doi:10.1006/jmbi.2000.3701 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 298, 971–984

JMB

A Breakdown of Symmetry in the Folding Transition State of Protein L

David E. Kim, Cindy Fisher and David Baker*

Department of Biochemistry
University of Washington
Seattle, WA 98195, USA

The 62 residue IgG binding domain of protein L consists of a central α-helix packed on a four-stranded β-sheet formed by N and C-terminal β-hairpins. The overall topology of the protein is quite symmetric: the β-hairpins have similar lengths and make very similar interactions with the central helix. Characterization of the effects of 70 point mutations distributed throughout the protein on the kinetics of folding and unfolding reveals that this symmetry is completely broken during folding; the first β-hairpin is largely structured while the second β-hairpin and helix are largely disrupted in the folding transition state ensemble. The results are not consistent with a ''hydrophobic core first'' picture of protein folding; the first β-hairpin appears to be at least as ordered at the rate limiting step in folding as the hydrophobic core.

Keywords: protein folding; folding kinetics; transition state; β-hairpin formation; protein L

*Corresponding author

Introduction

Understanding the folding mechanisms of small proteins which fold without well populated intermediates requires the determination of the distribution of structure in the folding transition state ensemble, and the features of the sequence and structure responsible for this distribution. There has been much activity in this area over the past several years. On the theoretical side, the degree of heterogeneity in the transition state ensemble has been the topic of considerable debate (Pande *et al.*, 1998; Shakhnovich, 1998; Thirumailai & Klimov, 1998). On the experimental side, the folding transition state ensembles of a number of small proteins have been characterized by determining the effects of mutations on the kinetics of folding (Burton *et al.*, 1997; Chiti *et al.*, 1999; Fulton *et al.*, 1999; Itzhaki *et al.*, 1995; Kragelund *et al.*, 1999; Martinez & Serrano, 1999; Milla *et al.*, 1995; Riddle *et al.*, 1999; Sosnick *et al.*, 1996; Villegas *et al.*, 1998). In some proteins, the folding transition state ensemble appears to be quite polarized, with one portion of the protein largely structured, and the remainder, largely unstructured, while in others, the majority of the protein appears to be partially ordered in the transition state ensemble. Recent results suggest that these differences arise at least in part from differences in native state topology (Alm & Baker, 1999). The study of proteins whose native structures contain considerable symmetry is thus of interest because any breakdown of this symmetry in the folding transition state ensemble has the potential to highlight determinants of the folding mechanism beyond native state topology.

We have chosen the B1 IgG binding domain of peptostreptococcal protein L as a model system for understanding the folding process in detail (Scalley *et al.*, 1997; Gu *et al.*, 1997; Kim *et al.*, 1998b). The structure of this domain (referred to as protein L throughout this paper) can be separated into three secondary structural elements (Figure 1(a), Wikstrom *et al.*, 1994): the first β-hairpin (residues 4 to 23), the α-helix (residues 26 to 40), and the second β-hairpin (residues 46 to 63). The two β-hairpins make up a four-stranded β-sheet that packs with the α-helix to form the core of the protein. Both β-hairpins are connected to the α-helix by short loop segments, have nearly symmetrical side-chain contact distributions (Figure 1(b), lower right triangle), bury comparable amounts of surface area (the first and second β-hairpins bury 1053 Å^2 and 979 Å^2, respectively), and have similar numbers of backbone hydrogen bonds (Figure 1(b), upper left triangle). NMR characterization of peptide fragments representing each of the three secondary structural elements indicates that no segment has well-defined structure in isolation (Ramirez-Alvarado *et al.*, 1997). Despite the overall

Abbreviations used: GuHCl, guanidine hydrochloride.

E-mail address of the corresponding author: dabaker@u.washington.edu

0022-2836/00/050971–14 $35.00/0

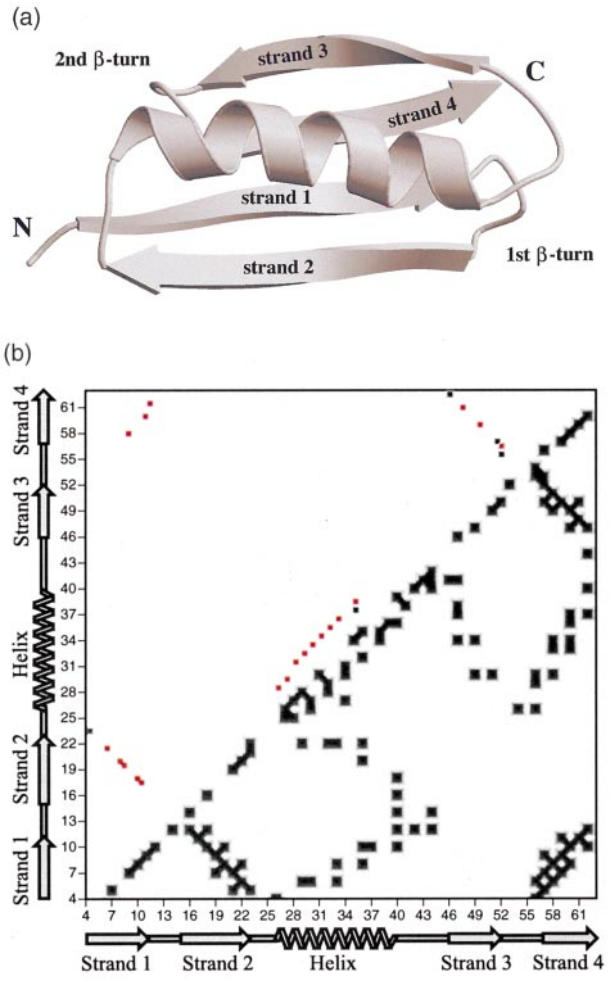

Figure 1. (a) Backbone ribbon diagram of the protein L NMR structure (Wikstrom *et al.*, 1994) with the strands, β-hairpin turns, and N and C termini labeled. The image was created using Molscript (Kraulis, 1991) and Raster3d (Bacon & Anderson, 1998; Merritt & Murphy, 1994). (b) Backbone hydrogen bonds (upper left triangle) and side-chain contacts (lower right triangle). Hydrogen bonds involving residues that display slow amide proton exchange (Wikstrom *et al.*, 1993) are plotted (red). Hydrogen bonds were identified from the NMR solution structure (Wikstrom *et al.*, 1994). Side-chain contacts were determined using a Voronoi polyhedra method (Gerstein, 1995).

symmetry of the structure and the similarities in the two β-hairpins, previous studies have shown that mutations in the first β-turn slow the folding rate and have little effect on the unfolding rate, while mutations in the second β-turn increase the unfolding rate but have little effect on the folding rate (Gu *et al.*, 1997). To thoroughly characterize the folding transition state ensemble of protein L and to determine the degree to which symmetry is broken during protein L folding, we have determined the effects of mutations of all residues which make significant interactions in the native state on the kinetics of folding and unfolding. Here, we report the results of these experiments and describe the picture of the folding transition state ensemble that emerges from the data.

Results

Thermodynamics and kinetics of folding of point mutants

To determine the contribution of all residues that make significant interactions in the native state of protein L to thermodynamic stability and folding kinetics, point mutations were made at 54 of the 62 positions in the protein (with the exception of W47, the remaining residues are almost entirely solvent exposed and probably make little contribution to either stability or folding kinetics). The roles of entire side-chains were probed by alanine and glycine substitutions, and those of specific subsets of side-chain atoms, by partial side-chain truncations (I to V, F to L, F to V, and Y to L). The effects of decreasing helix propensity were probed by glycine substitutions on the solvent exposed side of the helix (several such mutations have been described by Kim *et al.* (1998b)). The β-turns were probed by mutations that disrupt or increase turn propensity (A13P, A13V, N14A, G15A, G15V, G15A/N14A, G55A; several of these were described earlier (Gu *et al.*, 1997, 1999)).

The mutagenesis, protein expression and protein purification required to prepare the mutants were carried out using standard methods (see Materials and Methods). The changes in the free energy of folding resulting from the mutations ($\Delta\Delta G$) were determined using standard equilibrium guanidine hydrochloride (GuHCl) denaturation experiments (Figure 2(a)) taking care to avoid long extrapolations (Table 1, see Materials and Methods). Two different estimates ($\Delta\Delta G^{2M}_{F\text{-}U}$ and $\Delta\Delta G^{Cm}_{F\text{-}U}$) are listed in Table 1; for most of the mutants the two estimates are quite consistent. The free energies of folding and their denaturant dependencies (the m values) of the most destabilized mutants could not be determined accurately because of the inability to accurately determine the folded baseline (for example, F62L; Figure 2(a) (◇)), and as a result, the two estimates of $\Delta\Delta G$ are less consistent.

