistry*, **36**, 9200–9210.
Wikström, M., Drakenberg, T., Forsén, S., Sjöbring, U. & Björck, L. (1994). Three-dimensional solution structure of an immunoglobulin light chain-binding domain of protein L. Comparison with the IgG-binding domains of protein G. *Biochemistry*, **33**, 14011–14017.
Wikström, M., Sjöbring, U., Drakenberg, T., Forsén, S. & Björck, L. (1995). Mapping of the immunoglobulin light chain-binding site of protein L. *J. Mol. Biol.* **250**, 128–133.
Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H. & Dyer, R. B. (1996). Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry*, **35**, 691–697.
Yi, Q., Scalley, M. L., Simons, K. T., Gladwin, S. T. & Baker, D. (1997). Characterization of the free energy spectrum of peptostreptococcal protein L. *Fold. Des.* **2**, 271–280.

Appendix: Origins of Partial Φ Values for Mutations at Solvent-exposed Sites

As discussed in the main text and by previous authors (Fersht *et al.*, 1994; Itzhaki *et al.*, 1995a; Sosnick *et al.*, 1996), partial Φ values at solvent exposed sites could reflect partial ordering of the local structural element containing the mutated residue in the transition state ensemble. Partial Φ values could arise if only one of two free energy barriers (either in parallel (Fersht *et al.*, 1994) or in series) of roughly equal height are affected by the mutation. For example, letting H represent a local structural element such as a β-turn or α-helix, C represent the unfolded form of the structural element, and P the rest of the (unfolded) protein, consider a simple kinetic scheme in which both forms of the local structural element can associate productively with the remainder of the protein, but with different rates:

$$H + P \xrightarrow{k_1} N, \quad C + P \xrightarrow{k_2} N, \quad C \overset{K_{eq}}{\leftrightarrow} H$$

(The use of an equilibrium constant to characterize the C ↔ H transition is clearly a crude approximation.) Then:

$$k_f = (K_{eq}/(1 + K_{eq}))k_1 + (1/(1 + K_{eq}))k_2 \qquad \text{(A1)}$$

and, provided that the helix is not significantly populated in the denatured state ($K_{eq} < 1$), the Φ value for a mutation at a solvent-exposed site that changes the free energy of folding exclusively by changing K_{eq} is:

$$\Phi = \Delta\Delta G^{\ddagger}_{-U}/\Delta\Delta G_{F-U} = (d \ln k_f)/(d \ln K_{eq}) = K_{eq}/(k_2/k_1 + K_{eq}) - K_{eq}/(1 + K_{eq}) \quad \text{(A2)}$$

For $k_1 > k_2$, the Φ value is positive and approaches 1 for $k_1 \gg k_2/K_{eq}$; for $k_1 = k_2$, $\Phi = 0$ (the folding rate is independent of the helix population), and for $k_1 < k_2$ the Φ value is negative (the protein folds faster with the helix not formed, as perhaps for the T39G mutation). This model differs from the parallel pathway model described by Fersht *et al.* (1994), in that the folding rate depends on the fraction of helix in the unfolded state rather than directly on the equilibrium constant for helix formation; the two models are identical for $K_{eq} \ll 1$, but equation (A1) predicts smaller Φ values when K_{eq} approaches 1.

Distinguishing this parallel pathway model from the partial ordering model requires mutations that produce a wide range of stabilities (Fersht *et al.*, 1994). Even for the large number of mutations characterized at solvent-exposed sites in the helix in CI2, it is difficult to distinguish between the two models (Figure A1); it appears that a range of more than 4 kcal is required for the parallel pathway model to produce enough curvature to clearly distinguish it from the partial ordering model (in contrast, the Brönsted behavior observed over a more than 4 kcal range for the CI2 minicore clearly rules out parallel pathway models (Fersht *et al.*, 1994)). A range of >4 kcal/mol is not easily achievable for protein L since mutations producing such destabilizations would also greatly destabilize the native state (the free energy of unfolding is ~5–6 kcal/mol). However, very large destabilizations have been obtained for S peptide binding to S protein (J.M.G. and R.L.B., unpublished results), in this case the Φ values remain constant, supporting the partial ordering mechanism. It should be noted that the actual folding transition state ensemble for a protein is likely to lie somewhere between these two extremes; the simple kinetic models are useful for illustrating the consequences of differences in the breadth of the transition state ensemble on experimentally measurable quantities.

