Dynamic Similarity of the Unfolded States of Proteins L and G[†,‡]

Vijay R. Singh,[§,||] Michaela Kopka,[§,||] Yujie Chen,[§] William J. Wedemeyer,[§,⊥] and Lisa J. Lapidus*[,§]

Departments of Physics and Astronomy and Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824

Received February 7, 2007; Revised Manuscript Received June 15, 2007

ABSTRACT: The formation of specific intramolecular contacts has been studied under a range of denaturing conditions in single domains of the immunoglobulin-binding proteins L and G. Although they share no significant sequence similarity and have dissimilar folding pathways, the two domains have a similar native fold. Our measurements show that the rates of forming corresponding contacts in the unfolded states of both proteins are remarkably similar and even exhibit similar dependence on denaturant concentration. The unfolded proteins were modeled using Szabo, Schulten, and Schulten (SSS) theory as wormlike chains with excluded volume; when combined with our experimental data, the SSS analysis suggests that the unfolded state becomes uniformly more compact and less diffusive (i.e., rearranges more slowly) with decreasing denaturant concentrations.

The unfolded state of proteins is often characterized as a completely random, non-interacting polymer (*1*). However, Levinthal proposed the paradox that if a polypeptide chain were allowed to randomly sample conformations with a residence time of less than a nanosecond, it would take astronomically long for any protein to fold (*2*). Initially researchers responded to this paradox by hypothesizing that parts of the protein would exhibit "flickering" local structure that would become stabilized upon associating with other such parts (*3*−*5*) and that the native fold would be quickly locked in once such assembly had led to a roughly correct topology (*6*). Later work on energy landscape theory suggested that the top of the folding funnel is flat and featureless, implying that the unfolded protein has no such "flickering" structure (*7*). Recent experimental work with a variety of techniques has challenged the idea that the unfolded states of proteins are generic non-interacting polymers. Much of the experimental investigations of unfolded proteins have focused on residual structure measured by NMR under highly denaturing conditions (*8*, *9*) which may give the misleading picture that these chains are not highly flexible. On the other hand, small-angle X-ray scattering (SAXS[1]) and fluorescence resonance energy transfer (FRET) have been used to measure the overall radius of gyration of the unfolded protein, which generally increases with denaturant and can be modeled by a random walk chain (*10*−*12*). Thus a picture is emerging of the unfolded state that is still quite random but is biased by transient interactions (*13*). What has been missing is the time scale on which this random, transiently interacting chain reorganizes.

New spectroscopic techniques have been developed to probe the internal dynamics of the unfolded state (*14*−*16*). In particular, the rate of forming a specific intramolecular contact can be monitored by the triplet quenching of a tryptophan side chain by a cysteine side chain (*17*). Unstructured peptides of $(Ala\text{-}Gly\text{-}Gln)_n$ repeats can be well modeled as wormlike chains with persistence lengths of 1−2 peptide bonds and an excluded volume radius of 2−4 Å (*18*, *19*). Comparing measurements of the same peptides in 0 and 6 M GdnHCl, the average chain dimension was smaller and the intramolecular diffusion coefficient was higher in aqueous solutions. A similar study of the cold shock protein of *Thermotoga maritima* also suggested that the unfolded state gradually became more compact as the denaturant concentration was reduced, consistent with FRET observations (*20*); however, the degree of compaction was not well quantified (*21*). Another study using fluorescence quenching coupled with fluorescence correlation spectroscopy investigated intramolecular diffusion in intestinal fatty acid binding protein (IFABP) under a variety of solvent conditions and also observed a similar compaction (*22*). More recently, a single molecule FRET study on the natively unfolded yeast prion also observed compaction with decreasing denaturant con-

[†] M.K. was supported by Michigan State University start-up funds to W.J.W. V.R.S. and Y.C. were supported by Michigan State University start-up funds to L.J.L. The research of L.J.L. was supported in part by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund.

[‡] L.J.L. and W.J.W. designed the research; M.K. carried out the protein mutagenesis, expression, and purification; M.K. and V.R.S. measured the intramolecular contact rates and analyzed the data to obtain k_R and k_D; V.R.S., M.K., and Y.C. took the equilibrium unfolding data and assessed the mutant folding stability; L.J.L., V.R.S., and W.J.W. carried out the computational simulations using GPL software written by W.J.W. All the authors contributed significantly to writing the paper.

* Corresponding author. Phone: (517) 355-9200 x2211. Fax: (517) 353-4500. E-mail: lapidus@msu.edu.

[§] Departments of Physics and Astronomy.

[||] These authors contributed equally to the research and can be considered coauthors.

[⊥] Departments of Biochemistry and Molecular Biology.

[1] Abbreviations: SAXS, small-angle X-ray scattering; FRET, fluorescence resonance energy transfer; Tris, tris(hydroxymethyl)aminomethane hydrochloride; GdnHCl, guanidinium hydrochloride; CD, circular dichroism; IPTG, isopropyl-*β*-D-thiogalactopyranoside; SSS, Szabo, Schulten, and Schulten.

10.1021/bi700270j CCC: $37.00 © 2007 American Chemical Society
Published on Web 08/08/2007

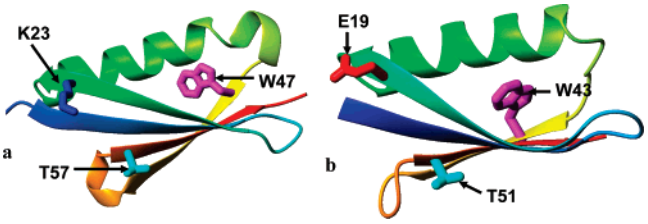

FIGURE 1: Structure and locations of mutations for the (a) protein L and (b) protein G domains.

centration and measured very rapid configuration times (10−100 ns) in water; however, the loop sizes could not be well quantified because of the presence of many quenching residues in the sequence (*23*). There have also been a few studies on contact formation in small proteins or peptides that fold on the microsecond time scale. Because these proteins fold very fast, the internal dynamics of the unfolded state could not be investigated separately, but the extracted dynamics saw little change with temperature, pH, or denaturant (*24*−*26*). The various fluorescence quenching methods used in these earlier studies have not been confirmed to be diffusion-limited, so the reported rates may also reflect the photophysics of the particular probe.

Here, the Trp/Cys triplet quenching method is used to investigate the unfolded states of two domains of immunoglobulin-binding proteins L and G under various denaturant concentrations. These two domains share no significant sequence similarity, but both have about 60 residues and adopt a similar native fold, a central α-helix resting on a four-stranded β-sheet composed of N- and C-terminal β-hairpins (Figures 1a and b). Both domains have a single tryptophan residue near the beginning of the C-terminal hairpin, Trp43 of the wild-type protein G B1 domain and Trp47 of the well-studied Y47W mutant of the protein L B1 domain (*27*). Both domains have been mutated to have a cysteine in one of two positions: (1) near the end of the N-terminal hairpin (K23C and E19C in proteins L and G, respectively), forming a 24-residue loop with the tryptophan; or (2) in the final strand of the C-terminal hairpin (T57C and T51C in protein L and G, respectively), forming a 10- and 8-residue loop. In the native conformation, contact between the tryptophan and cysteine is impossible, being separated by at least 10 Å; hence, cysteinyl quenching of the tryptophan triplet can occur *only* through a non-native contact in the unfolded state (Figure 2).