The folding and unfolding kinetics were characterized for each mutant using stopped-flow fluorescence experiments (see Materials and Methods). To avoid long extrapolations, the folding rate constants are reported in 0.4 M GuHCl, and unfolding rate constants in 2 M and 4 M GuHCl (Table 2). A third estimate of the change in the free energy of folding ($\Delta\Delta G^{kin}_{F\text{-}U}$) was obtained using the kinetic data (Table 1), and was found to correlate well with the estimates from the equilibrium experiments (slope $= 0.91(\pm 0.02)$, $R = 0.98$; Figure 2(b)) as expected for a two state folding reaction where the free energy of folding $\Delta G_{F\text{-}U} = -RT\ln K_{eq} = -RT(\ln k_f - \ln k_u)$ (K_{eq} is the equilibrium constant for folding, and k_f and k_u are the folding and unfolding rates, respectively). Repre-

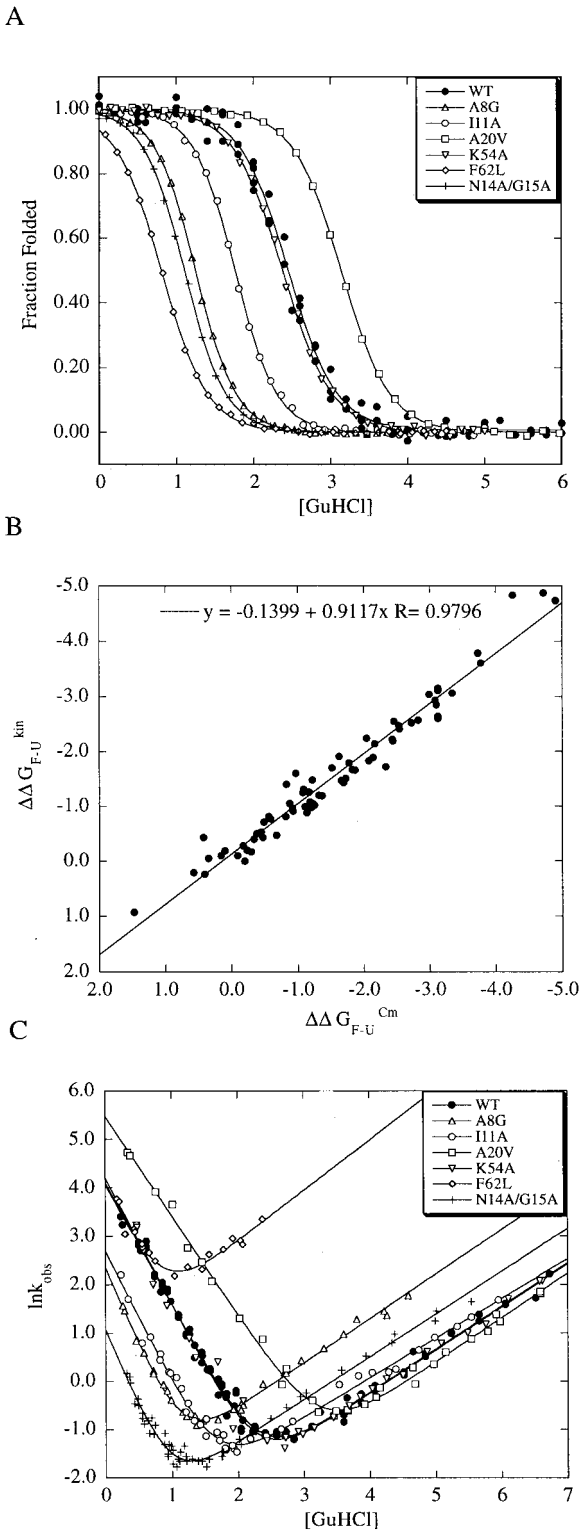

Figure 2. Representative data from the thermodynamic and kinetic experiments. (a) Equilibrium denaturation data normalized as the fraction of folded protein for mutants A8G (△); I11A (○); A20V (□); K54A (▽); F62L (◇); the double mutant, N14A/G15A (+); and wild-type (●). (b) Plot of the difference in the free energy of folding between mutant and wild-type determined by equilibrium denaturation, $\Delta\Delta G^{Cm}_{F-U}$, *versus* the difference in free energy calculated from kinetics, $\Delta\Delta G^{kin}_{F-U}$. The thermodynamics and kinetics data correlate well as expected for a two-state model for folding (slope = 0.91 ± 0.02, $R = 0.98$). (c) The GuHCl dependence of the logarithm of the observed folding and unfolding relaxation rates ($\ln k_{obs}$) for mutants A8G (△); I11A (○); A20V (□); K54A (▽); F62L (◇); the double mutant, N14A/G15A (+); and wild-type (●).

sentative kinetic data are shown in Figure 2(c); some mutations affected only the folding rate (Figure 2(c), (○) and (□)), only the unfolding rate (Figure 2(c), (◇)), or both the folding and unfolding rates (Figure 2(c); (△) and (+)) (Table 2).

Φ Value analysis

In a simple transition state theory based model of protein folding kinetics, where $k_f = \exp[-(\Delta G^{\ddagger}_{-U}/RT)]$ and $\Delta G^{\ddagger}_{-U}$ is the activation energy for folding, the change in the free energy of the transition state brought about by a mutation is $\Delta\Delta G^{\ddagger}_{-U} = -RT\ln k_f^{wt}/k_f^{mut}$, where k_f^{wt} and k_f^{mut} are the folding rates of the wild-type and mutant, respectively. The distribution of structure in the folding transition state ensemble can thus be deduced from the effect of mutations on the kinetics of folding: the greater the decrease in the folding rate brought about by a mutation, the more important the removed interactions are in stabilizing the transition state ensemble. To account for the differences in the size of the perturbations caused by different mutations, it is convenient to normalize by dividing by the effect of the mutation on the free energy of folding. The quantity $\Phi = \Delta\Delta G^{\ddagger}_{-U}/\Delta\Delta G_{F-U}$, introduced by Fersht and co-workers (Matouschek *et al.*, 1989), is thus a convenient measure of the extent of structure in the transition state ensemble. A value of 1 indicates the residue makes similar interactions in the transition state and in the native state; interactions removed by mutation that stabilize the native state equally stabilize the transition state. A value of 0 indicates that the interactions removed by mutation in the native state are not present in the transition state. To guard against possible artifacts due to changes in denatured state structure and/or folding mechanism, we draw conclusions only from results that are consistent among a number of neighboring residues.

Mutations that destabilize the protein by less than 0.3 kcal mol^{-1} determined from either $\Delta\Delta G^{Cm}_{F-U}$, $\Delta\Delta G^{2M}_{F-U}$, or $\Delta\Delta G^{kin}_{F-U}$ (Table 3) were not considered because of the large errors which can result from division by small numbers. To provide an indication of the magnitude of the errors, three different estimates were obtained for the Φ value of each mutation using the different estimates of $\Delta\Delta G$ in the denominator and either the folding or unfolding kinetic data to estimate the numerator (see Materials and Methods). The three different estimates of the Φ values are in general quite consistent (Table 3). The Φ_F values from kinetic data