There has recently been considerable progress in determining the rate of formation of isolated helices, and it is instructive to consider models in which the helix formation rate enters directly. For example, consider a simple sequential kinetic mechanism in which the first step corresponds to the formation of conformations with a critical subset of the low resolution topological features of the native state, and the second step, to helix formation in the context of this overall topological order. Helix formation in this context is assumed to lead to very rapid completion of folding. The expression for the rate constant of a sequential reaction of this type is given by Cleland (1975); reasonable rate constants for refolding are obtained when the free energy of the critical subset is 7 kcal/mol more than that of unfolded protein, and the rate constants for helix formation and breakdown are $\sim 10^7$ s^{-1} (Gilmanshin *et al.*, 1997; Thompson *et al.*, 1997; Williams *et al.*, 1996) at $K_{eq} = 1$. In this scenario, the rate of folding depends on the rate of helix formation, not the equilibrium population of helix, and hence very different Φ values will be obtained for mutations that produce the same net destabilization of helix but affect the helix formation and disruption rates differently: mutations that exclusively slow the rate of helix formation, increase the rate of helix disruption, or slow the formation rate and increase the disruption rate by the same factor will have Φ values of 1.0, 0.0 and 0.5, respectively. Interpretation of partial Φ values in this scenario would clearly be facilitated by data on the effects of

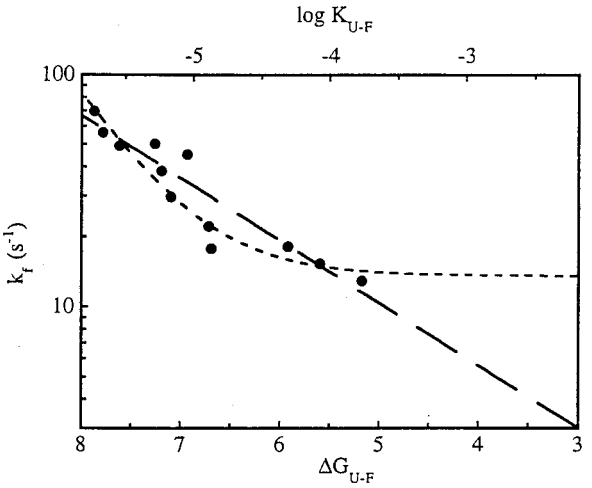


Figure A1. Fit of parallel pathway model to experimental data. Equation (A1) was fit to the refolding rate constants and free energies of unfolding [$-RT \ln (k_u/k_f)$] for the CI2 mutants expected to affect the stability of the helix but not packing interactions on the interior of the protein (S12G, S12A, E14A/E15A, S12G/E14A/E15A, S12A/E14A/E15A, K17A, K17G, K18A, K18G, Q22A and Q22G). K_{eq} for the helix in wild-type CI2 was estimated from experimental data on the isolated 12-24 segment of CI2 (9.2% helical; Itzhaki *et al.*, 1995b); the changes in free energy were assumed to result entirely from changes in K_{eq}. The broken curve is the fit to equation (A1) with $k_1 = 600$ s^{-1} and $k_2 = 14$ s^{-1}; the broken straight line, the fit to the partial ordering (Brönsted) model. Both models are consistent with these experimental data, but the partial ordering model has one less free parameter and is much more consistent with the nucleation-condensation mechanism deduced from the extensive characterization of the kinetic consequences of mutations throughout the entire protein (Fersht *et al.*, 1994; Itzhaki *et al.*, 1995a).

mutations on the rate of formation and disruption of isolated helices.

References

Cleland, W. W. 1975 Partition analysis and the concept of net rate constants as tools in enzyme kinetics *Biochemistry* **14** 3220–3224

Fersht, A. R., Itzhaki, L. S., El Masry, N. F., Matthews, J. M. & Otzen, D. E. 1994 Single versus parallel pathways of protein folding and fractional formation of structure in the transition state *Proc. Natl Acad. Sci. USA* **91** 10426–10429

Gilmanshin, R., Williams, S., Callender, R. H., Woodruff, W. H. & Dyer, R. B. 1997 Fast events in protein folding: relaxation dynamics of secondary and tertiary structure in native apomyoglobin *Proc. Natl Acad. Sci. USA* **94** 3709–3713

Itzhaki, L. S., Neira, J. L., Ruiz-Sanz, J., de Prat Gay, G. & Fersht, A. R. 1995b Search for nucleation sites in smaller fragments of chymotrypsin inhibitor 2 *J. Mol. Biol.* **254** 289–304

Thompson, P. A., Eaton, W. A. & Hofrichter, J. 1997 Laser temperature jump study of the helix $\Longleftrightarrow$ coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model *Biochemistry* **36** 9200–9210

Williams, S, Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H. & Dyer, R. B. 1996 Fast events in protein folding: helix melting and formation in a small peptide *Biochemistry* **35** 691–697

Edited by P. E. Wright

(Received 30 April 1998; received in revised form 31 August 1998; accepted 1 September 1998)