1.67-Å X-ray Structure of the B2 Immunoglobulin-Binding Domain of Streptococcal Protein G and Comparison to the NMR Structure of the B1 Domain

Aniruddha Achari,[‡,§] Stephen P. Hale,[‡,∥] Andrew J. Howard,[‡] G. Marius Clore,[⊥] Angela M. Gronenborn,[⊥] Karl D. Hardman,[‡,ℱ] and Marc Whitlow[*,‡]

Protein Engineering Department, Enzon, Incorporated,[#] 16020 Industrial Drive, Gaithersburg, Maryland 20877, and Laboratory of Chemical Physics, Building 2, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

ABSTRACT: The structure of the B2 immunoglobulin-binding domain of streptococcal protein G has been determined at 1.67-Å resolution using a combination of single isomorphous replacement (SIR) phasing and manual fitting of the coordinates of the NMR structure of B1 domain of streptococcal protein G [Gronenborn, A. M., et al. (1991) *Science 253*, 657–661]. The final *R* value was 0.191 for data between 8.0 and 1.67 Å. The structure described here has 13 residues preceding the 57-residue Ig-binding domain and 13 additional residues following it, for a total of 83 residues. The 57-residue binding domain is well-determined in the structure, having an average *B* factor of 18.0. Only residues 8–77 could be located in the electron density maps, with the ends of the structure fading into disorder. Like the B1 domain, the B2 domain consists of four β-strands and a single helix lying diagonally across the β-sheet, with a −1, +3x, −1 topology. This small structure is extensively hydrogen-bonded and has a relatively large hydrophobic core. These structural observations may account for the exceptional stability of protein G. A comparison of the B2 domain X-ray structure and the B1 domain NMR structure showed minor differences in the turn between strands and two and a slight displacement of the helix relative to the sheet. Hydrogen bonds between crystallographically related molecules account for most of these differences.

Numerous infectious microorganisms have developed protective proteins that interact with immunoglobulins in order to reduce the effectiveness of the hosts' immune response. The most familiar of these proteins is staphylococcal protein A, which interacts with the Fc portion of mammalian IgG's, IgA's, and similar proteins with high affinity and fairly low specificity. Streptococcal protein G performs a similar function, but it binds only to IgG's (Fahnestock et al., 1986). As produced in the microorganism, protein G is a multidomain entity consisting of two or three Fc-binding domains, B1, B2, and B3, and a domain that binds to human serum albumin; these globular domains are connected by extended linkers (Fahnestock et al., 1990). Protein G offers several advantages over protein A for the binding of IgG immunoglobins. Protein G binds to all subclasses of human IgG, whereas protein A does not bind the IgG3 subclass (Reis et al., 1984). Protein G binds a number of animal IgG's to which protein A binds only weakly.

Protein G has a number of useful applications, including the treatment of diseases and the detection and purification of IgG's. Protein G could be used in the removal of IgG's from plasma, using an extracorporeal column (Tenan et al., 1981). The detection of IgG's is important in screening for hybridoma clones, in quantitative studies of immune response in animals, and in detection of antigens by competition-binding assay. The detection of IgG is more sensitive and less prone to interference than other immunoglobulin detection methods (Boyle et al., 1984). The simplicity and high yields of affinity chromatography make protein G the preferred method for purifying IgG's in our laboratory.

One of the more interesting features of the B domains of protein G is their unique topology. The B domains are topologically similar to ubiquitin (Kraulis, 1991; Gronenborn & Clore, 1991), but are more symmetrical. The symmetrical nature of this topology may contribute to the remarkable stability of this small domain. The single B1 and B2 domains of protein G have melting temperatures of 87.5 and 79.4 °C, respectively (Alexander et al., 1992), and the B1 domain remains native in 8 M urea (Gronenborn et al., 1991). This stability is particularly notable given that the 57-residue B domains of protein G lack disulfide bonds. The small size and high stability of protein G make it ideal for study by various methods including differential scanning calorimetry, site-directed mutagenesis, NMR,[1] and molecular mechanics.

In an effort to understand the properties of this protein, including its stability and its interaction with immunoglobulins, we have undertaken a structural study of the Ig-binding domains of protein G, B1, and B2. The form ("type 7") described here has 13 residues preceding the 57-residue B2 Ig-binding domain and 13 additional residues following it, for a total of 83 residues. We have recently described the NMR structure of the 56-residue B1 domain. The B1 and B2 Ig-binding domains differ at six positions: 6, 7, 19, 24, 29, and

---

*To whom reprint requests should be addressed; FAX (301) 926–1221.

[‡] Enzon, Inc.

[§] Current address: Wellcome Research Laboratories, Langley Court, South Eden Park Rd., Beckenham BR3 3BS, Kent, England.

