Structure, Vol. 13, 1677–1684, November, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2005.08.006

An Experimental Investigation of Conformational Fluctuations in Proteins G and L

Richard B. Tunnicliffe, Joe L. Waby, Ryan J. Williams, and Mike P. Williamson*
Department of Molecular Biology and Biotechnology
University of Sheffield
Firth Court
Western Bank
Sheffield S10 2TN
United Kingdom

Summary

The B1 domains of streptococcal proteins G and L are structurally similar, but they have different sequences and they fold differently. We have measured their NMR spectra at variable temperature using a range of concentrations of denaturant. Many residues have curved amide proton temperature dependence, indicating that they significantly populate alternative, locally unfolded conformations. The results, therefore, provide a view of the locations of low-lying, locally unfolded conformations. They indicate approximately 4–6 local minima for each protein, all within ca. 2.5 kcal/mol of the native state, implying a locally rough energy landscape. Comparison with folding data for these proteins shows that folding involves most molecules traversing a similar path, once a transition state containing a β hairpin has been formed, thereby defining a well-populated pathway down the folding funnel. The hairpin that directs the folding pathway differs for the two proteins and remains the most stable part of the folded protein.

Introduction

The NMR chemical shifts of amide protons in proteins have approximately linear dependence on temperature, the gradient of which can be used as a measure of hydrogen bonding (Baxter and Williamson, 1997). However, a significant minority of protons have curved temperature dependence, which has been shown to arise from the presence of local alternative states (Baxter et al., 1998). The curvature essentially arises from the fact that the observed chemical shift is a population-weighted average of the shift in the ground state and the shift in any excited state, which has to become significantly populated at higher temperature in order to cause observable chemical shift curvature. The curvature can be calculated by using a simple model that assumes a difference in the chemical shift and/or chemical shift gradient between the two states, and also a difference in entropy between the two states (Williamson, 2003). Horse cytochrome c was studied in detail with this method (Williamson, 2003), and it was shown that the least stable region (i.e., the region with the greatest density of alternative states) corresponds to the part of the protein previously shown to unfold first, and that the most stable region (containing no observable alternative states) corresponds to the region of the protein previously shown to be the most stable in both kinetic folding (Roder et al., 1988) and hydrogen exchange experiments (Bai et al., 1995). We further showed (Williamson, 2003) that the alternative states had chemical shifts and energetics characteristic of locally unfolded conformations. The method therefore identifies conformationally mobile regions of the protein that can access locally unfolded states under ambient conditions.

Here, we apply the method to two related proteins: the streptococcal IgG binding B1 domains from proteins G and L. These proteins have been a productive testbed for protein folding. The full-length proteins are multidomain proteins, whose function is to bind IgG via small and structurally homologous domains consisting of an α helix packed against a four-stranded β sheet (made up of two symmetrical β hairpins), which wraps around the helix in a "hot dog" fold (Gallagher et al., 1994; Gronenborn et al., 1991). Although the domains are structurally and functionally similar, they have no sequence similarity (Wikström et al., 1994), and they have been shown to fold with different folding transition states: in the case of protein G, this consists of the C-terminal hairpin, while for protein L, it is the N-terminal hairpin. It is therefore useful to measure conformational fluctuations in the two proteins, to compare fluctuations from the native state with what is known of the folding pathway. The data described here allow us to characterize the bottom half of the folding funnel, which we show to be locally rough, containing at least four local minima for each protein. Nevertheless, the region of the proteins that is most stable in the unfolded state remains the most stable in the folded state, implying that for these two proteins (which have been shown to fold via two-state mechanisms), folding is best described as a smooth or nonfrustrated folding pathway, via a transition state that consists of a β hairpin.

Results

Overexpression and purification of the His-tagged proteins was straightforward and produced a good yield of both proteins, in pure form, as judged by SDS-PAGE and NMR. NMR spectra were assigned by comparison to literature values (Gronenborn et al., 1991, Wikström et al., 1993), and they were checked by using ^{15}N-separated TOCSY (total correlation spectroscopy) and NOESY (nuclear Overhauser enhancement spectroscopy) 3D spectra. For protein L, residue 47 could not be assigned. Two-dimensional ^{15}N HSQC (heteronuclear single quantum coherence) spectra of both proteins were sharp and were almost completely resolved (Figure 1). Spectra of both proteins were acquired in 8° increments from 288 through 328 K (protein G) or from 280 through 320 K (protein L), and concentrations of guanidinium hydrochloride (GdmHCl) from 0 to 1.15 M (protein G) or from 0 to 1.5 M (protein L) were used. These conditions are well below what is needed for global denaturation for both proteins. The proteins

*Correspondence: m.williamson@sheffield.ac.uk

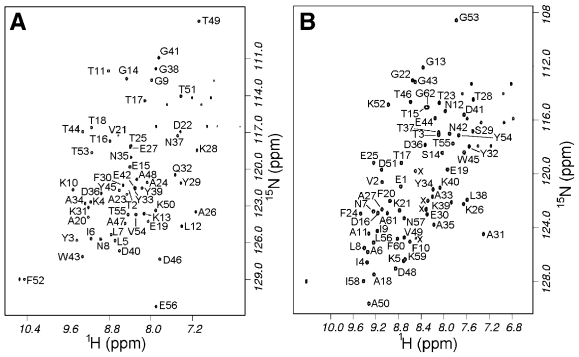

Figure 1. ^{15}N-^{1}H HSQC Spectra of Protein G and Protein L

(A and B) (A) Protein G and (B) protein L, 39ºC, zero GdmHCl. Peaks are marked by their assignments. Unmarked peaks are from side chains. Peaks marked "X" (in the protein L spectrum) come from the N-terminal His tag.