Although their folded conformations are similar, the protein L and G domains have structurally different folding transition states (*28*, *29*); moreover, the protein G domain has a folding intermediate, whereas the protein L domain does not (*27*, *30*−*32*). This work demonstrates, however, that the conformational dynamics of their unfolded states vary similarly with denaturant concentration, and differently from the dynamics of the *T. maritima* cold shock protein. Application of Szabo, Schulten, and Schulten (SSS) theory (*19*, *33*) suggests that the effective intramolecular diffusion coefficient of the unfolded state in the transition region is ∼6-fold smaller than that of the fully denatured state. We conclude that, under conditions that favor folding, the unfolded protein is compact and viscous.

MATERIALS AND METHODS

Materials. Ultrapure tris(hydroxymethyl)aminomethane hydrochloride (Tris), urea, and guanidinium hydrochloride (GdnHCl) were purchased from Invitrogen (Carlsbad CA), along with precast SDS-PAGE gels (NuPAGE, 4−12% bis-Tris). Imidazole was purchased from Sigma (St. Louis, MO), and sodium monobasic phosphate was purchased from J. T. Baker (Phillipsburg, NJ). All other reagents were of the highest grade commercially available. pH readings were made with a Fisher Accumet AR50 pH meter. Primer synthesis, mass spectrometry, and N-terminal protein sequencing were carried out in the Macromolecular Structure, Sequencing and Synthesis Facility at Michigan State University.

Protein Mutagenesis. Plasmids for the protein L and G domains were generous gifts from Drs. David Baker and Angela Gronenborn, respectively. The genes for the protein L domain (the Y47W mutant of the wild-type B1 domain of protein L) and protein G domain had been previously cloned into pET14b and pET11a expression vectors, respectively (Novagen, Madison, WI). Cysteine mutants were prepared using the Quikchange kit (Catalog No. 200518) from Stratagene (La Jolla, CA), using the following primers:

protein L (K23C): 5′-CA CAA ACT GCA GAA TTC TGC GGA ACA TTT GAA AAA GC-3′

protein L (T57C): 5′-AC GTT GCA GAT AAA GGT TAT TGC TTA AAT ATT AAA TTT GCT GG-3′

protein G (E19C): 5′-GGT GAA ACC ACC ACC TGC GCT GTC GAC GCT GCT-3′

protein G (T51C): 5′-CGA CGC TAC CAA ATG CTT CAC GGT AAC CG-3′

The resulting mutants were confirmed by DNA sequencing.

Protein Expression and Purification. The mutant plasmids were transformed into BL21(DE3) *Escherichia coli* cells for expression. A single colony was picked from the transformants, grown in 5.0 mL starter cultures, inoculated into 1 L cultures, and incubated for ∼17 h. Protein expression was induced by 1.0 mM IPTG, followed by incubation for an additional ∼6 h. The cells were harvested by centrifugation for 20 min at 6000 rpm and 4 °C, and the pellets were resuspended in NiA buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol, 100 mM leupeptin, and 1 mM phenylmethylsulfonylfluoride) at pH 8.0, 4 °C; the β-mercaptoethanol helped to prevent dimerization of the cysteine mutants during subsequent purification steps. The cells were lysed with a Misonix 3000 Sonicator (Farmingdale, NY), and pelleted at 13,000 rpm for 20 min at 4 °C. The supernatant was heated for 10 min in a 70 °C water bath and recentrifuged at 13,000 rpm for 20 min.

The final purification step differed for the protein L and G domains. The protein L domain, having a N-terminal hexa-His tag, was purified by Ni-affinity chromatography; after binding, the column was washed with three column volumes of Ni-A buffer, Ni-A with 6 M urea, and Ni-A with 600 mM NaCl to eliminate nonspecifically bound proteins. The protein was eluted using a linear gradient of Ni-A to Ni-B buffer (Ni-B buffer differs from Ni-A only in having 300 mM imidazole). Since the protein G domain did not have a histidine tag, it was purified by anion exchange chromatography. The protein L and G domains were buffer-exchanged into 15 mM Tris-HCl, 100 mM NaCl at pH 7.5 and purified

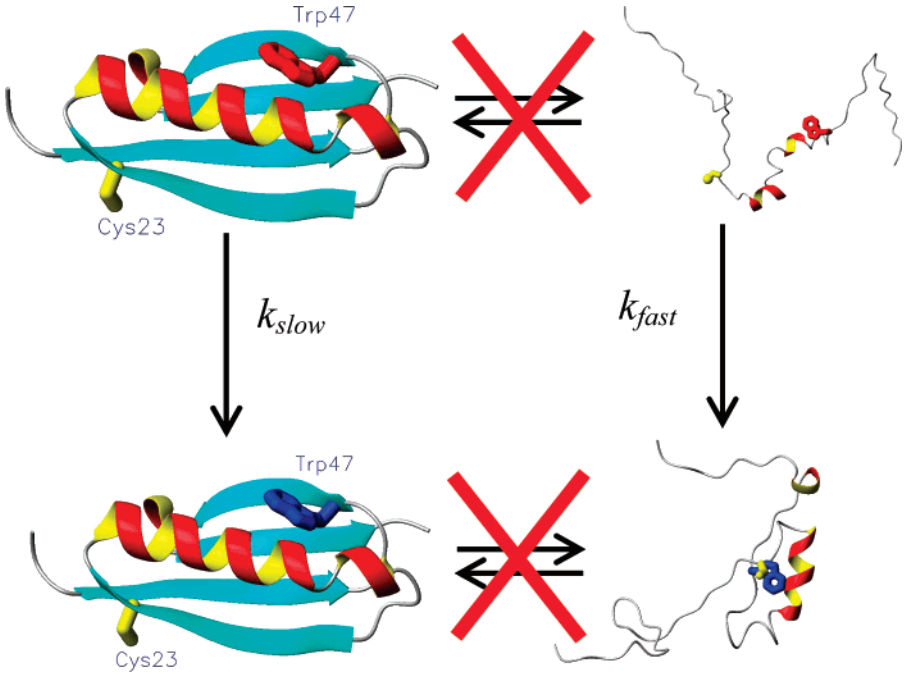

FIGURE 2: Model of the contact formation measurement. The tryptophan residue is excited to the triplet state (red) and can decay to the ground state (blue) either via natural decay in the folded state or via contact formation in the unfolded state. Since these processes occur much faster than folding or unfolding, the observed populations cannot interchange during the experiment (red ×'s).

in a final polishing step by gel filtration on a Hi-Prep 26/60 Sephacryl S100 column. The purified proteins were verified by N-terminal sequencing and MALDI-TOF mass spectrometry. Protein concentrations were calculated by absorbance at 280 nm. The pure proteins were stored at −20 °C in phosphate buffer (100 mM K_2HPO_4 and KH_2PO_4 in 3:2 ratio, with 1 mM Tris(2-carboxyethylphosphine at pH 7.0) with 1 mM TCEP.

The solution conditions for the contact formation experiments have been described previously (*18*). The measurements were generally made at pH 7 in 100 mM potassium phosphate buffer; however, all measurements of the Trp-only protein L (no cysteines) and some measurements of the mutants were taken at pH 8.5 as controls to assess the quenching ability of the hexa-His tag.

