Structure
Article

Mutational Tipping Points for Switching Protein Folds and Functions

Yanan He,[1] Yihong Chen,[1] Patrick A. Alexander,[1] Philip N. Bryan,[1,2] and John Orban[1,3,*]
[1]Institute for Bioscience and Biotechnology Research
[2]Department of Bioengineering
[3]Department of Chemistry and Biochemistry
University of Maryland College Park, 9600 Gudelsky Drive, Rockville, MD 20850, USA
*Correspondence: jorban@umd.edu
DOI 10.1016/j.str.2011.11.018

SUMMARY

While disordered to ordered rearrangements are relatively common, the ability of proteins to switch from one ordered fold to a completely different fold is generally regarded as rare, and few fold switches have been characterized. Here, in a designed system, we examine the mutational requirements for transitioning between folds and functions. We show that switching between monomeric 3α and 4β+α folds can occur in multiple ways with successive single amino acid changes at diverse residue positions, raising the likelihood that such transitions occur in the evolution of new folds. Even mutations on the periphery of the core can tip the balance between alternatively folded states. Ligand-binding studies illustrate that a new immunoglobulin G-binding function can be gained well before the relevant 4β+α fold is appreciably populated in the unbound protein. The results provide new insights into the evolution of fold and function.

INTRODUCTION

Some proteins can adopt more than one folded state and have been termed "metamorphic" (Murzin, 2008). Prions are a classic example of the malleability of polypeptide chains, where conformational change from a benign, predominantly α-helical form to an infectious, β-strand-rich state is driven by multimerization (Weissmann, 2005). However, there are a small but growing number of other naturally occurring examples such as lymphotactin (Tuinstra et al., 2008), Mad2 (Luo and Yu, 2008; Mapelli and Musacchio, 2007), and CLIC1 (Littler et al., 2004), suggesting that the phenomenon of fold switching may be more general. In these proteins, the equilibrium is shifted from one fold topology to another by changes in environmental factors such as salt conditions, the presence of a ligand, and redox state. Other studies, such as those on the Cro family of repressors (Roessler et al., 2008) and RfaH (Belogurov et al., 2009), support the idea that some folds may have resulted from switching an existing structure rather than evolving independently. Common features of such switchable folds are (1) flexible regions and diminished stability to allow large-scale changes, (2) a fair degree of mutual exclusivity in the core regions, and (3) the generation of a new binding surface that stabilizes the alternative fold and expands function (Bryan and Orban, 2010). Furthermore, theoretical studies predict that the sequences encoding certain protein folds are switchable to numerous other folds and that protein fold space may be more interconnected than previously considered (Meyerguz et al., 2007).

In addition to natural examples of fold switches, protein design has been used to investigate the question of how high the amino acid sequence identity of two proteins can be while maintaining different fold topologies (Ambroggio and Kuhlman, 2006; Rose and Creamer, 1994). Some of the earlier studies in this area showed that sequence identities of 50% or more could be achieved but that aggregation became a problem at higher identities, making biophysical characterization difficult (Blanco et al., 1999; Dalal and Regan, 2000). More recently, a binary system was designed where different fold topologies were obtained with very high (>85%) sequence identities (Alexander et al., 2007, 2009; He et al., 2008). The starting points were two small 56-amino-acid domains, termed G_A and G_B, from the multidomain *Streptococcus* cell surface protein G (Fahnestock et al., 1986). The G_A domain adopts a 3-α helix bundle (3α) structure (Johansson et al., 1997) and binds human serum albumin (HSA) (Falkenberg et al., 1992), whereas the G_B domain has a 4β+α fold (Gronenborn et al., 1991) and binds immunoglobulin G (IgG) (Myhre and Kronvall, 1977). The albumin- and IgG-binding epitopes (Sauer-Eriksson et al., 1995; Lejon et al., 2004) were engineered into both domains, creating latent binding sites that could be exposed on fold switching. The G_A and G_B sequences were then co-evolved with site-directed mutagenesis and phage display, increasing identity at mutation-tolerant sites using binary sequence space (i.e., only G_A or G_B amino acids) (Alexander et al., 2007). Nuclear magnetic resonance (NMR) structures were determined for 88% (He et al., 2008) and 95% (Alexander et al., 2009) identity protein pairs while still maintaining different folds and functions.

The ability of proteins to switch folds is generally regarded as rare and relatively few have been characterized structurally. One possible reason why there are not more reports is that the sequences of many such proteins may be inherently transient and rapidly evolve to their new functions and folds. The designed G_A/G_B system therefore provides an opportunity to examine the mutational requirements for transitioning between folds and functions. Here, we describe the three-dimensional (3D)