unfold thermally at over 80ºC, and at GdmHCl concentrations of 3.0 and 2.6 M (proteins G and L, respectively) (McCallister et al., 2000; Scalley et al., 1997; Yi et al., 1997). Peaks for each temperature and GdmHCl concentration were picked into a database, and they were sorted into temperature series, by residue, for each GdmHCl concentration. For visual inspection, each temperature series was fitted to a straight line (Figure 2A),

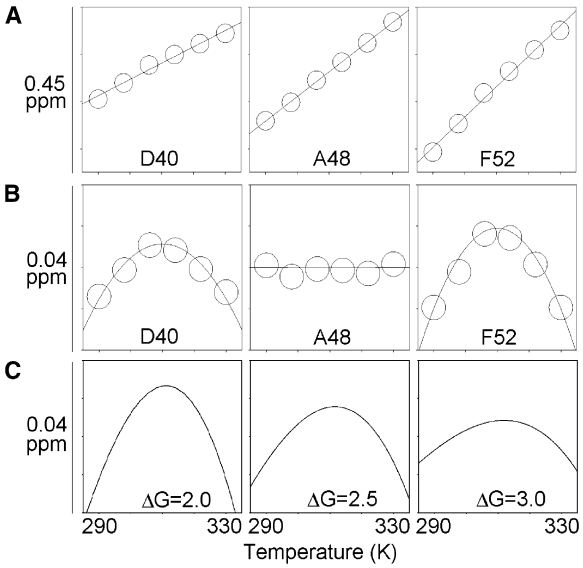

Figure 2. Chemical Shift Changes as a Function of Temperature

Data are shown for protein G, zero GdmHCl.

(A) Experimental shifts for residues D40, A48, and F52 of protein G, zero GdmHCl, as a function of temperature. The line shows the best-fit straight line. The temperature range is from 285 to 335 K, and the total chemical shift range is 0.45 ppm in each plot.

(B) Residual shifts after subtraction of the best-fit lines. The temperature range is as in (A), and the total chemical shift range is 0.04 ppm in each plot. The lines show the best-fitted parabolas. After the statistical test, A48 is not curved at $p = 0.05$; D40 is curved at $p < 0.05$, but not at $p < 0.01$; and F52 is curved at $p < 0.01$ (i.e., in Table 1, A48 has no symbol, D40 has an open circle, and F52 has a filled circle).

(C) Calculated curves, with parameters as in the text, using (left to right) $\Delta H = 7.1$, 7.6, and 8.1 kcal/mol (corresponding to $\Delta G = 2.0$, 2.5, and 3.0 kcal/mol at 298 K). Same plot ranges as in (B).

and the residuals were then plotted, to identify curved temperature series (Figure 2B). Curved data were evaluated by using a statistical test. This process was repeated for each GdmHCl concentration, to generate a table of residues that display curved temperature dependence at each concentration (Tables 1 and 2). There are a large number of residues showing curved dependence, a higher proportion than found in our earlier studies (Baxter et al., 1998; Williamson, 2003). This difference may be related to the low folding free energy of proteins G and L, and to their large heat capacities, which result in high thermal stability and relatively large increases in mobility with temperature compared to other proteins (Seewald et al., 2000). It also may possibly be the cause of the high degree of motional correlation between different parts of the protein (Mayer et al., 2003).

Residues were classified into three categories: those that do not display curvature at any concentration of GdmHCl; those that display curvature at all concentrations; and those that only display curvature at some. It is clear that residues with curved temperature dependence at all concentrations are accessing an alternative state that is populated to an extent large enough to cause perturbations to the chemical shift. The free energy difference required for curvature was checked by simulations of chemical shift temperature dependence, by using the methods previously developed (Williamson, 2003). Simulations for different free energy differences are shown in Figure 2C. These simulations indicate that for the residues identified here with curved dependence, free energy differences between the ground state and the alternative state of between 2 and 2.5 kcal/mol (8–10 kJ/mol) are required at 298 K.

Residues with no curvature present no evidence of any alternative state. This does not necessarily mean that no such state is present, but since significant changes in chemical shift can be caused by small changes in structure, lack of curvature at any concentration of GdmHCl can be taken as indicative of a lack of alternative states. Therefore, in Figure 3, we present a summary of residues with and without temperature-dependent curvature, these being residues, respectively, with an alternative state within 2.5 kcal/mol of the ground state, and residues showing no evidence of any low-lying alternative state.