[∥] Current address: Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742.

[⊥] National Institute of Diabetes and Digestive and Kidney Diseases.

[ℱ] Current address: Du Pont Merck Pharmaceutical Co., Experimental Station, P.O. Box 80228, Wilmington, DE 19880.

[#] Formerly Genex Corp.

```
                                       .       .      10        .
B1                             M T Y K L I L N G K T L K G E T T
B2     M D P G D A S E L T P A V T T Y K L V I N G K T L K G E T T
       .           .       10      .           20        .           30

       20            .         30            .          40          .
B1     T E A V D A A T A E K V F K Q Y A N D N G V D G E W T Y D D
B2     T K A V D A E T A E K A F K Q Y A N D N G V D G V W T Y D D
                 .            40             .           50        .           60

       50          .56
B1     A T K T F T V T E
B2     A T K T F T V T E M V T E V P V A S K R K E D
            .           70        .           80    83
```
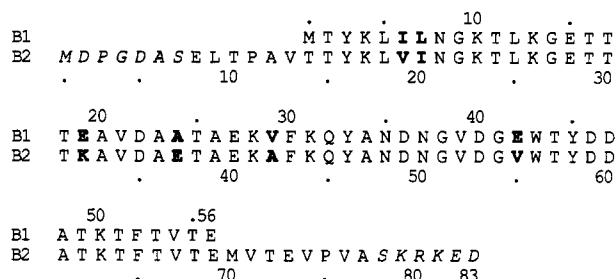
FIGURE 1: Primary sequences of protein G of the B1 (type 19) and B2 (type 7) domains. Differences are highlighted in bold-face type. Residues omitted from the X-ray crystallographic model are italicized.

42 (see Figure 1). We will describe the differences between the X-ray structure of the B2 domain and the NMR structure of the B1 domain. Finally, we will examine the structure for clues to protein G's binding to immunoglobins, including a comparison with the structure of the protein A–Fc fragment complex.

MATERIALS AND METHODS

Construction of the B2 Domain of Protein G. The cloning of the Lancefield group G species of streptococcal protein G gene (Fahnestock, et al., 1986) and the construction of the type 7 protein *Escherichia coli* expression plasmid (pGX5244) containing the B2 domain with 13-residue N- and C-terminal extensions have already been described (Fahnestock, 1990).

Expression of the B2 Domain of Protein G. The expression vector used for B2 protein expression in *E. coli* contains the hybrid λ-phage promoter O_L/P_R (Scandella et al., 1985). Plasmid pGX5244 was transformed into *E. coli* host GX6705, which has the mutant repressor gene cI^{857} integrated into the chromosome (Gottesman et al., 1980), to give the final expression strain GX8821. This expression vector/host strain combination provides a transcriptional regulation system where synthesis of the B2 protein is repressed at 30 °C culture growth and induction of the B2 protein occurs by raising the culture temperature to 42 °C.

Fermentation of the B2 Domain of Protein G. One 10-L fermentation was run on a casein digest–glucose–salts medium at 30 °C. At an optical density of 18–20 (600 nm), the B2 protein expression was induced by a 42 °C temperature shock for 1 h. The fermentation was continued for 3 h at 39 °C, to allow the B2 protein to accumulate. The B2 protein was expressed as a soluble protein at over 20% of the total cell protein. The fermentation was centrifuged using a Sorval RB-5 at 7000g at 10 °C for 15 min, and the supernatant was discarded. The cell paste was frozen at –20 °C.

Purification of the B2 Domain of Protein G. A 250-g sample of cell paste was suspended in 50 mM KH_2PO_4–50 mM NaCl (pH 7.4; PBS buffer) at ratio of 1.0 g of cell paste/5 mL of buffer. The cell suspension was heated in a water bath to 80 °C and maintained for 5 min. Immediately, the cell suspension was plunged into an ice–water bath until the temperature was below 10 °C. The suspension from the heat-released cells was centrifuged at 20000g at 10 °C for 30 min. The B2 protein-containing supernatant was kept between 4 and 7 °C until loaded onto an IgG–Sepharose 6B Fast Flow column. The IgG–Sepharose 6B column was equilibrated with five column volumes of PBS buffer, at a flow rate of 2.5 mL/min, prior to loading the supernatant. The supernatant from the heat-released cells was loaded onto the column at a flow rate of 2.0 mL/min. The column was washed with 15 column volumes of the PBS buffer at a flow rate of 5.0 mL/min. A second wash was done with 5 mM ammonium acetate (pH

5.2) until the pH of the effluent was 5.2, about 8–10 column volumes. The B2 protein was eluted with 0.5 M ammonium acetate (pH 3.0) at a flow rate of 0.5 mL/min. The purified material was collected in 2.5-mL fractions, and the protein peak fractions were subsequently pooled. The pooled protein was desalted through dialysis into 25 mM MES (pH 5.5). Initial crystallization conditions were obtained with this material.