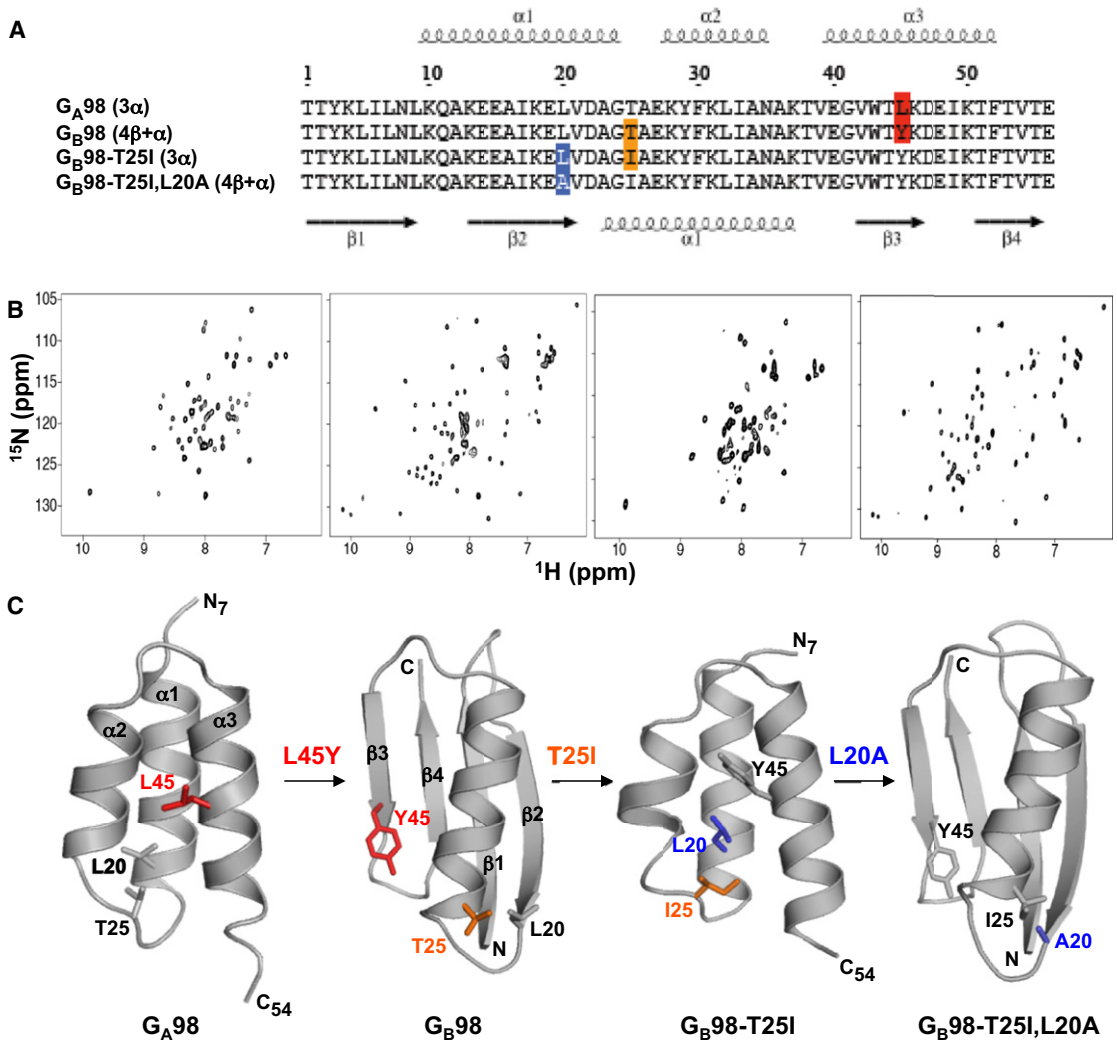

Figure 1. Single Amino Acid Mutations Leading to Fold Switching

(A) Alignment of amino acid sequences for the four proteins in this study, highlighting the positions at which changes lead to switching between 3α and 4β+α folds.

(B) 2D ^{15}N-HSQC spectra for G_A98 (left), G_B98 (center left), G_B98-T25I (center right), and G_B98-T25I,L20A (right). Viewing NMR spectra from left to right, large differences are observed from one spectrum to the next as the three successive single site mutations, L45Y, T25I, and L20A, are made (see also Figure S1).

(C) Representative structures from the NMR ensembles of G_A98, G_B98, G_B98-T25I, and G_B98-T25I,L20A. Residues mutated are highlighted.

structures for a series of high sequence identity G_A/G_B mutants, each being 98% identical to the next in the series. We show that the folds of these proteins switch between 3α and 4β+α with successive single amino acid changes at diverse residue positions and that there is a near-complete (≥95%) shift in the equilibrium between the two states. Thus, the pathway for fold switching is not unique, raising the probability of such events occurring. Moreover, ligand-binding studies on these high sequence identity mutants illustrate that changes in fold and function are not perfectly correlated. In our designed system, fold switching can be abrupt, occurring with a single amino acid mutation. However, the characteristics of a new binding function can be displayed well before the corresponding fold is appreciably populated in the unbound proteins. Overall, the results presented here provide insights into how different folds can be closely connected in sequence space and how new functions can evolve.

RESULTS

Description of Fold Topologies

The amino acid sequences of the four proteins described in this study, G_A98, G_B98, G_B98-T25I, and G_B98-T25I,L20A, are shown in Figure 1A, highlighting the single residue positions of nonidentity. Thus, G_A98 is converted to G_B98 by mutating L45 to Y45, a T25I mutation changes G_B98 to G_B98-T25I, and G_B98-T25I,L20A is generated through an amino acid change at position 20 of G_B98-T25I. Although the sequences of these proteins are nearly identical, the single amino acid mutation from one protein to the next leads to significant differences in NMR spectra, reflecting

Table 1. Summary of Structure Statistics

Variable	G_A98	G_B98	G_B98-T25I[a]	G_B98-T25I,L20A
Experimental restraints				
NOE restraints				
All NOEs	816	648		1,067
Intraresidue	507	395		627
Sequential (|i-j| = 1)	165	119		214
Medium range (1 < |i-j| ≤ 5)	78	43		57
Long range (|i-j| > 5)	66	91		169
Hydrogen bond restraints	50	62		62
Dihedral angle restraints	72	64		64
Total NOE restraints	938	774		1,193
CS-Rosetta input[a]				
$^{13}C^{\alpha}$ shifts			39	
$^{13}C^{\beta}$ shifts			18	
$^{13}C'$ shifts			43	
^{15}N shifts			47	
$^{1}H^{N}$ shifts			47	
$^{1}H^{\alpha}$ shifts			42	
RMSDs to the mean structure (Å)				
Over all residues[b]				
Backbone atoms	0.35 ± 0.10	0.81 ± 0.19	1.00 ± 0.27	0.61 ± 0.13
Heavy atoms	1.15 ± 0.16	1.66 ± 0.23	1.71 ± 0.35	1.44 ± 0.23
Secondary structures[c]				
Backbone atoms	0.31 ± 0.09	0.55 ± 0.14	0.96 ± 0.26	0.32 ± 0.07
Heavy atoms	1.12 ± 0.15	1.24 ± 0.16	1.72 ± 0.35	0.94 ± 0.13
Measures of structure quality				
Ramachandran distribution				
Most favored (%)	85.1 ± 3.9	75.6 ± 3.7	95.2 ± 2.1	77.7 ± 3.1
Additionally allowed (%)	12.5 ± 3.7	21.6 ± 4.5	4.8 ± 2.1	20.3 ± 3.7
Generously allowed (%)	0.6 ± 0.9	1.9 ± 1.4	0.0 ± 0.0	1.4 ± 1.4
Disallowed (%)	1.8 ± 1.3	0.9 ± 1.0	0.0 ± 0.0	0.6 ± 1.1
Bad contacts/100 residues	3.3 ± 1.1	1.0 ± 1.1	0.5 ± 0.7	1.5 ± 1.2
Overall dihedral G factor	0.08 ± 0.03	−0.04 ± 0.03	0.41 ± 0.03	0.01 ± 0.02