Discussion

The Nature of the Alternative States

This study has identified alternative conformational substates in proteins G and L. Such states are experimentally difficult to characterize because of their low populations, but they are a continuing area of interest because they have relevance to ideas of induced fit (Nevo et al., 2003) and allostery (Freire, 2000) and are important for understanding protein function, flexibility, and folding. Our previous studies have shown that the alternative states identified from curved temperature-dependent plots represent locally unfolded forms, typically consisting of approximately five amino acid residues centered on the residue that displays curved temperature dependence. Our results (Figure 3) demonstrate that there are at least 11 such residues in both

Table 1. Residues in Protein G Displaying Curved Temperature Dependence at Different Guanidinium Hydrochloride Concentrations

Res. No.	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56
0 Gdm		●	●	●	○	○			●	●					●		○		○	○		○		○	○			●	●	○		●	●	●	●		●	●	○		○											●	●		○	●	
0.28		●	●	●	●	●	●	●	●	●			○	○	●		●		●	●	●	●		●	●		○	●	●	●	○	○	●	●	●		●	●	●		●	●			○	○			○		●	●		●	○	○
0.51		●	●	●	○	●	●	○	●	●			○		●		○		●	●	●	●		●	○		○	●	●	●		○	●	●	●		●	●	●		○	●				○					●	●	○	●		
0.72	●	●	●	●		○	●	○	●	●			●	●	●		●		●	●	○	●		●	○		○	●	●	●		○	●	●	●	○	●	○	●		●	●									●	●		○	○	
0.90		○	●	○					●	●					○		○				○	○		○				●		○			○	○	●		○	○			○	●									○	●				
1.14			●	○	○	○	○		●	●			○		●	○			●			●		●				●		○			○	●	●		○	○	○		○	●									●			●	○	

Residues drawn as ● are curved at $p < 0.01$, and residues drawn as ○ are only curved at $p < 0.05$. In Figure 3, a "curved" residue is defined as one with ● at all [GdmHCl], although up to two ○ are allowed. A "not curved" residue is blank at all [GdmHCl], although one ● or ○ is allowed.

Table 2. Residues in Protein L Displaying Curved Temperature Dependence at Different Guanidinium Hydrochloride Concentrations

Res. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62
0 Gdm	●	●	●	●		●	●	●	●	○	●		○	●	●	●	●	●	○	●	●		●	○	●		●	○	●	●	●	○	●	●	●	●		●	●	●		●	●	●	●			●		●	●	●	●	○	●	●	●	●	●	●	●	●	●	●
0.3	●	●	○	●		●	●	●	●	●	●			●		●	○	●		●	○		●	●	●		●	○	○	●		○	●	●	●		●	●	●		○	●	●	●	●			●		●	●	●	●	○	●	●	●	●	●	●	●
0.6	●	○	○	●		●	●	○	●	●	●			○		●		●		●			●	●			●		○	○			○	○	●		●	●	○			●	●	○		○		●	●	●			●	●	○	●	●	○	●	●
1.2	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	○	●	●	○	●	●	●	○	●	●	●	●	●	●	●	●	●	●	●	●	●		●	●	●	●	●	●	○	●	●	●	●	●	●	●	●	●	●	●	●	●
1.5		○					●	○			●			○		●		●					●	●			●		○				●	●			●	●	○				●	●						●			●			○	●	●	
1.8	●	●	○	○						○	●		○	○		●	○	●		○			●	●													○	○			○	●						○	●	●	●				●		○	●	○

Residues drawn as ● are curved at $p < 0.01$, and residues drawn as ○ are only curved at $p < 0.05$. In Figure 3, a "curved" residue is defined as one with ● at all [GdmHCl], although one ○ is allowed. A "not curved" residue is blank at all [GdmHCl], although one ● or ○ is allowed.

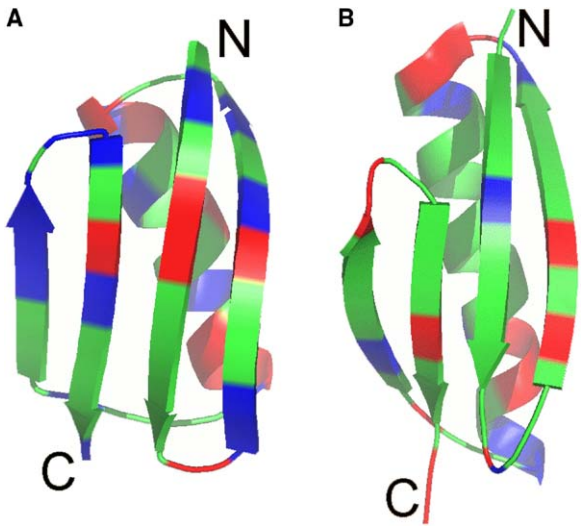

Figure 3. Structures of Protein G and Protein L

(A and B) (A) Protein G and (B) protein L, showing amino acids with consistent patterns of amide proton chemical shift curvature. Residues curved at all GdmHCl concentrations are shown in red, residues not curved at any GdmHCl concentration are shown in blue, and all other residues are shown in green. See Tables 1 and 2 for detailed interpretation.

protein G and protein L. We also note that the results displayed in Tables 1 and 2 indicate that there are many more residues that show curvature at some concentrations of GdmHCl, implying that probably many more residues than these 11 have alternative states accessible. However, these states are less clearly defined, and they are therefore presumably less populated than those observed here.