Examination of the IgG–Sepharose-purified B2 protein on isoelectric focusing gel shows the material to contain two bands, with isoelectric points of 4.2 and 4.5. These two isoelectric forms of the B2 protein were separated on an LKB TSK-525 DEAE 21.5 mm × 15 cm anion-exchange HPLC column. The anion-exchange column was equilibrated with 40 mM sodium acetate (NaOAc; pH 5.6). After the sample was loaded on the column, it was eluted with a 40-min linear gradient of 70% buffer A (40 mM NaOAc, pH 5.6), 30% buffer B (40 mM NaOAc, pH 3.26), and 100% buffer B. Both peaks were dialyzed into the crystallization buffer, 25 mM MES (pH 5.5).

Crystallization. Crystals were grown by the vapor diffusion, hanging-drop method (Wlodawer & Hodgson, 1975) in Linbro culture plates (Gilliland & Davies, 1984). Type 7 protein G was dissolved at 15–20 mg/mL in 30 mM sodium citrate (pH 5.5) containing 10 mM $CaCl_2$, 10% poly(ethylene glycol) (PEG-3350), and (in some cases) 10 mM octyl β-glucoside. Droplets (5 μL) of this solution were suspended over wells containing 20–30% PEG-3350, at room temperature. Poorly diffracting crystals were initially grown from the IgG–Sepharose-purified material. Crystals were grown from both isoelectric forms of the protein, but the lower p*I* material (peak 2) produced higher quality crystals. Crystals grew in 3–8 weeks to sizes up to 0.5 × 0.4 × 0.3 mm.

Data Collection. X-ray diffraction data were collected using a Siemens (formerly Xentronics) electronic area detector mounted on a Supper oscillation camera controlled by a PCS 9650 microcomputer. During data collection the area detector chamber was mounted 9 cm from the crystal. The carriage angle (the angle between the normal to the detector face and the direct beam) was varied from 0° to 35°, enabling the detector to intercept data from ∞ to 1.6 Å, depending upon its position. Diffraction data collected by the area detector are recorded as a series of discrete frames or electronic images, each comprising a 0.20–0.40° oscillation counted for between 30 and 110 s, depending upon the carriage angle. A single data run typically comprised 250–480 data frames, corresponding to 100° of crystal rotation about a vertical axis. Data were transferred by an Ethernet link to a Digital Equipment Corp. VAX 11/780 or a Silicon Graphics Personal Iris computer for processing. For each crystal several data runs were recorded, each at a different value of a goniostat φ motor set 45° from vertical.

The X-ray source used to generate Cu Kα radiation was an Elliott GX-21 rotating anode, operating at 70 mA and 40 kV with a 0.3 × 3.0 mm focal spot and a 0.25-mm collimator. Monochromatization was provided by a Huber graphite monochromator. All data collection was performed at well-controlled room temperature (16–20 °C). The crystals were very stable in the X-ray beam, allowing collection of an extensively replicated data set to 1.60 Å from a single crystal. The XENGEN software package (Howard et al., 1987) was used for data reduction, and several advances in the development of XENGEN were first applied in the analysis of the protein G structure.

Table I: Summary of Crystallographic Refinement

resolution (Å)		1.67[a]
no. of reflcns[b]		
measd		7,172
used ($F \geq 2\sigma(F)$)		6,365
R value[c]		0.191
distances (Å)	(target)	obsd
bonds (1–2)	(0.03)	0.021
angles (1–3)	(0.04)	0.028
intraplanar (1–4)	(0.05)	0.032
planar group (Å)	(0.03)	0.018
chiral centers (Å³)	(0.30)	0.210
nonbonding contacts (Å)		
single torsion	(0.20)	0.164
multiple torsion	(0.20)	0.141
possible H bonds	(0.20)	0.133
torsion angles (deg)		
planar (ω)	(6.0)	3.2
staggered	(10.0)	13.5
orthonormal	(10.0)	22.2

[a] Resolution range at which the mean $I/\sigma(I)$ ratio dropped below 2. [b] From 8.0 to 1.66 Å. [c] $R = \sum|F_o - F_c|/\sum F_o$, from 8.0 to 1.66 Å.

