BIOCHEMISTRY

 Volume 44, Number 43 November 1, 2005

Articles

Directed Evolution of Highly Homologous Proteins with Different Folds by Phage Display: Implications for the Protein Folding Code†

Patrick A. Alexander, David A. Rozak, John Orban, and Philip N. Bryan*

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

Received June 27, 2005; Revised Manuscript Received August 26, 2005

ABSTRACT: To better understand how amino acid sequences specify unique tertiary folds, we have used random mutagenesis and phage display selection to evolve proteins with a high degree of sequence identity but different tertiary structures (homologous heteromorphs). The starting proteins in this evolutionary process were the IgG binding domains of streptococcal protein G (G_B) and staphylococcal protein A (A_B). These nonhomologous domains are similar in size and function but have different folds. G_B has an α/β fold, and A_B is a three-helix bundle (3-α). IgG binding function is used to select for mutant proteins which retain the correct tertiary structure as the level of sequence identity is increased. A detailed thermodynamic analysis of the folding reactions and binding reactions for a pair of homologous heteromorphs (59% identical) is presented. High-resolution NMR structures of the pair are presented by He et al. [(2005) *Biochemistry 44*, 14055−14061]. Because the homologous but heteromorphic proteins are identical at most positions in their sequence, their essential folding signals must reside in the positions of nonidentity. Further, the thermodynamic linkage between folding and binding is used to assess the propensity of one sequence to adopt two unique folds.

A major difficulty in deciphering the protein folding code is that folding information is diffuse. That is, not every amino acid in a protein has equal information content toward specifying the native fold (*1*). In a typical protein, many amino acid positions are almost neutral in determining the free energy of the native fold. At other positions, a single mutation will cause unfolding. The diffuse nature of folding information coupled with the low stability of most natural proteins ($\Delta G_{unfolding}$ between 5 and 15 kcal/mol) results in some proteins with similar backbone topologies having no discernible sequence identity. Thus, ways to determine how the essential folding signal of a protein is distributed throughout the sequence would be helpful.

To do this, we describe here the directed evolution of a pair of monomeric proteins with a high degree of sequence identity but different tertiary structures. Because the homologous but heteromorphic proteins are identical at most positions in their sequence, their essential folding signals must reside in the positions of nonidentity. The IgG[1] binding domains of streptococcal protein G (G_B) and staphylococcal protein A (A_B) provided the starting point in the evolutionary process. Protein G is a multidomain component of the cell wall of streptococcal species from Lancefield group G. Protein G contains two homologous domains of 45 amino acids, which bind to human serum albumin, and two to three homologous domains of 56 amino acids, which bind to all

† This work was supported by NIH Grant GM62154.

* To whom correspondence should be addressed. E-mail: bryan@umbi.umd.edu. Telephone: (240) 314-6220. Fax: (240) 314-6255.

10.1021/bi051231r CCC: $30.25
Published on Web 10/06/2005

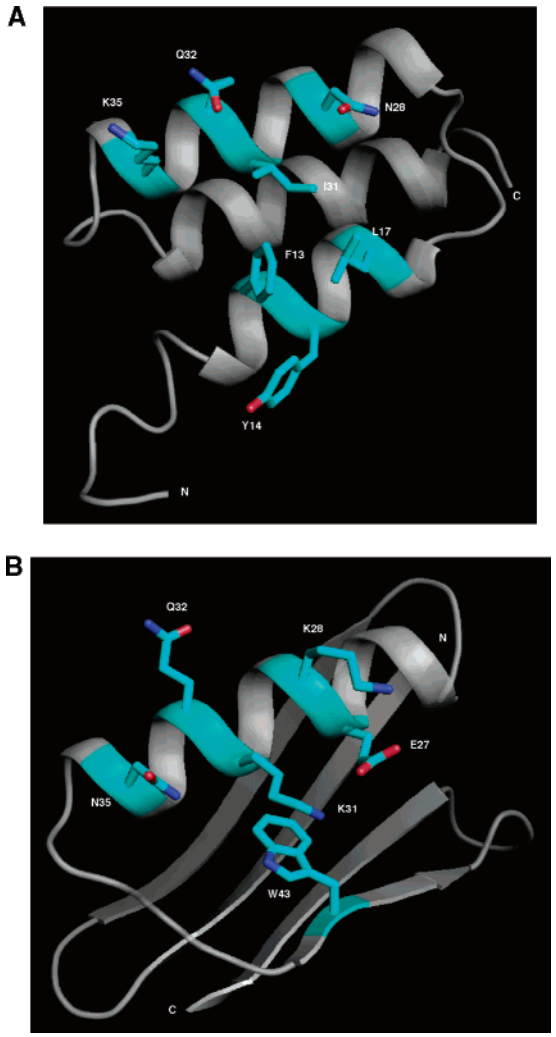

FIGURE 1: IgG binding epitopes of A_B and G_B. (A) Cartoon depiction of the secondary structure of A_B. Amino acids making direct contacts with IgG are colored cyan [from Protein Data Bank (PDB) entry 1edi]. (B) Cartoon depiction of the secondary structure of G_B. Amino acids making direct contacts with IgG are colored cyan (from PDB entry 1pga).

subclasses of human IgG by the Fc region (*2*, *3*). Protein A is a multidomain, cell wall component of *Staphylococcus aureus* and contains five homologous domains of 58 amino acids which bind to most subclasses of human IgG by the Fc region (*4*). Both solution and crystal structures have been determined for G_B (*5*−*9*). In addition, crystal structures exist for the A_B−Fc complex (*10*) and the G_B−Fc complex (*11*).

A_B and G_B compete for binding at a site on the C_{H2}−C_{H3} interface of Fc (*12*, *13*). A_B is a three-helix bundle (Figure 1A) (*14*, *15*), whereas G_B is a four-stranded β-sheet spanned by an α-helix (Figure 1B) (*5*, *16*, *17*). The central α-helix occurs in about the same position in both domains. Amino acid identities occur at 5 of 15 positions in the central helix, but otherwise, neither sequence nor structural homology between the domains is discernible.

The A_B and G_B domains are at the lower limit in size for a stable unique protein structure and achieve their stability without disulfide bonds or tight ligand binding. In the most general features, the energetics of folding are similar to those of globular proteins much larger in size (*18*−*21*). Their small size facilitates the application of numerous biophysical techniques in studying folding and stability, in particular, NMR methods.

We will describe the directed evolution by phage display of a pair of highly homologous (59% identical) but heteromorphic proteins: one with the 3-α A_B tertiary structure and one with the α/β G_B fold. The homologous heteromorphs have been analyzed by differential scanning calorimetry, titration calorimetry, and circular dichroism to determine the thermodynamics of the folding and IgG binding reactions. The accompanying paper describes the high-resolution structures of the two proteins determined by NMR.

METHODS

Library Construction. A library for screening the largely surface exposed amino acids in island 1 of the G_B structure (residues 2, 4, 6, 8, 19, 42, 44, 46, 51, 53, and 55) was synthesized using the polymerase chain reaction (PCR). Oligonucleotides encompassing amino acids 9−41 served as a template, allowing a 30 bp overlap for annealing the 5′ and 3′ oligonucleotides which encoded unique PstI and NotI restriction sites, respectively, for cloning into the phagemid pHEN vector. A 50 μL PCR mixture containing 0.5 pmol of template oligonucleotides, 50 pmol of 5′ and 3′ oligonucleotides, 200 μM dNTPs, 1× Vent DNA polymerase buffer (New England Biolabs) supplemented with additional 2 mM $Mg(SO_4)_2$, 0.1 mg/mL BSA (New England Biolabs), and 1.5 units of Vent DNA polymerase (NEB) was run for 30 cycles, each cycle consisting of a 30 s 94 °C melt, a 30 s 63 °C annealing, and a 20 s 72 °C extension step.