The residues with curved temperature dependence are distributed all over the proteins: in the strands of both β hairpins, in the β turns, and in the α helix. There is some spatial clustering of these residues: there are identifiable groups at each end of the helices, in the β turns, and at the C terminus of protein L. Interestingly, there is a suggestion of clustering in the β sheet regions also. In protein G (Figure 3A), there are four residues (K4, L5, T16, and F52) displaying consistent temperature-dependent curvature in the center of strands 1, 2, and 4. These residues are linked by hydrogen bonds: L5 HN hydrogen bonds to T16 O, T16 HN binds to L5 O, and F52 HN binds to K4 O. Similarly, in protein L (Figure 3B), residues Q16 and I58 in strands 2 and 4, respectively, are linked by their HN hydrogen bonding to strand 1 residues L8 and N7, respectively. Thus, the data suggest that there may be a concerted structural rearrangement involving residues in strands 1, 2, and 4 in both proteins. By clustering residues together as detailed above, we estimate 4–6 locally unfolded states in each protein. The curvatures are similar across all parts of the proteins, showing no obvious sequence or secondary structure dependence. We therefore conclude that each location is accessing alternative states independently from other locations. In other words, there is no sequential or hierarchical order of local unfolding: rather, local states are unfolding and refolding stochastically, based on local stabilities.

Relevance to Folding Energy Landscape

The states characterized here are within approximately 2.5 kcal/mol of the ground state, as indicated by simulations. They are all in fast exchange with the ground state on a chemical shift timescale, implying exchange rates of at least 1000 s^{-1}. The lowest possible energy barrier implied by such an exchange rate would correspond to the maximum possible "local folding" rate. This is determined by the preexponential factor in Kramers' equation, which has been estimated to be approximately 1 μs (Fersht, 1999; Kubelka et al., 2004), implying a maximum limiting energy barrier for refolding of approximately 4 kcal/mol. The implication of these calculations is that the energy barrier for local folding of the alternative state back to the ground state is very low, being in most cases less than 1.5 kcal/mol (i.e., 4–2.5). If the barrier were zero, there would be no population of the alternative state, and it would not be observable. The barriers for the states seen here are therefore expected to be in the range from approximately 0.5 kcal/mol (i.e., kT) to 1.5 kcal/mol. The folding funnel therefore contains at least 4–6 small, defined local minima for both proteins: it is locally rough. There is thus considerable conformational disorder in the "folded" state.

The reason for using a range of concentrations of GdmHCl is that it destabilizes the native state (and by implication all of the locally unfolded variants of the native state) relative to the unfolded state. One might therefore expect that at higher GdmHCl concentrations, other alternative states might become observable, as they become closer in energy to the native state. This was true for cytochrome c, where some residues displayed either less or more curvature with increased GdmHCl (Williamson, 2003). However, this does not occur for proteins G and L, where there is little variation in curvature with GdmHCl. It is tempting to rationalize this observation by noting that proteins G and L have been suggested to fold by a pure two-state mechanism: there is, for example, no suggestion of rollover on the hydrogen exchange chevron plots at low GdmHCl concentration (Gu et al., 1997; Kim et al., 2000; McCallister et al., 2000; Nauli et al., 2001; Scalley et al., 1997; Yi and Baker, 1996; Yi et al., 1997). By contrast, cytochrome c has a more typical behavior, with rollover at low GdmHCl concentration (Bai et al., 1995), a phenomenon that has been attributed to the presence of folding intermediates (Chamberlain et al., 1996; Oliveberg, 2001). We therefore suggest that the uniformity of curvature with GdmHCl is a further indication that there are no folding intermediates for proteins G and L.

The two proteins have very high stability against thermal denaturation: protein G unfolds at 87.5ºC under our solution conditions (Alexander et al., 1992). However, their free energies of folding are not unusually large, being 6.6 and 6.5 kcal/mol for proteins G and L, respectively, at 25ºC (Alexander et al., 1992; Orban et al., 1995; Scalley et al., 1997; Yi et al., 1997). The alternative states identified here are therefore midway in energy between the ground state and the unfolded state: they are not well described as "near the native structure" energetically, even though they are near structurally. We may therefore represent the energy landscape of both proteins as shown in Figure 4A. Other experiments (Ansari et al., 1985; Frauenfelder et al., 1991) and simulations

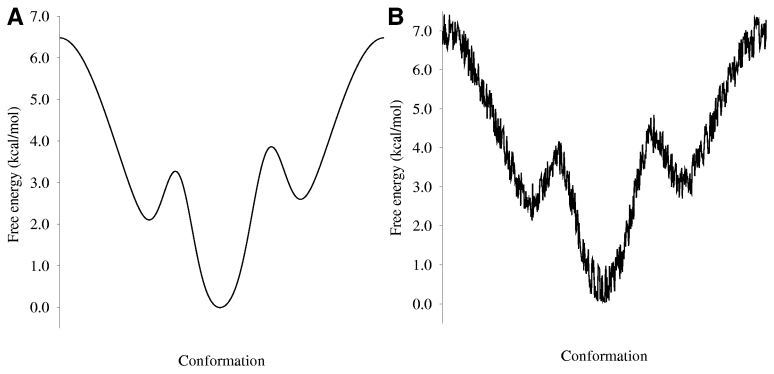

Figure 4. The Energy Landscape of Protein G

(A and B) Drawn for conditions favoring the folded form (i.e., 100 mM sodium acetate [pH 5.4]). The vertical scale is free energy (with reference to the native state). The figure depicts two alternative states, with energies and energy barriers representative of those characterized here (see text). Protein L has a very similar landscape. (A) Outline landscape as derived from the experiments described here. (B) Rougher landscape expected from other experiments and calculations (Ansari et al., 1985; Elber and Karplus, 1987; Frauenfelder et al., 1991).