[a] CS-Rosetta model based on assigned chemical shifts.
[b] Residues 1–56 for G_B98 and G_B98-T25I,L20A. Residues 9–51 for G_A98 and G_B98-T25I.
[c] The secondary structure elements used were as follows: G_B98 and G_B98-T25I,L20A, residues 1–8, 13–20, 23–36, 42–46, and 51–55; G_A98, residues 9–23, 27–34, and 39–51; and G_B98-T25I, residues 9–23, 28–34, and 39–52.

large-scale alterations in 3D structure. Spectra in the first and third panels of Figure 1B have cross-peak patterns typical of other 3α folded proteins in this series, whereas the spectra in the second and fourth panels are characteristic of a 4β+α fold topology. To more quantitatively describe the effect of these mutations, we obtained NMR assignments for these proteins using heteronuclear triple resonance NMR spectroscopy and 3D structures were determined. All four proteins are monomeric under NMR conditions from size exclusion chromatography and light-scattering measurements. G_A98, G_B98, G_B98-T25I, and G_B98-T25I,L20A have a melting temperature (T_m) of 37°C, 35°C, 36°C, and 46°C, respectively. Because of the relatively low stabilities, NMR spectra were recorded at 5°C. Figure 1C shows representative structures from each NMR ensemble, highlighting the amino acid differences between proteins. Complete structure statistics are summarized in Table 1, and the NMR ensembles for all four proteins are shown in Figure 2. The Protein Data Bank/BioMagResBank accession codes for G_A98, G_B98, G_B98-T25I, and G_B98-T25I,L20A are 2LHC/17839, 2LHD/17840, 2LHG/17843, and 2LHE/17841, respectively.

G_A98 (3α)

The G_A98 fold has a 3α-helical bundle topology similar to previously determined 3D structures of the original parental G_A sequence (He et al., 2006), G_A88 (He et al., 2008), and G_A95 (Alexander et al., 2009). Therefore, as the G_A sequences are co-evolved with G_B sequences to high identity levels, the 3α structures of the designed proteins do not veer significantly from the wild-type fold topology. A comparison of the pairwise backbone root-mean-square deviations (RMSDs) is shown in

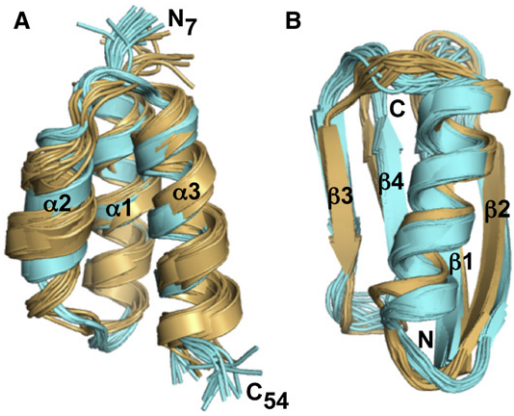

Figure 2. NMR Structures of the Designed Proteins G_A98, G_B98, G_B98-T25I, and G_B98-T25I,L20A

(A) Comparison of the 3α structures for G_A98 (blue) and G_B98-T25I (gold). The NMR ensemble for G_A98 consists of 20 final structures. The G_B98-T25I structure was determined with CS-Rosetta and main-chain chemical shift assignments and is represented by an ensemble of 10 final structures (see also Table S1).

(B) Comparison of the 4β+α structures for G_B98 (blue) and G_B98-T25I,L20A (gold). Both NMR ensembles are of 20 final structures (see also Table S2).

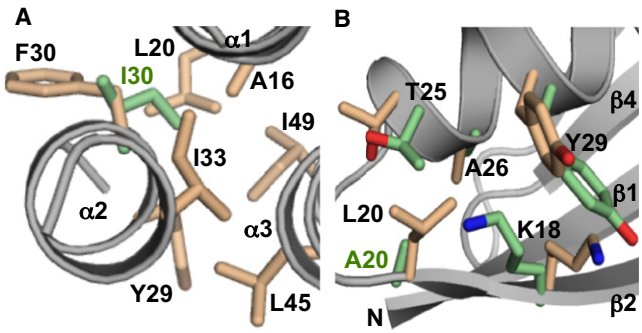

Figure 3. Structural Changes in Going from 95% to 98% Sequence Identity

(A) G_A95 to G_A98. A representative structure from the NMR ensemble for G_A98, highlighting some of the hydrophobic core residues (pale orange) and the position of F30 relative to the core. By comparison, the position of I30 in G_A95 is also shown (green). The main chain is in gray.

(B) G_B95 to G_B98. A representative NMR structure of G_B98 highlighting the A20L mutation site region in going from G_B95 to G_B98. G_B98 side-chain positions (pale orange) are compared with the corresponding G_B95 conformations (green).

Table S1 (available online). The N terminus from residues 1–8 and the C terminus from residues 52–56 are disordered with helices from residues 9–23, residues 27–34, and residues 39–51. Overall, the hydrophobic core interactions in G_A98 are similar to those in G_A95 and G_A88. Core residues in G_A98 are as follows: A12, A16, and A20 (α1-helix); I33 and A36 (α2-helix); and V42, K46, and I49 (α3-helix). The only difference in amino acid sequence in going from G_A95 to G_A98 is mutation of I30 to F30. Where I30 contributes to the core of G_A95 and is ~17% solvent accessible, F30 has limited interactions with the G_A98 core through its β-methylene group and has a solvent-exposed aromatic ring (Figure 3A). The I30F mutation is therefore expected to be destabilizing based on the G_A98 structure. Indeed, the structural data is consistent with stability measurements, which show that an I30F mutation destabilizes the 3α fold by ~1.5 kcal/mol (Alexander et al., 2009).

G_B98 (4β+α)

A single amino acid change in G_A98, L45Y, generates the G_B98 sequence, and the fold of this polypeptide chain is 4β+α rather than 3α. The differences between the G_A98 and G_B98 structures are striking. Residues that were at the unstructured ends of the G_A98 sequence form the central β strands, β1 and β4, of the G_B98 fold. Residues in the α1- and α3-helices of G_A98 correspond with the loop-β2-loop and loop-β3-loop-early β4 regions of G_B98, respectively. Comparison of the structure of G_B98 with previous structures of the parent G_B sequence (Gallagher et al., 1994), G_B88 (He et al., 2008), and G_B95 (Alexander et al., 2009), indicates very similar overall 4β+α fold topologies (Table S2). A single amino acid change from an alanine at position 20 to a leucine increases the identity from G_B95 to G_B98. The sterically more demanding leucine at position 20 is accommodated in the 4β+α fold with slight adjustments of the backbone and neighboring side-chain conformations (Figure 3B). Nevertheless, the incorporation of the branched side chain at L20 does lead to increased steric clashes with adjacent residues, consistent with a decrease in T_m of G_B98 by 12°C relative to G_B95 (Alexander et al., 2009).

