NMR structures of two designed proteins with high sequence identity but different fold and function

Yanan He, Yihong Chen, Patrick Alexander, Philip N. Bryan, and John Orban*

Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850

Edited by David Baker, University of Washington, Seattle, WA, and approved July 22, 2008 (received for review June 17, 2008)

How protein sequence codes for 3D structure remains a fundamental question in biology. One approach to understanding the folding code is to design a pair of proteins with maximal sequence identity but retaining different folds. Therefore, the nonidentities must be responsible for determining which fold topology prevails and constitute a fold-specific folding code. We recently designed two proteins, G_A88 and G_B88, with 88% sequence identity but different folds and functions [Alexander *et al.* (2007) *Proc Natl Acad Sci USA* 104:11963–11968]. Here, we describe the detailed 3D structures of these proteins determined in solution by NMR spectroscopy. Despite a large number of mutations taking the sequence identity level from 16 to 88%, G_A88 and G_B88 maintain their distinct wild-type 3-α and α/β folds, respectively. To our knowledge, the 3D-structure determination of two *monomeric* proteins with such high sequence identity but different fold topology is unprecedented. The geometries of the seven nonidentical residues (of 56 total) provide insights into the structural basis for switching between 3-α and α/β conformations. Further mutation of a subset of these nonidentities, guided by the G_A88 and G_B88 structures, leads to proteins with even higher levels of sequence identity (95%) and different folds. Thus, conformational switching to an alternative monomeric fold of comparable stability can be effected with just a handful of mutations in a small protein. This result has implications for understanding not only the folding code but also the evolution of new folds.

conformational switching | evolution | folding

Understanding the relationship between protein sequence and 3D structure (1) remains the fundamental unresolved problem in structural biology. There are several reasons why the protein folding problem is so difficult. The large number of conformations available to even a short polypeptide chain makes it difficult to calculate which conformation is most preferred. Also, amino acids in a polypeptide sequence contribute to different extents in coding for a particular fold (2–4). Mutations at some positions will have negligible effect on protein stability whereas other residues cannot be altered without resulting in complete unfolding. In this sense, most natural folds can be considered to be only marginally stable. The combination of these factors results in the general observation that many sequences with little or no discernible homology can frequently have the same overall fold, making prediction of 3D structure from sequence highly problematic in such cases.

A different way of looking at this problem was posed by Rose and Creamer (5). They suggested that one could gain insights into the folding code by determining the minimum number of amino acids required to specify one fold over another. The basic idea was to design a pair of proteins with maximal sequence identity but retaining their different wild-type folds. The nonidentities between these two amino acid sequences would then be responsible for coding one fold topology over the other and, thus, represent a fold-specific folding code. Several groups responded to this challenge (6–10) and the most successful of these studies resulted in a protein pair that had 80% sequence identity (11). Some of these proteins were characterized by circular dichroism (CD) spectroscopy under acidic conditions but had a propensity to aggregate near neutral pH. Further, solubility was limited and detailed 3D structures were not determined. Other groups were not able to obtain protein pairs with sequence identities >50% that were sufficiently stable and soluble for detailed structural studies. The tendency to aggregate is perhaps not surprising because it is well known that multimerization can accompany conformational switching and in fact can drive the change from one folded state to another (12–14).

We recently described the design and preliminary characterization of two small proteins, G_A88 and G_B88, with 88% sequence identity but different monomeric folds and functions (15). Here, we present their high-resolution 3D structures determined in solution by using NMR spectroscopy. We show that, despite a large number of mutations taking the sequence identity level from 16% to 88% (Fig. 1), the two proteins maintain their distinct wild-type folds. G_A88 has a 3-α helical structure whereas G_B88 is an α/β fold and each is shown to have high similarity to the structure of the parent protein, from which it was derived. Analysis of the geometries for the seven nonidentical residues suggests that higher levels of identity may be possible without switching folds. In particular, mutation of some nonidentities that are solvent accessible in at least one of the folds leads to proteins with even higher sequence identity.

In previous articles, we have referred to these protein pairs as "homologous." In evolutionary biology, this word implies that the sequence similarity was inherited from a common ancestor of the two proteins. To avoid confusion, we refer to these protein pairs as having "high sequence identity" in this article.

Results and Discussion

Protein Design. The design process is described in detail in ref. 15, and the made mutations are summarized in Fig. 1. The starting points were the human serum albumin (HSA)- and IgG-binding domains of streptococcal protein G, G_A, and G_{B1}, respectively (16–22). A total of 24 mutations were made in the 56-residue parent G_A protein sequence, PSD-1 (23), to generate G_A88 (Fig. 1*A*). Seventeen amino acid modifications were made in G_{B1} to produce G_B88 (Fig. 1*B*). Alignment of parent and designed sequences is shown in Fig. 1*C*.

NMR Assignment and Structure of G_A88. G_A88 was shown to be monomeric based on elution through G25 and G75 gel filtration columns. Linewidths in NMR spectra were also consistent with a monomeric state. ^{15}N HSQC spectra were recorded over a range of temperatures from 2 to 30°C to determine the optimum conditions for data collection with no evidence for exchange broadening of

Author contributions: P.A., P.N.B., and J.O. designed research; Y.H., Y.C., and P.A. performed research; P.N.B. contributed new reagents/analytic tools; Y.H. and J.O. analyzed data; and J.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The NMR assignments for G_A88 and G_B88 have been deposited in BioMagResBank, www.bmrb.wisc.edu (accession no. BMRB-15535 and BMRB-15537, respectively). The coordinates for the G_A88 and G_B88 structures have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 2JWS and 2JWU, respectively).