(Elber and Karplus, 1987; Frauenfelder et al., 1991) imply that, actually, the landscape is also locally rough, leading to a noisier representation (Figure 4B). However, it is important to stress that the energy barriers from the alternative states back to the ground state appear to be small, implying rapid collapse back to the ground state.

Comparison of Proteins G and L

The reason for comparing proteins G and L is that the two proteins have very similar tertiary folds, but different kinetic folding pathways. In protein G, only the C-terminal β hairpin is formed in the folding transition state, whereas in protein L, only the N-terminal hairpin is formed in the folding transition state (Kim et al., 2000; McCallister et al., 2000). This asymmetry in folding is reflected in the structures of fragments of the proteins: the isolated C-terminal region of protein G forms a hairpin, whereas the N-terminal region does not (Blanco et al., 1994; Blanco and Serrano, 1995). Isolated hairpin sequences from protein L do not form folded structures (Ramirez-Alvarado et al., 1997). However, a study of the denatured state of protein L suggested the presence of a preferred residual hairpin structure in the N-terminal hairpin (Yi et al., 2000), while a glycine-to-alanine mutation in the N-terminal hairpin of protein L slowed the rate of folding almost 10-fold without affecting the unfolding rate (Gu et al., 1997), implying a role for the N-terminal hairpin in driving folding. Correspondingly, an aspartate-to-alanine mutation in the C-terminal hairpin of protein G slowed the rate of folding 20-fold without affecting the unfolding rate (McCallister et al., 2000), implicating the C-terminal hairpin in folding. Most tellingly, replacement of the residues forming the N-terminal hairpin in protein G with residues designed to produce a preferred hairpin turn generated a protein that folded 100 times faster than wild-type, with only the N-terminal hairpin being formed at the rate-limiting step in folding (Nauli et al., 2001, 2002). Theoretical modeling agrees with these experimental results, demonstrating, for example, that it is the interactions present within the N-terminal hairpin of protein L that cause it to fold first (Karanicolas and Brooks, 2002). In summary, previous experiments have shown that the reason for the differing folding pathways is likely to lie in the preferred residual structures within the unfolded ensemble. Similar considerations hold for other proteins (Chamberlain et al., 1996).

It is therefore of interest to compare the locations of alternative locally unfolded structures in the two proteins. Study of Figure 3 shows that the differing local stabilities of the two hairpin turns are retained in the folded proteins under native conditions. In protein G, the C-terminal hairpin is dominated by residues showing no consistent evidence of curvature, while there are several residues in the N-terminal hairpin with consistent curvature (particularly the two residues at the tip of the loop). In contrast, the C-terminal hairpin of protein L contains many more residues showing curvature, including four residues near the tip of the loop. We therefore conclude that the relative stabilities of the two hairpins in isolated fragments and in the denatured proteins are retained in the native state. Although this result appears intuitively obvious, it is far from inevitable, and amounts to saying that the local interactions that control the start of the folding pathway (as far as the folding transition state) remain dominant in the folded state and, by implication, all the way through the folding process (Raschke et al., 1999). This result agrees with the recent conclusion that folding transition state ensembles have native-like topologies, although they may differ in other respects (Lindorff-Larsen et al., 2004). In the language of folding energy landscapes (Dill, 1999), this amounts to saying that in the cases of these two proteins, the majority of the micropaths (i.e., the folding trajectories followed by individual folding molecules) do resemble the macropath (the average folding pathway). For these proteins, it is thus reasonable to speak of a "folding pathway," at least from the folding transition state onward. In a downhill skiing model of folding, it suggests that most molecules travel down the same piste, rather than each molecule finding its own way down the hillside (Figure 5). It is clear from our results that this also implies that the preferred folding pathway is relatively smooth and "nonfrustrated," compared to other possible pathways (Figure 5). The fact that the same region of the protein is the most stable both in the unfolded and folded states also implies that the folding funnel is approximately "vertical": the "hole" at the top is centered on the same conformational region as the energy minimum.

Comparison with Other Motional Parameters

Table 3 presents statistics on comparisons with other data that can be used to characterize fast and slow motions in proteins G and L. The ^{15}N relaxation order parameter S^2 carries information about rapid motions in

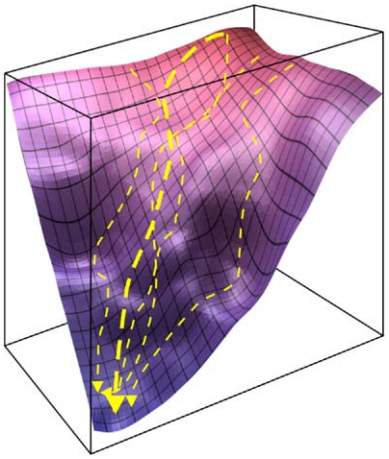

Figure 5. Part of a Three-Dimensional Model of the Energy Landscape of Protein G

Although some molecules will fold via local minima (left-most and right-most route), most will fold via smooth nonfrustrated paths, close to the average macroscopic folding pathway, shown by the thick arrow.