Oligonucleotide sequences were as follows: template oligo P131, 5′ggtaaaacattgttttatgaaacactgactaWggctgttgatgctgaaaccgcagaaaaagcgttcaaacaatacgctaacgacaacggtgttgacggt3′; template oligo P132, 5′accgtcaacaccgttgtcgttagcgtattgtttgaacgctttttctgcggtttcagcatcaacagccWtagtcagtgtttcataaaacaatgttttacc3′; 5′ oligo P126, 5′aatggactgcagatgNHttacNHttaNHtcttRaWggtaaaacattgttttatgaaacactgact3′ (PstI underlined); and 3′ oligo P127, 5′ttcacgagtccttgcggccgcttccKKaacgKYaaagtagRYccacRcaccgtcaacaccgttgtcgttagcgtattg3′ (NotI underlined). The ambiguity code is as follows: R = ag, W = at, Y = ct, K = gt, H = act, N = acgt, M = ac, and S = cg (in equimolar mixes). This degenerate library encoded approximately 1.4×10^7 possible protein sequences.

A similar strategy was employed to synthesize a second library, which incorporated the screened and selected residues from the previous surface library. The residues screened in the second library, (island 2, residues 5, 7, 9−12, 33, 34, 39, 41, 54, and 56) are mainly amino acids that are interacting partners in the tertiary structure of the native G_B domain but, again, should not eliminate IgG binding. A 50 μL PCR cocktail containing 0.5 pmol of template oligonucleotide, 50 pmol of 5′ and 3′ oligonucleotides, 200 μM dNTPs, 1× Vent DNA polymerase buffer (NEB), 0.1 mg/mL BSA, and 1.5

[1] Abbreviations: aa, amino acid; CD, circular dichroism; C_{ex}, excess specific heat as measured by calorimetry; ΔH_{cal}, calorimetric enthalpy for unfolding; ΔH_{vH}, van't Hoff enthalpy for unfolding; DSC, differential scanning calorimetry; EDTA, disodium salt of ethylenediaminetetraacetic acid; Fab, variable region of immunoglobulin; Fc, constant region of immunoglobulin; IgG, immunoglobulin G; IPTG, isopropyl β-D-thiogalactopyranoside; ITC, isothermal titration calorimetry; NMR, nuclear magnetic resonance; [P], protein concentration; PCR, polymerase chain reaction; PMSF, phenylmethanesulfonyl fluoride; *T*, temperature; T_m, temperature of thermal melting; Tris, tris(hydroxymethyl)aminomethane.

units of Vent DNA polymerase was run for 30 cycles, each cycle consisting of a 30 s 94 °C melt, a 30 s 80 °C annealing, and a 20 s 72 °C extension step. Oligonucleotide sequences were as follows: template oligo P138, 5′gatgctgaaaccgcagaaaaagcgttcaaacaatMtSYtaacgacaacggtKYtgacRgtgtgtggacctacgatgatgcgactaag3′; template oligo P139, 5′cttagtcgcatcatcgtaggtccacacacYgtcaRMaccgttgtcgttaRSaKattgtttgaacgctttttctgcggtttcagcatc3′; 5′ oligo P135, 5′aatggactgcagatgtattacctgNHtgtgNHtaaaSRaMagaMtSYtttttatgaaacactgactaaagctgttgatgctgaaaccgcagaaaaagcgttc3′ (PstI underlined); 3′ oligo P140, 5′tcacgtagtccttgcggccgctKcctgaRcggtaaaggtcttagtcgcatcatcgtaggtccacac3′ (NotI underlined). This degenerate library encoded approximately 2.4×10^6 possible protein sequences.

Cloning the PCR Library. The PCR products were purified on QIAquick (Qiagen) PCR spin columns according to the manufacturer's instructions, and digested with PstI and NotI using standard molecular biology techniques. After digestion, the library was again purified over spin columns, eliminating the short termini which may interfere with efficient ligation of the library. A typical ligation mixture of 200 μL with 1 μg of dephosphorylated PstI- and NotI-digested pHEN and a 5−10-fold molar excess of library DNA was incubated at 15 °C overnight, precipitated with ethanol, washed, dried in vacuo, and resuspended in 10 μL of sterile water. One microliter (100 ng) of ligation mixture was added to ten 40 μL electrocompetent TG1 (Stratagene) aliquots in 0.1 cm gapped cuvettes (Bio-Rad) that were electroporated at 1.8 kV with a 5 ms time constant. Cells were washed from the cuvette with 1 mL of SOC and placed into Falcon (2059) tubes and then outgrown at 37 °C for 1 h. The resulting 10 mL outgrown cells were diluted into 1 L of LB medium supplemented with 100 μg/mL ampicillin and 1% glucose. Samples of this culture were plated on LB 1% glucose plates with 100 μg/mL ampicillin to titer the transformation efficiency. The library was amplified by overnight growth at 37 °C on a shaking platform.

Phagemid Production. A portion (0.25 mL) of the amplified library was added to 25 mL of LB, 1% glucose, and 100 μg/mL ampicillin and grown to log phase at 37 °C. Dilution of 1 mL of the log culture into 25 mL of YT containing 100 μg/mL ampicillin and 10^8 pfu of M13K07 helper phage/mL resulted in a multiplicity of infection of less than 10. The culture was grown at 37 °C with vigorous shaking for 1.5 h when 70 μg/mL kanamycin was added. Growth continued overnight for less than 16 h.

Phagemid Purification. Cultures were centrifuged at 10000*g* to remove cells, and the supernatant containing phage was decanted and recentrifuged to remove any residual cells; 0.2 mL of 20% PEG 8000 and 2.5 M NaCl was added per milliliter of supernatant and incubated on ice for 20 min to precipitate phage particles. Centrifugation at 20000*g* pelleted phage, and the supernatant was carefully aspirated off. The phage pellet was resuspended in TE and reprecipitated twice as before. The resulting pellet was resuspended in TE and clarified by centrifugation at 13000*g*.

Titering Colony-Forming Units (cfu) *and Plaque-Forming Units (pfu).* One hundred microliters of stationary phase TG1 cells was infected with 1 μL of appropriate serial dilutions of the phagemid library at room temperature for 10 min and then plated on LB 1% glucose plates containing 100 μg/mL ampicillin. Plates were incubated at 37 °C overnight. Plaque-forming units were titered as described above except infected cells were overlaid onto YT plates with the addition of 3 mL of molten 0.7% YT top agar.

Purification and Biotinylation of Rabbit IgG. Normal rabbit serum was delipidated by centrifugation for 1 h at 100000*g*, and was purified by affinity chromatography on a protein G resin (gift from T. Lee). The delipidated serum was loaded and washed in PBS (pH 7), and the column washed to baseline with 5 mM $NH_4(CH_3COO^-)$ (pH 5) and eluted with 0.5 M $NH_4(CH_3COO^-)$ (pH 3). Pooled fractions were neutralized with NH_4OH, and dialyzed against PBS (pH 7). IgG (10 mg) in 5 mL of PBS was incubated for 1 h at room termperature (RT) with a 10-fold molar excess of biotin *N*-hydroxysuccinimide ester (Calbiochem). The reaction mixture was then dialyzed overnight at RT against 1 L of 0.15 M NaCl, with several changes of dialysis buffer to eliminate the unincorporated label. The labeled protein was then dialyzed into PBS (pH 7).

Bioselection of Phage Libraries. One milligram of streptavidin paramagnetic beads (Dynal Biotech) was blocked overnight at 4 °C in 50 mM Tris (pH 7.4), 150 mM NaCl, 0.5% Tween 20, and 0.1% BSA to minimize nonspecific binding. A mixture of 1 mL of 50 mM Tris (pH 7.4), 150 mM NaCl, and 0.5% Tween 20 (TBS Tween) containing 1.5 μg of biotinylated rabbit IgG (10 nM) and approximately 10^{11} cfu was incubated on a rocking platform at RT for 3 h. The blocked streptavidin beads were concentrated on a magnetic particle concentrator, and the supernatant was drawn off with a pipet. After an aliquot had been removed from the binding reaction mixture for titering, it was added to the blocked beads and incubated for 30 min at RT rocking. The beads were concentrated as before, and the supernatant was drawn off. Washing the beads thoroughly entailed resuspension in 1 mL of TBS Tween, rocking for 5 min on a platform at RT, concentration of beads, and removal of the supernatant. This was repeated seven times, each time in a fresh 1.5 mL tube. Elution of bound phage was achieved by resuspending the beads in 0.2 mL of 0.1 M glycine (pH 2) with 1 mg/mL BSA and incubation for 10 min at RT on a rocking platform, concentration of beads, and removal of the supernatant. The resulting phage supernatant was neutralized with 10 μL of 2 M Tris base.