*To whom correspondence should be addressed. E-mail: orban@umbi.umd.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0805857105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0805857105

Petitioner Merck, Ex. 1040, p. 14412

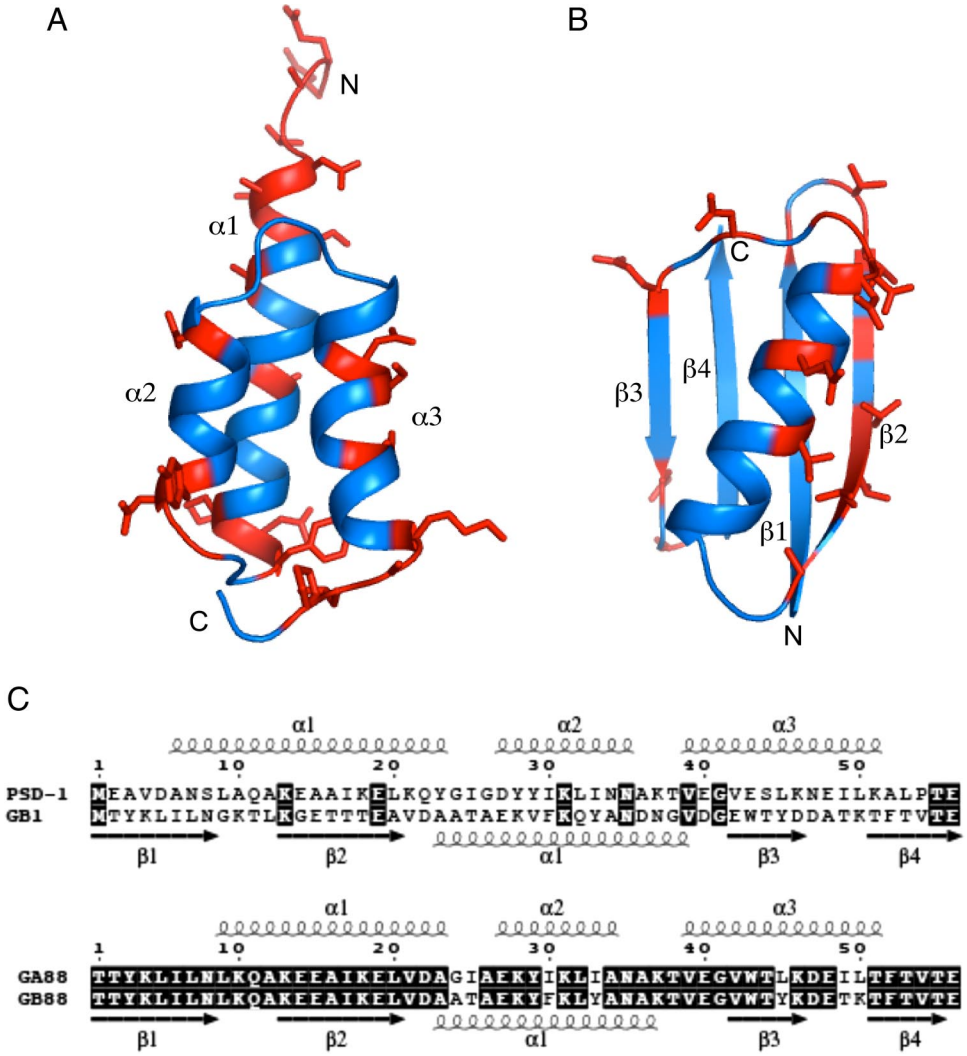

Fig. 1. Summary of mutations made in the parent proteins and sequence alignments. (*A*) Amino acid changes in PSD-1 [Protein Data Bank (PDB) code 2fs1] to generate G_A88 are shown in red. (*B*) Mutations in G_{B1} (PDB entry 1PGB) to generate G_B88 are shown in red. (*C*) Alignment of amino acid sequences for the parent proteins PSD-1 and G_{B1} (*Upper*), and the designed proteins G_A88 and G_B88 (*Lower*). Secondary-structure elements are displayed adjacent to the relevant sequences. The nine identities between the parent proteins PSD-1 and G_{B1} and the 49 identities between the designed proteins G_A88 and G_B88 are indicated. Sequence alignments are displayed with ESPript (http://espript.ibcp.fr/).

amide resonances. All main-chain signals were assigned and extensive assignments were also made for side-chain resonances. Chemical shift index (24) and NOE data analysis indicated that G_A88 contains three α-helices, $\alpha1$ from residues 9–23, $\alpha2$ from residues 27–34, and $\alpha3$ from residues 39–51. These helices pack against each other such that $\alpha2$ is antiparallel to both $\alpha1$ and $\alpha3$ with relatively short three- and four-residue loops connecting the $\alpha1$–$\alpha2$ and $\alpha2$–$\alpha3$ helices, respectively. The N-terminal residues 1–8 are disordered with sparse interresidue NOEs in this region. Similarly the C-terminal residues 54–56 are not well defined. The average backbone rmsd for the structured region of the polypeptide chain (residues 9–53) is 0.24 ± 0.06 Å with the loops less well defined than the helices. The complete structure statistics for G_A88 are summarized in Table 1, and the ensemble of 20 final calculated structures based on experimental data are shown in Fig. 2*A*.