Heavy Atom Derivatives and Phasing. Heavy atom derivatives were prepared by conventional soaking techniques. Each survey for a useful derivative consisted of collecting a complete 3.3- or 2.8-Å data set from a crystal soaked in a heavy atom solution and then calculating a merging R value between the derivative and the native data. Data from four derivatives (thallium chloride, uranyl nitrate, methylmercury chloride, and potassium platinocyanate) were judged to be promising, and multiple isomorphous phases were determined with the program packages PHARE (Otwinowski, 1990) and HEAVY (Terwilliger & Eisenberg, 1987).

Fitting and Refinement. A 2.5-Å single isomorphous replacement map based on the thallium derivative showed electron density indicative of four β-strands per asymmetric unit. Four strands and a helix were traced on the map using FRODO (Jones, 1985). Delineation of molecular boundary was difficult because of the low solvent content and because the β-strands of symmetry-related molecules combine with those of the central molecule to form a continuous β-sheet. After the initial chain tracing, a polyalanine model was refined. More side chains were added and the single tryptophan of the structure was located on a β-strand. Restrained-parameter least-squares procedures (Hendrickson, 1985) modified by Finzel (1987) to incorporate the fast Fourier algorithms of Ten Eyck and Agarwal (1978, 1980) were used to refine the structure and to extend the resolution. In later refinement stages the dynamics-aided refinement package X-PLOR (Brünger, 1988) was employed. Several cycles of alternate least-squares refinement and dynamics-aided refinement and manual refitting were done. The refinement had plateaued at an R value of 0.30 at 1.80-Å resolution.

During this time the NMR structure of the B1 domain of protein G had been determined. The coordinates of the NMR structure were manually fit to the single isomorphous replacement (SIR) coordinates, using Phe 43 and Tyr 46 as guides. Several cycles of alternate least-squares or dynamics-aided refinement and manual refitting of the electron density were done. The final refinement was done with PROFFT, with no restraints on the B factors. Table I contains statistics on the final refinement. These coordinates have been submitted to the Protein Data Bank (Bernstein et al., 1977) and have been assigned the designation 1PGX.

Comparison of the NMR and X-ray Structures. The coordinates of B2 domain residues 14–69 of the X-ray structure were superimposed on B1 domain residues 1–56 of the restrained minimized mean NMR structure of protein G, Protein Data Bank entry 2GB1, using the program X-PLOR

(Brünger et al., 1986; Brünger, 1988). Both the rms coordinate shifts and the shifts in backbone torsion angles were calculated. Each of the 60 NMR solutions to the B1 domain, Protein Data Bank entry 1GB1, was superimposed on the X-ray structure and the average rms shift was calculated. The coordinate shifts were calculated with and without the NMR structure's disordered side chains (Met 1 from C_β position outward; Lys 10, Glu 15, Glu 19, Lys 28, and Gln 32 from C_γ outward; Lys 4, Lys 13, Lys 31, Glu 42, and Glu 56 from C_δ outward). The residues that differ between the B1 and B2 domains (Ile–Val19, Leu7–Ile20, Glu 19–Lys32, Ala24–Glu37, Val29–Ala42, and Glu42–Val55; see Figure 1) were omitted from these analyses.

RESULTS AND DISCUSSION

The construction of the gene for recombinant 83-residue protein containing the B2 domain of protein G (type 7) has already been described (Fahnestock, 1990). The B2 domain gene was expressed as a soluble protein at over 20% of total cell protein, in *E. coli* using a hybrid λ-promotor O_L/P_R. After the cells were made leaky by briefly heating them to 80 °C, the B2 protein was initially purified over a IgG–Sepharose 6B column. Two forms of the B2 protein were identified on an isoelectric focusing gel with pI's of 4.2 and 4.5. These two forms are likely due to incomplete processing of the N-terminal Met by *E. coli* methionylaminopeptidase (Ben-Bassat et al., 1987; Tsunasawa et al., 1985). They were individually purified on a DEAE anion-exchange HPLC column.

The B2 domain of protein G crystallized in 30 mM sodium citrate buffer at pH 5.5 with 10 mM octyl β-glucoside (β-OG) and 10 mM $CaCl_2$ using 30% PEG-3350 as a precipitant. The initial crystallographic conditions were worked out using the IgG–Sepharose affinity-purified material. After separation of the two forms by anion-exchange HPLC, the lower isoelectric form (pI = 4.2) was found to produce high-quality crystals. These crystals belong to the space group $P2_1$, with cell dimensions $a = 26.04$, $b = 34.50$, and $c = 35.95$ Å and $\beta = 100.84°$, and contain one molecule in the asymmetric unit. The mean $I/\sigma(I)$ ratio of these crystals drops below 2 at a resolution of 1.67 Å in a typical diffraction experiment. Native data were collected at 90% completeness to 1.66 Å with $R_{sym}(w) = 0.0313$ and $R_{sym} = 0.0327$.