Equilibrium Fluorescence and Circular Dichroism. Protein fluorescence spectra were recorded on a Jobin Yvon Spex Fluorolog-3 spectrofluorometer with a 1 s integration time and 3 nm bandwidth. The typical concentration was 10 μM. Circular dichroism (CD) spectra were recorded with Applied Photophysics Chirascan with 0.5−4 s adaptive integration time and 1 nm bandwidth. The average concentration was 10 μM, and the path length was 10 mm. Singular value decomposition was performed using Matlab (The Mathworks).

Contact Formation Rates. The instrument to measure tryptophan triplet lifetimes is essentially the same as that described previously (*17*), with the following exceptions. The excitation beam is shifted from 266 to 289 nm using a 1 m Raman cell (Light Age) filled with ∼150 psi of methane gas (conversion efficiency ∼25%). The triplet population is probed at 442 nm by a HeCd laser (Kimmon), which reduces the relative amplitude of long-lived radical species created by the UV pulse (*34*) compared to previous measurements (*17*). The temperature and viscosity were controlled as described previously (*19*). The viscosity of each solvent at every temperature was measured independently using a cone−cup viscometer (Brookfield Engineering).

Kinetic Modeling. The kinetics of intramolecular contact formation in the unfolded state (k_{fast}) can be modeled using a two-state model (*17*). Intramolecular diffusion brings the excited tryptophan and the cysteine residues together with a forward rate, k_{D+}, and the two residues diffuse apart at a backward diffusion rate, k_{D-}. When the two residues are in close contact, irreversible quenching can occur with a rate k_q. Therefore, the observed rate can be expressed as

$$k_{obs} = \frac{k_{D+}k_q}{k_q + k_{D-}} \tag{1}$$

If $k_q \gg k_{D-}$, $k_{obs} = k_{D+}$. Since previous studies on the Trp-Cys quenching system have shown that diffusion and quenching rates are comparable (*17*), k_{obs} can be rewritten as

$$\frac{1}{k_{obs}} = \frac{1}{k_q K} + \frac{1}{k_{D+}} = \frac{1}{k_R(T)} + \frac{1}{k_{D+}(T,\eta)} \tag{2}$$

where $K \equiv k_{D+}/k_{D-}$ is the equilibrium constant for forming the Trp-Cys contact and $k_R \equiv k_q K$. We assume that the reaction limited rate, k_R, depends only on temperature and k_{D+} depends on both temperature and viscosity (*19*). Therefore, a plot of $1/k_{obs}$ vs viscosity (η) at a constant temperature is a line with the intercept equal to $1/k_R$ and the slope equal to $1/\eta k_{D+}$ (see Figure 6). A data set of k_{obs} at a particular denaturant concentration, 5 temperatures, and 4 concentrations of sucrose is modeled as

$$k_R(T) = k_{R0} \exp\left(\frac{E_0(T - T_0)}{RTT_0}\right) \tag{3}$$

$$k_{D+}(T) = \frac{k_{D+0}T}{\eta T_0} \exp(\gamma(T - T_0)) \quad (4)$$

where T is the absolute temperature, $T_0 = 273$ K, η is the solution viscosity and k_{R0}, k_{D+0}, E_0, and γ are fitting parameters. For the lowest concentrations of GdnHCl in which the unfolded state was observed, the temperature dependence of k_R disappeared so E_0 was constrained to be zero in all mutants except T57C.

Computational Modeling. The proteins studied experimentally were modeled with wormlike chains using excluded volume following the method similar to that in ref *18*. Briefly, chains with a persistence length, l_p, were randomly generated by sequentially adding random links to the C-terminus. Each link in the chain is 0.38 Å long, or 1/10 of a peptide bond. However, instead of discarding all chains that had intrachain interactions that were less than d_α when an excluded volume clash was detected, the chain was truncated 3 persistence lengths before the clash and the chain was regenerated. This change of procedure significantly improved the speed of the algorithm, allowing for the simulation of millions of chains in less than an hour of CPU time. Buscaglia et al. found that reaction limited rates were depressed by about a factor of 1.7 for each tail due to excluded volume and that the effect was very insensitive to the length of the tail (*18*), so the simulated tails were always set to 5 residues. For each loop length (8, 10, and 24 residues), 2×10^7 chains were generated for each persistence and contour length, and a normalized histogram (0.1 Å binning) of Trp−Cys distances was used as $P(r)$ in eqs 5 and 6 which were numerically integrated to find k_R and k_{D+}. As a cross-check k_R was estimated from the mean quenching rate averaged over all generated chains.

RESULTS

Tryptophan Solvent Exposure in the Folded and Partially Folded States. The tryptophan triplet may decay by two mechanisms, electron transfer resulting from close molecular contact with a cysteine sulfhydryl and a slower "natural" decay possibly due to solvent quenching (Figure 2). Because the cysteine mutations were located far from the tryptophan in the folded structure, intramolecular contact is possible only when the molecule is unfolded. However, the unquenched tryptophan triplet decays at a slower rate that depends on the local environment of the amino acid. Therefore, at intermediate concentrations of denaturant two kinetic phases are observed as can be seen in Figure 3. We attribute the fast rate to intramolecular contact with cysteine in the unfolded state and the slow rate to natural decay in the folded or partially folded state. Figures 4a and 4b show the fraction of unfolded molecules as measured by the fast phase amplitude. The equilibrium stability of both proteins is essentially the same, regardless of mutant. These equilibrium unfolding curves are generally in agreement with that reported by Park et al. for protein G (*30*, *31*) but are significantly different from those reported by Scalley et al. for protein L (*32*), both of which measured equilibrium fluorescence. However, folding equilibria determined by a variety of spectroscopic methods yields a wide range of thermodynamic parameters for protein L (see Table 1), which may mean that a two-state model is not appropriate for protein L.

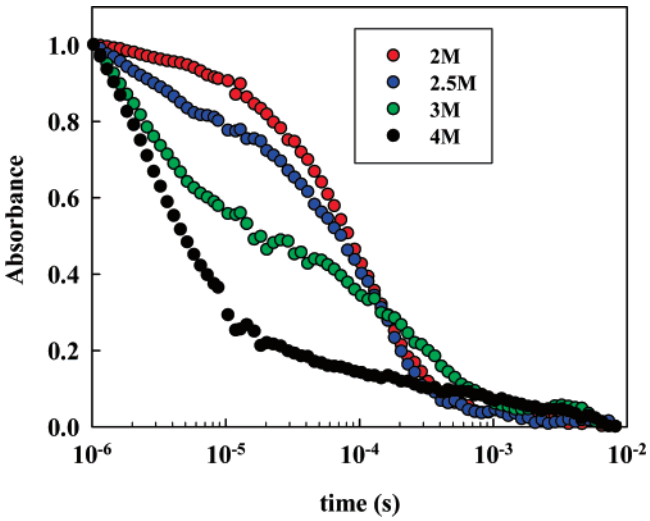

FIGURE 3: Normalized absorbance of the tryptophan triplet state of the T51C mutant as a function of time for various concentrations of GdnHCl.

Figures 4c and 4d show the fast and slow rates of both proteins grouped by loop length. Whereas, for protein G, the slow rate is the same at all concentrations of denaturant, for protein L, the rate decreases from ~22000 s^{-1} to ~5000 s^{-1} in an apparently cooperative transition. This is observed in both double mutants and in the control protein, Trp-only protein L (gray squares in Figure 4c). The higher protein L rate is approximately equal to that measured for *N*-acetyltryptophan amide (gray diamonds in Figure 4d) in water whereas the lower rate is consistent with rates for tryptophan measured in the hydrophobic core of a protein (*35*). Neither rate is fast enough to reflect intramolecular contact. These results suggest that protein L has a compact non-native state in which the tryptophan on hairpin 2 is solvent exposed.