G_B98-T25I (3α)

For G_B98-T25I, de novo structure determination with extensive Nuclear Overhauser Effect (NOE) restraints was not possible because of low sample solubility (<100 μM), which prevented complete assignments particularly of side-chain resonances. However, assignment of a significant proportion of main-chain ^{15}N, $^{1}H^{N}$, $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, and $^{13}C'$ resonances (Table 1), as well as some $^{13}C^{\beta}$ signals, showed three distinct helical regions by consensus chemical shift index analysis (Wishart and Sykes, 1994). Comparison of backbone $^{1}H^{N}$ and ^{15}N shifts showed a much closer match with those of G_A98 rather than G_B98 (Figure S1), indicating that fold switching from a 4β+α to 3α conformation had occurred with the T25I mutation. Furthermore, a 3D structure was calculated from these experimental chemical shifts using the CS-Rosetta algorithm (Shen et al., 2008). We previously used CS-Rosetta to determine structures for G_A88, G_B88, G_A95, and G_B95 and found that these structures compared well (backbone RMSDs ~1.0–1.5 Å) with the structures calculated from mostly NOE restraint data (Shen et al., 2010). The CS-Rosetta structure of G_B98-T25I shows a 3α fold with α-helices at residues 9–23, 28–34, and 39–52 and disordered regions at the N and C termini similar to G_A98 (Figures 1C and 2A). The main differences from G_A98, therefore, are a slightly shorter α2-helix and an α3-helix that is longer by one residue.

At low contouring of the G_B98-T25I ^{15}N HSQC spectrum, a minor species (~5%) was also detected that has cross-peaks consistent with the alternative 4β+α fold. The minor component was detected reproducibly in different sample preparations that were purified on fresh columns, free of possible contaminants. The NMR peaks of the minor species could not be assigned

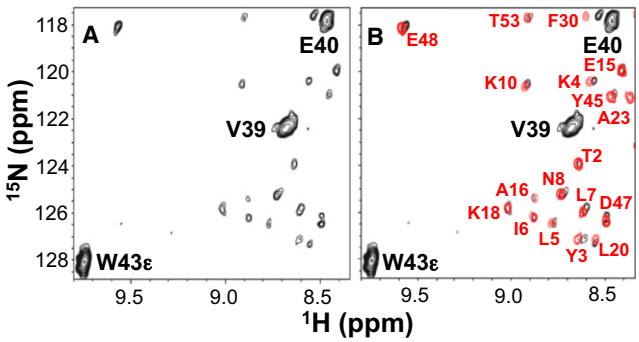

Figure 4. ^{15}N HSQC Spectra of G_B98-T25I at Low Contour Level
(A) Peaks due to the major 3α state of G_B98-T25I are labeled in black. Other unlabeled low-intensity peaks indicate the presence of a minor species.
(B) Overlay with the ^{15}N HSQC spectrum of G_B98 (red).

directly because of the low intensity levels of these signals. However, the pattern of ^{15}N HSQC peaks observed is very similar (but not identical) to that of G_B98 (Figure 4). Indeed the differences from the G_B98 spectrum are consistent with the incorporation of the T25I mutation. Two main lines of evidence indicate that the minor 4β+α species is in equilibrium with the major 3α species. Notably, changing the buffer from 100 mM KPi, pH 7.0, to 50 mM Tris, pH 8.0, leads to only the major 3α conformer within NMR detection limits - this process is reversible. Thus the minor 4β+α state cannot be an impurity or result from a covalent modification. Other evidence of equilibrium between the major and minor components comes from ligand-binding experiments (described later). The data therefore indicate that both the 3α and 4β+α folds are detectably populated for the G_B98-T25I sequence and that these two species are in slow exchange on the NMR time scale. This is similar to the case of lymphotactin where two different folds with identical amino acid sequences are in equilibrium and can be detected simultaneously in HSQC spectra (Tuinstra et al., 2008). Attempts to observe crosspeaks due to exchange between the two G_B98-T25I conformations either in NOE spectroscopy (NOESY) or zz-exchange experiments were unsuccessful, presumably because of the low relative population of the minor 4β+α state as well as the low solubility of the protein.

G_B98-T251/L20A (4β+α)

We determined the structure of G_B98-T25I,L20A, the final mutant in this series, using an extensive set of NOE and torsion angle restraints. The NMR structures show that mutation of a single amino acid, L20A, in the G_B98-T25I sequence leads to switching from a predominantly (~95%) 3α fold in G_B98-T25I to a 4β+α fold in G_B98-T25I,L20A. The G_B98-T25I,L20A structure is similar to other 4β+α structures in this series (Figure 2B; Table S2).

Ligand Binding

NMR spectroscopy was used to measure IgG and HSA binding to G_A and G_B high-identity proteins, and the results are summarized in Table 2, Figure 5, and Figure S2. The dissociation constant (K_D) obtained by NMR for binding between G_A88 and HSA was also compared with the K_D from isothermal titration calorimetry and found to be in good agreement (Figure S3).

Table 2. Ligand-Binding Data

Mutant	G_A88	G_A95	G_A98	G_B98	G_B98-T25I	G_B98-T25I,L20A
Fold	3α	3α	3α	4β+α	3α (95%) 4β+α (5%)	4β+α
K_D (HSA)	37 (31)[a]	30	244	n.b.	n.b.	n.b.
K_D (IgG)	n.b.	n.b.	62	<1[b]	15	<1[b]

Dissociation constants are in micromolar units. The experimental error is estimated to be ± 30%. All binding measurements were done at 15°C except those for G_B98-T25I, which were done at 5°C. n.b., no binding detected.
[a] Value in parentheses is from isothermal titration calorimetry (see also Figure S3).
[b] All G_B protein was bound even at the lowest IgG concentrations used.

The 4β+α folded proteins, G_B98 and G_B98-T25I,L20A, bind to the wild-type ligand, IgG, with high affinity (K_D < 1 μM). The other designed G_B proteins with 4β+α folds, ranging from G_B77 to G_B95, also bind to IgG tightly. None of the G_B proteins, including G_B98 and G_B98-T25I,L20A, show any detectable binding to the G_A ligand, HSA.