Data from four derivatives (thallium chloride, uranyl chloride, methylmercury chloride, and potassium platinocyanate) were judged to be promising, and multiple isomorphous phases were determined with the program package PHARE (Otwinowski, 1990). All four derivatives were located at the same site, $x = 0.55$ and $z = 0.89$ (y is arbitrary in space group $P2_1$). Subsequent examination of the thallium data shows this site to be close to the carboxyls of Glu 40, Glu 37, and possibly Glu 8 on a symmetry-related molecule. Phasing powers for uranyl, mercury, platinum, and thallium derivatives were 1.02, 1.05, 1.30, and 2.30, respectively, for data between 8.0 and 2.5 Å. Thus, only the thallium derivative proved to have more than marginal phasing power.

A 2.5-Å single isomorphous replacement map using the thallium derivative showed electron density indicative of four β-strands and an α-helix. No solvent leveling was done because of low solvent content of the unit cell. It was initially difficult to determine the molecular bounds of the B2 domain, because the four-stranded β-sheet was stacked on top of a screw-related β-sheet, resulting in a continuous β-sheet running through the crystal (see Figure 3). The positioning, direction, and sequence of the α-helix were clear due to Phe 43 and Tyr 46. We knew that there must be two strands on either side of the helix's
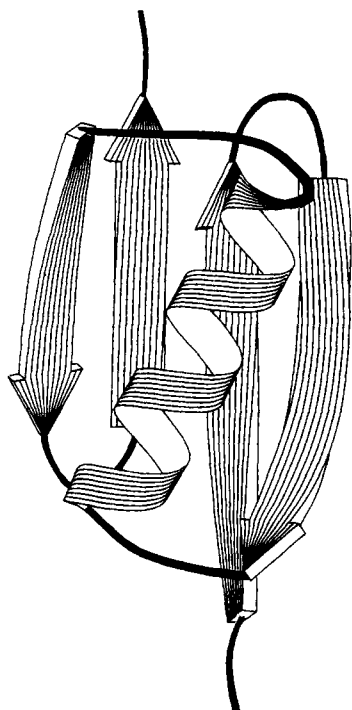
FIGURE 2: Schematic drawing (in the style of Jane Richardson) of the chain trace of B2 domain of protein G.

sequence due to its position. The β-sheet was initially fit as an antiparallel four-stranded β-sheet, with the strand order 1, 2, 4, and 3. After refinement with PROFFT and X-PLOR and extensive refitting, the R value had plateaued at 0.30 at 1.8-Å resolution, and the electron density of some of the loops and strand 2 continued to be difficult to interpret. Close proximity of the N-terminal end of a molecule and the C-terminal end of its symmetry-related mate was another source of difficulty in interpreting the map. We suspected that the chain tracing was incorrect.

During this time, the NMR structure of the B1 domain of protein G had been determined (Gronenborn, et al., 1991). The NMR structure clearly showed an α-helix and a four-stranded β-sheet. It also showed that the two inner strands of the β-sheet in the NMR structure were parallel to one another and the two outer strands were antiparallel to the two inner strands. The structure could be described as a four-stranded β-sheet composed of two β-hairpins, and joined by an α-helix. We manually superimposed the coordinates of the B1 domain NMR structure on the SIR coordinates of the B2 domain, using Phe 43 and Tyr 46 as guides (see Figure 4). An initial rigid-body refinement of the positioned NMR model was done against the X-ray data. The structure was refined with both X-PLOR and PROLSQ. Several refittings were needed to change the six residues that differ between B1 and B2 and to add the N- and C-terminal extensions. During the final refinement the structure was examined for bad φ–ψ angles, resulting in the final structure having no bad φ–ψ angles (see Figure 5). The final structure contained residues 8–77 and 61 water molecules. Two of the waters have alternative positions. The final R value to 1.66-Å resolution was 0.191 (see Table I).

This R value is higher than that of many structures determined at this resolution, but we were only able to locate 68 of the possible 83 residues or 82% of the protein G type 7 in the electron density maps. The electron density at each end of the protein faded into disorder and was difficult to interpret (see Figure 6A and B), unlike the clearly defined

electron density at the core of the 57-residue B2 domain (see Figure 6C). The N- and C-terminal peptide extensions that are observed in the electron density (residues 8–13 and 71–77) interact with symmetry-related molecules. Leu 9, Thr 10, and Pro 11 of the N-terminal extension interact with Glu 37, Glu 40, Trp 56, and Tyr 58 of a symmetry-related molecule. Thr 72 of the C-terminal extension interacts with Thr 64 in a second symmetry-related molecule. The rest of the C-terminal extension (Glu 73–Ala 77) interacts with the α-helix in a third symmetry-related molecule. The fading of the electron density can clearly be seen in the B factors of the structure (see Figure 7). The core 57-residue B2 domain has a mean B factor of 18.0 and the N- and C-terminal extensions have mean B factors of 37.3 and 38.8, respectively. The B factors increase as one moves away from the core domain, as shown in Figure 7.