Further evidence of a compact non-native state of protein L can be seen in Figure 5, which shows the two-state fits of various spectroscopic data: fluorescence intensity, wavelength of peak fluorescence, relative amplitude of intramolecular contact formation, and the magnitude of the unquenched tryptophan triplet lifetime. The difference in the fluorescence intensity and contact formation curves suggests that these data are sensitive to two different structural states. The relative amplitude of the fast rate gives the fraction of unfolded molecules whereas fluorescence intensity or maximum wavelength changes are usually sensitive to the fraction of folded molecules. The presence of at least one state between the folded and unfolded states could account for the difference between the fluorescence intensity and contact formation curves. This state could be described by the fluorescence wavelength and the rate of the slow process equilibrium curves, which are in quite good agreement. Vanderkooi and co-workers have shown that the tryptophan triplet lifetime depends on the solvent exposure of the residue and correlates with the fluorescence spectrum: the longer the lifetime of the triplet, the shorter the maximum fluorescence wavelength (*35*).

The slow rate observed in protein G does not change significantly with GdnHCl concentration, implying that the tryptophan remains hydrophobically buried until the protein is fully unfolded. Although protein G acts as a 2-state protein in our experiments, Park et al. have observed a kinetic

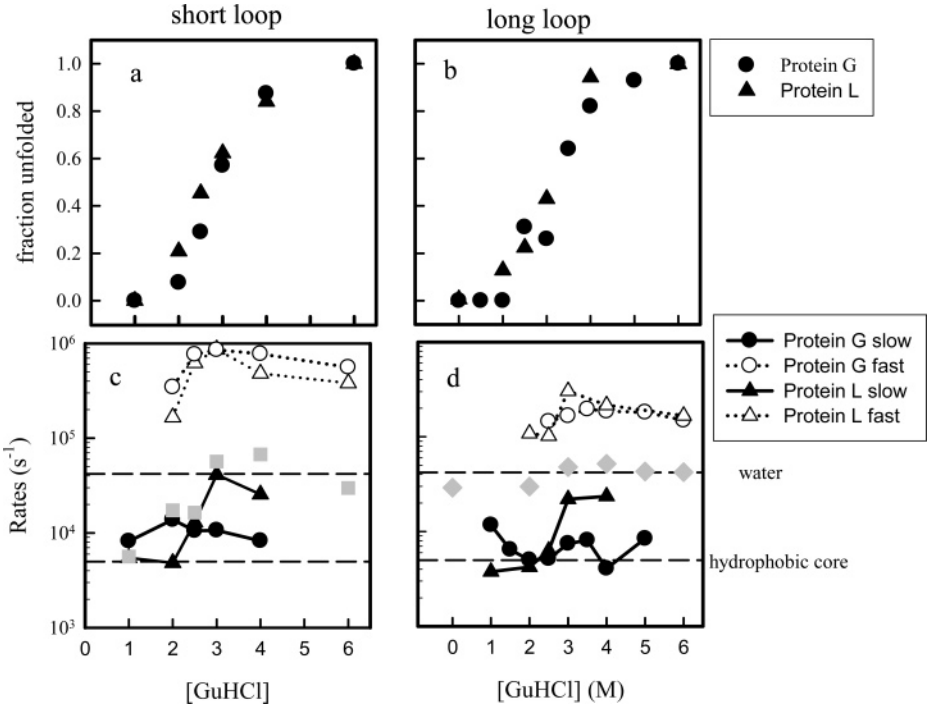

FIGURE 4: (a, b) Fraction of unfolded molecules as measured by the relative amplitude of the fast phase. (c, d) Fast and slow rates measured at various concentrations of GdnHCl. Graphs on the left are for the short loop mutants (T57C and T51C), and graphs on the right are for the long loop mutants (K23C and E19C). The gray points are measured tryptophan triplet rates in the control protein L Trp-only mutant (squares), which has no cysteine, and *N*-acetyltryptophan amide (diamonds).

Table 1: Thermodynamic Parameters for Protein L Wildtype (Trp Only) and T57C Mutant

type	method[a]	ΔG_U (kcal/mol)	m (kcal/mol/M)
WT	Fluor, $\lambda_{ex} = 280$, $\lambda_{em} = 320$[b]	4.6	1.85
WT	Fluor, $\lambda_{ex} = 280$, $\lambda_{em} = 334$	3.86 ± 1.01	1.73 ± 0.44
WT	Fluor, $\lambda_{ex} = 297$,[c] $\lambda_{em} = 334$	1.26 ± 0.66	0.69 ± 0.26
WT	Fluor, max wavelength[d]	5.15 ± 1.23	1.78 ± 0.43
WT	CD $\lambda_{ex} = 220$	2.24 ± 0.42	0.88 ± 0.16
T57C	Fluor, $\lambda_{ex} = 297$, $\lambda_{em} = 334$	1.18 ± 0.95	0.82 ± 0.41
T57C	Fluor, max wavelength	5.13 ± 1.01	2.12 ± 0.41
T57C	CD $\lambda_{ex} = 220$	1.75 ± 0.33	1.02 ± 0.14
T57C	intramolecular contact[e]	2.61 ± 0.33	0.96 ± 0.12

[a] All measurements were made at room temperature in 0.1 M potassium phosphate buffer (pH 7). [b] From Scalley et al. (*32*). [c] The difference between exciting fluorescence at 280 and 297 nm is that the 297 nm excites only tryptophan while 280 nm also excites tyrosine emission. [d] The peak fluorescence intensity wavelength, $\lambda_{ex} = 297$ nm. [e] Data from Figure 4a (triangles).

intermediate of protein G using a rapid mixer and fluorescence detection (*30*, *31*). This suggests that protein L and G each have an intermediate, but these intermediates are structurally different from one another. In protein G, hairpin 2 forms a stable intermediate that effectively buries the tryptophan at position 43. The rate-limiting step is the formation of hairpin 1 (*29*), but the contact formation measurement is insensitive to this step. In contrast, in protein L hairpin 1 forms first (*27*, *28*), which creates a compact, stable intermediate but leaves the tryptophan at position 47 solvent exposed. The rate-limiting step is the formation of hairpin 2 which hydrophobically buries the tryptophan, slowing the unquenched triplet lifetime. The contact rate measurements cannot prove conclusively whether these intermediates lead directly to the folded state, but Park et al. concluded that the kinetic intermediate for protein G was

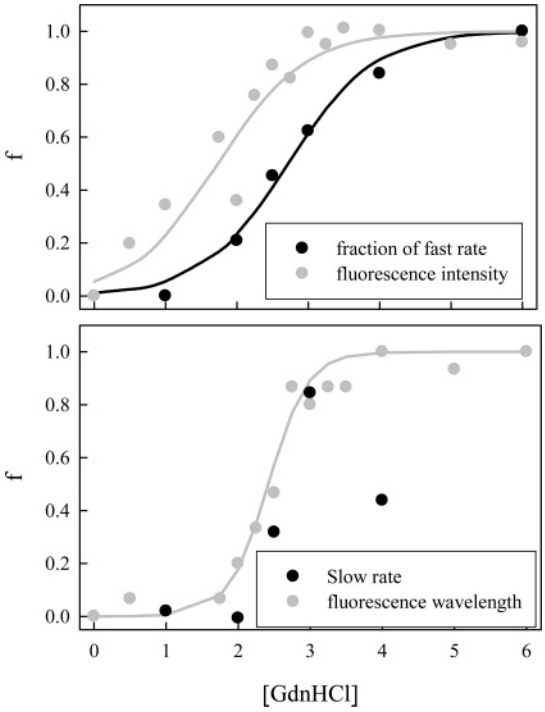

FIGURE 5: The fraction of unfolded molecules (circles) of the T57C mutant of protein L as measured by various spectroscopic probes. The lines are fits to two-state models (see Table 1 for parameters).

on pathway (*31*), so the intermediate for L may be on pathway as well.