In contrast, the G_A proteins do not bind their wild-type ligand, HSA, as robustly, with binding affinity to HSA decreasing as identity to G_B is increased. Thus, G_A88 binds HSA with a dissociation constant of 31–37 μM, whereas G_A98 has a considerably weaker affinity with a K_D of 244 μM. In addition to decreased affinity for HSA, the G_A proteins acquire affinity for the G_B ligand, IgG, when the sequence identity is increased to 98%. Therefore, G_A98 is bifunctional with weak binding affinity to its cognate ligand HSA and approximately fourfold tighter binding to IgG (K_D 62 μM).

The G_B98-T25I mutant is similar to G_A98 in that it also binds IgG, despite adopting a predominantly 3α fold in the unbound state. However, binding to IgG is approximately fourfold tighter than seen for G_A98, and there is no detectable binding to HSA. The tighter binding to IgG corresponds with a higher level of the alternative 4β+α state in G_B98-T25I than in G_A98. In G_B98-T25I, the minor 4β+α state is detectable in the HSQC spectrum (Figure 4) at a level of ~5%, whereas the 4β+α conformer cannot be observed in the HSQC spectrum of G_A98. Based on the relative IgG-binding constants for G_A98 (K_D ~62 μM) and G_B98-T25I (K_D ~15 μM), the 4β+α state is therefore estimated to be at a level of approximately 1% in G_A98. Thus, the binding results support the assumption that IgG binding only occurs through the 4β+α fold, where the engineered IgG-binding epitope is revealed. Ligand-induced fold switching would then presumably occur through IgG binding to the weakly populated 4β+α state, driving the equilibrium from the 3α to the 4β+α conformer. Further experiments to test this assumption are in progress.

DISCUSSION

Structural Basis for Fold Switching

The three mutation sites described here that are responsible for fold switching are not localized in one specific part of the 3α or 4β+α structures but rather are spread over several different structural elements. In the 3α fold, the switching loci correspond to amino acid changes in the α3-helix (L45Y), the α1-α2 loop

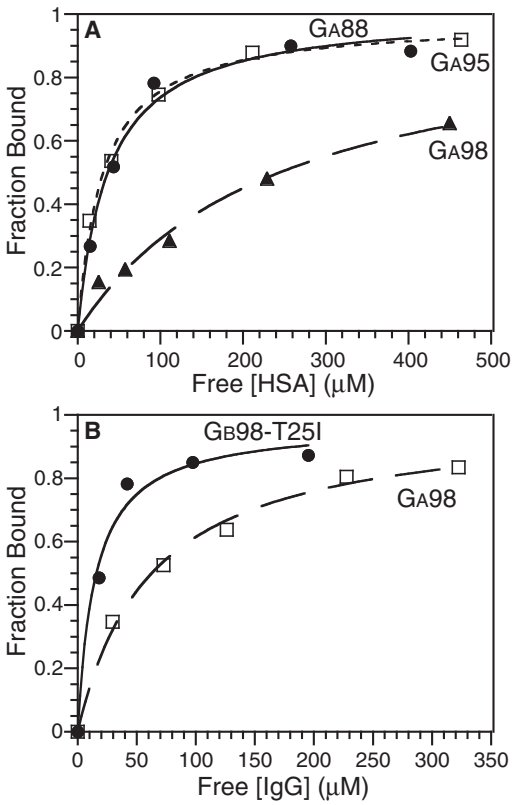

Figure 5. Ligand-Binding Curves

(A) HSA-binding curves for G_A88 (filled circles), G_A95 (open squares), and G_A98 (filled triangles).

(B) IgG-binding curves for G_B98-T25I (filled circles) and G_A98 (open squares). (See also Figure S2.)

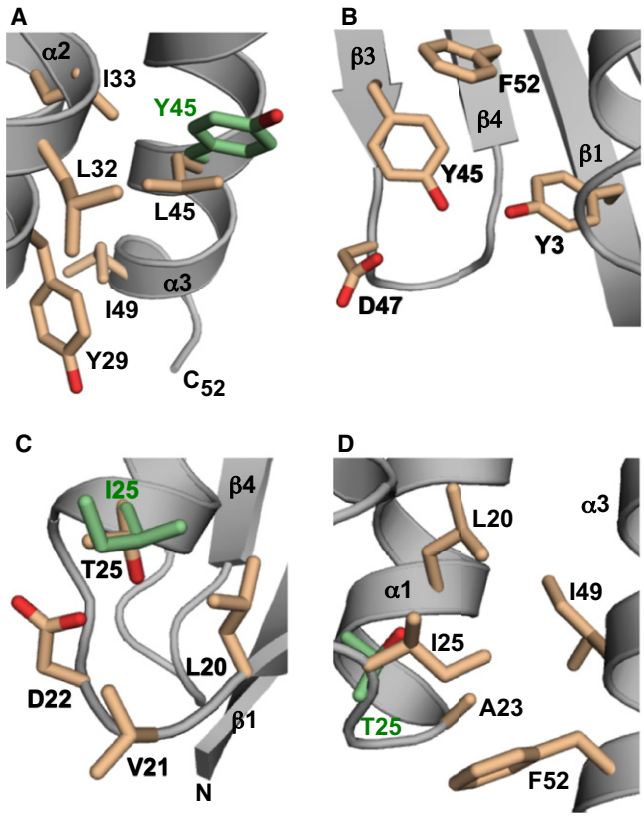

Figure 6. Structural Basis for Switching between 3α and 4β+α Folds

(A) A representative NMR structure of G_A98 showing L45 and nearest neighbors (pale orange) described in the text. The side chain conformation of Y45 (green) in the G_B98-T25I CS-Rosetta structure is superimposed for comparison purposes.

(B) G_B98 NMR structure highlighting Y45 and adjacent amino acids.

(C) NMR structure of G_B98 showing T25 and surrounding residues (pale orange). The I25 side chain (green) in the NMR structure of G_B98-T25I,L20A is superimposed for comparison.

(D) CS-Rosetta structure of G_B98-T25I highlighting I25 and neighboring hydrophobic contacts (pale orange). The corresponding position of T25 in G_A98 (green) is superimposed.