Library Amplification. Eluted phage were added to 0.5 mL of stationary phase TG1 cells and incubated for 10 min at RT. This infected culture was used to inoculate 50 mL of LB containing 1% glucose and 100 μg/mL ampicillin and was shaken overnight at 250 rpm and 37 °C. Dilutions of the culture were plated on LB agar plates containing 1% glucose and 100 μg/mL ampicillin and incubated at 37 °C overnight for growth of individual colonies. Amplified cultures were also used to produce phagemid for subsequent bioselection as described above. Typically, three rounds of bioselection were conducted.

Plasmid Purification and Sequencing. Isolated colonies were grown in 4 mL of LB, 1% glucose, and 100 μg/mL ampicillin at 37 °C overnight. Plasmid was purified using Wizard SV plus minipreps (Promega) according to the manufacturer's instructions. Plasmids were sequenced using dye terminator sequencing.

Site-Directed Mutagenesis. Site-directed mutants were made in plasmids using QuikChange (Stratagene) mutagenesis kits according to the manufacturer's instructions.

Expression of Selected Sequences. A fusion protein vector, pG58, was constructed so that selected sequences could be expressed and purified. The creation plasmid pG58 from pG5 and proR9 has been described previously (*22*). This vector encodes an engineered subtilisin pro sequence as the N-terminus of the fusion protein. Expression is regulated by an IPTG inducible phage T7 promoter.

Minimal medium (*18*) was used for ^{15}N labeling, but for ^{15}N- and ^{13}C-labeled samples, the [^{13}C]glucose concentration was 4 g/L. Fresh G311 BL21 DE3 transformants were grown at 37 °C overnight in LB 100 μg/mL ampicillin; 500 mL of minimal medium was inoculated with 8 mL of the overnight culture and grown at 37 °C until an A_{600} of 1 was attained, upon which 1 mM IPTG was added. Induced cells were grown for 90 min at 37 °C, and cells were harvested by centrifugation. A219 cells were inoculated, grown, and induced similarly, but induction continued for 3 h.

Protein Purification. G311 cells were resuspended in 2 mL of 25% (w/v) sucrose, 50 mM Tris (pH 8), 1 mM EDTA, and 5 mM $MgCl_2$, and then 1 mg/mL lysozyme and 30 μg/mL DNase I were added. After incubation at 37 °C for 10 min, 2 volumes of lysis buffer containing 50 mM Tris (pH 8), 0.15 M NaCl, 1% (v/v) Triton X-100, and 1% (w/v) sodium deoxycholic acid was added. The solution was mixed by inversion and then frozen solid in a dry ice/acetone bath. After being thawed at RT, inclusion bodies were pelleted by centrifugation 10000*g*. The pellet was thoroughly washed by resuspension in approximately 15 mL of 50 mM Tris (pH 8), 0.15 M NaCl, 0.5% Triton X-100, and 1 mM EDTA, and inclusion bodies were centrifuged at 5000*g*. This was repeated until 250 mL of washing buffer was consumed. The inclusion body pellet was extracted into 1 mL of 0.1 M KP_i (pH 7.0) and 6 M GuHCl at RT. The soluble protein was diluted to 10 mL with 0.1 M KP_i (pH 7.0), yielding a final GuHCl concentration of ~0.6 M. This solution was clarified by centrifugation at 10000*g* before application to a fusion protein processing column.

A219 cells were resuspended in 5 mL of PBS (pH 7) containing 30 μg/mL DNase I, 50 μg/mL RNase A, 5 mM $MgCl_2$, and 1 mM PMSF. Lysozyme was added to a final concentration of 0.3 mg/mL, and the solution was incubated at 37 °C for 10 min and then placed on ice. The cooled solution was sonicated on ice to disrupt the cells. Debris was removed by centrifugation at 20000*g*, and the supernatant was carefully decanted and delipidated by centrifugation at 125000*g* for 1 h. The supernatant was drawn off and applied to the fusion protein processing column.

Soluble cell extract of the pro domain fusion protein was injected onto a 5 mL S189 column at 5 mL/min for 3 min to allow binding and then washed at 10 mL/min for 3 min to remove impurities (*22*). To cleave and elute the purified target protein, 6 mL of 100 mM KF and 100 mM KPO_4 (pH 7.2) was injected at a rate of 0.5 mL/min. The purified protein was then dialyzed into 2 mM ammonium bicarbonate buffer (pH 7.0) and lyophilized.

Circular Dichroism (CD). CD measurements were performed with a Jasco spectropolarimeter, model J-720, using water-jacketed quartz cells with path lengths of 1 cm, 1 mm, or 0.1 mm depending on protein concentrations which ranged from 5 to 170 μM. Temperature control was provided by a Neslab RTE-110 circulating water bath interfaced with a MTP-6 temperature programmer. Far-UV wavelength scans were recorded at 20 °C from 250 to 200 nm. The average for five CD spectra is presented. The ellipticity results were expressed as mean residue ellipticity, [θ], in degrees square centimeter per decimole.

Temperature-induced unfolding of the A219 and G311 cells was performed in between −5 and 90 °C in 1 cm, 1 mm, and 0.1 mm cuvettes. Ellipticities at 222 nm were continuously monitored at a scanning rate of 1 deg/min. The fraction native is determined by subtracting the unfolded baseline from the experimental CD signal and then dividing by the total CD difference between 100% folded and 0% folded at that temperature. Reversibility of the denaturations was confirmed by comparing the CD spectra at 20 °C before melting and after heating to 90 °C and cooling.

Differential Scanning Calorimetry (DSC). DSC measurements were taken on a VP-DSC Micro Calorimeter (MicroCal Inc., Northampton, MA). During each run, the cell temperatures were increased from 15 to 125 °C at a rate of 1 °C/min in 100 mM KPO_4 (pH 7.0). Scans were first run with the buffer in both the sample and reference cells. Then a quantity of G311, which had been dialyzed into the same buffer, was introduced into the sample cell, and at least two more scans were run. The G311 cell concentration was 200 μM as determined by UV spectra. In many cases, the second consecutive G311 melting curve showed a significantly weakened signature. Buffer effects were canceled by either subtracting the melting curves of known protein concentrations to determine the heat contribution of their molar difference or using the depleted curve produced by the second consecutive protein/buffer run to estimate and subtract out buffer effects.

Isothermal Titration Calorimetry (ITC). Binding constants were measured at 25 °C using a Microcal VP ITC apparatus. Protein solutions were dialyzed into 0.1 M KP_i (pH 7.2) and 0.1 mM EDTA in the same graduated cylinder, exactly matching the buffer to minimize heat effects upon titration. The reference cell was loaded with the dialysis buffer, and 5 μL injections of 6.9×10^{-4} M A219 spaced 210 s apart were injected at 310 rpm into 12.5 μM rabbit IgG loaded in the experimental cell. The differential power upon binding was measured until past the saturation point, and data were fit using Origin (Microcal).