Intramolecular Contact Rates of the Unfolded State. The contact formation (fast) rate of the unfolded state for both proteins increases slightly with decreasing denaturant but then decreases significantly below 3 M GdnHCl (white symbols for Figures 4c and 4d). This contrasts sharply with similar measurements of the cold shock protein of *Thermotoga*

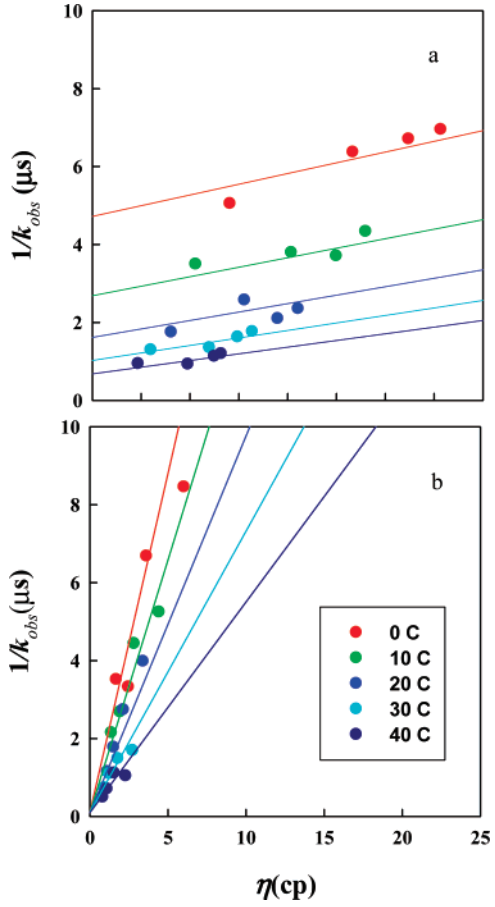

FIGURE 6: Temperature and viscosity dependence of observed quenching rates of protein L T57C at 6 M (a) and 2.3 M (b) GdnHCl. Errors of these measurements are typically less than 10%. Lines are fits to eqs 2, 3, and 4.

maritime in which the fast rate increased monotonically as denaturant decreased to at least 1 M GdnHCl (*21*). Since the observed fast rate is composed of a reaction-limited rate, $k_R(T)$, and a diffusion-limited rate, $k_{D+}(T,\eta)$, measurements were made at varying viscosities and temperatures, but the same denaturant concentration. Figure 6 shows such a data set for 2.3 and 6 M GdnHCl. As described in Materials and Methods, the intercept of these plots gives $1/k_R$, while the slope gives $1/\eta k_{D+}$. These two rates have opposing trends; as the denaturant concentration is decreased, the reaction limited rate, k_R, increases and the diffusion-limited rate, k_{D+}, decreases (Figure 7). For the lowest denaturant concentrations at which the unfolded state is measurable (2.3 M GdnHCl), the temperature dependence of k_R (*y*-intercepts in Figure 6) appears to be suppressed. As shown in eq 2, k_R depends on two quantities, the instantaneous quenching rate k_q, and the equilibrium rate $K \equiv k_{D+}/k_{D-}$. Based on the known bimolecular quenching rate of tryptophan by cysteine of 2×10^8 M^{-1} s^{-1}, at a quenching radius of 4 Å, k_q is 7×10^6 s^{-1}, which is approximately the reaction-limited rate for the short loops at 2.3 M GdnHCl. Thus, we may conclude that the equilibrium rate $K \geq 1$ at low concentrations of GdnHCl and that most of the temperature dependence observed in k_R is due to *K*. Furthermore since k_{D+} exhibits a similar temperature dependence, k_{D+} is likely the primary temperature contributor to *K*.

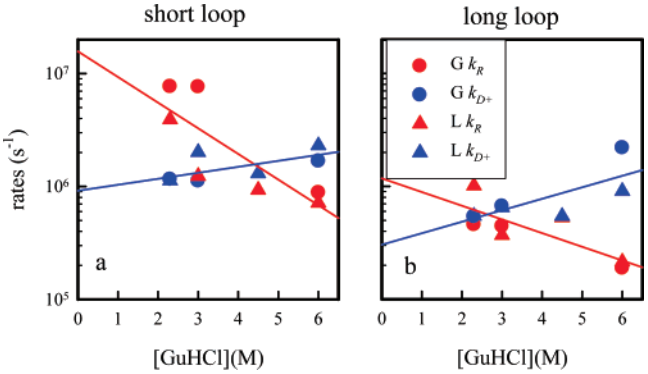

FIGURE 7: Reaction-limited and diffusion-limited rates for the short loop (a) and long loop (b) in proteins L and G determined by eqs 2, 3, and 4. Based on the sum of squares of the fit, the error of these rates is typically less than 10%. The lines are added as an aid to the reader.

DISCUSSION

In this study we sought to investigate the unfolded state of two structurally similar but sequentially nonhomologous proteins under folding conditions. However, our technique has also yielded information on a hitherto unseen partially folded state of protein L. Our mutations placed a cysteine either 8, 10, or 23 residues apart in sequence but at least 1 nm apart in the folded structure such that the tryptophan triplet is not quenched by contact with the cysteine. Therefore, the observed lifetime of the folded state only depends on the local environment of the tryptophan, and the relative amplitude of the cysteine quenching rate is a good probe of the equilibrium transition of the unfolded state. The value of the slow rate undergoes an apparently cooperative transition in protein L. This implies that there exists a compact intermediate in which W47 on the C-terminal hairpin is solvent exposed. Similar behavior of the slow rate is not observed in protein G, but previous studies have shown that the C-terminal hairpin is native-like in the intermediate state (*30*, *31*).