The B2 domain extends from Thr 14 to Met 70, for a total of 57 residues. Both Thr 14 and Met 70 are involved in central strands of the β-sheet. Thr 14 makes two backbone to backbone hydrogen bonds to Ala 33, and Met 70 makes a single backbone to backbone hydrogen bond with the carbonyl of Asp 53. In addition, the $O_\gamma 1$ of Thr 14 is hydrogen-bonded to a water molecule, which in turn is hydrogen-bonded to the carbonyl of Val 34. The overall fold of the 57-residue B2 IgG-binding domain of protein G is identical to the previously described B1-binding domain (Gronenborn et al., 1991). Both the B1 and B2 domains have a four-stranded β-sheet composed of two β-hairpins and joined by an α-helix (see Figure 2). In Richardson's (1981) notation, the β-sheet has unique −1, +3x, −1 topology. This structure has a large number of hydrogen bonds, including 44 backbone to backbone (1 backbone to backbone hydrogen bond is found in the N-terminal extension), 9 backbone to side chain, and 5 side chain to side chain. Both ends of the 16-residue α-helix were capped by hydrogen bonds (Asp 35 OD2 to Thr 38 N and Tyr 46 O to Asn 50 ND2). The hydrophobic core of the B domains lies between the β-sheet and the α-helix and includes residues Tyr 16, Leu 18, Ile 20, Leu 25, Thr 29, Thr 31, Val 39, Phe 43, Tyr 46, Val 52, Trp 56, Tyr 58, Phe 65, and Val 67 (see Figure 8). Of the six residues in the hydrophobic core with hydrogen-bonding capabilities, two were hydrogen-bonded to water (Tyr 16 OH and Trp 56 NE) and Tyr 58 was hydrogen-bonded to OD2 of Asp 60. A water molecule that bridges between the amide hydrogen of Tyr 46 (33)[2] and the carboxyl of Ala 42 (Val 29) in the solution structure (Clore & Gronenborn, 1992) is also observed in this structure.

The 55 residues of the B2 domain from the X-ray structure have been compared to the B1 domain structure determined by NMR, residues 15–69 and residues 2–56, respectively. The rms coordinate shift between the X-ray structure and the restrained minimized mean NMR structure (⟨SA⟩) is 1.13 Å and between the X-ray structure and all 60 NMR structure solutions (⟨SA⟩r) is 1.17 Å for the C_α atoms and slightly higher for all the atoms (see Table II). The largest differences between the X-ray and NMR structures are seen around residues 24 (11), 49 (36), and 61 (48) (see Figure 9). Thr 24 (11) is in the turn between strand 1 and strand 2. Asp 49 (36) is the next to last residue in the α-helix and is in close proximity to Thr 24 (11) (see Figure 10). Ala 61 (48) is in the turn between strand 3 and strand 4. All these residues are close to five intermolecular hydrogen bonds between strand 2 of one molecule and strand 3 of a symmetry-related molecule,