Table 1. Thermodynamic parameters

	m (kcal mol^{-1} M^{-1})	Cm (M)	$\Delta\Delta G_{F-U}^{Cm}$ (kcal mol^{-1})	$\Delta\Delta G_{F-U}^{2\,M}$ (kcal mol^{-1})	$\Delta\Delta G_{F-U}^{kin}$ (kcal mol^{-1})
Wt	1.9	2.42			
V4A	2.0	1.83	−1.22	−1.15	−1.48
T5A	2.2	1.63	−1.63	−1.61	−1.91
I6A	1.8	0.05	−4.90	−4.26	−4.73
I6V	1.9	2.15	−0.56	−0.51	−0.81
V6A	1.8	0.05	−4.34	−3.75	−4.05
K7A	1.9	1.97	−0.92	−0.85	−0.94
A8G	2.2	1.24	−2.43	−2.48	−2.22
N9A	2.1	1.51	−1.87	−1.83	−1.66
L10A	1.8	0.91	−3.12	−2.78	−2.60
I11A	2.2	1.76	−1.37	−1.32	−1.19
I11V	1.9	2.19	−0.47	−0.42	−0.43
V11A	2.2	1.76	−0.90	−0.89	−0.76
F12A	2.2	0.91	−3.12	−3.17	−2.64
F12L	1.9	2.09	−0.68	−0.63	−0.47
L12A	2.2	0.91	−2.44	−2.54	−2.17
A13P	1.9	2.47	0.10	0.09	−0.19
A13V	2.2	2.02	−0.83	−0.75	−1.40
N14A	1.9	1.56	−1.78	−1.65	−1.79
G15A	2.1	1.69	−1.52	−1.46	−1.70
G15V	2.6	1.19	−2.53	−2.92	−2.47
N14A(G15A)[a]	2.2	1.10	−0.94	−1.11	−0.73
G15A(N14A)[a]	2.2	1.10	−1.20	−1.29	−0.82
N14A/G15A	2.2	1.10	−2.72	−2.76	−2.52
S16A	1.9	2.27	−0.30	−0.27	−0.17
T17A	2.0	1.86	−1.17	−1.08	−1.26
T19A	2.0	1.88	−1.11	−1.03	−0.99
A20G	2.2	1.37	−2.17	−2.16	−2.14
A20V	1.8	3.14	1.47	1.30	0.93
E21A	1.9	2.14	−0.59	−0.54	−0.77
F22A	3.0	0.36	−4.25	−5.67	−4.83
F22L	2.4	0.91	−3.12	−3.43	−3.10
L22A	3.0	0.36	−1.13	−2.24	−1.73
K23A	2.0	1.99	−0.88	−0.81	−1.05
G24A	2.1	1.41	−2.08	−2.04	−1.83
T25A	2.0	1.82	−1.25	−1.17	−1.02
F26G	2.4	0.92	−3.08	−3.37	−2.93
F26L	1.9	2.24	−0.38	−0.34	−0.50
L26G	2.4	0.92	−2.70	−3.03	−2.43
K28G	1.7	2.50	0.16	0.06	−0.10
A29G	2.2	1.19	−2.54	−2.57	−2.41
T30A	2.1	1.89	−1.09	−1.02	−1.31
S31A	2.0	2.62	0.41	0.44	0.24
S31G	2.0	2.03	−0.82	−0.75	−0.81
A31G	2.0	2.03	−1.23	−1.17	−1.05
E32G	1.9	1.85	−1.19	−1.09	−1.08
E32I	2.0	1.90	−1.08	−1.01	−1.25
A33G	2.8	0.92	−3.10	−3.83	−2.85
Y34A	2.4	1.05	−2.82	−3.04	−2.57
A35G	2.0	1.78	−1.32	−1.23	−1.20
Y36A	2.1	1.23	−2.46	−2.44	−2.54
A37G	2.4	0.91	−3.12	−3.44	−3.14
D38A	1.9	1.84	−1.21	−1.12	−0.98
D38G	2.0	1.38	−2.14	−2.01	−1.89
A38G	2.0	1.38	−0.93	−0.90	−0.91
T39G	1.9	2.34	−0.17	−0.17	−0.28
L40A	2.0	1.24	−2.44	−2.31	−2.19
E32G/A35G/T39G	2.1	0.97	−2.99	−2.99	−3.03
K41A	2.0	2.70	0.58	0.59	0.21
K42A	1.9	2.59	0.35	0.32	−0.05
N44A	1.9	2.26	−0.34	−0.32	−0.40
G45A	1.9	1.29	−2.23	−2.12	−1.72
E46A	2.2	2.31	−0.23	−0.11	−0.20
T48A	2.1	1.95	−0.97	−0.90	−1.60
V49A	2.0	1.98	−0.92	−0.85	−0.96
D50A	2.0	2.33	−0.20	−0.15	0.00
V51A	2.1	1.87	−1.14	−1.07	−0.88
A52G	2.0	2.18	−0.49	−0.43	−0.71
K54A	1.8	2.38	−0.09	−0.12	−0.10
G55A	2.1	1.43	−2.04	−2.00	−2.24
Y56A	2.0	1.62	−1.66	−1.54	−1.47

Table 1. (*continued*)

Y56L	1.8	2.63	0.43	0.36	−0.43
L56A	2.0	1.62	−2.08	−1.90	−1.04
T57A	2.0	1.53	−1.83	−1.74	−1.67
L58A	2.1	0.59	−3.77	−3.72	−3.60
N59A	2.0	1.58	−1.73	−1.62	−1.51
I60A	2.1	0.13	−4.72	−4.64	−4.87
I60V	2.1	1.60	−1.69	−1.64	−1.43
V60A	2.1	0.13	−3.03	−3.00	−3.62
K61A	1.8	2.20	−0.45	−0.43	−0.53
F62L	1.9	0.80	−3.34	−3.13	−3.05
F62V	2.0	0.61	−3.73	−3.62	−3.78

All parameters are described in Materials and Methods.
[a] N14A(G15A) is the effect of the N14A mutation made in the effect of the G15A background and G15A(N14A) is the effect of the G15A mutation made in the N14A background.

were used in the following analysis of the structure of the folding transition state since they require little or no extrapolation of either the folding or unfolding data.

The structure of the folding transition state determined by Φ_F value

The hydrophobic core

Mutations of side-chains involved in the hydrophobic core of protein L can be broken into two classes, those that have intermediate Φ_F values (0.2-0.8) and those with values close to zero. These two classes of mutations cluster dramatically in the three-dimensional structure (compare Figure 3(b) and (c) to Figure 3(d) and (e)). The first class, in which the average Φ_F value is 0.34, is contained largely in the first β-hairpin and the portion of the helix that contacts the first β-hairpin. The second class, in which the average Φ_F value is 0.07, consists primarily of mutations in the second β-hairpin and the portion of the helix that contacts the second β-hairpin. These results clearly indicate that the core is not uniformly formed in the folding transition state; the residues in and contacting the first β-hairpin make more interactions in the transition state than those in and contacting the second β-hairpin.

Multiple mutations were made at several sites to probe interactions in the transition state in more detail. Mutation of F26 in the loop connecting the second strand and the helix to leucine removes interactions mainly with K54 and Y56, both of which are in the second β-hairpin turn, and produces a low Φ_F value of 0.08 (Table 3). Further truncation by the L26G mutation removes more local interactions within the helix and first β-hairpin (V4, E27, and T30) and produces a higher Φ_F value of 0.30. These results suggest that the nonlocal interactions that F26 makes with the second β-turn are not conserved in the transition state while the more local interactions are partially maintained. F12L is the only core mutation in the first β-hairpin that has a Φ_F value less than 0.20 and does not significantly effect the folding rate (Tables 2 and 3). This mutation removes interactions mostly with residues that also have low Φ_F values (L40, N44, and F62) and has a Φ_F value of −0.07. Interestingly, further truncation by the L12A mutation, which removes local interactions within the first β-hairpin, reduces the folding rate and has a higher Φ_F value of 0.26. Taken together, the overall clustering of residues with higher Φ_F values in and adjacent to the first β-hairpin, the contrast between the helix residues that contact the first β-hairpin and the helix residues that contact the second β-hairpin, and the lower Φ_F values associated with removing atoms that interact primarily with residues in the second β-hairpin for both F26 and F12, suggest that the network of interactions among the core residues in the first β-hairpin are similar in the native and transition states.

First β-hairpin

The formation of the first β-hairpin was probed by 16 point mutations distributed throughout the β-turn and solvent exposed positions in strands 1 and 2. Five of the point mutations, A13P, A13V, N14A, G15A, and G15V have been previously studied (Gu *et al.*, 1997, 1999) and are included in this analysis. Strands 1 and 2 are connected by a type I β-turn from F12 to G15. While several of the mutations increase the size of the side-chain, which potentially can complicate interpretation of the Φ_F values, the consistency of the results suggests that changes in folding mechanism and/or denatured state structure are quite unlikely. The high Φ_F values in the different positions in the turn (A13, N14, and G15; Table 3) strongly suggest that the turn is largely formed in the folding transition state ensemble. To determine whether the first β-turn is formed in the folding transition state even after destabilization with the N14A or G15A mutations, the double mutant (N14A/G15A) was characterized. The high Φ_F values of the double mutant (0.78), the N14A mutation in the G15A background (0.88), and the G15A mutation in the N14A background (0.72), suggest that the formation of the turn remains rate limiting even after