the ns–ps range and is very significantly negatively correlated with the log of the amide proton exchange rate ($R = -0.49$, $p < 0.0005$), implying that amide exchange is providing information on rapid local fluctuations, presumably large enough to cause transient hydrogen bond opening. These two parameters show significant correlation with the crystallographic B factor, as expected, since the B factor in these high-resolution structures is mainly indicative of local dynamic disorder, and is therefore dominated by rapid motions rather than slow conformational exchange. By contrast, the ^{15}N T_2 parameter R_{ex} carries information on slow (ms–μs) backbone motions, albeit somewhat imprecise (see legend to Table 3), and shows no correlation with these rapid-motion parameters, and little with the Φ value.

Table 3. Statistical Significance of Correlations between Parameters Previously Measured for Proteins G and L

	Φ Value	S^{2a}	$R_{ex}{}^{b}$	Log k^{c}	B Factor^{d}	Φ Value^{e}
R_{ex}	0.15	0.79	—	—	—	—
log k	0.76	0.00	0.88	—	0.01	0.22
B factor	0.83	0.24	0.70	0.00	—	0.39

The data are the statistical significance of a correlation between pairs of data sets (carried out by using SPSS). Thus, any figure less than 0.05 can be regarded as statistically correlated. Data on the lower-left side of the table relate to protein G, and those on the upper-right relate to protein L.

[a] ^{15}N relaxation order parameter (Seewald et al., 2000). Other groups have also measured this parameter, with essentially similar results (e.g., Idiyatullin et al., 2003).

[b] Chemical exchange contribution to ^{15}N T_2 relaxation. Seewald et al. (2000) report values at six different temperatures, which show little correlation in position or magnitude. The data used here are therefore the number of occurrences of non-zero R_{ex} values by residue for the six temperatures.

[c] Logarithm of the HN exchange rate, taken from Orban et al. (1995) (25ºC) and Yi and Baker (1996) for proteins G and L, respectively.

[d] Crystallographic temperature factor for the backbone nitrogen, from the crystal structures 1igd and 1hz6 for proteins G and L, respectively.

[e] The change in free energy in going from the denatured to the transition state divided by the change in free energy in going from the denatured to the folded state, resulting from a mutation at a specific amino acid (Fersht, 1999). A Φ value of 0 implies that the protein structure is unfolded at the site of the mutation in the transition state as much as it is in the denatured state, while a value of 1 implies that the structure is folded at the site of the mutation in the transition state as much as it is in the native state. Values for proteins G and L were taken from McCallister et al. (2000) and Kim et al. (2000), respectively.

Table 4. Statistical Significance for Comparison of Motional Parameters with Amide Proton Temperature-Dependent Shift Curvature in Protein G

	S^2	Log k	B Factor	R_{ex}	Φ
t test[a]	0.13	0.85	0.67	0.16	0.08
ANOVA[b]	0.20	0.92	0.83	0.22	0.09

The meanings of significance and the motional parameters are as for Table 3.

[a] Comparison of means of motional parameters for amide protons showing either no or consistent temperature-dependent curvature.

[b] Comparison of means for the three groups of amide protons showing no, some, or consistent curvature.

Table 4 shows how these measures compare to the chemical shift curvature described here. The best correlation, though still not good enough to be statistically significant, is with the Φ value. This is encouraging, because Φ is the parameter that most directly probes the unfolding pathway. There is also some correlation with R_{ex}, as might be expected because R_{ex} relates to slow timescale fluctuations, presumably similar in nature to those required for transitions to alternative states. Of particular interest is the observation that the residue in protein G that most consistently displays significant R_{ex} values is T11, in the first β hairpin, which is one of the regions showing the most amide proton temperature-dependent curvature. There is no statistical correlation with the fast-motion parameters B and log k, although, surprisingly, there is some correlation with S^2. It is however noteworthy that amide proton exchange experiments indicate that the N-terminal hairpin of protein G is less stable (Orban et al., 1995), whereas in protein L, it is marginally more stable (Yi and Baker, 1996), in agreement with the curvatures. We therefore suggest that the correlation with S^2 may indicate that regions of lower stability are also more prone to fast internal motion, and therefore that there is a coupling of fast and slow internal motions. These comparisons thus demonstrate consistency between our results and previously measured slow conformational exchange and folding intermediates, but little or none with rapid internal motions.

Finally, it is of interest to compare our results to a recent study on protein G (Ding et al., 2004), which identifies "melting hot spots" from analysis of residual dipolar couplings in a destabilized triple mutant that are suggested to consist of primarily T11, L12, and K13, followed by a second cluster of A24 and T25. These match well to two of the four locally unfolded clusters identified here for protein G. The three mutations in the destabilized protein G were L5M, L7V, and V54I, which overlaps with a third cluster identified here.

In summary, our results demonstrate several interesting features of the energy landscape for proteins G and L. However, these two proteins are relatively unusual in

that they have rapid, two-state folding. It would therefore be of interest to carry out similar studies on proteins with more complex folding pathways in order to investigate the generality of these results and to examine the cooperativity of local unfolding in different regions. Such studies are under way.