The hydrophobic core of G_A88 contains residues K13, A16, and L20 from $\alpha1$, I33 from $\alpha2$, and V42, K46, and I49 from $\alpha3$. These are all well ordered in the ensemble with average heavy atom rmsds to the mean structure of 0.36 ± 0.10 Å. The aliphatic side chains of K13 and K46 contribute to packing of the hydrophobic core whereas their ammonium groups are solvent exposed. Additionally, the ε-ammonium group of K13 is proximal to the main-chain carbonyl groups of I33 and A34 in $\alpha2$, providing a likely C-terminal capping interaction that stabilizes the $\alpha2$-helix. Similarly, the side-chain ammonium group of K46 is adjacent to the acidic groups of E15 and E19 in $\alpha1$, thereby further stabilizing the interaction between $\alpha3$ and $\alpha1$. A network of adjacent acidic and basic residues is present in this highly charged region of the structure and also includes K10, E14, K18, D22, and D47. Some boundary residues between the core and surface are also ordered in the ensemble including A12, E19, A23, A26, L32, and L50. At the C terminus of the $\alpha3$-helix, F52 is partially ordered (heavy atom rmsd 0.83 ± 0.56 Å) and forms stabilizing hydrophobic interactions with I25 in the adjacent $\alpha1$–$\alpha2$ loop and I49 in the $\alpha3$-helix (Fig. 3*A*).

NMR Assignment and Structure of G_B88. G_B88 was found to be monomeric based on its mobility during size-exclusion gel chromatography. ^{15}N HSQC spectra recorded over the temperature range 2–25°C indicated no exchange broadening of amide resonances and backbone and side-chain resonance assignments were made at 22°C. The NMR backbone chemical shift data indicated that G_B88 contains four β-strands and one α-helix with the secondary structures arranged in the order $\beta1$–$\beta2$–α–$\beta3$–$\beta4$. Detailed NOE analysis showed that the β-strands form a four-stranded β-sheet with $\beta1$ (residues 1–8) and $\beta4$ (residues 51–55) as the central strands in a parallel arrangement and $\beta2$ (residues 13–20) and $\beta3$ (residues 42–46) as the outer strands, antiparallel to $\beta1$ and $\beta4$, respectively. The α-helix (residues 23–36) runs diagonally across the sheet and is connected to $\beta2$ by a two-residue loop and to $\beta3$ by a five-residue loop. The main chain is generally well ordered with an average backbone rmsd of 0.52 ± 0.10 Å. The hydrophobic core of G_B88 is well defined and consists of residues Y3, L5, and L7 in $\beta1$, A26, F30,

BIOPHYSICS

Petitioner Merck, Ex. 1040, p. 14413

Table 1. Statistics for the G_B88 and G_A88 ensembles of 20 structures

	G_B88	G_A88
Experimental restraints		
NOE restraints		
All NOEs	918	931
Intraresidue	544	525
Sequential ($\|i - j\| = 1$)	171	190
Medium-range ($1 < \|i - j\| \leq 5$)	88	164
Long-range ($\|i - j\| > 5$)	115	52
Hydrogen bond restraints	62	48
Dihedral angle restraints	76	70
Total restraints	1,056	1,049
Rmsds to the mean structure, Å		
Over all residues*		
Backbone atoms	0.52 ± 0.10	0.24 ± 0.06
Heavy atoms	1.30 ± 0.15	1.23 ± 0.20
Secondary structures†		
Backbone atoms	0.39 ± 0.08	0.19 ± 0.05
Heavy atoms	1.15 ± 0.15	1.25 ± 0.22
Measures of structure quality		
Ramachandran distribution		
Most favored, %	81.2 ± 2.9	74.0 ± 2.9
Additionally allowed, %	17.6 ± 2.8	17.9 ± 3.7
Generously allowed, %	1.3 ± 1.4	6.7 ± 2.7
Disallowed, %	0.0 ± 0.0	1.4 ± 1.6
Bad contacts/100 residues	1.7 ± 1.1	3.6 ± 1.3
Overall dihedral *G* factor	0.05 ± 0.02	−0.05 ± 0.03

*Residues 1–56 for G_B88. Residues 9–53 for G_A88.
†The secondary elements used were as follows: G_B88, residues 1–8, 13–20, 23–36, 42–46, and 51–55; G_A88, residues 9–23, 27–34 and 39–51.

and A34 in the α-helix, contributions from W43 and Y45 in β3, and F52 and V54 in β4. These residues have average heavy atom rmsds of 0.73 ± 0.17 Å. A number of boundary residues pack against this core providing further stabilizing interactions. These include L9 in β1, K18 and L20 in β2, and E27, Y29, K31, Y33, and T38 in the α-helix. The structure statistics for G_B88 are summarized in Table 1 and an overlay of the 20 final structures is shown in Fig. 2*B*.

Comparison of G_A88 and Parent Structures. A superposition of the G_A88 and PSD-1 (25) 3D structures is shown in Fig. 4*A*. The backbone rmsd for residues 9–53 is 1.0 ± 0.43 Å between the mean structures indicating that the global folds are very similar for these two proteins despite the large number of mutations introduced. Although most surface residue changes have little impact on the global fold, some differences in local main-chain structure are immediately apparent. First, the α1-helix is one turn shorter at its N terminus in G_A88 than in PSD-1. This stems directly from the mutations A6I, N7L, and S8N, which result in a net decrease in helical propensity for this region. Second, the C-terminal unstructured tail (residues 54–56) in G_A88 goes in the opposite direction to that seen in PSD-1. The mutational reasons for this feature are not so obvious but the α3-helix is somewhat more tightly wound at the C-terminal end in G_A88 than in PSD-1 and the rmsds in this region are the largest of the three helices [supporting information (SI) Fig. S1]. Also, the C-terminal capping interaction that occurs between L53 and I25/L50 in PSD-1 is now between F52 and I25/I49 in G_A88 (Fig. 3). Thus, the A52F mutation results in the reorientation of the surface alanine in PSD-1 to a capping phenylalanine in G_A88 and this difference, and the L53T mutation, leads in part to the altered trajectory of the polypeptide chain for the last few residues. The A52F mutation causes an ≈10-fold decrease in HSA-binding affinity (data not shown), which may be partly due to these small structural changes. Previous work showed that the