RESULTS

Creation of Highly Homologous Proteins with Different Folds. We evolved the G_B protein so that it was more similar in sequence to A_B without changing its basic fold. To do this, we created a scheme to select for mutations in G_B which increase its level of identity to A_B with a minimal impact on its conformational stability. The selection is based on the fact that binding of G_B to IgG is thermodynamically linked to folding. That is, the correct presentation of contact amino acids depends on the correct tertiary structure (Figure 1B). If mutations are introduced in regions of G_B which do not directly contact IgG, their effects on binding will be linked to whether they preserve the native conformation (*23*, *24*). Therefore, G_B mutants which are fully folded in the unbound form generally will bind to IgG with the highest affinity. We previously have studied the relationship between stability and IgG binding using H−D exchange experiments on the IgG-bound and free forms of G_B (*25*).

The equilibrium for binding of G_B to IgG is

$$G_B(\text{unfolded}) \leftrightarrow G_B + \text{IgG} \leftrightarrow G_B{-}\text{IgG}$$

where K_f is the equilibrium constant for folding G_B and $K_{binding}$ is the association constant of folded G_B for IgG. The observed binding constant is then

$$K_{GB(folded+unfolded)} = [K_f/(1 + K_f)]K_{binding} \qquad (1)$$

As the fraction of folded protein approaches 1, the observed association constant will approach its maximum. Since the $K_{binding}$ of wild-type G_B for IgG is 10^8 M^{-1} (0.1 M KP_i, pH 7.0, and 25 °C), the $K_{binding}$ for a fully folded G_B mutant presumably will be $\leq 10^8$ M^{-1}. This thermodynamic linkage allows us to find mutants with a certain threshold of stability by using the monovalent phage display selection procedure (*23*, *26*, *27*).

The IgG-binding epitopes of A_B and G_B are known from X-ray crystallographic studies of the complexes (*10*, *11*), mutagenesis studies (*28*, *29*), and peptide mapping studies (*30*). The IgG-binding interface of A_B comprises amino acids F13, Y14, and L17 in helix 1 and N28, I31, Q32, and K35 in helix 2 (Figure 1A). The binding interface of G_B for IgG comprises amino acids E27, K28, K31, Q32, and N35 on the central helix, W43 in strand β3, and main chain contacts in the intervening turn (Figure 1B). Thus, positions 28, 31, 32, and 35 in the central α-helix are critical to binding of both A_B and G_B. Because of their different binding modes, however, different amino acids occur at three of the four positions in the central helix. We began the mutagenesis process by introducing the A_B binding residues F13, Y14, and L17 into G_B to create G_B-1. The solubility, stability, and IgG−Fc binding properties of G_B-1 are similar to those of G_B (100 mM NaP_i, pH 7.0). The mutations create half of the A_B binding epitope, albeit cryptically displayed, in the α/β fold. As will be described later, this will aid in the measurement of the propensity of mutants to switch from the α/β to the 3-α fold. Also relevant is the fact that these mutations increase the level of homology between G_B and A_B without affecting G_B stability, and they eliminate the secondary Fab binding interface present in G_B (*9*).

The most straightforward way to select for the G_B fold with the highest level of identity would be to simultaneously randomize amino acids at all positions which do not comprise the IgG binding epitope of G_B. This would involve randomizing 50 positions. Even restricting the randomization to binary sequence space would generate 10^{15} combinations, a number far exceeding the selection capability of phage display. Binary sequence space is defined as either of the two naturally occurring amino acids at each equivalent position of the protein pair. Since the vast majority of mutants are expected to be unstable, it is important to conduct a thorough search of the sequence space. A complete search of sequence space being impractical, we decided to limit the search to the most relevant sequence space by identifying "topological islands". An island is a subset of amino acid positions, which can be randomized in a manner that is independent of other islands. The designation of the topological island requires that the mutational choices in one island be largely independent of mutational choices on another island. For example, the first island chosen comprised

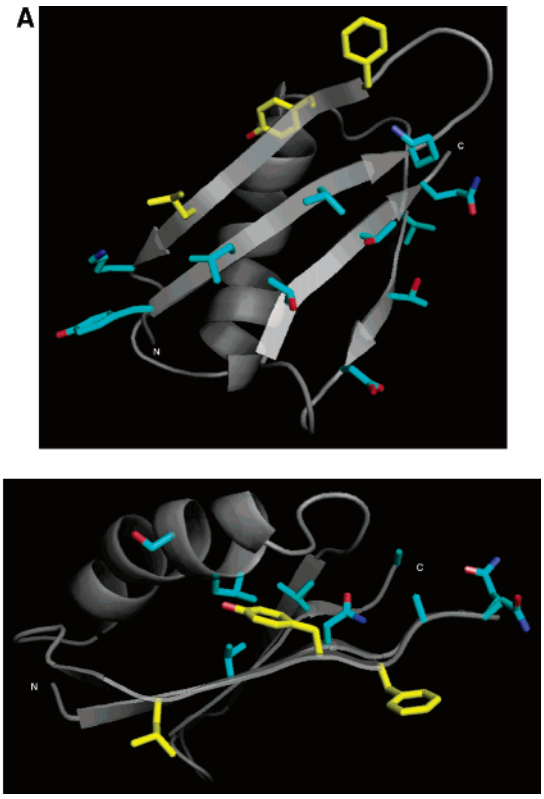

FIGURE 2: Regions of random mutagenesis of G_B. (A) Main depiction of the secondary structure of G311 with amino acids randomized in island 1 colored cyan. A_B amino acids F13, Y14, and L17 are colored yellow (from the NMR structure of G311 from the accompanying paper). (B) Main depiction of the secondary structure of G311 with amino acids randomized in island 2 colored cyan. A_B amino acids F13, Y14, and L17 are colored yellow (from the NMR structure of G311 from the accompanying paper).

all nonidentical residues on the solvent-exposed surface of the β-sheet (Figure 2A).

This included 11 amino acid positions in G_B-1 which were randomized to introduce A_B amino acids into the G_B-1 sequence. None of these positions directly alter the G_B binding epitope (Figure 2A). The randomized sites were positions 2, 4, 6, and 8 in the β1 strand, position 19 in β2, positions 42, 44, and 46 in β3, and positions 51, 53, and 55 in β4 (Table 1). Positions 2, 4, and 6 are not structured in A_B and were randomized with degenerate codons of the sequence N, [not G], T. Codons of this sequence encode 12 amino acids with equal probability and exclude amber and umber termination codons, glycine, cysteine, tryptophan, glutamine, and charged residues. Amino acids which are poor at forming β-sheet were avoided with this codon choice (*31*−*33*). At other positions, the strategy was to shuffle the sequence with either the A_B or G_B amino acid. Because of the nature of the amino acid code and expediency in gene synthesis using degenerate oligonucleotides, however, we chose to allow for four (rather than two) possible amino acids at positions 8, 44, 51, 53, and 55. This created a library with 1.4×10^7 possible combinations. The G_B-1 mutants in the library were synthesized as fusion proteins with the gp3 coat protein of M13 phage (*26*, *27*). The phagemid library comprised 7.5×10^6 independent clones. M13 phage particles tagged with G_B-1 mutants were selectively retained by binding to biotinylated IgG. The biotinylated IgG was in turn bound to streptavidin-coated magnetic beads which were

Table 1: Random Mutagenesis and Selection of Island 1[a]

	A/N/H/I/L/M F/P/S/T/Y/V			N/E K/D	K/M	V/A	T/V A/I	D/A	T/L M/P	T/D A/N	T/Q K/P
	aa 2	aa 4	aa 6	aa 8	aa 19	aa 42	aa 44	aa 46	aa 51	aa 53	aa 55
G_B	T	K	V	N	K	V	T	D	T	T	T
A_B	D	K	N	E	M	A	V	A	L	D	Q
round 2											
1025 (1)	I	V	H	E	M	V	A	A	L	T	P
1026 (3)	Y	L	Y	K	K	V	T	D	T	T	Q
1027 (1)	D	Y	I	K	K	V	T	D	T	T	T
1029 (3)	**Y**	**L**	**V**	**K**	**K**	**A**	**T**	**D**	**T**	**T**	**Q**
1033 (1)	T	H	I	K	K	A	T	D	T	T	T
1041 (1)	Y	D	A	D	M	V	A	D	L	T	Q
round 3											
1048 (1)	F	L	I	D	K	V	T	D	T	T	Q
1049 (1)	P	Y	I	K	M	V	V	D	T	T	T
1050 (15)	Y	L	Y	K	K	V	T	D	T	T	Q
1051 (4)	Y	L	I	K	K	V	T	D	T	T	T
1056 (1)	T	H	I	K	K	A	T	D	T	T	T
1069 (1)	Y	l	V	K	K	V	A	D	T	T	T
1072 (3)	Y	L	V	K	K	A	T	D	T	T	Q

[a] Sequences observed after two and three rounds of phagemid selection for IgG binding. The number of occurrences of a sequence is shown in parentheses. Clone 1029 (bold) was selected for further mutagenesis.

collected on a magnetic particle concentrator as described in Methods. Twenty-four of the mutant genes from selected phage were sequenced after each of three rounds of selection.