Table 2. Kinetics parameters

	$-m_f$ (kcal mol^{-1} M^{-1})	$k_f^{0\ M}$ (s^{-1})	$k_f^{0.4\ M}$ (s^{-1})	m_u (kcal mol^{-1} M^{-1})	$k_u^{2\ M}$ (s^{-1})	$k_u^{4\ M}$ (s^{-1})
Wt	1.5	60.60	21.72	0.50	0.11	0.61
V4A	1.4	14.94	5.92	0.61	0.27	2.12
T5A	1.6	26.63	9.12	0.59	0.91	6.87
I6A	1.6	3.38	1.10	0.77	19.47	270.93
I6V	1.5	27.88	10.32	0.56	0.17	1.17
V6A	1.6	3.38	1.10	0.77	19.47	270.93
K7A	1.4	20.10	7.92	0.50	0.20	1.13
A8G	1.9	10.35	2.87	0.54	0.59	3.67
N9A	1.8	51.73	15.46	0.56	1.12	7.57
L10A	1.9	11.54	3.12	0.54	1.21	7.72
I11A	1.6	14.71	4.90	0.48	0.21	1.08
I11V	1.6	55.35	19.03	0.49	0.21	1.13
V11A	1.6	14.71	4.90	0.48	0.21	1.08
F12A	1.9	31.72	8.75	0.56	3.43	23.20
F12L	1.5	62.80	23.02	0.52	0.24	1.45
L12A	1.9	31.72	8.75	0.56	3.43	23.20
A13P	1.5	82.77	30.17	0.56	0.17	1.17
A13V	1.7	20.10	6.20	0.59	0.26	1.95
N14A	1.7	7.38	2.32	0.55	0.22	1.42
G15A	1.8	7.09	2.10	0.58	0.15	1.11
G15V	1.8	4.26	1.25	0.58	0.45	3.24
N14A(G15A)[a]	1.9	2.85	0.77	0.52	0.28	1.66
G15A(N14A)[a]	1.9	2.85	0.77	0.52	0.28	1.66
N14A/G15A	1.9	2.85	0.77	0.52	0.28	1.66
S16A	1.6	66.49	22.38	0.51	0.15	0.84
T17A	1.6	26.58	9.20	0.54	0.35	2.27
T19A	1.6	44.17	15.22	0.58	0.33	2.36
A20G	1.7	19.06	6.06	0.58	0.92	6.78
A20V	1.2	239.51	105.00	0.54	0.09	0.59
E21A	1.4	20.72	8.00	4.53	0.14	0.85
F22A	1.7	2.28	0.70	0.81	14.55	232.66
F22L	1.8	15.64	4.45	0.69	4.69	49.78
L22A	1.7	2.28	0.70	0.81	14.55	232.66
K23A	1.5	25.67	9.19	0.51	0.28	1.59
G24A	1.7	29.47	9.20	0.58	0.84	6.03
T25A	1.4	27,61	10.27	0.53	0.27	1.66
F26G	1.9	21.39	5.72	0.63	2.86	24.92
F26L	1.4	51.90	20.26	0.54	0.21	1.34
L26G	1.9	21.39	5.72	0.63	2.86	24.92
K28G	1.3	47.16	18.79	0.50	0.11	0.63
A29G	1.6	25.31	8.32	0.52	1.80	14.79
T30A	1.4	47.02	18.03	0.61	0.62	4.89
S31A	1.5	87.78	32.34	0.55	0.09	0.60
S31G	1.7	57.26	18.50	0.55	0.33	2.10
A31G	1.7	57.26	18.50	0.55	0.33	2.10
E32G	1.5	48.29	16.75	0.57	0.43	3.01
E32I	1.5	55.11	20.34	0.63	0.58	4.96
A33G	2.0	24.51	6.27	0.60	3.10	23.82
Y34A	2.1	72.08	17.50	0.60	5.27	41.13
A35G	1.5	33.73	12.17	0.50	0.48	2.71
Y36A	1.5	18.88	6.62	0.61	1.86	14.86
A37G	1.8	40.93	11.97	0.69	13.70	146.73
D38A	1.6	129.16	42.45	0.54	1.04	6.49
D38G	1.7	81.18	25.28	0.56	2.77	18.56
A380	1.7	81.18	25.28	0.56	2.77	18.56
T39G	1.5	108.50	37.76	0.60	0.22	1.72
L40A	1.7	43.88	13.40	0.52	2.81	16.45
E32G/A35G/T39G	1.8	24.99	7.40	0.67	6.93	69.20
K41A	1.4	68.90	26.66	0.47	0.10	0.52
K42A	1.4	61.06	23.21	0.54	0.11	0.71
N44A	1.5	57.91	20.64	0.49	0.22	1.16
G45A	1.7	89.80	28.83	0.45	3.34	15.83
E46A	1.5	60.67	22.07	0.54	0.14	0.87
T48A	1.4	27.48	10.55	0.67	0.48	4.69
V49A	1.5	34.83	12.70	0.54	0.29	1.88
D50A	1.5	83.58	29.61	0.52	0.14	0.83
V51A	1.5	46.67	16.23	0.53	0.34	2.07
A52G	1.3	58.90	23.63	0.54	0.35	2.26
K54A	1.5	66.44	23.39	0.52	0.13	0.78
G55A	1.8	35.59	10.63	0.63	1.65	14.25
Y56A	1.8	50.77	14.97	0.50	0.97	5.31

Table 2. (*continued*)

Y56L	1.4	37.30	14.67	0.47	0.17	0.86
L56A	1.8	50.77	14.97	0.50	0.97	5.31
T57A	1.7	48.24	14.98	0.53	1.20	7.44
L58A	1.5	11.00	3.94	0.69	9.97	106.12
N59A	1.6	42.54	13.89	0.50	0.94	5.29
I60A	0.8	9.39	5.43	0.80	121.24	1870.17
I60V	1.6	37.43	12.65	0.59	0.55	4.20
V60A	0.8	9.39	5.43	0.80	121.24	1870.17
K61A	1.5	52.79	18.73	0.50	0.24	1.32
F62L	1.6	56.44	18.54	0.61	18.12	146.18
F62V	1.5	66.86	23.97	0.74	82.65	1046.97

All parameters are described in Materials and Methods.
[a] N14A(G15A) is the effect of the N14A mutation in the G15A background and G15A(N14A) is the effect of the G15A mutation made in the N14A background.

the turn is destabilized by the N14A or G15A mutations.

Mutations made at solvent exposed positions in strands 1 and 2 that are located on the backside of the β-sheet (Figure 4(a)-(c)) include T5A, K7A, N9A, I11A, I11V, V11A, T17A, T19A, E21A, and K23A, and have an average Φ_F value of 0.47. Mutations made at the end of the hairpin, T5A,

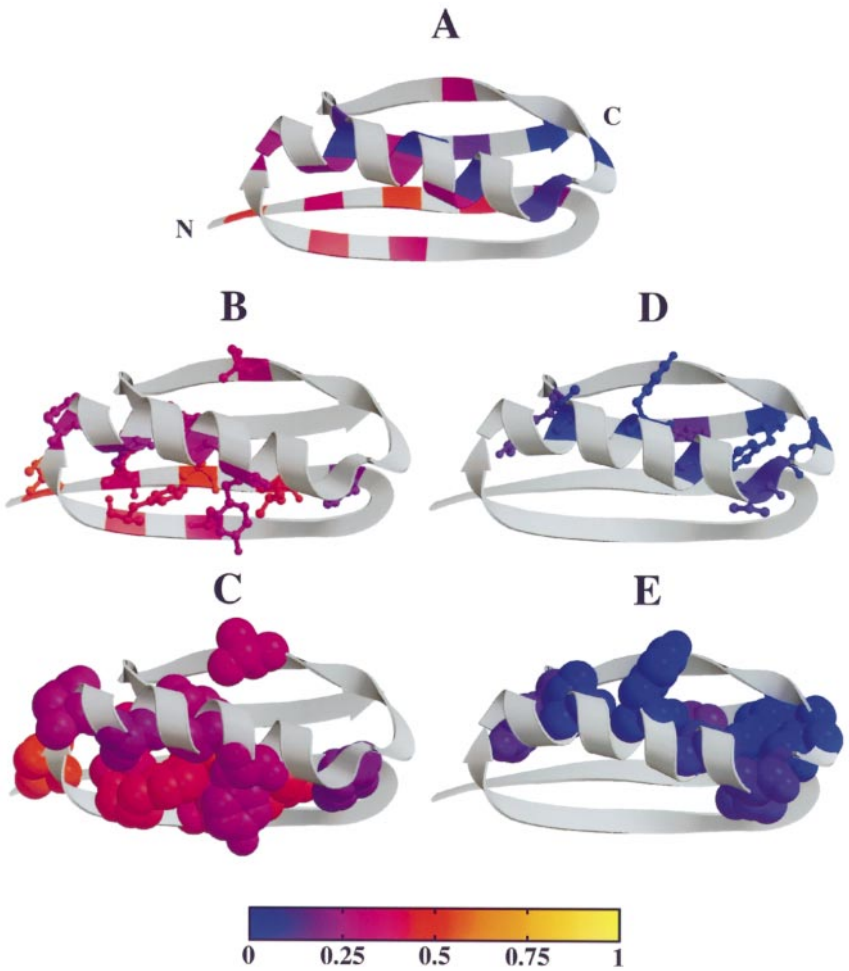

Figure 3. The formation of the core in the transition state of protein L is not uniform. (a) Structure of protein L with residues involved in the core colored by Φ_F from a scale of 1.0 (yellow) to 0.5 (red) to 0.0 (blue). Side-chains with intermediate Φ_F values (0.2-0.8) are displayed in (b) ball-and-stick and (c) spacefill representations. These mutations include V4A, I6A, A8G, L10A, F12A, A20G, F22A, F26G, A29G, A33G, Y36A, V49A, and L58A, and are mostly located in the first β-hairpin. Side-chains with Φ_F values less than 0.2 (T30A, Y34A, A37G, L40A, N44A, Y56A, I60A, and F62L) are also displayed in (d) ball-and-stick and (e) spacefill representations. The images were created using Molscript (Kraulis, 1991) and Raster3d (Bacon & Anderson, 1998; Merritt & Murphy, 1994).