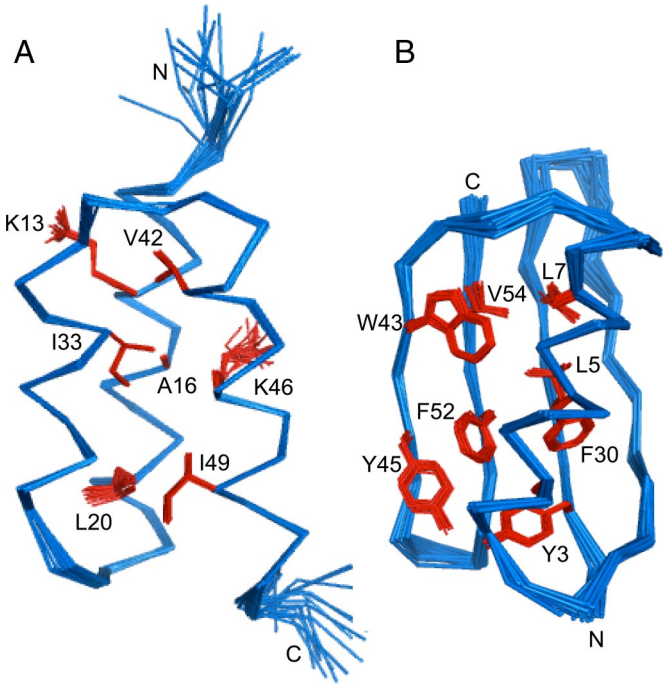

Fig. 2. NMR structures of designed proteins G_A88 and G_B88. (*A*) NMR ensemble of the 20 final structures for G_A88 (residues 6–55) in ribbon representation. (*B*) NMR ensemble of 20 final structures for G_B88 (residues 1–56). The main chain is shown in blue, whereas core and other key side chains are shown in red for both structures.

affinity of G_A modules for albumins can be affected by relatively small changes in helical arrangement (25). Most core residues are situated similarly in both proteins apart from those in the α3-helix that are displaced 2–3 Å from their wild-type PSD-1 positions.

Comparison of G_B88 and Parent Structures. The design of G_B88 involved the introduction of 17 mutations and the impact of these changes was evaluated by comparing its 3D structure with that of the wild-type precursor, G_{B1} (26). A superposition of G_B88 with G_{B1} is shown in Fig. 4*B*. The average backbone rmsd between these structures is 0.83 ± 0.36 Å indicating a high degree of similarity in the overall folds. Further, the core residues have very similar positions to those in the wild-type G_B structure. The small differences are presumably due to changes at adjacent residues just outside the core.

One of the main observations is that large changes in amino acid size can be incorporated in regions between the core and surface

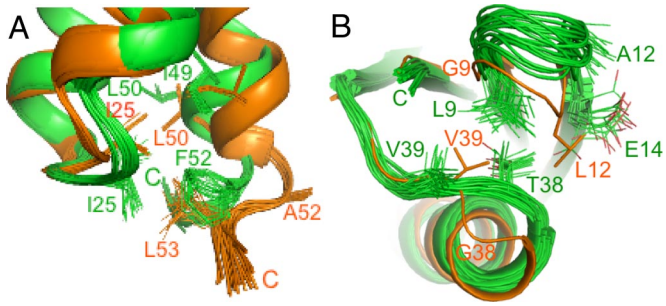

Fig. 3. Altered local structure in designed proteins. (*A*) Packing of F52 in G_A88 (green) and comparison with A52 in the parent structure, PSD-1 (orange). (*B*) Backbone and side-chain orientations in the α–β3 loop and β1–β2 loop regions of G_B88 (green) compared with the parent G_{B1} (orange).

Petitioner Merck, Ex. 1040, p. 14414

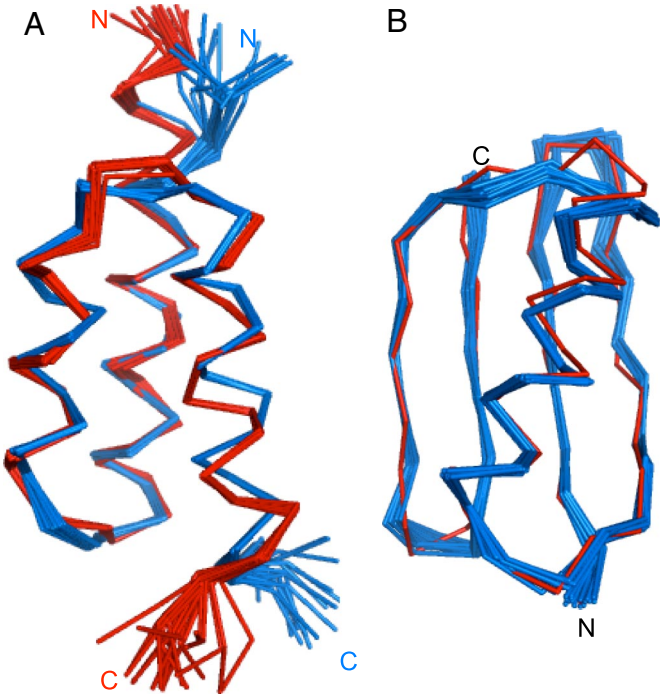

Fig. 4. C_α trace backbone superposition of parent and designed protein structures. (*A*) Structural alignment of G_A88 (20-structure ensemble, blue) with PSD-1 (20-structure ensemble, red). (*B*) Overlay of G_B88 (20-structure ensemble, blue) with the x-ray structure of G_{B1} (red).