After three rounds of selection, only six different amino acid sequences were observed among the 24 clones that were sequenced (Table 1). Five of these proteins from the third round (1049, 1050, 1051, 1069, and 1072) and one from the second round (1029) were purified and analyzed using circular dichroism (CD). The CD spectra of all six are very similar to that of G_B. The thermal denaturation profile of each of the six was measured by CD in 100 mM NaP_i at pH 7.1.

Because of the difficulties in extrapolating folding free energies to lower temperatures without detailed thermodynamic characterization (described below), we chose to make a preliminary comparison of the stability of mutants at 75 °C, a temperature at which the equilibrium constant could be determined from CD data for all mutants. The $\Delta G_{unfolding}$ of G_B is −1.1 kcal/mol at 75 °C. One of the mutants (1050) is similar in stability to G_B. The other four have higher T_m's but aggregate at some point during the heating cycle, complicating an assessment of reversibility. Thermal denaturation experiments were repeated at pH 2.0 to increase the solubility of the unfolded state. Mutant 1029 unfolded reversibly under these conditions and was more stable than G_B. These positions suggest that new amino acids were selected which form favorable interactions with the A_B amino acids F13, Y14, and L17 that were introduced in step 1 (Figure 2A).

Mutant 1029 was selected to proceed into the next round of random mutagenesis. The second topological island comprised 12 positions: 5, 7, 9−12, 33, 34, 38, 41, 54, and 56 (Figure 2B). These amino acids occur at interior positions between β1 and β2, the C-terminal end of the α-helix, and β4 (Figure 2B). Positions 5 and 7 were randomized with degenerate codons of the sequence N, [not G], T encoding 12 amino acids as described above. Other positions were randomized as specified in Table 2, creating a library with 2.4×10^6 possible combinations. A phagemid library of 3×10^6 independent clones was screened. Mutants sequenced from the second and third rounds of selection are listed in Table 2. Mutant 2024 was chosen for further analysis because it was highly expressed in *E. coli* in soluble form and because of its homology to A_B. Mutant 2024 melts at a temperature ∼20 °C lower than G_B, but has a highly reversible unfolding reaction at pH 7.0 (0.1 M KP_i). It was labeled with ^{15}N and purified using IgG-Sepharose. It exhibited a well-dispersed HSQC spectrum.

Randomization of a third topological island (3, 4, 6, 16, 37, 46, 49, and 51) was carried out but did not identify any IgG-binding mutants with an increased level of identity to A_B. This result may indicate that subsequent mutations decrease $\Delta G_{unfolding}$ significantly below the threshold of 2 kcal/mol and precipitously decrease the level of IgG binding.

Having reached an apparent end point in the phage selection process, we sought to create additional identities between heteromorphs by making the following mutations in A_B, amino acids 2−8 (DNKFNKE) replaced with YYLVVNK and G29A. The resulting mutant is denoted A219. Because the N-terminal eight amino acids of A_B are disordered, we expected that A219 would be similar in stability to A_B. As described below, however, the N-terminal substitutions resulted in a loss of ∼3 kcal/mol in the free energy of unfolding. The G29A mutation had little effect on stability.

At this point, the remainder of the A_B IgG binding epitope was introduced into mutant 2024 by making the following mutations: K28N, K31I and N35K, N37D, and D40Q. These mutations destroy the G_B binding epitope in the α/β fold but, together with the K13F, G14Y, and T17L mutations introduced earlier, encode a complete A_B binding epitope in cryptic form. The resulting mutant (G311) has 27 residues that are identical with those of A_B and is similar in stability to 2024. A219 and G311 (59% identical) were then analyzed by CD melting and differential scanning calorimetry to determine the energetics of folding. ITC was use to determine the energetics of IgG binding. A flowchart of the mutagenesis and selection process is shown in Figure 3.

Thermodynamic Analysis of the Temperature Unfolding Profile. Circular dichroism spectra indicate that both A219

Table 2: Random Mutagenesis and Selection of Island 2[a]

	A/N/H/I/L/MF/P/S/T/Y/V		G/Q R/E	K/Q	T/N	L/A V/P	Y/S	A/L V/D	V/S A/F/	G/S	V/A	A/E
	aa 5	aa 7	aa 9	aa 10	aa 11	aa 12	aa 33	aa 34	aa 39	aa 41	aa 54	aa 56
G_B	L	L	G	K	T	L	Y	A	V	G	V	E
A_B	F	K	Q	Q	N	A	S	L	S	S	A	A
round 2												
2011	L	V	E	K	N	V	Y	A	F	G	V	A
2012	N	D	G	K	N	V	Y	L	V	G	V	A
2016	H	D	G	K	N	V	Y	A	V	G	V	A
2017	T	V	G	K	N	L	Y	A	V	G	V	A
2018	V	I	E	K	T	V	Y	A	F	G	V	A
2020	T	T	G	K	T	L	Y	A	F	G	V	A
2021	I	D	G	K	N	V	Y	A	V	G	V	A
2022	V	Y	R	K	T	P	Y	A	V	G	V	A
2024	**V**	**N**	**G**	**Q**	**N**	**A**	**S**	**L**	**V**	**G**	**V**	**A**
round 3												
2025	D	Y	G	Q	T	A	Y	L	V	G	V	A
2026	V	S	G	K	T	L	Y	A	V	G	V	E
2027	L	N	G	Q	N	A	Y	A	V	G	V	A
2028	N	V	G	K	T	A	Y	L	V	G	V	A
2030	V	D	G	K	N	V	Y	L	V	G	V	A
2033	V	A	G	Q	T	L	Y	A	F	G	V	E
2034	A	I	G	K	T	A	Y	A	V	G	V	A
2035	V	N	G	H	T	A	Y	A	V	G	V	A
2038	A	Y	G	K	T	V	Y	A	V	G	V	A

[a] Sequences after two and three rounds of selection for IgG binding. The number of occurrences of a sequence is shown in parentheses. Clone 2024 (bold) was selected for further mutagenesis.