(T25I), and the α1-helix (L20A), whereas in the 4β+α fold, these mutations are in the β3 strand, the α-helix, and the β2 strand, respectively. Most of the mutation sites, with the exception of L20 in the 3α fold, occur on the periphery of the 3α or 4β+α cores. As such, the resulting changes in stability are generally small (in an absolute sense) but nonetheless can have a significant impact on the fold outcome because of the low stability (ΔG_u ~1.5-2.5 kcal/mol) of the high-identity proteins. The extent to which each state is populated will depend on the relative energy levels of the 3α and 4β+α conformations. The large shifts in equilibrium between these two states can best be understood by considering the effects of the mutations on the two folds, destabilizing one structure while simultaneously stabilizing the alternative conformer in this binary system.

In G_A98, the L45Y mutation destabilizes the 3α conformation in the following way. The L45 side chain is not buried in the core but rather packs against it, making stabilizing hydrophobic contacts with the core residues I33 and I49 as well as with the more exposed L32 and Y29 side chains (Figures 3A and 6A). When L45 is mutated to Y45, these relatively small stabilizing interactions are mostly lost, as the more rigid tyrosine side chain with its fewer rotational degrees of freedom cannot pack as efficiently against the 3α core as a leucine. At the same time, the L45Y mutation also stabilizes the 4β+α conformation of G_B98 through a favorable hydrophobic interaction between the aromatic rings of Y45 in the β3 strand and the core F52 residue in the β4 strand (Figure 6B). Further stabilizing hydrogen bond interactions from Y45-ηOH to the D47-carboxylate and to the hydroxyl group of Y3 in the β1 strand are also likely based on the G_B98 NMR ensemble of structures. Thus, the destabilization of the 3α conformation is small but significant because of the low stability of G_A98. The relative gain in stability of the 4β+α fold is enough to shift the equilibrium to this state almost completely (~99%).

Introduction of a T25I mutation into G_B98 produces G_B98-T25I, which populates mostly (~95%) the 3α state with a small amount (~5%) of the 4β+α conformer. Inspection of the G_B98 structure gives insights into the destabilizing influence of the T25I mutation on the 4β+α fold. The T25 residue in G_B98 is located near the N terminus of the α-helix and has closest proximity to the L20, V21, and D22 side chains (Figure 6C), with a likely H-bonding interaction between the T25-γOH group and the carboxylate of D22. Loss of this H-bond in a T25I mutant will

therefore destabilize the 4β+α fold. This is consistent with earlier stability studies on G_B77, which showed that a T25I mutation decreased the T_m by 4.4°C (Alexander et al., 2009). However, the destabilizing effect of the T25I mutation is probably larger in G_B98 because of additional unfavorable steric interactions that are likely to exist between the adjacent branched side chains of I25 and L20.

In contrast, the T25I mutation stabilizes the 3α fold of G_B98-T25I. While CS-Rosetta only models the positions of the side chains shown in Figure 6D, their conformations are similar to those obtained in related NMR structures and provide a useful guide for discussing mutational effects. In the CS-Rosetta structure of G_B98-T25I, the side chains in the region of the T25I mutation have an average heavy atom RMSD of 1.6 ± 0.5 Å in the ensemble. These side-chain arrangements have average RMSDs of 1.5–1.9 Å when compared with their corresponding positions in both NMR (He et al., 2008; Alexander et al., 2009) and CS-Rosetta (Shen et al., 2010) structures of G_A88 and G_A95. Thus, I25 is located in the loop between the α1- and α2-helices of the G_B98-T25I 3α fold, and examination of the CS-Rosetta structure suggests that I25 contributes more extensively to the hydrophobic core than a threonine residue. A similar stabilizing conformation is also seen for I25 in G_A88 (He et al., 2008). The putatively stabilizing hydrophobic interactions involving I25 and neighboring residues L20, A23, I49, and F52 (Figure 6D) counteract the known destabilizing influence of having Y45 in the α3-helix so that this sequence adopts a predominantly 3α conformation. The G_B98-T25I mutant is the only case so far where both folds are detectable by NMR; therefore, the 4β+α and 3α conformations of this amino acid sequence must be the closest in energy of the four mutants in this series.

The third fold switch involves an L20A mutation in the α1-helix of G_B98-T25I. L20 is completely buried in the hydrophobic core of the 3α conformation of G_B98-T25I (Figure 6D). Conversion to alanine at this position decreases packing interactions with other neighboring residues contributing to the core such as A16, I25, and I49, thereby destabilizing the 3α fold. Indeed this is the most destabilizing of the three mutations in this study because the L20A mutation cannot be tolerated even in more stable 3α mutants such as G_A77 (T_m 77°C) and leads to an unfolded protein (Alexander et al., 2009). In the 4β+α conformer, mutation to the smaller A20 residue lowers the energy of this state by removing unfavorable steric contacts that would otherwise exist between the proximal L20 and I25 side chains. Thus, the equilibrium is shifted almost completely (>95%) to the 4β+α fold with an L20A substitution. G_B98-T25I,L20A is the most stable mutant in this series (T_m ~46°C) and therefore must have the largest energy gap between 4β+α and 3α states of any of the proteins studied here.

Gain and Loss of Function

The ligand-binding study provides insights into how folds and functions can evolve. In particular, a new IgG-binding function is gained in G_A98 before complete loss of the original HSA-binding function. Bifunctional mutants such as G_A98 therefore serve as transitory species between distinct functional states. Moreover, the new IgG-binding function of G_A98 is detectable before there is significant population of the corresponding 4β+α fold. This stems from the intrinsically tight binding of IgG to G_B sequences adopting the 4β+α conformation. In this way, even low levels of the 4β+α fold in the unbound states of G_A98 (~1%) and G_B98-T25I (~5%) can lead to IgG binding, with K_D values of 62 and 15 μM, respectively (Table 2). Thus, increased equilibrium amounts of the 4β+α state in samples with predominantly 3α conformers correlate with a gain of IgG-binding function.

In contrast, the baseline HSA affinity of G_A88 is at least 30-fold weaker than IgG affinity to the 4β+α proteins G_B98 and G_B98-T25I,L20A. Much of this loss in affinity is due to the mutation of A52F, which apparently alters the contact with HSA. Further loss of HSA-binding function occurs in G_A98 and G_B98-T25I, even though 3α levels are still high. The drop in HSA affinity from G_A95 to G_A98 is primarily due to a decrease in protein stability, as the only amino acid change (I30F) is located away from the HSA-binding epitope (Lejon et al., 2004; He et al., 2007) at the α2-α3 surface (Figure 3A). In the case of G_B98-T25I, the complete loss of HSA-binding function is mainly due to the presence of a tyrosine residue at position 45 (Figure 1C). Even in other more stable 3α mutants, changing leucine to tyrosine at residue 45 was found to abolish HSA binding. This is consistent with the observation that L45 is centrally located in the binding interface between HSA and a variant of wild-type G_A (He et al., 2007).