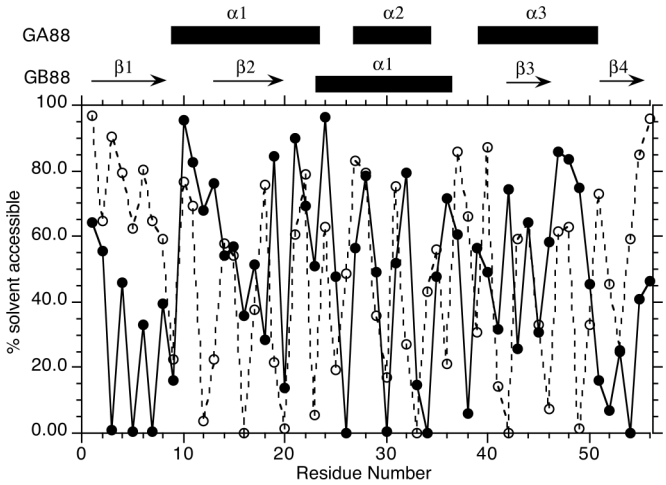

Fig. 5. Solvent accessibility in G_A88 and G_B88. Average values per residue were obtained from 20-structure ensembles for G_A88 (open circles) and G_B88 (filled circles). Standard deviations are omitted for clarity but were in the following ranges: ±4–20% for highly exposed residues (>50% accessible), ±2–11% for boundary residues (20–50% accessible), and ±0–6% for buried residues (<20% accessible).

with essentially no effect on the global fold. However, some mutations do lead to local structural differences compared with wild-type G_{B1}. For example, introduction of the G9L, L12A, G14E, and G38T mutations results in side-chain repacking and some changes in backbone conformation in the α–β3 loop and adjacent β1–β2 loop (Fig. 3*B*). Overall, this region of G_B88 has the largest backbone rmsds with the structure of native G_{B1} (Fig. S2). In particular, residue G38 is solvent exposed in G_{B1}, whereas in G_B88, T38 packs as a boundary residue adjacent to the core region. Consequently, the main chain of T38 is displaced from the wild-type position with an rmsd of 1.7 Å. Also, the T38 side chain in G_B88 partially occupies the space of V39 in its G_{B1} conformation. As a result, V39 is shifted into a more solvent-accessible orientation in G_B88. Regarding other mutations nearby, G9L is accommodated without a significant change in the main chain at this position and, together with T38, improves packing in this region compared with G_B (Fig. 3*B*). Preliminary stability studies indicate that a G9L/L12A mutant is stabilizing (data not shown). The more open conformation at residue 12 in G_B88 is likely to be due to the introduction of the G14E mutation. The bulkier glutamate side chain would have unfavorable steric interactions with both the main chain and side chain of residue 12 in the parent G_B x-ray conformation.

Amino Acid Identities. A key question is how can so many identities be tolerated in either the 3-α fold or the α/β fold. Presumably, part of the reason this is possible is because most of the core residues do not overlap in G_A88 and G_B88. Significantly overlapping core residues would be difficult to repack in alternate folds while maintaining sequence identity. If one considers the identities first, residues that contribute appreciably to the core of G_A88 (K13, A16, L20, V42, and K46; Fig. 2*A*) are more solvent accessible and have fewer interactions with other amino acids in G_B88 (Fig. 5). Thus, most of these residues can be accommodated in the alternate α/β fold without major repacking requirements. An exception is L20, which is completely buried in G_A88 but also partially buried in G_B88. In the wild-type α/β fold, residue 20 is an alanine, and so mutation to leucine produces increased steric contacts and slight repositioning of the N-terminal part of the helix, which leads to some loss of stability ($\Delta T_m \approx -12°C$) but, nevertheless, this mutation can be tolerated in G_B88. Similarly, identities that contribute significantly to the core of G_B88 (Y3, L5, L7, A26, A34, W43, F52, and V54; Fig. 2*B*) are markedly more solvent accessible in G_A88 (Fig. 5). Therefore, these residues can be tolerated in the alternate 3-α fold without any significant change in packing of the hydrophobic core.

Amino Acid Nonidentities. The geometries of the seven nonidentities in each structure are indicated in Fig. 6. These residues, or a subset of these residues, are responsible for shifting the equilibrium between the 3-α and α/β-folds. It should be noted that no signals due to the other fold, or due to the unfolded state, are observable within the detection limits of the NMR spectra for either G_A88 or G_B88 and so these equilibria lie strongly (>95%) to the respective folds. The extents of these equilibria are also supported by CD melting data, which estimate that these folds are populated at 99.9% for 3-α and 97% for α/β at 20°C. To better understand which of the nonidentities are most likely to shift the equilibrium from one conformation to the other, their contributions to each structure were analyzed in terms of packing interactions/solvent accessibility and intrinsic secondary-structure propensity (27, 28).