---

[2] The residue numbering and/or residue type of the B1 domain solution structure of protein G is shown in parentheses.
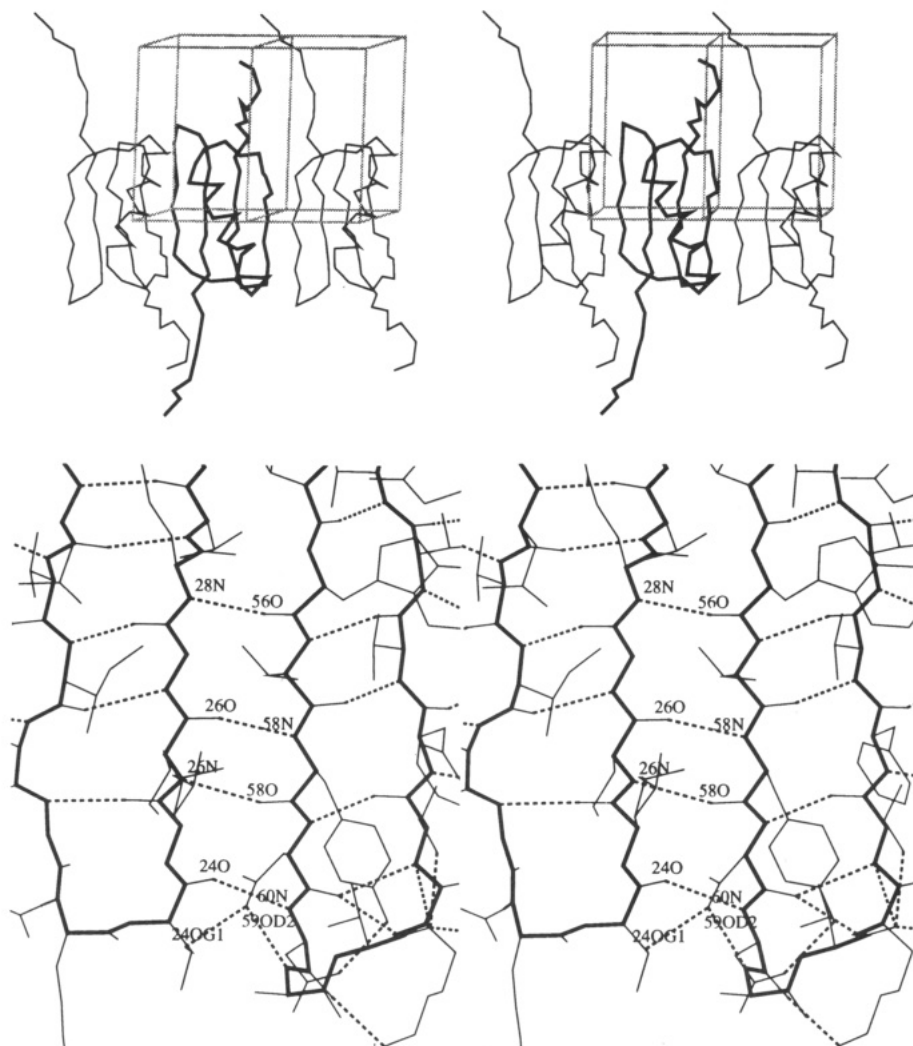
FIGURE 3: Continuous β-sheet running through the molecule. (A, top) C$_\alpha$ diagram with two asymmetric units of the unit cell shown. The primary molecule is shown in heavy lines and the symmetry-related molecules, related by the 2-fold screw, are shown as light lines. (B, bottom) Hydrogen bonding involved in forming the continuous β-sheet. Hydrogen bonds are shown as dashed lines.

which form an extended β-sheet running through the entire crystal (see Figure 3B). The five intermolecular hydrogen bonds are as follows: Thr 24 (11) O$_\gamma$1 to Asp 59 (46) O$_\delta$2; Thr 24 (11) O to Asp 60 (47) N; Lys 26 (13) N to Tyr 58 (45) O; Lys 26 (13) O to Tyr 58 (45) N; and Glu 28 (15) N to Trp 56 (43) O.

The α-helix in the X-ray structure is rotated 10° from the NMR structure, such that it lies more parallel to the β-sheet strands (see Figure 10A). This rotation affects residues 32 (19) to 50 (37) and is centered at Ala 42 (Val 29). Three of the six residues that differ between the B2 and B1 domains lie in this range (Lys 32–Glu 19, Glu 37–Ala 24, and Ala 42–Val 29). Of the other three differences (Val 19–Ile 6, Ile 20–Leu 7, and Val 55–Glu 42) only the conservative change of Ile 20 to Leu 7 lies in the hydrophobic core between the β-sheet and the α-helix and could possibly effect the rotation of the α-helix. Both Ile 20 (Leu 7) and Ala 42 (Val 29) are at the center of the rotation and thus are unlikely to be the cause. Lys 32 in the X-ray structure and Glu 19 in the NMR structure are not involved in any interaction that would effect the rotation of the α-helix. Glu 37 (Ala 24) could play a role in the helix rotation, for the O$_\epsilon$1 of Glu 37 makes a hydrogen bond to the O$_\gamma$1 of Thr 10 on a symmetry-related molecule, pushing the helix away from the symmetry-related molecule (see Figure 10B). As with the other major differences between the B1 NMR and B2 X-ray structures, the α-helix rotation

can be explained by an intermolecular hydrogen bond, i.e., by a crystal-packing interaction.

The side chains of Lys 17, Leu 25, and Met 70 are modeled with alternative conformations in the X-ray structure of the B2 domain. No electron density was found for the N$_\zeta$ of Lys 32 and Lys 41, for the C$_\epsilon$ and N$_\zeta$ of Lys 26, or for the N and C termini of the peptide extensions. The NMR structure had more disordered side chains than did the X-ray structure. In addition to those side chains found in the X-ray structure to be disordered, Lys 23 (10), Glu 28 (15), Gln 44 (32), Glu 55 (42), and the N-terminal Met 14 (1) and C-terminal Glu 69 (56) had disordered side chains in the NMR structure.