Table 3. Φ Values[a] and side-chain interactions

	Φ_F	Φ_F [a]	$1-\Phi_U^{2M}$	Structure[b]	Burial %	Interactions with the wild-type side-chain[c]
V4A	0.51	0.67	0.55	s1	72	F26,Y56
T5A	0.26	0.30	0.24	s1	41	K7,E21,K23
I6A	0.37	0.34	0.29	s1	100	F22,A29,T30,A33,Y56,L58
I6V	0.53	0.82	0.50	s1	100	F22,A29,T30,A33,Y56,L58
V6A	0.32	0.28	0.27	s1	100	F22,A29,T30,A33,Y56,L58
K7A	0.62	0.70	0.59	s1	43	T5,N9,T19,E21,T57,N59
A8G	0.53	0.43	0.61	s1	100	L10,A20,F22,A33,Y36,L58,I60
N9A	0.12	0.05	0.26	sl	73	K7,I11,T19,T57,N59
L10A	0.43	0.31	0.50	sl	98	A8,F12,Q18,A20,Y36,A37,L40,I60,F62
I11A	0.72	0.59	0.72	s1	69	N9,T17,N59,K61
I11V	0.18	0.11	0.12	s1	69	N9,T17,N59,K61
V11A	1.00	0.86	1.00	s1	69	N9,T17,N59,K61
F12A	0.20	0.12	0.37	t1	88	L10,N14,S16,Q18,L40,K42,N44,F62
F12L	−0.07	−0.03	0.26	t1	88	L10,N14,S16,Q18,L40,K42,N44,F62
L12A	0.26	0.16	0.39	t1	88	L10,N14,S16,Q18,L40,K42,N44,F62
A13V	0.52	0.78	0.34	t1	37	
N14A	0.85	0.67	0.86	t1	40	F12,S16
G15A	0.77	0.86	0.86	t1	15	
G15V	0.67	0.61	0.72	t1	15	
N14A(G15A)[d]	0.88	0.58	0.86	t1	40	F12,S16
G15A(N14A)[d]	0.72	0.44	0.73	t1	15	
N14A/G15A	0.78	0.65	0.81	t1	40/15	F12,S16
T17A	0.40	0.42	0.37	s2	32	I11
T19A	0.21	0.17	0.38	s2	41	K7,N9,E21
A20G	0.35	0.31	0.43	s2	96	A8,L10,F22,Y36
A20V	0.98	0.54	0.93	s2	96	A8,L10,F22,Y36
E21A	0.75	1.08	0.73	s2	24	T5,K7,T19,K23
F22A	0.41	0.45	0.50	s2	84	I6,A8,A20,A29,E32,A33,Y36
F22L	0.30	0.25	0.37	s2	84	I6,A8,A20,A29,E32,A33,Y36
L22A	0.62	0.99	0.71	s2	84	I6,A8,A20,A29,E32,A33,Y36
K23A	0.47	0.57	0.32	s2	37	T5,E21
G24A	0.27	0.20	0.42	l	65	
T25A	0.43	0.37	0.55	l	49	K28
F26G	0.26	0.20	0.44	h	65	V4,E27,T30,K54,Y56
F26L	0.08	0.24	−0.12	h	65	V4,E27,T30,K54,Y56
L26G	0.30	0.19	0.50	h	65	V4,E27,T30,K54,Y56
A29G	0.23	0.20	0.37	h	97	I6,F22,K28,E32,Y56
T30A	0.08	0.14	0.02	h	82	I6,F26,E27,S31,Y34,Y56,L58
S31G	0.11	0.04	0.16	h	36	T30,Y34
A31G	0.31	0.20	0.37	h	36	T30,Y34
E32G	0.11	0.11	0.24	h	46	F22,K28,A29,A35,Y36
E32I	0.05	0.05	0.06	h	46	F22,K28,A29,A35,Y36
A33G	0.25	0.17	0.49	h	100	I6,A8,F22,L58,I60
Y34A	0.05	−0.04	0.26	h	54	T30,S31,D38,W47,V49,L58,I60
A35G	0.28	0.25	0.33	h	61	E32,D38
Y36A	0.27	0.27	0.33	h	59	A8,L10,A20,F22,E32,T39,L40,I60
A37G	0.11	0.07	0.19	h	99	L10,W47,I60,F62
D38A	−0.39	−0.42	−0.33	h	43	Y34,A35,K41,W47
D38G	−0.05	−0.08	0.07	h	43	Y34,A35,K41,W47
A38G	0.33	0.25	0.45	h	43	Y34,A35,K41,W47
L40A	0.13	0.08	0.19	h	69	L10,F12,Y36,T39,K42,N44,F62
E32G/A35G/T39G	0.21	0.18	0.15	h	46/61/49	F22,K28,A29,Y34,Y36,D38
N44A	0.07	0.08	−0.25	l	95	F12,L40,K41,K42,F62
G45A	−0.10	−0.10	0.10	l	24	
T48A	0.26	0.48	0.04	s3	27	K61
V49A	0.32	0.35	0.33	s3	59	Y34,W47,V51,L58,I60
V51A	0.19	0.13	0.40	s3	19	V49,L58
A52G	−0.07	0.04	−0.70	s3	48	D50,D53,T57,N59
G55A	0.17	0.18	0.00	t2	53	
Y56A	0.15	0.06	0.18	t2	96	V4,I6,F26,A29,T30,D53,K54,T57
L56A	−0.01	−0.09	0.47	t2	96	V4,I6,F26,A29,T30,D53,K54,T57
T57A	0.13	0.07	0.21	s4	79	K7,N9,A52,D53,Y56,N59
L58A	0.27	0.26	0.30	s4	94	I6,A8,T30,A33,Y34,V49,V51,I60
N59A	0.17	0.12	0.23	s4	65	K7,N9,I11,D50,A52,T57,K61
I60A	0.17	0.23	0.12	s4	100	A8,L10,A33,Y34,Y36,A37,W47,V49,L58,F62
I60V	0.22	0.17	0.43	s4	100	A8,L10,A33,Y34,Y36,A37,W47,V49,L58,F62
V60A	0.14	0.26	−0.04	s4	100	A8,L10,A33,Y34,Y36,A37,W47,V49,L58,F62
K61A	0.16	0.18	−0.07	s4	42	I11,T48,D50,N59
F62L	0.03	0.01	0.06	s4	96	L10,F12,A37,L40,N44,W47,I60
F62V	−0.02	−0.02	−0.06	s4	96	L10,F12,A37,L40,N44,W47,I60

[a] Φ Values were calculated as described in Materials and Methods.
[b] s1, s2, s3, s4, h, l, t1 and t2 correspond to strands 1-4, the helix, loops, and turns 1 and 2, respectively.
[c] Side-chain contacts were determined using the Voronoi polyhedra method (Gerstein, 1995).
[d] N14A(G15A) is the effect of the N14A mutation in the G15A background and G15A(N14A) is the effect of the G15A mutation in the N14A background.

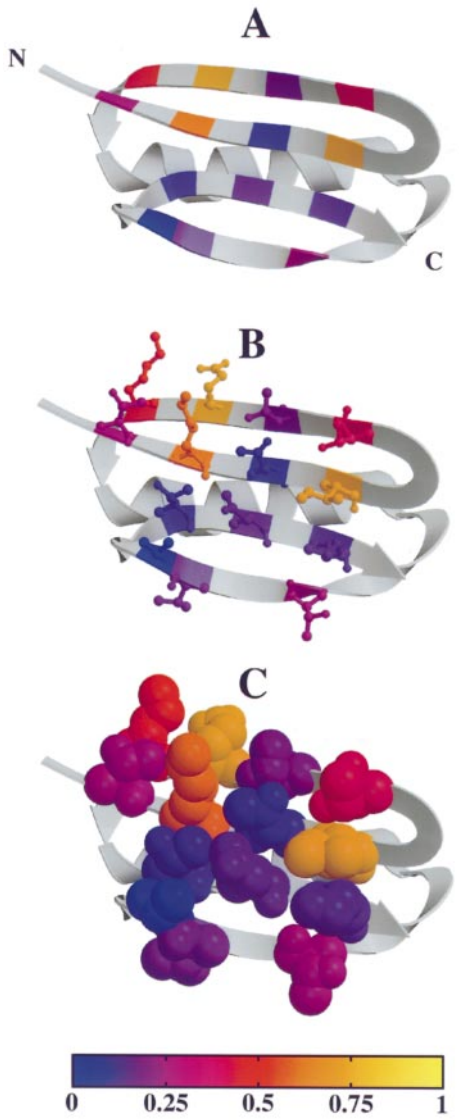

Figure 4. Solvent exposed residues in the β-sheet have higher Φ_F values in the first β-hairpin. (a) Solvent exposed residues colored by Φ_F from a scale of 1.0 (yellow) to 0.5 (red) to 0.0 (blue), and displayed in (b) ball-and-stick and (c) spacefill representations. Mutations include T5A, K7A, N9A, I11A, T17A, T19A, E21A, K23A, T48A, V51A, A52G, T57A, N59A, and K61A. The images were created using Molscript (Kraulis, 1991) and Raster3d (Bacon & Anderson, 1998; Merritt & Murphy, 1994).