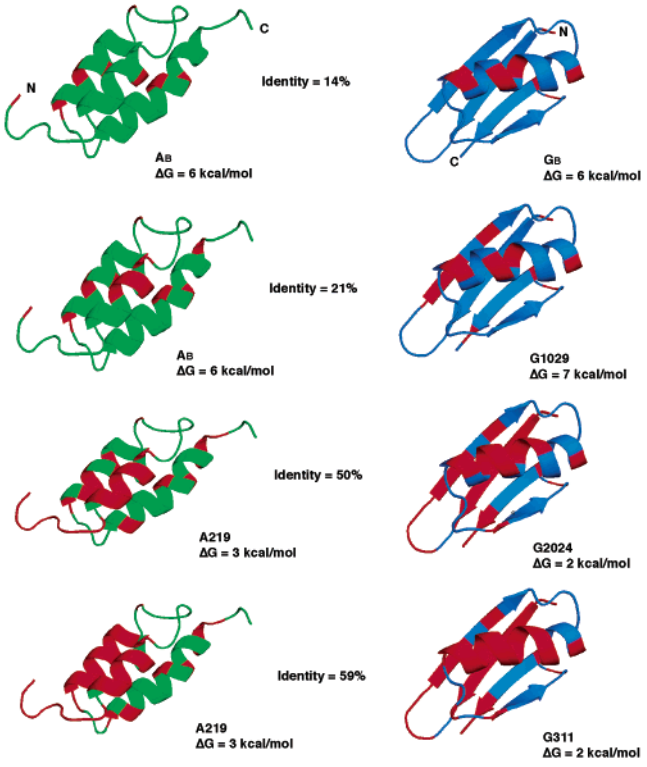

FIGURE 3: Co-evolution of homologous heteromorphs. Regions of identity between heteromorphic pairs are colored red.

and G311 are predominantly folded at 20 °C (Figure 4A). The shape and intensity of the spectral profiles are consistent with their expected secondary structure. This was subsequently confirmed by high-resolution structure determination (Y. He et al., accompanying paper). To characterize the stability of the proteins, thermal denaturation was studied by both CD and DSC (Figures 4B and 5A). The use of both techniques allowed us to perform unfolding experiments over a wide range of protein concentrations (4−200 μM). The observed invariance in the thermal denaturation profile over this concentration range shows that intermolecular associations in neither the folded nor unfolded states affected the energetics of unfolding. The monomeric states of G311 and A219 were also confirmed by gel filtration (data not shown).

The melting temperature of G311 is 56 °C and that of A219 55 °C, as determined by CD melting. These values are in good agreement with T_m values determined by DSC. The unfolding reactions for both proteins were completely reversible after scanning to 90 °C.

The unfolding reaction of G311 determined by DSC followed a two-state model for unfolding, as expressed in the van't Hoff equation, $d(\ln K)/dT = \Delta H_{vH}/(RT^2)$, with ΔH_{vH}, the van't Hoff enthalpy, or apparent enthalpy, equal to the calorimetric, ΔH_{cal}, or true enthalpy ($\Delta H_{329} = 48.8$ kcal/mol, $\Delta S_{329} = 0.144$ kcal deg^{-1} mol^{-1}, Figure 5A).

The equilibrium constant for folding can only be determined directly at temperatures close to the T_m, where the fraction of both folded and unfolded forms can be accurately measured. The $\Delta G_{unfolding}$ can be extrapolated to temperatures beyond the experimentally observable range, however, using the Gibbs−Helmholtz equation:

$$\Delta G = \Delta H_o - T\Delta S_o + \Delta C_p(T - T_o - T \ln T/T_o) \quad (2)$$

where ΔH_o and ΔS_o are the enthalpy and entropy of unfolding evaluated at a reference temperature T_o, respectively, and ΔG is the free energy of unfolding at a temperature T (*34*−*38*). By fitting the G311 melting data to eq 2, we estimated its ΔC_p to be 0.8 kcal deg^{-1} mol^{-1} from the curvature in the free energy profile (Figure 5B). This is within experimental error of the ΔC_p determined previously for G_B unfolding (0.7 kcal deg^{-1} mol^{-1}) (*18*). As a result, the shape of the ΔG versus temperature functions for G_B and G311 are similar, with the G311 reaching a maximum ΔG of ∼2 kcal/mol.

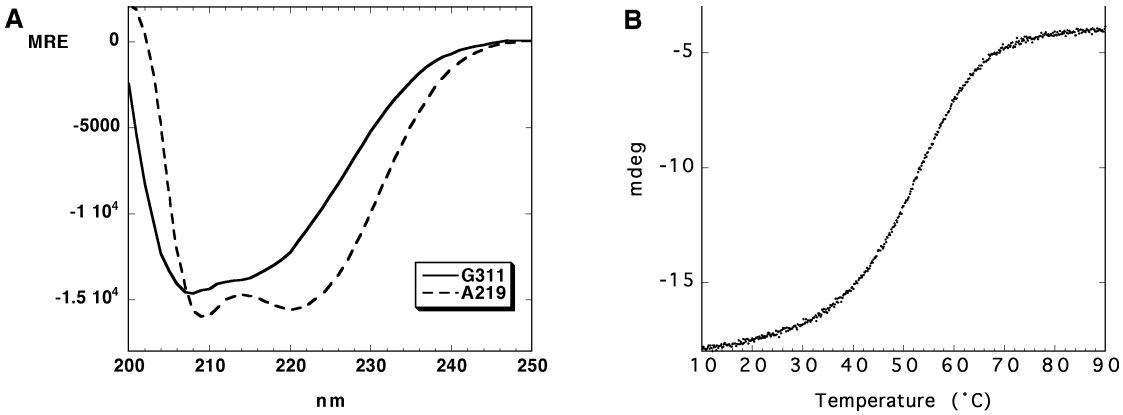

FIGURE 4: Circular dichroism of A219 and G311. (A) The mean residue ellipticity (degrees square centimeter per decimole) is plotted vs wavelength for A219 and G311. Spectra were measured in 0.1 M potassium phosphate buffer (pH 7.0) using a 1 mm cylindrical cuvette at 25 °C with 100 μM protein. (B) Millidegrees at 222 nm for G311 is plotted vs temperature in the range from 10 to 90 °C. The temperature profile was recorded using a 1 cm cylindrical cuvette with a protein concentration of 26 μM in 0.1 M potassium phosphate buffer (pH 7.0).

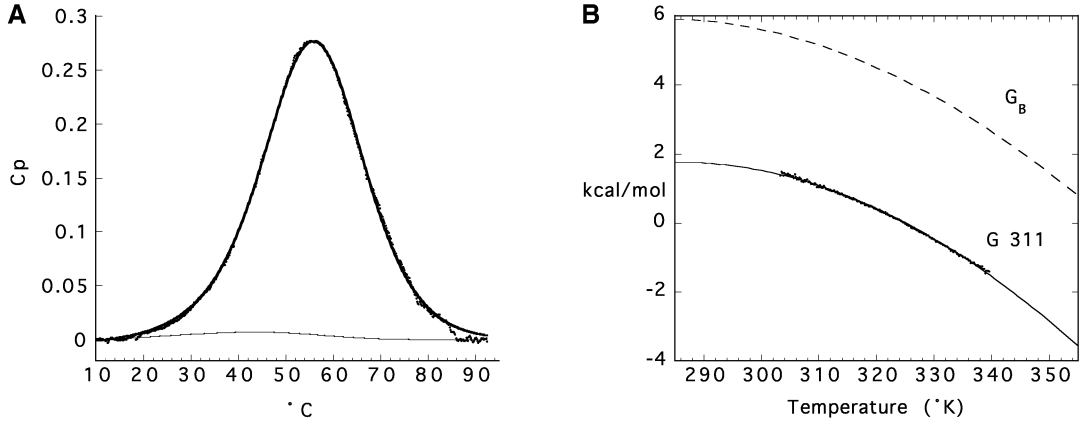

FIGURE 5: Thermodynamics of G311 and G_B unfolding. (A) Calorimetric data and a theoretical two-state fit are shown for G311 (T_m = 55.7 °C). The experimentally measured excess heat capacity was fit to a computer-derived theoretical curve that simulates a two-state unfolding process not involving association or dissociation (*44*). Excess heat capacity is in units of millicalories per degree Celsius. The calorimeter ampule contained 1.6 mg of protein. (B) The temperature unfolding profile measured by far-UV CD for G311 and G_B (*18*) was converted to an apparent $\Delta G_{unfolding}$ (points). The solid line is a theoretical curve calculated using the Gibbs−Helmholtz equation: $\Delta G_{unfolding}$ for G311 = $\Delta H_o - T\Delta S_o + \Delta C_p(T - T_o - T \ln T/T_o)$, where T_o = 329 K, ΔH_o = 48.8 kcal/mol, ΔS_o = 0.144 kcal deg^{-1} mol^{-1}, and ΔC_p = 0.83 kcal deg^{-1} mol^{-1}.

Similar analysis carried out for A219 yielded a $\Delta G_{unfolding}$ of ~3 kcal/mol at 20 °C (data not shown).