Nonidentical residues that are solvent accessible in either the 3-α or α/β fold and, therefore, have few packing interactions with other amino acids may be mutation tolerant and not have a major role in conformational switching. Solvent accessibility calculations of the G_A88 and G_B88 structures (Fig. 5) indicated that residues 24, 25, 49, and 50 may fall in this category. Residue 24 is highly solvent exposed at the N terminus of the α-helix in G_B88 in the α1–α2 loop of G_A88, and has limited contacts to other residues in both folds. Likewise, T25 is solvent accessible in the N terminus of the α-helix in G_B88 whereas I25 is partially buried in the α1–α2 loop of G_A88. T49 is a solvent-exposed surface residue in the β3–β4 loop of G_B88 with few interactions to other side chains whereas I49 is buried in the core of G_A88. Last, K50 has a disordered side chain in the β3–β4 loop of G_B88 and L50 is partly buried in the α3-helix of G_A88. The intrinsic secondary-structure propensities of the corresponding G_A/G_B residues at these positions are similar for the most part, and so packing interactions are likely to be the major influence. An

BIOPHYSICS

Petitioner Merck, Ex. 1040, p. 14415

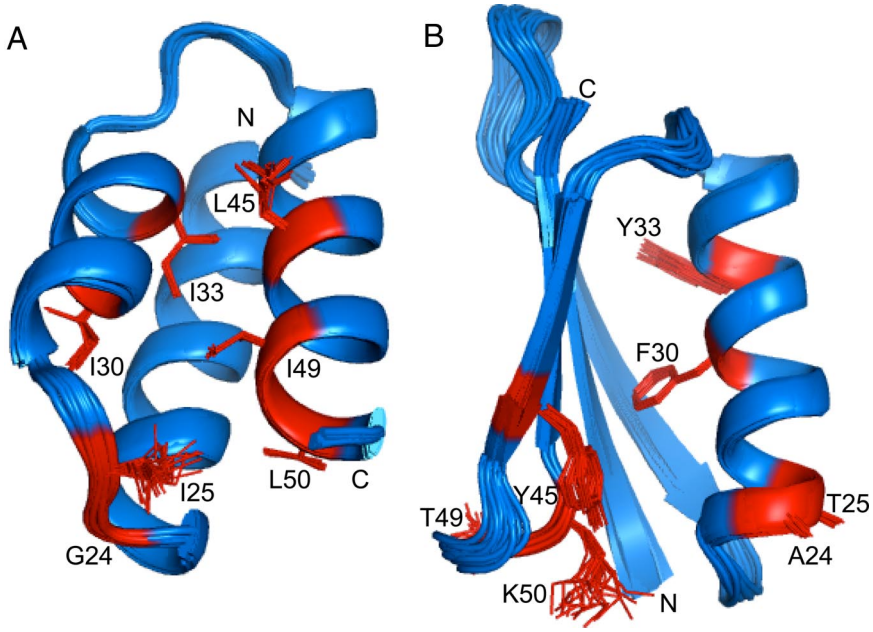

Fig. 6. Geometries of nonidentical residues. (*A*) NMR structure of G_A88 highlighting the positions of the seven aa differences (red) between the two sequences. (*B*) NMR structure of G_B88 showing the locations of the seven nonidentities (red). Residues at the disordered N- and C-terminal tails of G_A88 have been omitted for clarity. The blue regions represent sequence identity between G_A88 and G_B88.

exception is the G/A24 pair, where alanine has a high helical propensity while glycine is destabilizing. However, these propensities are for internal helical positions. At the ends of helices, a preference for glycine over alanine has been shown in other systems (29).

Mutations were made at the following positions based on analysis of the structures: I25T and L50K in G_A88, A24G and T49I in G_B88. Further mutations were also made at Y33I and L20A in the α/β fold. The latter mutation decreases identity but was made to regain some stability in the α/β fold. The resulting proteins, G_A95 and G_B95, have 95% sequence identity and are monomeric by gel filtration. The thermal denaturation midpoint is 52.0°C for G_A95 and 48.5°C for G_B95. The HSQC spectra of these proteins are consistent with the distinct 3-α and α/β folds of the parent proteins (Fig. 7). Backbone assignments were made by using $^{13}C/^{15}N$-labeled samples and consensus chemical shift index analysis (Table S1 and Table S2) indicates that the secondary-structure elements of the parent folds are maintained in the designed proteins. Detailed structures of G_A95 and G_B95 and stabilities of mutants will be described elsewhere.

Thus, the structures of G_A88 and G_B88 lead to the design of proteins with even higher identities. Only three or fewer amino acids are responsible for shifting the equilibrium from one stable monomeric fold to the other. The remaining three nonidentities are L20, I30, and L45 in G_A95 and A20, F30, and Y45 in G_B95. The destabilizing effect of an A20L mutation in G_B88 has already been discussed above. The I30/F30 residues have similar secondary-structure propensities while F30 is completely buried in the α/β fold and I30 has limited solvent accessibility in the 3-α fold. Given the large number of contacts to these amino acids in both folds, mutation of either may be sufficient to switch conformations or unfold the protein. In contrast, the L45/Y45 residues have similar