The results here demonstrate that the mode for fold switching is not unique but can occur in multiple ways, thus increasing the probability for such events. These large-scale structural changes can occur through a series of single amino acid substitutions, once a suitable high-sequence identity background has been reached. The present study uses only binary sequence space (either G_A or G_B amino acids) and is not exhaustive, so it is likely that other switch mutants also exist. Expansion to the complete range of amino acids may further increase the number of single amino acid switch mutants and could also potentially lead to other folds and functions. Indeed, recent theoretical studies suggest that the high-identity G_A/G_B sequences may be capable of adopting numerous other fold topologies (Cao and Elber, 2010).

It is possible that the in vitro-directed evolution of the G_A/G_B system may reflect some aspects of the in vivo evolution of the G_A and G_B domains in protein G. Protein G is a multidomain protein with two to three copies of each of the G_A and G_B domains. One plausible hypothesis based on our results is that a duplicated HSA-binding G_A domain evolved the IgG-binding function through fold switching. In such a multidomain system, the likelihood that this could occur seems high, because functionality would be gained with no loss of fitness (due to the multiple copies). Other multidomain proteins may therefore provide further examples where fold switching has occurred.

EXPERIMENTAL PROCEDURES

Mutagenesis

We prepared mutants with a QuikChange (Stratagene) kit using the manufacturer's protocol.

Protein Expression and Purification

G_A and G_B mutants were cloned into a vector encoding an N-terminal subtilisin-prodomain fusion tag system (Profinity eXact, Bio-Rad) described previously (Ruan et al., 2004). *Escherichia coli* BL21(DE3) cells were transformed

with this vector, and cells were grown at 25°C to a density of 0.6–0.8 OD_{600} in M9 minimal media for ^{13}C and ^{15}N labeling. Protein expression was induced with 1 mM IPTG, and the cells were grown a further 6 hr at 25°C. The culture was then centrifuged, and the cells were suspended in 100 mM KPi buffer (pH 7.0) and then sonicated. The cleared cell extract was loaded on to a 5-ml eXact column at 5 mL/min and then washed extensively with 100 mM potassium phosphate buffer (pH 7.0). The pure target protein was cleaved and eluted with an injection of 6 ml of a solution containing 10 mM sodium azide and 100 mM potassium phosphate (pH 7.0) at 0.5 mL/min. Purified samples were then concentrated for NMR analysis.

NMR Spectroscopy

Isotope-labeled samples were prepared at concentrations of 0.05–0.3 mM for NMR analysis in 100 mM potassium phosphate buffer (pH 7.0) containing 5% D_2O. NMR spectra were acquired on a Bruker AVANCE 600 MHz spectrometer with cryoprobe. Spectra were processed with NMRPipe (Delaglio et al., 1995) and analyzed with Sparky (Goddard and Kneller, 2004). Backbone resonances were assigned with HNCA, HN(CO)CA, HN(CA)CO, HNCO, HNCACB, and CBCA(CO)NH experiments. Aliphatic side-chain assignments were obtained with (H)C(CO)NH-TOCSY and H(CCO)NH-TOCSY spectra, while aromatic resonances were assigned with two-dimensional (2D) CBHD/CBHE and NOESY spectra. We measured NOEs using 3D ^{15}N NOESY and aliphatic and aromatic 3D ^{13}C NOESY spectra.

Ligand Binding

Free and bound states were in slow exchange on the NMR time scale; hence, peaks due to the high molecular weight bound states were broadened beyond detection. Therefore, binding was determined by measuring the decay in peak intensity of amide protons in ^{15}N HSQC spectra of G_A or G_B mutants as a function of ligand concentration. In a typical binding experiment, the ^{15}N-labeled protein was approximately 50 μM, and ligand concentrations ranging from 0.1 to 8.0 molar equivalents were used depending on binding affinity. Control experiments were also carried out to determine how much of the peak intensity decay was due to an increase in solution viscosity from added IgG or HSA. This was done by adding increasing concentrations of IgG or HSA to a known nonbinder and measuring the decrease in amide peak intensities. Thus, the viscosity affect of IgG was determined by adding IgG to the nonbinder, ^{15}N-labeled PSD-1, a variant of wild-type G_A (He et al., 2006). Similarly, HSA was added to the nonbinder, ^{15}N-labeled wild-type G_B, to determine the HSA viscosity affect. Amide intensity decay curves due to binding were then corrected for the viscosity affect.

Structure Calculations and Analysis

We determined NMR structures for G_A98, G_B98, and G_B98-T25I,L20A using CNS 1.1 (Brünger et al., 1998). Assignment of NOEs was assisted with an in-house program, NOEID. Interproton NOE distance restraints were classified as strong (1.8–3.0 Å), medium (1.8–4.0 Å), weak (1.8–5.0 Å), and very weak (2.8–6.0 Å). TALOS (Cornilescu et al., 1999) was used to provide backbone dihedral restraints from chemical shift data. Hydrogen bond restraints were incorporated in the latter stages of refinement. The final ensemble of 20 structures was chosen based on low total energy, no NOE distance violations >0.3 Å, no dihedral angle violations >5°, and other parameters shown in Table 1. Structures were analyzed with PROCHECK-NMR (Laskowski et al., 1996), PyMol (Delano, 2002), and MOLMOL (Koradi et al., 1996). We used the standard CS-Rosetta3.2 protocol (Shen et al., 2008) to determine the G_B98-T25I structure based on chemical shifts. One thousand CS-Rosetta models were generated, and the 10 lowest energy models clustered with a backbone RMSD of 1.00 ± 0.27 Å.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two tables and three figures and can be found with this article online at doi:10.1016/j.str.2011.11.018.

ACKNOWLEDGMENTS

We thank John Moult, Ron Elber, Nick Grishin, and Jeff Skolnick for helpful discussions; Yang Shen for assistance with CS-Rosetta; and Jane Ladner for laser light-scattering analysis. This work was supported by National Institutes of Health Grant GM62154 and a grant from the W.M. Keck Foundation.

Received: October 6, 2011
Revised: November 29, 2011
Accepted: November 30, 2011
Published: February 7, 2012

REFERENCES

Alexander, P.A., He, Y., Chen, Y., Orban, J., and Bryan, P.N. (2007). The design and characterization of two proteins with 88% sequence identity but different structure and function. Proc. Natl. Acad. Sci. USA *104*, 11963–11968.