K7A, E21A, and K23A, have values of 0.26, 0.62, 0.75, and 0.47, respectively. T5 and K7 are involved in cross-strand pair interactions with K23 and E21, respectively, and K7A and E21A may be involved in a salt-bridge. The intermediate to high Φ_F values of these four residues suggest that the end of the hairpin is significantly formed in the transition state for folding which is consistent with the core mutations made near this region. In contrast, N9 and T19, which are paired near the center of the hairpin, have Φ_F values of 0.12 and 0.21, respectively, indicating that the hairpin may be less structured at its center. Closer to the turn are the adjacent residues I11 and T17. The Φ_F values of I11A and T17A are 0.72 and 0.40, respectively, suggesting partial formation of the hairpin near the turn. Truncation of I11 to valine removes a methyl group (C^{δ}) that packs against N9 and N59, both of which have low Φ_F values (Table 3), and produces a Φ_F value of 0.18. Interestingly, further truncation by the V11A mutation gives a high Φ_F value of 1.0, suggesting that interactions of the gamma carbons with N9, T17, N59, and K61 are made in the folding transition state ensemble. In summary, the distribution of Φ_F values along the solvent exposed side (Figure 4) suggests that the hairpin is largely intact near the turn and at the opposite end in the folding transition state ensemble, but somewhat disrupted at its center.

The sequence of strand 2 is STQTAEFK and contains only one large hydrophobic residue, F22. A20V was made to probe the effect of increasing the strand propensity and the interactions with the first strand and helix. The A20V mutation produces nearly a fourfold increase in the folding rate and has a Φ_F value of 0.98, suggesting that the interactions introduced by adding two methyl groups stabilize the transition state for folding. It is difficult to determine whether the increase in the folding rate results from increasing the population of the β-hairpin, and/or increasing the size of the hydrophobic core. However, the high Φ_F value does suggest that this region of the protein is largely formed in the folding transition state and is consistent with the mutations made in the turn and at adjacent core positions.

Helix

The formation of the helix was probed by ten helix destabilizing point mutations made at solvent exposed positions along the helix and a triple mutant (E32G/A35G/T39G) that was designed to destabilize the helix along its entire length. Five of these mutations (K28G, E32G, E32I, A35G, and T39G) and the triple mutant have been previously studied (Kim *et al.*, 1998b) and are included in this analysis. Glycine substitutions are well suited for probing the consequences of reducing the population of the helix on the rate of folding. S31G, E32G, E32I, and D38G have low Φ_F values of 0.11, 0.11, 0.05, and −0.05, respectively, suggesting that the helix is largely disrupted in the transition state for folding. Interestingly, the triple mutant also has a low Φ_F value of 0.21, suggesting that no part of the helix needs to be intact in the folding transition state. In order to avoid possible complications of changes in tertiary interactions and solvation energy accompanying the above mutations, alanine to glycine mutations were also made at positions 31, 35, and 38. The Φ_F values of these mutations are slightly higher than those described above (0.31, 0.28, and 0.33, respectively). As noted in our previous study (Kim *et al.*, 1998b), it is difficult to determine whether partial Φ values of solvent

exposed residues represent partial ordering of the helix in the transition state for folding or multiple states with fully formed and disrupted helices since a large enough range in stabilities is not obtained to distinguish the two possibilities using a Brönsted analysis (Fersht *et al.*, 1994). A range of over 4 kcal mol^{-1} appears to be necessary but the mutations in the helix change the stability by only 1-2 kcal mol^{-1}. Nevertheless, the Φ values of the A to G mutations made in the helix are low, suggesting the helix is only marginally formed.

Second β-hairpin

To determine the extent of formation of the second β-hairpin in the transition state for folding, ten point mutations (E46A, T48A, D50A, V51A, A52G, K54A, G55A, T57A, N59A, and K61A) were made at solvent exposed positions along the hairpin (Figure 4). Strands 3 and 4 are connected by a β-turn with three consecutive residues with positive phi angles (residues D53, K54, and G55). A mutation of G55 to alanine has been previously studied (Gu *et al.*, 1997) and is included in this analysis. Φ Values for E46A, D50A, and K54A could not be determined because they were only slightly destabilized. With the exception of T48A, the Φ_F values for the remaining mutations (V51A, A52G, G55A, T57A, N59A, and K61A) are less than 0.2 with an average Φ_F value of 0.13. T48A has an intermediate value of 0.26. The low Φ_F values of the solvent exposed positions suggest that the hairpin is largely unstructured in the transition state for folding.

Loops

G24A and T25A probe the formation of the loop that connects the first β-hairpin to the helix and have Φ_F values of 0.20 and 0.37, respectively. T25 is partially exposed to solvent and interacts with adjacent residues in the helix. The intermediate Φ_F values suggest the partial formation of the loop in the transition state for folding. These results are consistent with the partial Φ_F values obtained for core residues located near this region of structure.

The loop that connects the C-terminal end of the helix with the second β-hairpin is longer than the previous loop and was probed by K41A, K42A, N44A, and G45A. The stabilities of K41A and K42A were not reduced and as a result Φ values could not be determined. The Φ_F values for N44A and G45A are 0.08 and −0.10, respectively. The value for N44A is less accurate because of its small change in stability (0.34 kcal mol^{-1}). These results suggest that the loop following the helix is not structured in the folding transition state and is consistent with the low Φ_F values obtained for the second β-hairpin and the core residues located near this loop (F12, L40, and F62).

Discussion

Distribution of structure in the folding transition state ensemble

The kinetic data presented in this paper provide a comprehensive picture of the distribution of structure in the protein L transition state ensemble. These data are conveniently summarized in the schematic of the structure of protein L colored by Φ_F values displayed in Figure 5(a)-(c). High Φ_F values (red to yellow) indicate regions largely formed in the folding transition state ensemble, while low Φ_F values (blue) indicate regions largely unstructured in the folding transition state. Almost all of the high Φ_F values are contained within the first β-hairpin, suggesting that this hairpin is largely structured in the folding transition state while the helix and the second hairpin are largely unstructured. A number of mutations within the first hairpin and between the first hairpin and the helix have intermediate Φ_F values suggesting that this part of the hydrophobic core is partially structured in the folding transition state. It is interesting that L58 and I60 in the fourth strand have Φ values of ∼0.25 suggesting that the basic topology of the protein is to some extent established in the folding transition state. The structural polarization of the folding transition state is also evident in the plot of $\Delta\Delta G_{\ddagger-U}$ *versus* $\Delta\Delta G^{kin}_{F-U}$ displayed in Figure 5(d); mutations in the first β-hairpin (open triangles) group closer to the line of slope = 1 (Φ = 1), while the remaining mutations (filled circles) are distributed closer to the line of slope = 0 (Φ = 0).

The details of the Φ value distribution suggest that the folding transition state is stabilized predominantly by native-like interactions. The Φ values are consistent with a simple picture in which the first β-turn and the base of the first β-hairpin are largely structured: the Φ values are high in these regions, intermediate in regions which contact them (the residues in the helix that contact the first β-hairpin and central residues in the last strand), and very low elsewhere. As described in Results, the differences in the Φ values of multiple mutations at the same site are also consistent with this native state based model. The strongest evidence for non-native structure in the transition state is in the helix: the pattern of Φ values in the core residues suggests that the orientation of the residues towards the first β-hairpin is preserved in the transition state (Φ values are higher for sidechains that interact with the first β-hairpin, Figures 3(b)-(e)), but the low Φ values on the solvent exposed side suggest the regular helix structure is largely disrupted (Kim *et al.*, 1998b).