The dynamics of the unfolded states of proteins G and L are remarkably similar. The most striking aspect of the dynamics is that the reaction-limited rate and the diffusion-limited rate vary in opposite ways as the solvent conditions become more favorable for folding. At 6 M GdnHCl, the diffusion-limited rate is 2−5 times faster than the reaction-limited rates for all mutants and is similar to rates measured for unstructured peptides under the same conditions. But while the reaction-limited rate *increases* as denaturant decreases, the diffusion-limited rate *decreases*. To understand this behavior quantitatively, we employ Szabo, Schulten, and Schulten (SSS) theory, which states that the reaction-limited and diffusion-limited rates are given by (*19*, *33*)

$$k_R = \int_a^\infty q(r)\, P(r)\, dr \qquad (5)$$

$$\frac{1}{k_{D+}} = \frac{1}{k_R^2 D}\int_a^{l_c} \frac{dr}{P(r)}\left\{\int_r^{l_c} (q(x) - k_R)P(x)\, dx\right\}^2 \qquad (6)$$

where *r* is the distance between the tryptophan and cysteine and *P(r)* is the probability density of finding the polypeptide at that distance. *D* is the effective intramolecular diffusion coefficient, *a* is the distance of closest approach (defined to

be 4 Å), and $q(r)$ is the distance-dependent quenching rate. The distance-dependent quenching rate for the Trp-Cys system has been determined experimentally and drops off very rapidly beyond 4 Å, so the reaction-limited rate is mostly determined by the probability of the shortest distances (*36*). Thus the only free parameters in the equations above are D and $P(r)$. For example, if the unfolded protein is modeled by a Gaussian chain and q is assumed to be zero for $r > a$, the reaction-limited rate is

$$k_R = \frac{4\pi q a^3}{[(2/3)\pi\langle r^2\rangle]^{3/2}} \exp\left(-\frac{3a^2}{2\langle r^2\rangle}\right) \quad (7)$$

where $\langle r^2\rangle$ is the average tryptophan-cysteine distance squared. If the average chain volume, $\langle r^2\rangle^{3/2}$, decreases, k_R increases, so a ratio of the measured reaction-limited rates at 2 denaturant concentrations gives an approximate change in the volume of the unfolded chain. For the long loops, E19C and K23C, k_R(2.3 M)/k_R(6 M) is 2.4 and 4.7, respectively. These results are in qualitative agreement with work by Sherman and Haran and by Merchant et al., who measured the end-to-end distance of the unfolded state of protein L at various concentrations of GdnHCl with single-molecule FRET (*11*, *37*).

The diffusion-limited rate of the Gaussian chain is

$$k_{D+} = \frac{4\pi D a}{[(2/3)\pi\langle r^2\rangle]^{3/2}} \quad (8)$$

k_{D+} is also inversely proportional to $\langle r^2\rangle^{3/2}$ and should also increase, as was seen for unstructured peptides (*18*). However k_{D+} is also directly proportional to D so the observed decrease in k_{D+} must correspond to a significant decrease in the diffusion coefficient. Taking the ratio $(k_R/k_{D+})|_{2.3M}(k_{D+}/k_R)|_{6M}$ we find that the effective diffusion coefficient decreases by 8−10-fold for the long loops and 15−30-fold for the short loops.

To further investigate the polymer dynamics of the unfolded state under folding conditions, we modeled the protein as a wormlike chain with excluded volume and used Szabo, Schulten, and Schulten (SSS) theory to estimate the effective persistence length and intramolecular diffusion constant at various concentrations of GdnHCl. We assume that the persistence length, l_p, is an intrinsic property of the chain and does not change with solvent conditions, but that the excluded volume diameter, d_α, and diffusion constant, D, depend on solvent and intramolecular interactions and therefore depend on the concentration of guanidine.

Simulation of 2 million chains results in a probability distribution, $P(r)$. Then, using $P(r)$ and the full description of $q(r)$ determined by Lapidus et al. (*36*) eq 4 was used to calculate the reaction-limited rate for a variety of persistence lengths and excluded volume diameters and compared to the measured rates at 6 M GdnHCl. We found that the calculated rates were much more sensitive to d_α than to l_p. Therefore the persistence length was kept fixed at 4 Å as was found for unstructured peptides (*18*), and the excluded volume radius was allowed to vary. As shown in Table 2, for each mutant at 6 M GdnHCl, the best fits of d_α are very close to 4 Å, indicating that the polymer properties of the fully denatured proteins are very similar to the completely unstructured peptides from ref *18* in 6 M GdnHCl. Chains

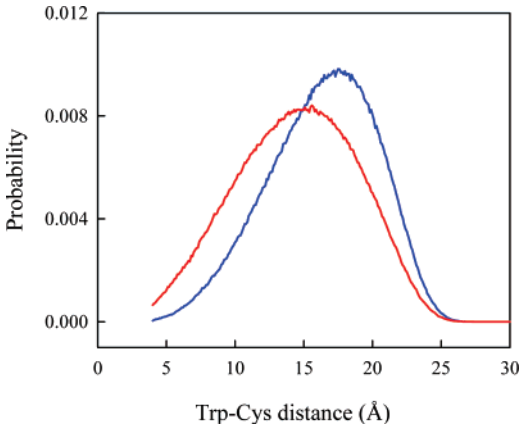

FIGURE 8: Distance distribution of an 8-residue wormlike chain with 5 residue tails. Thus the distance shown is between the 6th and 14th residues of a 19-residue chain. The chain parameters are (blue) $l_p = 4.0$ Å, $d_\alpha = 4.05$ Å and (red) $l_p = 4.0$ Å, $d_\alpha = 2.00$ Å.

Table 2: Parameters Used in Wormlike Chain Simulations and the Resulting k_R Calculated Using Eq 5

mutant	n	l_p (Å)	[GdnHCl] (M)	d_α (Å)	$\langle r^2\rangle^{1/2}$ (Å)	calc k_R (s^{-1})	meas k_R (s^{-1})
T51C	8	4.0	6.0	4.05	16.8	8.0×10^5	8.8×10^5
			2.3	2.0	15.1	8.0×10^6	7.7×10^6
T57C	10	4.0	6.0	3.9	19.2	7.4×10^5	7.2×10^5
			2.3	2.3	17.5	4.3×10^6	3.9×10^6
E19C	24	4.0	6.0	3.9	32.4	1.8×10^5	1.9×10^5
			2.3	3.3	30.7	4.0×10^5	4.6×10^5
K23C	24	4.0	6.0	3.8	32.1	2.0×10^5	2.1×10^5
			2.3	2.3	28.6	1.1×10^6	1.0×10^5

were then generated for smaller d_α to match the measured reaction-limited rates at 2.3 M GdnHCl (see Table 2). Column 6 of Table 2 gives the root mean squared Trp−Cys distance calculated using these probability distributions; between 6 and 2.3 M GdnHCl, the average Trp−Cys distance decreases by ~10% in reasonable quantitative agreement with Merchant et al. (*37*).

Figure 8 shows the probability of Trp−Cys distances for the 8-residue loop (T51C) that best fit the measured k_R at 6 M (blue) and 2.3 M (red) GdnHCl. Compared to 6 M GdnHCl, the probability distribution at 2.3 M is also broader. This trend of smaller intramolecular distances and broader distributions is in general agreement with the observations of Laurence et al. of single molecule FRET measurements with nanosecond resolution on unfolded CI2 and ACBP, who noted that this trend cannot be satisfied with either a Gaussian chain or a wormlike chain without excluded volume (*38*). In contrast, similar measurements by Hoffman et al. on unfolded cold shock protein show good agreement with a Gaussian chain with excluded volume for all concentrations of GdnHCl (*12*). Buscaglia et al. (*21*) measured intramolecular contact in the cold shock protein (though did not resolve reaction-limited and diffusion-limited rates) and observed a qualitatively different pattern of rates than seen in Figure 4, so the dynamics of its unfolded state may be significantly different than that for proteins L and G. Further measurements by Hoffman et al. using time-resolved circular dichroism following ultrafast mixing show that about one-third of the secondary structure is formed in the first 400 μs of folding. Additionally, measurements on the cold shock protein of *Bacillus caldolyticus* by Magg et al. showed that the initial collapse was neither entirely specific nor nonspe-

Table 3: Diffusion Coefficients Calculated from Eq 6 Using the Wormlike Chain Probability Distributions Given in Table 2

protein	mutant	[GdnHCl] (M)	meas k_{D+} (s^{-1}) ($\eta = 1$ cP)	$D \times 10^6$ ($cm^2\ s^{-1}$)	$D_{6M}/D_{2.3M}$
G	T51C	6	1.7×10^6	0.17	5.7
		2.3	1.2×10^6	0.03	
L	T57C	6	2.3×10^6	0.33	6.6
		2.3	1.1×10^6	0.05	
G	E19C	6	2.2×10^6	1.5	6.8
		2.3	5.4×10^5	0.22	
L	K23C	6	9.1×10^5	0.57	5.2
		2.3	5.4×10^5	0.11	

cific (*39*). Taken together, these results suggest that the unfolded states of different folds exhibit different dynamics. Future avenues of research could compare transient structure in the unfolded state for proteins of different structural classes.