ITC Results. Titration experiments of IgG were carried out with A219 and G311 as explained in Methods. The data points correspond to the negative heat of binding associated with each titration (Figure 6). The titration calorimeter is sensitive to changes in $K_{binding}$ under conditions at which the product of $K_{binding}$[P] is between 1 and 1000 (*39*). The IgG concentration was 12.5 μM, which corresponds to 25 μM binding sites. A219 bound IgG with a $K_{binding}$ of 5 × 10^6 M^{-1} and a $\Delta H_{binding}$ of −30.7 kcal/mol (25 °C). The stoichiometry of binding was two A219 molecules per molecule of IgG. The binding reaction is enthalpically driven, with a loss of entropy upon binding ($\Delta S_{binding}$ = −0.103 kcal deg^{-1} mol^{-1}).

As expected, no heat was detected upon titration of IgG with G311. The lack of observable binding reflected the disruption of the Fc binding epitope by the K28N, K31I and N35K, N37D, and D40Q mutations made in the transformation of 2024 into G311. We also note that the weak binding to the Fab portion of IgG observed for G_B is not seen. This was eliminated by the K13F, G14Y, and T17L mutations.

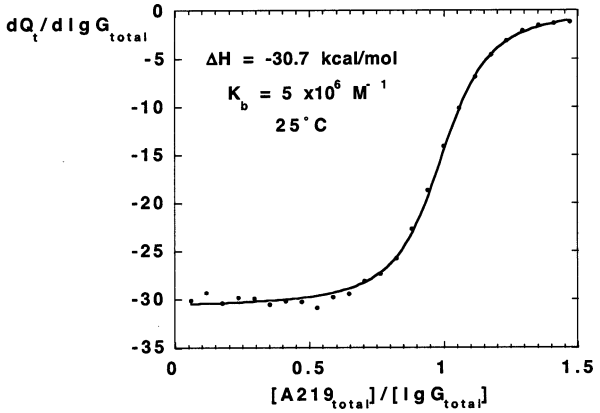

FIGURE 6: Titration calorimetry of A219 and IgG. The heats of binding for successive additions of A219 are plotted vs the ratio of [A219]/[IgG (binding sites)]. The molar concentration of IgG is given in terms of binding sites (two per molecule). Binding curves were determined by nonlinear least-squares minimization using eq 3 in the text (*39*). The temperature was 25 °C.

DISCUSSION

In this paper, we have been able to evolve a pair of monomeric proteins which are almost 60% identical but have

different folds. In our directed evolution system, retention of binding function is determined by the fraction folded. A $\Delta G_{unfolding}$ of >2 kcal/mol gives essentially full binding activity.

While other sets of homologous heteromorphic proteins have been created by design, this is the first example of which we are aware where both heteromorphs are monomeric (*40*−*43*). Mutations which increase the level of identity between heteromorphic proteins frequently expose the hydrophobic surface to the solvent and can result in dimerization and/or higher-order associations. The stability of the alternative conformation can be quite low in the monomeric state, but the energy of oligomer formation can be sufficient to drive the protein to the alternative conformation in a concentration-dependent reaction. In the case described here, both folds are stable independent of the binding energy contributed from the formation of multimeric complexes. By linking selection with IgG binding function, we were able to avoid the energetic sink of oligomerization and preserve the stability of the monomer.

A219 has mutations at 9 of 56 positions which collectively decrease the free energy of folding by 2 kcal/mol relative to that of the wild type (an average of 0.22 kcal/mol per mutation). G311 has mutations at 21 of 56 positions which collectively decrease the free energy of folding by 3.5 kcal/mol relative to that of wild-type G_B (an average of 0.17 kcal/mol per mutation). Most remaining mutations will decrease the free energy of unfolding by more than this average and result in a large change in the fraction of folded molecules. Thus, the essential folding information must reside within the 23 nonidentical residues.

Perhaps the more interesting problem, however, is defining the free energy of folding of the G311 sequence into the 3-α fold. The energy gap between the two alternative folds for G311 can be defined in terms of a three-state equilibrium among 3-α, α/*β*, and unfolded forms. The unfolded state (U) is defined as all conformations expect α/*β* and 3-α. The energy gap between the two conformations ($\Delta G_{\alpha/\beta,3-\alpha}$) is equal to $-RT \ln(K_{3-\alpha}/K_{\alpha/\beta})$, where $K_{3-\alpha} = [3\text{-}\alpha]/[U + \alpha/\beta]$ and $K_{\alpha/\beta} = [\alpha/\beta]/[U + 3\text{-}\alpha]$. We have determined that $K_{\alpha/\beta} = 65$ (0.1 M KP_i, pH 7.0, and 20 °C). We cannot directly measure $K_{3-\alpha}$ for G311 because it is so small. Because the entire A_B binding epitope is encoded cryptically within G311, however, we can determine the upper limit of $K_{3-\alpha}$ based on the linkage between folding and IgG binding. Assuming that the only way G311 can bind to IgG is to adopt the 3-α fold, then its binding constant will be

$$K_{bind-obs} = K_{bind}K_{3-\alpha}/(1 + K_{3-\alpha}) \quad (3)$$

where $K_{binding}$ is the binding affinity of the fully folded 3-α structure (5×10^6 M^{-1}). The cryptic site thus provides a means of measuring the propensity of G311 to adopt the 3-α fold. We detect no IgG binding of G311 by ITC. The threshold of detection in this experiment is $\sim 10^4$ M^{-1}. According to eq 3, we should be able to detect binding of IgG to a sequence which has a $K_{3-\alpha}$ of >0.002. On the basis of this, $\Delta G_{\alpha/\beta,3-\alpha}$ > 6 kcal/mol. This is the minimum energy required to force G311 into the 3-α fold. The remaining 23 nonidentical residues must create this energy gap. The energetic contribution of a mutation at a specific site toward stabilizing the 3-α fold is defined as $\Delta\Delta G_{3-\alpha} = -RT \ln(K_{3-\alpha}^{mut}/K_{3-\alpha}^{wt})$. The total information content of the mutation involves its effects on both 3-α and α/*β* conformations, however, and would be equal to $\Delta\Delta G_{3-\alpha} + \Delta\Delta G_{\alpha/\beta}$. The reason that certain nonidentical sites are preserved between homologous heteromorphs is because of high information content. High information content likely results from synergistic interactions among nonidentical sites. The 23 remaining nonidentities contain most of the information content.

Future analyses will try to identify small sets of highly synergistic mutations which may contribute several kilocalories per mole to spanning the energy gap. The object will be to determine the maximum amount of folding information which can be encoded with the fewest mutations. The binary sequence space separating A219 and G311 ($2^{23} \sim 8 \times 10^6$ sequences) has decreased by almost 8 orders of magnitude relative to those of A_B and G_B. This allows an exhaustive search of the information-enriched sequence space in a single phagemid library. We have preliminary evidence based on recent phage selections that significant IgG binding affinity can be generated in G311 with as few as five synergistic mutations. This result indicates that evolution of new folds, driven by selection for binding, can occur with few mutations.

ACKNOWLEDGMENT

We thank Fenhong Song for synthesizing the oligonucleotides used in site-directed mutagenesis and DNA sequencing and Kathryn Fisher and Biao Ruan for advice and helpful discussion.