It is interesting to compare the folding transition state structure of protein L with that of another small α/β protein whose folding transition state has been extensively characterized: CI2. The folding transition states of the two proteins are very different. The α-helix is the most ordered element in CI2. In contrast, the first β-hairpin in protein L

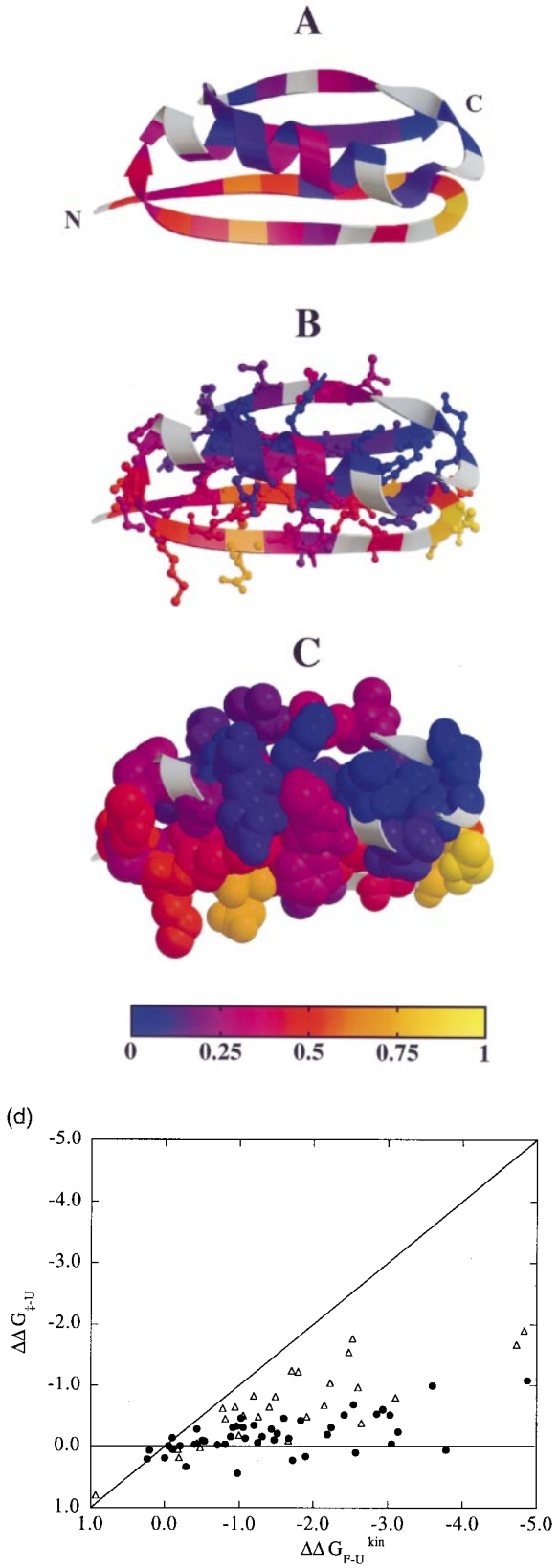

Figure 5. Structural polarization of the folding transition state ensemble. (a) Structure of protein L colored by Φ_F from a scale of 1.0 (yellow) to 0.5 (red) to 0.0 (blue), and displayed in (b) ball-and-stick and (c) space-fill representations. The Φ_F value of the mutation that makes the largest truncation of the wild-type side-chain was used at positions where multiple mutations were made. (d) Plot of $\Delta\Delta G_{\ddagger-U}$ *versus* $\Delta\Delta G^{kin}_{F-U}$. Mutations made in the first β-hairpin (open triangles) group closer to the line of slope = 1 ($\Phi = 1$). The remaining mutations are displayed as closed circles. The images were created using Molscript (Kraulis, 1991) and Raster3d (Bacon & Anderson, 1998; Merritt & Murphy, 1994).

is significantly structured while the α-helix is largely disrupted. In addition, the plot of $\Delta\Delta G_{U-\ddagger}$ *versus* $\Delta\Delta G_{U-F}$ for CI2 is linear with a slope of around 0.3 (Itzhaki *et al.*, 1995) which is equivalent to an average Φ value of 0.3. This uniform effect on the transition state is in contrast to the more dispersed effect displayed in Figure 5(d) for protein L. The distribution of Φ values for protein L is more similar to that of the significantly larger protein barnase (Itzhaki *et al.*, 1995).

Role of β-hairpin formation

A striking feature of the protein L results is the importance of the β-hairpin. Interestingly, recent studies have indicated that β-hairpin formation is also a critical step in folding of the SH3 domain (Riddle *et al.*, 1999). β-Hairpins may be favored in folding transition states since many favorable interactions can be formed without a great loss in chain entropy (the interactions are quite local). The Φ value distribution is clearly not consistent with a ''hydrophobic core first'' picture of folding; the β-hairpin appears to be at least as ordered at the rate limiting step in folding as the hydrophobic core.

The detailed effects of the mutations in the hairpin are interesting in light of recent discussions of the mechanism of β-hairpin formation (Blanco *et al.*, 1998; Dinner *et al.*, 1999; Munoz *et al.*, 1997, 1998; Pande & Rokhsar, 1999). Two alternative models have been proposed: first, that hairpins fold by zipping up from the β-turn, and second, that hairpins fold by a hydrophobic collapse followed by hydrogen bonding (Munoz *et al.*, 1997; Dinner *et al.*, 1999). Our results suggest an intermediate scenario for protein L: both the β-turn and hydrophobic interactions at the opposite end of the hairpin appear to be formed in the folding transition state, while side-chain interactions near the center of the hairpin appear to be disrupted.

Implications for models of folding

For proteins that fold in a two-state process, recent results have suggested that the shape of the folding landscape, and thus the folding process, is highly dependent upon the topology of the native state (Alm & Baker, 1999). Dramatic changes in sequence generated in phage display selection experiments have been found to have relatively little effect on protein folding rates (Riddle *et al.*, 1997; Kim *et al.*, 1998a), and proteins with the same

topology but with little sequence homology have been shown to have similar folding rates (Perl *et al.*, 1998). Additionally, folding rates have been shown to be highly correlated with the contact order (the average sequence separation of contacting residues), a property of the native topology (Plaxco *et al.*, 1998).

A simple model for folding free energy landscapes based on native topology reproduces the Φ value distribution of a number of experimentally characterized proteins with some success (Alm & Baker, 1999). However, the model fails for protein L because of the symmetry of the native structure (Figure 1(a) and (b)). The two β-hairpins make very similar contacts and bury similar amounts of surface area with each other and the helix, and therefore, the simple model treats both β-hairpins with equal importance. However, the experimental data clearly show that the structural elements that form in the folding transition state are the first β-hairpin and the adjoining hydrophobic cluster while the second β-hairpin and helix are largely disrupted. We consider it unlikely that the polarity of the chain is responsible for the asymmetry in folding, since the hairpins appear to fold in the opposite order in the structurally related IgG binding domain, protein G (E. McCallister & D.B., unpublished results).

The failure of the simple topology based model makes protein L an excellent case study for identifying factors beyond topology which determine the folding free energy landscape. There may include local sequence biases which favor particular local structure elements or heterogeneities in strength of the interresidue interactions, for example, differences in the side-chain:side-chain packing interactions in the two β-hairpins. For protein L in particular, conformational strain caused by the three consecutive positive Φ angles in the second β-turn may disfavor the formation of the second β-hairpin (the distortion in the region around the second β-turn makes possible non-local interactions with the N terminus of the helix which are likely to be realized only late in folding). On the other hand, the formation of the first β-hairpin may be favored by side-chain:main-chain hydrogen bonds in the β-turn (N14). Recent experimental evidence suggests that the first β-hairpin in protein L may be more populated than the second β-hairpin already in the denatured state ensemble (Scalley *et al.*, 1999). It is interesting that a similar consistency in the distribution of structure in the denatured state ensemble and the transition state ensemble is observed for spectrin SH3 (Serrano, personal communication), and the IgG binding domain of protein G, which has a structure very similar to that of protein L (E. McCallister & D.B., unpublished results). It is evident that the factors favoring one β-hairpin over the other already are operative in the denatured state in both protein L and protein G. Identification of these factors should considerably improve our understanding of the determinants of protein folding mechanisms.

Materials and Methods

Mutagenesis

Point mutants were made using the QuikChange site-directed mutagenesis kit (Stratagene), and were expressed and purified as described previously (Gu *et al.*, 1995, 1997). All mutants were verified by DNA sequencing and mass spectrometry.

Thermodynamic and kinetic analysis

For each experiment, protein solutions were made in 50 mM sodium phosphate, pH 7, and the temperature was kept at 295 K. The stability was determined for all mutants by equilibrium guanidine denaturation experiments using either CD or fluorescence as described previously (Scalley *et al.*, 1997). The folding and unfolding kinetics were measured by fluorescence using a Bio-Logic SFM-4 stopped flow instrument. The kinetic and equilibrium data were fit to a two state model and the data analysis was carried out as described (Scalley *et al.*, 1997).