Using eq 6, the effective diffusion coefficient, *D*, was calculated from the wormlike chain probability distributions given in Table 2 and the measured diffusion-limited rates. These values are given in Table 3. They vary quite substantially by mutant and, except for K23C, are significantly lower than that measured by Buscaglia et al. in 6 M GdnHCl. We conclude that the diffusion coefficient is a local property of the loop sequence and support this claim by two observations. First, the sequences measured by Buscaglia et al. are ~33% glycine, so intramolecular diffusion should be faster than the sequences in this study. Second, the diffusion coefficients for the T51C and T57C loops, which form β-strands, are much lower than for the K23C and E19C loops, which form mostly α-helix; this difference likely reflects the propensity for extended structure in the β-strand sequences.

The calculated effective diffusion coefficients in 2.3 M GdnHCl for all mutants *decrease* by about a factor of 6 relative to 6 M GdnHCl, as given in the last column of Table 3. This uniformity of scaling for each mutant suggests that this trend is a global property of the chain. Thus, the loss of denaturant in a real protein yields an unfolded state that is less diffusive than the fully denatured state and reflects transient interactions throughout the entire chain. A 6-fold decrease in *D* represents a very significant change in the internal dynamics of the proteins. As a point of reference, consider the fact that, empirically, bimolecular diffusion coefficients of folded proteins through water depend on the molecular weight as $D = 10^{-7} \exp(6.4 - 0.4 \ln m)$ $cm^2\ s^{-1}$ (*17*), so a decrease in *D* from 1×10^{-6} to 2×10^{-7} $cm^2\ s^{-1}$ is equivalent to an almost 1000-fold decrease in mass. Note that this decrease in *D* is completely different than the trend observed by Buscaglia et al. on unstructured peptides in which the coefficient *increases* by about a factor of 2 between 6 and 0 M GdnHCl (*18*). We attribute this qualitatively different behavior to the fact that the peptides measured previously were completely unstructured and contained no hydrophobic residues. The intramolecular diffusion rate of unstructured peptides has been used as a measure of the protein folding "speed limit" (*40*). Since *D* values measured in this work under folding conditions are up to 20 times lower than measured for unstructured peptides, we conclude that the low diffusivity of the unfolded state ultimately limits the folding process. Our extrapolation of k_{D+} to 0 M GdnHCl in Figure 4 indicates that this internal friction time is on the order of microseconds, in good agreement with measurements by Hagen et al. of fast folding protein rates as a function of viscosity and temperature (*41*).

Two other recent experiments have also observed a decrease in diffusion coefficients with decreasing denaturant. Nettles et al. measured intramolecular diffusion of the unfolded state of the cold shock protein using single molecule fluorescence correlation spectroscopy and observed a 2-fold decrease in *D* from 6 M to 2.3 M GdnHCl (*42*). This result is in good agreement with our earlier observation that the cold shock protein does not seem to collapse to as viscous a state as proteins L and G. Möglich et al. have measured intramolecular diffusion rates and average end-to-end distance in unstructured Gly-Ser peptides using bulk FRET (*43*). They also find that the average end-to-end distance and the diffusion coefficient decrease in water compared to high levels of denaturant, but to a much larger degree than is observed in this work or by Nettles et al., an observation Möglich et al. attribute to backbone−backbone interactions. Their results also contradict Ala-Gly-Gln measurements of Buscaglia et al., which use the same methodology as the present work. Furthermore, it seems unclear that the observed collapse is due to the formation of backbone−backbone hydrogen bonds because such interactions seem unlikely by themselves to condense an otherwise unstructured peptide (*44*).

Finally we note that the diffusion coefficient scaling for each protein is the same. Should this necessarily be the case? These proteins have the same fold, but very different folding transition states and sequences. However, they also have quite similar hydrophobicity patterns as measured by recurrence quantification analysis (*45*). If it is assumed that hydrophobic interactions predominate in the early steps of folding (*46*), these unfolded states may diffuse on energy landscapes similarly patterned by such transient accessible interactions which may ultimately determine the folding pathway and final structure.

ACKNOWLEDGMENT

We thank Dr. David Baker and Dr. Angela Gronenborn for kindly providing plasmids for the protein L and G domains, respectively. Dr. Jun Sun and Terry Ball provided assistance in protein expression.

SUPPORTING INFORMATION AVAILABLE

Supplemental Figure 1 shows the background corrected fluorescence spectra for Protein L T57C at various concentrations of GdnHCl. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Tanford, C. (1968) Protein Denaturation, *Adv. Protein Chem. 23*, 121−282.
2. Levinthal, C. (1969) In *Mossbauer Spectroscopy in Biological Systems* (De Brunner, P., Tsibris, J., and Munck, E., Eds.) pp 22−24, University of Illinois Press, Monticello, Illinois.
3. Anfinsen, C. B. (1973) Principles That Govern the Folding of Protein Chains, *Science 181*, 223−230.
4. Ptitsyn, O. B. (1973) Stagewise Mechanism Of Self-Organization Of Protein Molecules, *Dokl. Akad. Nauk SSSR 210*, 1213−1215.
5. Scheraga, H. A. (1971) Theoretical And Experimental Studies Of Conformations Of Polypeptides, *Chem. Rev. 71*, 195 ff.