Alexander, P.A., He, Y., Chen, Y., Orban, J., and Bryan, P.N. (2009). A minimal sequence code for switching protein structure and function. Proc. Natl. Acad. Sci. USA *106*, 21149–21154.

Ambroggio, X.I., and Kuhlman, B. (2006). Design of protein conformational switches. Curr. Opin. Struct. Biol. *16*, 525–530.

Belogurov, G.A., Mooney, R.A., Svetlov, V., Landick, R., and Artsimovitch, I. (2009). Functional specialization of transcription elongation factors. EMBO J. *28*, 112–122.

Blanco, F.J., Angrand, I., and Serrano, L. (1999). Exploring the conformational properties of the sequence space between two proteins with different folds: an experimental study. J. Mol. Biol. *285*, 741–753.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Bryan, P.N., and Orban, J. (2010). Proteins that switch folds. Curr. Opin. Struct. Biol. *20*, 482–488.

Cao, B., and Elber, R. (2010). Computational exploration of the network of sequence flow between protein structures. Proteins *78*, 985–1003.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR *13*, 289–302.

Dalal, S., and Regan, L. (2000). Understanding the sequence determinants of conformational switching using protein design. Protein Sci. *9*, 1651–1659.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Delano, W.L. (2002). The PyMOL Molecular Graphics System (San Carlos, CA: DeLano Scientific).

Fahnestock, S.R., Alexander, P., Nagle, J., and Filpula, D. (1986). Gene for an immunoglobulin-binding protein from a group G streptococcus. J. Bacteriol. *167*, 870–880.

Falkenberg, C., Björck, L., and Akerström, B. (1992). Localization of the binding site for streptococcal protein G on human serum albumin. Identification of a 5.5-kilodalton protein G binding albumin fragment. Biochemistry *31*, 1451–1457.

Gallagher, T.D., 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.

Goddard, T.D. and Kneller, D.G. (2004). SPARKY 3, University of California San Francisco.

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.

He, Y., Rozak, D.A., Sari, N., Chen, Y., Bryan, P., and Orban, J. (2006). Structure, dynamics, and stability variation in bacterial albumin binding modules: implications for species specificity. Biochemistry *45*, 10102–10109.

He, Y., Chen, Y., Rozak, D.A., Bryan, P.N., and Orban, J. (2007). An artificially evolved albumin binding module facilitates chemical shift epitope mapping of GA domain interactions with phylogenetically diverse albumins. Protein Sci. *16*, 1490–1494.

He, Y., Chen, Y., Alexander, P., Bryan, P.N., and Orban, J. (2008). NMR structures of two designed proteins with high sequence identity but different fold and function. Proc. Natl. Acad. Sci. USA *105*, 14412–14417.

Johansson, M.U., de Château, M., Wikström, M., Forsén, S., Drakenberg, T., and Björck, L. (1997). Solution structure of the albumin-binding GA module: a versatile bacterial protein domain. J. Mol. Biol. *266*, 859–865.

Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. *14*, 51–55, 29–32.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR *8*, 477–486.

Lejon, S., Frick, I.M., Björck, L., Wikström, M., and Svensson, S. (2004). Crystal structure and biological implications of a bacterial albumin binding module in complex with human serum albumin. J. Biol. Chem. *279*, 42924–42928.

Littler, D.R., Harrop, S.J., Fairlie, W.D., Brown, L.J., Pankhurst, G.J., Pankhurst, S., DeMaere, M.Z., Campbell, T.J., Bauskin, A.R., Tonini, R., et al. (2004). The intracellular chloride ion channel protein CLIC1 undergoes a redox-controlled structural transition. J. Biol. Chem. *279*, 9298–9305.

Luo, X., and Yu, H. (2008). Protein metamorphosis: the two-state behavior of Mad2. Structure *16*, 1616–1625.

Mapelli, M., and Musacchio, A. (2007). MAD contortions: conformational dimerization boosts spindle checkpoint signaling. Curr. Opin. Struct. Biol. *17*, 716–725.

Meyerguz, L., Kleinberg, J., and Elber, R. (2007). The network of sequence flow between protein structures. Proc. Natl. Acad. Sci. USA *104*, 11627–11632.

Murzin, A.G. (2008). Biochemistry. Metamorphic proteins. Science *320*, 1725–1726.

Myhre, E.B., and Kronvall, G. (1977). Heterogeneity of nonimmune immunoglobulin Fc reactivity among gram-positive cocci: description of three major types of receptors for human immunoglobulin G. Infect. Immun. *17*, 475–482.

Roessler, C.G., Hall, B.M., Anderson, W.J., Ingram, W.M., Roberts, S.A., Montfort, W.R., and Cordes, M.H. (2008). Transitive homology-guided structural studies lead to discovery of Cro proteins with 40% sequence identity but different folds. Proc. Natl. Acad. Sci. USA *105*, 2343–2348.

Rose, G.D., and Creamer, T.P. (1994). Protein folding: predicting predicting. Proteins *19*, 1–3.

Ruan, B., Fisher, K.E., Alexander, P.A., Doroshko, V., and Bryan, P.N. (2004). Engineering subtilisin into a fluoride-triggered processing protease useful for one-step protein purification. Biochemistry *43*, 14539–14546.

Sauer-Eriksson, A.E., Kleywegt, G.J., Uhlén, M., and Jones, T.A. (1995). Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. Structure *3*, 265–278.

Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J.M., Liu, G., Eletsky, A., Wu, Y., Singarapu, K.K., Lemak, A., et al. (2008). Consistent blind protein structure generation from NMR chemical shift data. Proc. Natl. Acad. Sci. USA *105*, 4685–4690.

Shen, Y., Bryan, P.N., He, Y., Orban, J., Baker, D., and Bax, A. (2010). De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. Protein Sci. *19*, 349–356.

Tuinstra, R.L., Peterson, F.C., Kutlesa, S., Elgin, E.S., Kron, M.A., and Volkman, B.F. (2008). Interconversion between two unrelated protein folds in the lymphotactin native state. Proc. Natl. Acad. Sci. USA *105*, 5057–5062.

Weissmann, C. (2005). Birth of a prion: spontaneous generation revisited. Cell *122*, 165–168.

Wishart, D.S., and Sykes, B.D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. J. Biomol. NMR *4*, 171–180.