Experimental Procedures

Protein G was expressed as a His-tagged protein (with N-terminal sequence MH_6AMD preceding the normal N-terminal sequence TY…). Protein L was expressed as a His-tagged protein (containing an N-terminal H_6AME sequence preceding the normal N-terminal sequence EVTIK…). The protein L construct used has a mutation of Y45W, which has very similar structure, stability, and amide hydrogen exchange to the wild-type protein (Scalley et al., 1997). This mutant was used in most of Baker's folding studies of the protein and is therefore the most appropriate form of the protein for this study. The genes for both proteins were separately coded into pET15b vectors. Both proteins were expressed by transformation into *E. coli* BL21 (DE3) pLysS, and they were grown in M9 minimal medium containing 2 g/l $^{15}NH_4Cl$ plus 34 μg/ml chloramphenicol and 50 μg/ml carbenicillin. Expression was induced by 1.5 mM IPTG when the cells reached an OD_{600} of 0.6, and cells were incubated overnight at 25ºC. Cells were harvested by centrifugation and lysed by freeze-thaw, and the protein was purified by using a nickel column (Amersham Biosciences), followed by a gel filtration step in 50 mM potassium phosphate (pH 6.0) for protein G. All samples were dialyzed and concentrated by using Amicon membranes (Millipore) to 1 mM in 100 mM acetate-d_3 (pH 5.4), 10% D_2O. For spectra in the presence of guanidinium hydrochloride, a guanidinium stock solution was prepared at 4 M (pH 5.4) and was added to the NMR tube as appropriate. The yield of protein was approximately 50 mg/l for protein G and 25 mg/l for protein L.

NMR spectra were acquired on a Bruker DRX-500. Temperatures were calibrated by using glycerol with a calibration curve supplied by Bruker. Each set of temperatures for the same sample was acquired in the same session, to produce the most consistent data. Spectra were processed by using Felix (Accelrys Inc., San Diego, CA), and data were analyzed by using home-written scripts running under Unix. Trimethyl silyl propionate was used as a chemical shift reference (0.0 ppm). Its chemical shift is independent of temperature (Wishart et al., 1995). As a measure of curvature, the experimental data (or, equivalently, the residuals from the linear fitting) were fitted to a parabola, by using a Marquardt nonlinear least-squares fitting based originally on a Numerical Recipes algorithm. All protons for which the second-order coefficient (c in the equation $y = a + bx + cx^2$) differed from zero at $p < 0.05$ (using a ^1H chemical shift error of 0.005 ppm) were treated as having curved temperature dependence. Simulations of chemical shift curvature were carried out as described previously (Williamson, 2003), by using the parameters $\Delta S = 17.2$ cal/mol K^{-1} ($T\Delta S = 5.1$ kcal/mol at 25ºC), $\delta_1 = 8.5$ ppm, $\delta_2 = 8.0$ ppm, $g_1 = -2.0$ ppb/K, and $g_2 = -7.0$ ppb/K.

Acknowledgments

We thank David Baker (University of Washington, Seattle) for providing plasmids for proteins G and L, Jeremy Craven (Sheffield) for extensive advice on software, and the Wellcome Trust and Biotechnology and Biological Sciences Research Council for grants for the purchase of equipment.

Received: June 13, 2005
Revised: July 28, 2005
Accepted: August 2, 2005
Published: November 8, 2005

References

Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (1992). Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2: why small proteins tend to have high denaturation temperatures. Biochemistry *31*, 3597–3603.

Ansari, A., Berendzen, J., Bowne, S.F., Frauenfelder, H., Iben, I.E.T., Sauke, T.B., Shyamsunder, E., and Young, R.D. (1985). Protein states and protein quakes. Proc. Natl. Acad. Sci. USA *82*, 5000–5004.

Bai, Y.W., Sosnick, T.R., Mayne, L., and Englander, S.W. (1995). Protein folding intermediates: native-state hydrogen exchange. Science *269*, 192–197.

Baxter, N.J., and Williamson, M.P. (1997). Temperature dependence of ^1H chemical shifts in proteins. J. Biomol. NMR *9*, 359–369.

Baxter, N.J., Hosszu, L.L.P., Waltho, J.P., and Williamson, M.P. (1998). Characterisation of low free-energy excited states of folded proteins. J. Mol. Biol. *284*, 1625–1639.

Blanco, F.J., and Serrano, L. (1995). Folding of protein G B1 domain studied by the conformational characterization of fragments comprising its secondary structure elements. Eur. J. Biochem. *230*, 634–649.

Blanco, F.J., Rivas, G., and Serrano, L. (1994). A short linear peptide that folds into a native stable β-hairpin in aqueous solution. Nat. Struct. Biol. *1*, 584–590.

Chamberlain, A.K., Handel, T.M., and Marqusee, S. (1996). Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. Nat. Struct. Biol. *3*, 782–787.

Dill, K.A. (1999). Polymer principles and protein folding. Protein Sci. *8*, 1166–1180.

Ding, K., Louis, J.M., and Gronenborn, A.M. (2004). Insights into conformation and dynamics of protein GB1 during folding and unfolding by NMR. J. Mol. Biol. *335*, 1299–1307.

Elber, R., and Karplus, M. (1987). Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. Science *235*, 318–321.

Fersht, A.R. (1999). Structure and Mechanism in Protein Science (New York: W.H. Freeman).