6. Levitt, M., and Warshel, A. (1975) Computer-Simulation Of Protein Folding, *Nature 253*, 694−698.
7. Onuchic, J. N., LutheySchulten, Z., and Wolynes, P. G. (1997) Theory of protein folding: The energy landscape perspective, *Annu. Rev. Phys. Chem. 48*, 545−600.
8. Dyson, H. J., and Wright, P. E. (2004) Unfolded proteins and protein folding studied by NMR, *Chem. Rev. 104*, 3607−3622.
9. Montelione, G. T., and Scheraga, H. A. (1989) Formation Of Local Structures In Protein Folding, *Acc. Chem. Res. 22*, 70−76.
10. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Hasan, M. Z., Pande, V. S., Ruczinski, I., Doniach, S., and Plaxco, K. W. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins, *Proc. Natl. Acad. Sci. U.S.A. 101*, 12491−12496.
11. Sherman, E., and Haran, G. (2006) Coil-globule transition in the denatured state of a small protein, *Proc. Natl. Acad. Sci. U.S.A. 103*, 11539−11543.
12. Hoffman, A., Kane, A., Nettles, D., Hertzog, D. E., Baumgartel, P., Lengefeld, J., Reichardt, G., Horsley, D. A., Seckler, R., Bakajin, O., and Schuler, B. (2006) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy, *Proc. Natl. Acad. Sci. U.S.A.* (in press).
13. McCarney, E. R., Kohn, J. E., and Plaxco, K. W. (2005) Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins, *Crit. Rev. Biochem. Mol. Biol. 40*, 181−189.
14. Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M., and Kiefhaber, T. (1999) The speed limit for protein folding measured by triplet-triplet energy transfer, *Proc. Natl. Acad. Sci. U.S.A. 96*, 9597−9601.
15. Hudgins, R. R., Huang, F., Gramlich, G., and Nau, W. M. (2002) A fluorescence-based method for direct measurement of submicrosecond intramolecular contact formation in biopolymers: An exploratory study with polypeptides, *J. Am. Chem. Soc. 124*, 556−564.
16. Neuweiler, H., Schulz, A., Bohmer, M., Enderlein, J., and Sauer, M. (2003) Measurement of submicrosecond intramolecular contact formation in peptides at the single-molecule level, *J. Am. Chem. Soc. 125*, 5324−5330.
17. Lapidus, L. J., Eaton, W. A., and Hofrichter, J. (2000) Measuring the rate of intramolecular contact formation in polypeptides, *Proc. Natl. Acad. Sci. U.S.A. 97*, 7220−7225.
18. Buscaglia, M., Lapidus, L. J., Eaton, W. A., and Hofrichter, J. (2006) Effects of Denaturants on the Dynamics of Loop Formation in Polypeptides, *Biophys. J. 91*, 276−288.
19. Lapidus, L. J., Steinbach, P. J., Eaton, W. A., Szabo, A., and Hofrichter, J. (2002) Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: Testing a 1-dimensional diffusion model for peptide dynamics, *J. Phys. Chem. B 106*, 11628−11640.
20. Schuler, B., Lipman, E. A., and Eaton, W. A. (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy, *Nature 419*, 743−747.
21. Buscaglia, M., Schuler, B., Lapidus, L. J., Eaton, W. A., and Hofrichter, J. (2003) Kinetics of Intramolecular Contact Formation in a Denatured Protein, *J. Mol. Biol. 332*, 9−12.
22. Chattopadhyay, K., Elson, E. L., and Frieden, C. (2005) The kinetics of conformational fluctuations in an unfolded protein measured by fluorescence methods, *Proc. Natl. Acad. Sci. U.S.A. 102*, 2385−2389.
23. Mukhopadhyay, S., Krishnan, R., Lemke, E. A., Lindquist, S., and Deniz, A. A. (2007) A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures, *Proc. Natl. Acad. Sci. U.S.A. 104*, 2649−2654.
24. Buscaglia, M., Kubelka, J., Eaton, W. A., and Hofrichter, J. (2005) Determination of ultrafast protein folding rates from loop formation dynamics, *J. Mol. Biol. 347*, 657−664.
25. Lapidus, L. J., Eaton, W. A., and Hofrichter, J. (2002) Measuring dynamic flexibility of the coil state of a helix-forming peptide, *J. Mol. Biol. 319*, 19−25.
26. Neuweiler, H., Doose, S., and Sauer, M. (2005) A microscopic view of miniprotein folding: Enhanced folding efficiency through formation of an intermediate, *Proc. Natl. Acad. Sci. U.S.A. 102*, 16650−16655.
27. Scalley, M. L., Nauli, S., Gladwin, S. T., and Baker, D. (1999) Structural transitions in the protein L denatured state ensemble, *Biochemistry 38*, 15927−15935.
28. Kim, D. E., Fisher, C., and Baker, D. (2000) A breakdown of symmetry in the folding transition state of protein L, *J. Mol. Biol. 298*, 971−984.
29. McCallister, E. L., Alm, E., and Baker, D. (2000) Critical role of beta-hairpin formation in protein G folding, *Nat. Struct. Biol. 7*, 669−673.
30. Park, S. H., ONeil, K. T., and Roder, H. (1997) An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core, *Biochemistry 36*, 14277−14283.
31. Park, S. H., Shastry, M. C. R., and Roder, H. (1999) Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing, *Nat. Struct. Biol. 6*, 943−947.
32. Scalley, M. L., Yi, Q., Gu, H. D., McCormack, A., Yates, J. R., and Baker, D. (1997) Kinetics of folding of the IgG binding domain of peptostreptoccocal protein L, *Biochemistry 36*, 3373−3382.
33. Szabo, A., Schulten, K., and Schulten, Z. (1980) 1st Passage Time Approach to Diffusion Controlled Reactions, *J. Chem. Phys. 72*, 4350−4357.
34. Bent, D. V., and Hayon, E. (1975) Excited-State Chemistry of Aromatic Amino-Acids and Related Peptides. 3. Tryptophan, *J. Am. Chem. Soc. 97*, 2612−2619.
35. Vanderkooi, J. M. (1992) in *Topics in Fluorescence Spectroscopy: Biochemical Applications* (Lakowicz, J. R., Ed.), Plenum Press, New York.
36. Lapidus, L. J., Eaton, W. A., and Hofrichter, J. (2001) Dynamics of intramolecular contact formation in polypeptides: Distance dependence of quenching rates in a room-temperature glass, *Phys. Rev. Lett.* 87.
37. Merchant, K. A., Best, R. B., Louis, J. M., Gopich, I. V., and Eaton, W. A. (2007) Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations, *Proc. Natl. Acad. Sci. U.S.A. 104*, 1528−1533.
38. Laurence, T. A., Kong, X. X., Jager, M., and Weiss, S. (2005) Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins, *Proc. Natl. Acad. Sci. U.S.A. 102*, 17348−17353.
39. Magg, C., Kubelka, J., Holtermann, G., Haas, E., and Schmid, F. X. (2006) Specificity of the initial collapse in the folding of the cold shock protein, *J. Mol. Biol. 360*, 1067−1080.
40. Hagen, S. J., Hofrichter, J., Szabo, A., and Eaton, W. A. (1996) Diffusion-limited contact formation in unfolded cytochrome c: Estimating the maximum rate of protein folding, *Proc. Natl. Acad. Sci. U.S.A. 93*, 11615−11617.
41. Hagen, S. J., Qiu, L. L., and Pabit, S. A. (2005) Diffusional limits to the speed of protein folding: fact or friction?, *J. Phys.: Condens. Matter 17*, S1503−S1514.
42. Nettels, D., Gopich, I. V., Hoffmann, A., and Schuler, B. (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics, *Proc. Natl. Acad. Sci. U.S.A. 104*, 2655−2660.
43. Moglich, A., Joder, K., and Kiefhaber, T. (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation, *Proc. Natl. Acad. Sci. U.S.A. 103*, 12394−12399.
44. Krantz, B. A., Srivastava, A. K., Nauli, S., Baker, D., Sauer, R. T., and Sosnick, T. R. (2002) Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects, *Nat. Struct. Biol. 9*, 458−463.
45. Zbilut, J. P., Mitchell, J. C., Giuliani, A., Colosimo, A., Marwan, N., and Webber, C. L. (2004) Singular hydrophobicity patterns and net charge: a mesoscopic principle for protein aggregation/folding, *Physica A: (Amsterdam) 343*, 348−358.
46. Dyson, H. J., Wright, P. E., and Scheraga, H. A. (2006) The role of hydrophobic interactions in initiation and propagation of protein folding, *Proc. Natl. Acad. Sci. U.S.A. 103*, 13057−13061.

BI700270J