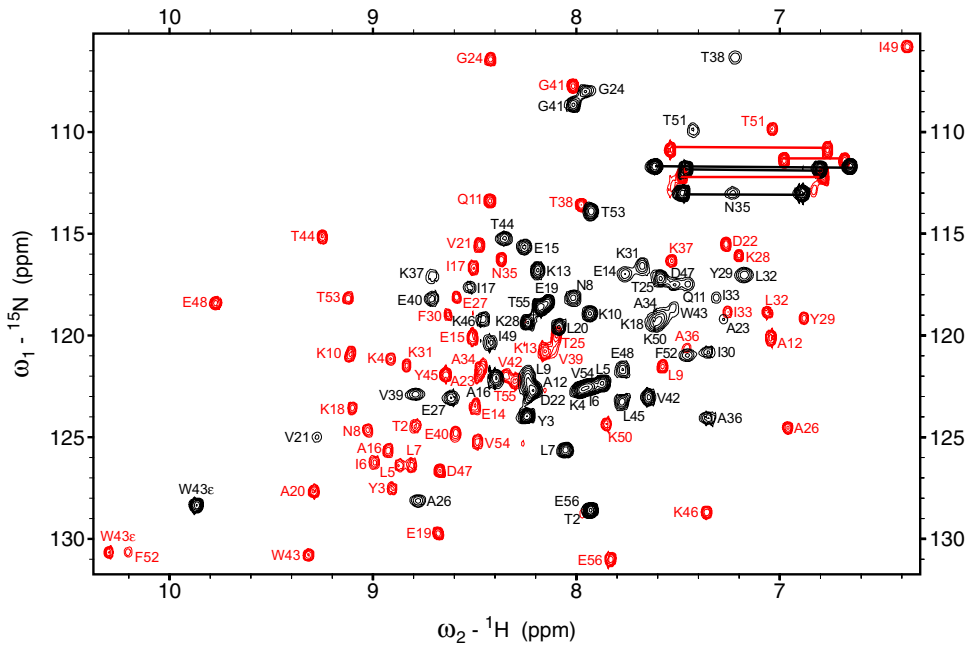

Fig. 7. Overlaid ^{15}N-HSQC spectra of G_A95 (black) and G_B95 (red). Main-chain assignments are indicated and chemical shift index data are included in Table S1 and Table S2. Horizontal lines connect side-chain amide signals. Fifty-three of 56 aa in these two proteins are identical but have different chemical environments reflecting the distinct folds.

Petitioner Merck, Ex. 1040, p. 14416

solvent accessibility but very different secondary-structure propensities. In G_A88, L45 is located in the α3-helix and has strong helical and moderate β-strand propensities. In G_B88, Y45 is located in the β3-strand and has strong β-strand and relatively weak helical propensities. Therefore, residue 45 may represent a tipping point in the sequences particularly when one considers that the nearest neighbors on either side, WT and KD, have net β-strand and α-helical preferences, respectively. Future studies will address the sequence dependence of conformational switching in these high-sequence identity proteins.

Methods

Protein Expression and Purification. G_A and G_B variants were cloned into the vector pG58, which encodes an engineered subtilisin prosequence as an N-terminal fusion domain, and the resulting fusion proteins were purified by using an affinity-cleavage tag system that we developed (30), essentially as described in ref. 15. The system enabled the rapid, standardized purification of mutant proteins, even of low stability. A commercial version of the purification system is available through Bio-Rad Laboratories (Profinity eXact Purification System). Minimal media (31) was used for ^{15}N and ^{13}C labeling. Soluble cell extract of prodomain (eXact tag) fusion protein was injected on a 5-ml S189 column at 5 ml/min to allow binding and then washed with 10 column volumes of 100 mM potassium phosphate, pH 7.2 to remove impurities. To cleave and elute the purified target protein, 6 ml of 10 mM sodium azide, 100 mM potassium phosphate, pH 7.2 was injected at 0.5 ml/min. The purified protein was then concentrated for NMR analysis.

NMR Spectroscopy. ^{15}N- and ^{13}C/^{15}N-labeled protein samples were prepared at concentrations in the range of 0.15–0.3 mM in 100 mM potassium phosphate buffer (pH 7.2) containing 10% D_2O. NMR spectra were collected on a Bruker AVANCE 600 MHz spectrometer fitted with a *z* axis gradient ^{1}H/^{13}C/^{15}N triple resonance cryoprobe. NMR spectra of G_A88 and G_B88 were recorded at 22°C, whereas spectra of G_A95 and G_B95 were acquired at 20°C. Processing was done by using nmrPipe (32) and spectra were analyzed with Sparky (33). NMR assignments were obtained by using standard triple resonance methods. Backbone resonances were assigned with HNCACB, CBCA(CO)NH, HBHA(CO)NH, and HNCO spectra. Aliphatic side-chain assignments were made by using a combination of (H)C(CO)NH-TOCSY and H(CCO)NH-TOCSY spectra. Aromatic assignments were obtained from 2D CBHD and CBHE experiments and NOESY spectra. Interproton NOEs were derived from 3D ^{15}N NOESY and aliphatic and aromatic 3D ^{13}C NOESY spectra with mixing times of 100 and 150 ms.

Structure Calculations and Analysis. Structures were calculated by using CNS 1.1 (34) with standard simulated annealing and torsion dynamics protocols. An extended polypeptide chain was used as the starting point. All prochiral groups were given floating assignments until they could be unambiguously assigned from the structure. An initial set of NOE restraints was generated automatically by using NOEID, an in-house NOE assignment program. Further assignments were obtained in a semiautomated mode by using NOEID and intermediate structures to narrow down ambiguous assignments iteratively. Interproton distance restraints were based on peak intensities and categorized 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). Backbone dihedral restraints were obtained from chemical shift data by using TALOS (35). Hydrogen bond restraints, 1.5–2.5 Å for $r_{HN\text{-}O}$ and 2.3–3.2 Å for $r_{N\text{-}O}$, were used only in the final stages of refinement. Final values for force constants were 1,000 kcal mol^{-1}Å^{-2} for bond lengths, 500 kcal mol^{-1} rad^{-2} for angles and improper torsions, 40 kcal mol^{-1}Å^{-2} for experimental distance restraints, 200 kcal mol^{-1} rad^{-2} for dihedral angle restraints, and 4.0 kcal mol^{-1}Å^{-4} for the van der Waals repulsion term. The final ensemble of 20 structures was chosen by using standard criteria including low total energy, no NOE distance violations >0.3 Å, no dihedral angle violations >5°, and other measures of structure quality shown in Table 1. Structures were analyzed by using PROCHECK-NMR (36), QUANTA (Molecular Simulations Inc.), MOLMOL (37), and PyMol (38). Solvent accessible surface areas were calculated with GETAREA (39).