REFERENCES

1. Lattman, E. E., and Rose, G. D. (1993) Protein folding: What is the question? *Proc. Natl. Acad. Sci. U.S.A. 90*, 439−441.
2. 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.
3. Reis, K. J., Ayoub, E. M., and Boyle, M. D. P. (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.
4. Åkerstrom, B., and Bjorck, L. (1986) A physicochemical study of protein G, a molecule with unique immunoglobulin G-binding properties, *J. Biol. Chem. 261*, 10240−10247.
5. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., and Clore, M. G. (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G, *Science 253*, 657−661.
6. Achari, A., Hale, S., Howard, A. J., Clore, G. M., Gronenborn, A. M., Hardman, K. D., and Whitlow, M. (1992) 1.67-.ANG. X-ray structure of the *β*2 immunoglobulin-binding domain of streptococcal protein G and comparison to the NMR structure of the *β*1 domain, *Biochemistry 31*, 10449−10457.
7. Gallagher, T. D., Alexander, P., Bryan, P., and Gilliland, G. (1994) Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR, *Biochemistry 33*, 4721−4729.
8. Lian, L.-Y., Derrick, J. P., Sutcliffe, M. J., Yang, J. C., and Roberts, G. C. K. (1992) Determination of the solution structures of domains II and III of protein G from *Streptococcus* by ^{1}H nuclear magnetic resonance, *J. Mol. Biol. 228*, 1219−1234.
9. Derrick, J. P., and Wigley, D. B. (1994) The third IgG-binding domain from streptococcal protein G: An analysis by X-ray crystallography of the structure alone and in a complex with Fab, *J. Mol. Biol. 243*, 906−918.
10. Deisenhofer, J. (1981) Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment

B of protein A from *Staphylococcus aureus* at 2.9- and 2.8-Å resolution, *Biochemistry 20*, 2361−2370.
11. Sauer-Eriksson, A. E., Keywegt, G. J., Uhlen, 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.
12. Wright, C., Willan, K., Sjödahl, J., Burton, D. R., and Dwek, R. (1977) The interaction of protein A and Fc fragment of rabbit immunoglobulin G as probed by complement-fixation and nuclear magnetic resonance studies, *Biochem. J. 167*, 661−668.
13. Stone, G. S., Sjöbring, U., Björck, L., Sjöquist, J., Barber, C. V., and Nardella, F. A. (1989) The Fc binding site for streptococcal protein G is in the Cγ2−Cγ3 interface region of IgG and is related to the sites that bind staphylococcal protein A and human rheumatoid factors, *J. Immunol. 143*, 565−570.
14. Gouda, H., Torigoe, H., Saito, A., Sato, M., Arata, Y., and Shimada, I. (1992) Three-dimensional solution structure of the B domain of staphylococcal protein A: Comparisons of the solution and crystal structures, *Biochemistry 31*, 9665−9672.
15. Tashiro, M., Tejero, R., Zimmerman, D. E., Celda, B., Nilsson, B., and Montelione, G. T. (1997) High-resolution solution NMR structure of the Z domain of staphylococcal protein A, *J. Mol. Biol. 272*, 573−590.
16. Orban, J., Alexander, P., and Bryan, P. (1992) Sequence-specific ^{1}H NMR assignments and secondary structure of the streptococcal protein G B2-domain, *Biochemistry 31*, 3604−3611.
17. Lian, L. Y., Yang, J. C., Derrick, J. P., Sutcliffe, M. J., Roberts, G. C. K., Murphy, J. P., Goward, C. R., and Atkinson, T. (1991) Sequential ^{1}H NMR assignments and secondary structure of an IgG-binding domain from Protein G, *Biochemistry 30*, 5335−5340.
18. 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.
19. Bottomley, S. P., Popplewell, A. G., Scawen, M., Wan, T., Sutton, B. J., and Gore, M. G. (1994) The stability and unfolding of an IgG binding protein based upon the B domain of protein A from *Staphylococcus aureus* probed by tryptophan substitution and fluorescence spectroscopy, *Protein Eng. 7*, 1463−1470.
20. Bai, Y., Karimi, A., Dyson, H. J., and Wright, P. E. (1997) Absence of a stable intermediate on the folding pathway of protein A, *Protein Sci. 6*, 1449−1457.
21. Johansson, M. U., de Chateau, M., Bjorck, L., Forsen, S., Drakenberg, T., and Wikstrom, M. (1995) The GA module, a mobile albumin-binding bacterial domain, adopts a three-helix-bundle structure, *FEBS Lett. 374*, 257−261.
22. 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.
23. Gu, H., Qian, Y., Bray, S. T., Riddle, D. S., Shiau, A. K., and Baker, D. (1995) A phage display system for studying the sequence determinants of protein folding, *Protein Sci. 4*, 1108−1117.
24. Ruan, B., Hoskins, J., Wang, L., and Bryan, P. N. (1998) Stabilizing the subtilisin BPN′ pro-domain by phage display selection: How restrictive is the amino acid code for maximum protein stability? *Protein Sci. 7*, 2345−2353.
25. Orban, J., Alexander, P., and Bryan, P. (1994) Hydrogen−deuterium exchange in the free and immunoglobulin G-bound Protein G B-domain, *Biochemistry 33*, 5702−5710.
26. Scott, J. K., and Smith, G. P. (1990) Searching for peptide ligands using an epitope library, *Science 249*, 386−390.
27. Smith, G. P., and Scott, J. K. (1993) Libraries of peptides displayed on filamentous phage, *Methods Enzymol. 217*, 228−257.
28. Cedergren, L., Andersson, R., Jansson, B., Uhlen, M., and Nilsson, B. (1993) Mutational analysis of the interaction between staphylococcal protein A and human IgG1, *Protein Eng. 6*, 441−448.
29. Sloan, D. J., and Hellinga, H. W. (1999) Dissection of the protein G B1 domain binding site for human IgG Fc fragment, *Protein Sci. 8*, 1643−1648.
30. Frick, I. M., Wikstrom, M., Forsen, S., Drakenberg, T., Gomi, H., Sjobring, U., and Bjorck, L. (1992) Convergent evolution among immunoglobulin G-binding bacterial proteins, *Proc. Natl. Acad. Sci. U.S.A. 89*, 8532−8536.
31. Kim, C. A., and Berg, J. M. (1993) Thermodynamic β-sheet propensities measured using a zinc-finger host peptide, *Nature 362*, 267−270.
32. Minor, D. L., Jr., and Kim, P. S. (1994) Measurement of the β-sheet-forming propensities of amino acids, *Nature 367*, 660−663.
33. Smith, C. K., Withka, J. M., and Regan, L. (1994) A thermodynamic scale for the β-sheet forming tendencies of the amino acids, *Biochemistry 33*, 5510−5517.
34. Brandts, J. F. (1964) Thermodynamics of protein denaturation. II. A model of reversible denaturation and interpretations regarding the stability of chymotrypsinogen, *J. Am. Chem. Soc. 86*, 4302−4314.
35. Pace, C. N., and Tanford, C. (1968) Thermodynamics of the unfolding of β-lactoglobulin A in aqueous urea solutions between 5 and 55°, *Biochemistry 7*, 198−208.
36. Privalov, P. L., and Khechinashvili, N. N. (1974) A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study, *J. Mol. Biol. 86*, 665−684.
37. Privalov, P. L. (1979) Stability of proteins, small globular proteins, *Adv. Protein Chem. 33*, 167−241.
38. Becktel, W. J., and Schellman, J. A. (1987) Protein stability curves, *Biopolymers 26*, 1859−1877.
39. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L.-N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter, *Anal. Biochem. 179*, 131−137.
40. Jones, D. T., Moody, C. M., Uppenbrink, J., Viles, J. H., Doyle, P. M., Harris, C. J., Pearl, L. H., Sadler, P. J., and Thornton, J. M. (1996) Towards meeting the Paracelsus Challenge: The design, synthesis, and characterization of paracelsin-43, an α-helical protein with over 50% sequence identity to an all-β protein, *Proteins 24*, 502−513.
41. 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.
42. Dalal, S., Balasubramanian, S., and Regan, L. (1997) Transmuting α helices and β sheets, *Folding Des. 2*, R71−R79.
43. Yuan, S. M., and Clarke, N. D. (1998) A hybrid sequence approach to the paracelsus challenge, *Proteins 30*, 136−143.
44. Schwarz, F. P., and Kirchhoff, W. H. (1988) Biological thermodynamic data for the calibration of differential scanning calorimeters: Heat capacity data on the unfolding transition of ribonuclease A in solution, *Thermochim. Acta 128*, 267−295.

BI051231R