Frauenfelder, H., Sligar, S.G., and Wolynes, P.G. (1991). The energy landscapes and motions of proteins. Science *254*, 1598–1603.

Freire, E. (2000). Can allosteric regulation be predicted from structure? Proc. Natl. Acad. Sci. USA *97*, 11680–11682.

Gallagher, T., Alexander, P., Bryan, P., and Gilliland, G.L. (1994). Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. Biochemistry *33*, 4721–4729.

Gronenborn, A.M., Filpula, D.R., Essig, N.Z., Achari, A., Whitlow, M., Wingfield, P.T., and Clore, G.M. (1991). A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. Science *253*, 657–661.

Gu, H.D., Kim, D., and Baker, D. (1997). Contrasting roles for symmetrically disposed β-turns in the folding of a small protein. J. Mol. Biol. *274*, 588–596.

Idiyatullin, D., Nesmelova, I., Daragon, V.A., and Mayo, K.H. (2003). Comparison of $^{13}C_\alpha H$ and ^{15}NH backbone dynamics in protein GB1. Protein Sci. *12*, 914–922.

Karanicolas, J., and Brooks, C.L. (2002). The origins of asymmetry in the folding transition states of protein L and protein G. Protein Sci. *11*, 2351–2361.

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.

Kubelka, J., Hofrichter, J., and Eaton, W.A. (2004). The protein folding 'speed limit'. Curr. Opin. Struct. Biol. *14*, 76–88.

Lindorff-Larsen, K., Vendruscolo, M., Paci, E., and Dobson, C.M. (2004). Transition states for protein folding have native topologies despite high structural variability. Nat. Struct. Mol. Biol. *11*, 443–449.

Mayer, K.L., Earley, M.R., Gupta, S., Pichumani, K., Regan, L., and Stone, M.J. (2003). Covariation of backbone motion throughout a small protein domain. Nat. Struct. Biol. *10*, 962–965.

McCallister, E.L., Alm, E., and Baker, D. (2000). Critical role of β-hairpin formation in protein G folding. Nat. Struct. Biol. *7*, 669–673.

Nauli, S., Kuhlman, B., and Baker, D. (2001). Computer-based redesign of a protein folding pathway. Nat. Struct. Biol. *8*, 602–605.

Nauli, S., Kuhlman, B., Le Trong, I., Stenkamp, R.E., Teller, D., and Baker, D. (2002). Crystal structures and increased stabilization of the protein G variants with switched folding pathways NuG1 and NuG2. Protein Sci. *11*, 2924–2931.

Nevo, R., Stroh, C., Kienberger, F., Kaftan, D., Brumfeld, V., Elbaum, M., Reich, Z., and Hinterdorfer, P. (2003). A molecular switch between alternative conformational states in the complex of Ran and importin β1. Nat. Struct. Biol. *10*, 553–557.

Oliveberg, M. (2001). Characterisation of the transition states for protein folding: towards a new level of mechanistic detail in protein engineering analysis. Curr. Opin. Struct. Biol. *11*, 94–100.

Orban, J., Alexander, P., Bryan, P., and Khare, D. (1995). Assessment of stability differences in the protein G B1 and B2 domains from hydrogen-deuterium exchange: comparison with calorimetric data. Biochemistry *34*, 15291–15300.

Ramirez-Alvarado, M., Serrano, L., and 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.

Raschke, T.M., Kho, J., and Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: a protein engineering study. Nat. Struct. Biol. *6*, 825–831.

Roder, H., Elöve, G.A., and Englander, S.W. (1988). Structural characterization of folding intermediates in cytochrome *c* by H-exchange labeling and proton NMR. Nature *335*, 700–704.

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.

Seewald, M.J., Pichumani, K., Stowell, C., Tibbals, B.V., Regan, L., and Stone, M.J. (2000). The role of backbone conformational heat capacity in protein stability: temperature dependent dynamics of the B1 domain of streptococcal protein G. Protein Sci. *9*, 1177–1193.

Wikström, M., Sjöbring, U., Kastern, W., Björck, L., Drakenberg, T., and Forsén, 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.

Wikström, M., Drakenberg, T., Forsén, S., Sjöbring, U., and Björck, L. (1994). Three-dimensional solution structure of an immunoglobulin light chain-binding domain of protein L. Comparison with the IgG-binding domains of protein G. Biochemistry *33*, 14011–14017.

Williamson, M.P. (2003). Many residues in cytochrome c populate alternative states under equilibrium conditions. Proteins *53*, 731–739.

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., and Sykes, B.D. (1995). ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. J. Biomol. NMR *6*, 135–140.

Yi, Q., and Baker, D. (1996). Direct evidence for a two-state protein unfolding transition from hydrogen-deuterium exchange, mass spectrometry, and NMR. Protein Sci. *5*, 1060–1066.

Yi, Q., Scalley, M.L., Simons, K.T., Gladwin, S.T., and Baker, D. (1997). Characterization of the free energy spectrum of peptostreptococcal protein L. Fold. Des. *2*, 271–280.

Yi, Q., Scalley-Kim, M.L., Alm, E.J., and Baker, D. (2000). NMR characterization of residual structure in the denatured state of protein L. J. Mol. Biol. *299*, 1341–1351.