Since our analysis depends on accurate measurements of free energy changes, we use three independent methods and avoid extrapolation whenever possible. The three estimates are:

$$\Delta\Delta G^{\mathrm{Cm}}_{\mathrm{F-U}} = \langle m\rangle(Cm^{\mathrm{mut}} - Cm^{\mathrm{wt}}) \quad (1)$$

where $\langle m\rangle$ is the average m value for all the mutants (2.06(±0.22) kcal mol^{-1} M^{-1}), and Cm^{wt} and Cm^{mut} are the concentrations of GuHCl at which 50 % of the wild-type and mutant proteins are unfolded, respectively:

$$\Delta\Delta G^{\mathrm{2M}}_{\mathrm{F-U}} = m^{\mathrm{mut}}(Cm^{\mathrm{mut}} - 2) - m^{\mathrm{wt}}(Cm^{\mathrm{wt}} - 2) \quad (2)$$

where m^{mut} and m^{wt} are the m values for mutant and wild-type, respectively:

$$\Delta\Delta G^{\mathrm{kin}}_{\mathrm{F-U}} = RT(\ln(k_{\mathrm{f}}^{\mathrm{mut(0.4M)}}/k_{\mathrm{u}}^{\mathrm{mut}(*)}) - \ln(k_{\mathrm{f}}^{\mathrm{wt(0.4M)}}/k_{\mathrm{u}}^{\mathrm{wt}(*)})) \quad (3)$$

where $k_{\mathrm{f}}^{\mathrm{wt(0.4\ M)}}$ and $k_{\mathrm{f}}^{\mathrm{mut\ (0.4\ M)}}$ are the folding rates in 0.4 M GuHCl of wild-type and mutant, respectively, and $k_{\mathrm{u}}^{\mathrm{wt}(*)}$ and $k_{\mathrm{u}}^{\mathrm{mut}(*)}$ are the unfolding rates of wild-type and mutant, respectively. $k_{\mathrm{u}}^{\mathrm{wt}(*)}$ and $k_{\mathrm{u}}^{\mathrm{mut}(*)}$ were determined in 2 M GuHCl for mutants that were significantly destabilized (I6A, G15V, F22A, F22L, A37G, E32G/A35G/T39G, L58A, I60A, F62L, and F62V), and in 4 M GuHCl for the others. We use equation (3) because it requires little or no extrapolation of either the folding or unfolding kinetic data, and the implicit assumption that $\Delta\Delta G^{\mathrm{kin}}_{\mathrm{F-U}}$ is independent of the guanidine concentration is supported by the relatively small changes in m_{f} and m_{u} in most of the mutants (Table 2).

Three different Φ value estimates were obtained using:

$$\Phi_{\mathrm{F}} = -RT\ln(k_{\mathrm{f}}^{\mathrm{wt(0.4M)}}/k_{\mathrm{f}}^{\mathrm{mut(0.4M)}})/\Delta\Delta G^{\mathrm{kin}}_{\mathrm{F-U}} \quad (4)$$

$$\Phi^{\wedge}_{\mathrm{F}} = -RT\ln(k_{\mathrm{f}}^{\mathrm{wt}}/k_{\mathrm{f}}^{\mathrm{mut}})/\Delta\Delta G^{\mathrm{Cm}}_{\mathrm{F-U}} \quad (5)$$

$$\Phi^{\mathrm{2M}}_{\mathrm{U}} = -RT\ln(k_{\mathrm{u}}^{\mathrm{wt(2M)}}/k_{\mathrm{u}}^{\mathrm{mut(2M)}})/\Delta\Delta G^{\mathrm{2M}}_{\mathrm{U-F}} \quad (6)$$

where $k_{\mathrm{f}}^{\mathrm{wt}}$ and $k_{\mathrm{f}}^{\mathrm{mut}}$ are the folding rates in the absence of denaturant for wild-type and mutant, respectively,

$k_u^{wt\ (2\ M)}$ and $k_u^{mut\ (2\ M)}$ are the unfolding rates in 2 M GuHCl for wild-type and mutant, respectively, and $\Delta\Delta G^{2\ M}_{U-F} = -\Delta\Delta G^{2\ M}_{F-U}$. In a two-state model, $\Phi_F = 1 - \Phi_U$. The fluctuations in these values for a given mutant provide more reliable error estimates than those obtained from the fitting of the kinetic and thermodynamic data.

Acknowledgments

We thank Kim Matulef, Ben McFarland and Matt Kennedy for assisting in the characterization of the mutants, Jerry Tsai for calculating the side-chain contact distribution for Figure 1(b) and Table 3, Eric Alm for calculating the buried surface areas for Table 3, and members of the Baker laboratory for useful comments on the manuscript. This work was supported by a grant from the NIH and young investigator awards to D.B. from the NSF and the Packard Foundation.

References

Alm, E. & Baker, D. (1999). Matching theory and experiment in protein folding. *Curr. Opin. Struct. Biol.* **9**, 189-196.

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

Blanco, F., Ramirez-Alvarado, M. & Serrano, L. (1998). Formation and stability of beta-hairpin structures in polypeptides. *Curr. Opin. Struct. Biol.* **8**, 107-111.

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.

Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. (1999). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.* **6**, 1005-1009.

Dinner, A. R., Lazaridis, T. & Karplus, M. (1999). Understanding beta-hairpin formation. *Proc. Natl Acad. Sci. USA,* **96**, 9068-9073.

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.

Fulton, K. F., Main, E. R., Daggett, V. & Jackson, S. E. (1999). Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J. Mol. Biol.* **291**, 445-461.

Gerstein, M., Tsai, J. & Levitt, J. (1995). The volume of atoms on the protein surface: calculated from simulation, using Voronoi poyhedra. *J. Mol. Biol.* **249**, 955-966.

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 beta-turns in the folding of a small protein. *J. Mol. Biol.* **274**, 588-596.

Gu, H., Doshi, N., Kim, D. E., Simons, K. T., Santiago, J. V., Nuali, S. & Baker, D. (1999). Robustness of protein folding kinetics to surface hydrophobic substitutions. *Protein Sci.* **8**, 2734-2741.

Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). 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.

Kim, D. E., Gu, H. & Baker, D. (1998a). The sequences of small proteins are not extensively optimized for rapid folding by natural selection. *Proc. Natl Acad. Sci. USA,* **95**, 4982-4986.

Kim, D. E., Yi, Q., Gladwin, S. T., Goldberg, J. M. & Baker, D. (1998b). The single helix in protein L is largely disrupted at the rate-limiting step in folding. *J. Mol. Biol.* **284**, 807-815.

Kragelund, B. B., Osmark, P., Neergaard, T. B., Schiodt, J., Kristiansen, K., Knudsen, J. & Poulsen, F. M. (1999). The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP. *Nature Struct. Biol.* **6**, 594-601.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Martinez, J. C. & Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Struct. Biol.* **6**, 1010-1016.

Matouschek, A., Kellis, J. T., Jr., 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.

Milla, M. E., Brown, B. M., Waldburger, C. D. & Sauer, R. T. (1995). P22 Arc repressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry,* **34**, 13914-13919.

Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. (1997). Folding dynamics and mechanism of beta-hairpin formation. *Nature,* **390**, 196-199.

Munoz, V., Henry, E. R., Hofrichter, J. & Eaton, W. A. (1998). A statistical mechanical model for beta-hairpin kinetics. *Proc. Natl Acad. Sci. USA,* **95**, 5872-5879.

Pande, V. S. & Rokhsar, D. S. (1999). Molecular dynamics simulations of unfolding and refolding of a beta-hairpin fragment of protein G. *Proc. Natl Acad. Sci. USA,* **96**, 9062-9067.

Pande, V. S., Grosberg, A. Y., Tanaka, T. & Rokhsar, D. (1998). Pathways for protein folding: is a new view needed? *Curr. Opin. Struct. Biol.* **8**, 68-79.

Perl, D., Welker, C., Schindler, T., Schroder, K., Marahiel, M. A., Jaenicke, R. & Schmid, F. X. (1998). Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nature Struct. Biol.* **5**, 229-235.

Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985-994.

Ramirez-Alvarado, M., Serrano, L. & Blanco, F. J. (1997). Conformational analysis of peptides corresponding to all the secondary structure elements of protein L B1 domain: secondary structure propensities are not

conserved in proteins with the same fold. *Protein Sci.* **6**, 162-174.

Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Yi, Q. & Baker, D. (1997). Functional rapidly folding proteins from simplified amino acid sequences. *Nature Struct. Biol.* **4**, 805-809.

Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. & Baker, D. (1999). Experiment and theory highlight role of native state topology in SH3 folding. *Nature Struct. Biol.* **6**, 1016-1024.

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.

Scalley, M. L., Nauli, S., Gladwin, S. T. & Baker, D. (1999). Structural transitions in the protein L denatured state ensemble. *Biochemistry,* **38**, 15927-15935.

Shakhnovich, E. I. (1998). Folding nucleus: specific or multiple? Insights from lattice models and experiments. *Fold. Des.* **3**, R108-R111.

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 alpha-helical coiled coil. *Proteins: Struct. Funct. Genet.* **24**, 427-432.

Thirumalai, D. & Klimov, D. K. (1998). Fishing for folding nuclei in lattice models and proteins. *Fold. Des.* **3**, R112-R118.

Villegas, V., Martinez, J. C., Aviles, F. X. & Serrano, L. (1998). Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J. Mol. Biol.* **283**, 1027-1036.

Wikstrom, M., Sjobring, U., Kastern, W., Bjorck, L., Drakenberg, T. & Forsen, S. (1993). Proton nuclear magnetic resonance sequential assignments and secondary structure of an immunoglobulin light chain-binding domain of protein L. *Biochemistry,* **32**, 3381-3386.

Wikstrom, M., Drakenberg, T., Forsen, S., Sjobring, U. & Bjorck, 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.

Edited by C. R. Matthews

(Received 18 November 1999; received in revised form 8 March 2000; accepted 14 March 2000)