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grant GM62154 and a grant from the W. M. Keck Foundation.

1. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181:223–230.
2. Lattman EE, Rose GD (1993) Protein folding - what's the question? *Proc Natl Acad Sci USA* 90:439–441.
3. Dahiyat BI, Mayo SL (1997) Probing the role of packing specificity in protein design. *Proc Natl Acad Sci USA* 94:10172–10177.
4. Cordes MH, Walsh NP, McKnight CJ, Sauer RT (1999) Evolution of a protein fold in vitro. *Science* 284:325–328.
5. Rose GD, Creamer TP (1994) Protein folding: Predicting predicting. *Proteins* 19:1–3.
6. Jones DT, *et al.* (1996) Towards meeting the Paracelsus Challenge: The design, synthesis, and characterization of paracelsin-43, an alpha-helical protein with over 50% sequence identity to an all-beta protein. *Proteins* 24:502–513.
7. Dalal S, Balasubramanian S, Regan L (1997) Transmuting alpha helices and beta sheets. *Fold Des* 2:R71–79.
8. Dalal S, Balasubramanian S, Regan L (1997) Protein alchemy: Changing beta-sheet into alpha-helix. *Nat Struct Biol* 4:548–552.
9. Yuan SM, Clarke ND (1998) A hybrid sequence approach to the paracelsus challenge. *Proteins* 30:136–143.
10. Blanco FJ, Angrand I, 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.
11. Dalal S, Regan L (2000) Understanding the sequence determinants of conformational switching using protein design. *Protein Sci* 9:1651–1659.
12. Kirsten FM, Dyda F, Dobrodumov A, Gronenborn AM (2002) Core mutations switch monomeric protein GB1 into an intertwined tetramer. *Nat Struct Biol* 9:877–885.
13. Kuloglu ES, McCaslin DR, Markley JL, Volkman BF (2002) Structural rearrangement of human lymphotactin, a C chemokine, under physiological solution conditions. *J Biol Chem* 277:17863–17870.
14. Weissmann C (2005) Birth of a prion: Spontaneous generation revisited. *Cell* 122:165–168.
15. Alexander PA, He Y, Chen Y, Orban J, Bryan PN (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.
16. Myhre EB, 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.
17. Reis KJ, Ayoub EM, Boyle MDP (1984) Streptococcal Fc receptors. II. Comparison of the reactivity of a receptor from a group C streptococcus with staphylococcal protein A. *J Immunol* 132:3098–3102.
18. Fahnestock SR, Alexander P, Nagle J, Filpula D (1986) Gene for an immunoglobulin-binding protein from a Group G Streptococcus. *J Bacteriol* 167:870–880.
19. Falkenberg C, Bjorck L, Akerstrom 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.
20. Frick IM, *et al.* (1992) Convergent evolution among immunoglobulin G-binding bacterial proteins. *Proc Natl Acad Sci USA* 89:8532–8536.
21. de Chateau M, Bjorck L (1994) Protein PAB, a mosaic albumin-binding bacterial protein representing the first contemporary example of module shuffling. *J Biol Chem* 269:12147–12151.
22. de Chateau M, Holst E, Bjorck L (1996) Protein PAB, an albumin-binding bacterial surface protein promoting growth and virulence. *J Biol Chem* 271:26609–26615.
23. Rozak DA, *et al.* (2006) Using offset recombinant polymerase chain reaction to identify functional determinants in a common family of bacterial albumin binding domains. *Biochemistry* 45:3263–3271.
24. Wishart DS, Sykes BD (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.
25. He Y, *et al.* (2006) Structure, dynamics, and stability variation in bacterial albumin binding modules: Implications for species specificity. *Biochemistry* 45:10102–10109.
26. Gallagher TD, Alexander P, Bryan P, Gilliland G (1994) Two crystal structures of the B1 Immunoglobulin-binding domain of Streptococcal Protein G and comparison with NMR. *Biochemistry* 33:4721–4729.
27. Munoz V, Serrano L (1994) Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices: Comparison with experimental scales. *Proteins* 20:301–311.
28. Street AG, Mayo SL (1999) Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proc Natl Acad Sci USA* 96:9074–9076.
29. Serrano L, Neira JL, Sancho J, Fersht AR (1992) Effect of alanine versus glycine in alpha-helices on protein stability. *Nature* 356:453–455.
30. Ruan B, Fisher KE, Alexander PA, Doroshko V, Bryan PN (2004) Engineering subtilisin into a fluoride-triggered processing protease useful for one-step protein purification. *Biochemistry* 43:14539–14546.
31. Alexander P, Fahnestock S, Lee T, Orban J, 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.
32. Delaglio F, *et al.* (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293.
33. Goddard TD, Kneller DG (2001) *SPARKY* (University of California, San Francisco) Version 3.
34. Brunger AT, *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D* 54:905–921.
35. Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289–302.
36. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477–486.
37. Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graphics* 14:51–55.
38. DeLano WL (2002) *The PyMOL Molecular Graphics System.* (DeLano Scientific, San Carlos, CA).
39. Fraczkiewicz R, Braun W (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comp Chem* 19:319–333.

BIOPHYSICS

Petitioner Merck, Ex. 1040, p. 14417