The B1 and B2 domains of protein G are very stable. The stability of protein G allows one to release the protein from the cytoplasm of *E. coli* cells by heating the cells to 80 °C for 30 min. The single B1 and B2 domains of protein G have melting temperatures of 87.5 and 79.4 °C, respectively, as determined by differential scanning calorimetry (Alexander, et al., 1992). This unusual stability for a small protein lacking disulfide bonds is likely due to a number of factors. First, the overall fold stabilizes the protein. The B domains of protein G cannot be unfolded by simply unraveling protein from one end, because both ends of the polypeptide chain comprise the central strands of the β-sheet. One can imagine an unfolding pathway in which the hydrogen bonds between the central strands of the β-sheet were first broken, and the structure
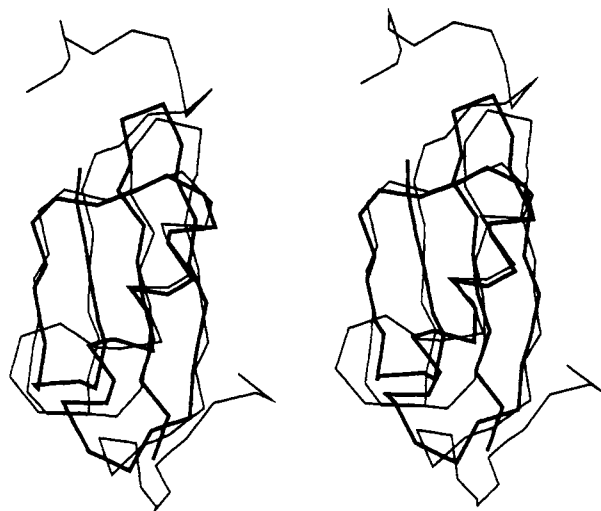
FIGURE 5: Ramachandran plot of protein G B2 domain crystallographic structure: (□) glycyl residues; (+) other residues.

FIGURE 4: Superposition of the B1 domain NMR structure on the SIR structure of the B2 domain of protein G: (A, top) superposition of the C_α's; (B, bottom) region around Phe 43 and Tyr 46 in the α-helix.

breaks up into two β-hairpins and an α-helix, which further unfold. The second postulated reason for the stability of protein G makes this unfolding pathway unlikely. The IgG-binding domains of protein G have a very tightly packed hydrophobic core (see Figure 8). The negative solvation-free energies (Eisenberg & McLachlan, 1986) of the B1 and B2 domains are −55 ± 2 and −47 ± 2 kcal/mol, respectively. Third, over 95% of the residues in the 57-residue core of the B domains are involved in regular secondary structure elements. Finally, the B2 domain is extensively hydrogen-bonded, with 38 of the 57 backbone carbonyl oxygen atoms (67%) and 40 of the 57 backbone amide nitrogen atoms (70%) participating in hydrogen bonds to other protein atoms.

The relative contribution of the overall folding, hydrogen bonding, and hydrophobic interactions could be examined. Using site-directed mutagenesis, residues comprising the hydrophobic core [Tyr 16 (3), Leu 18 (5), Ile 20 (Leu 7), Gly 22 (9), Leu 25 (12), Thr 29 (16), Thr 31 (18), Ala 39 (26), Phe 43 (30), Tyr 46 (33), Ala 47 (34), Val 52 (39), Trp 56 (43), Tyr 58 (45), Phe 65 (52), and Val 67 (54)] could be replaced. Likewise, hydrogen bonds could be removed, such as those involving the helix-capping Asp 35 and Asn 50.
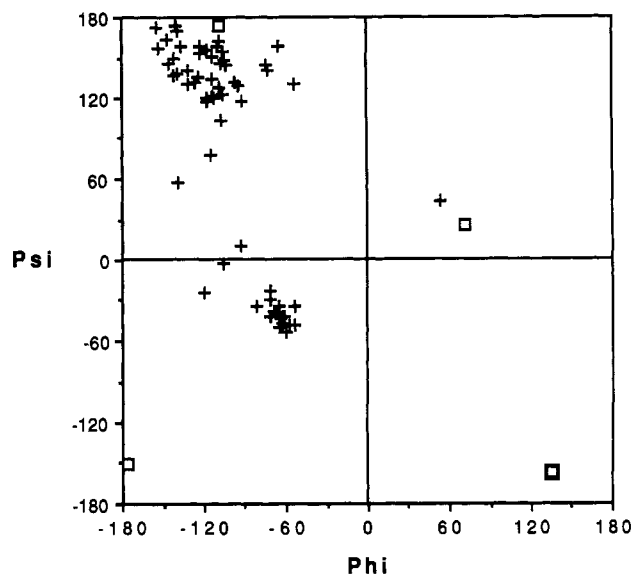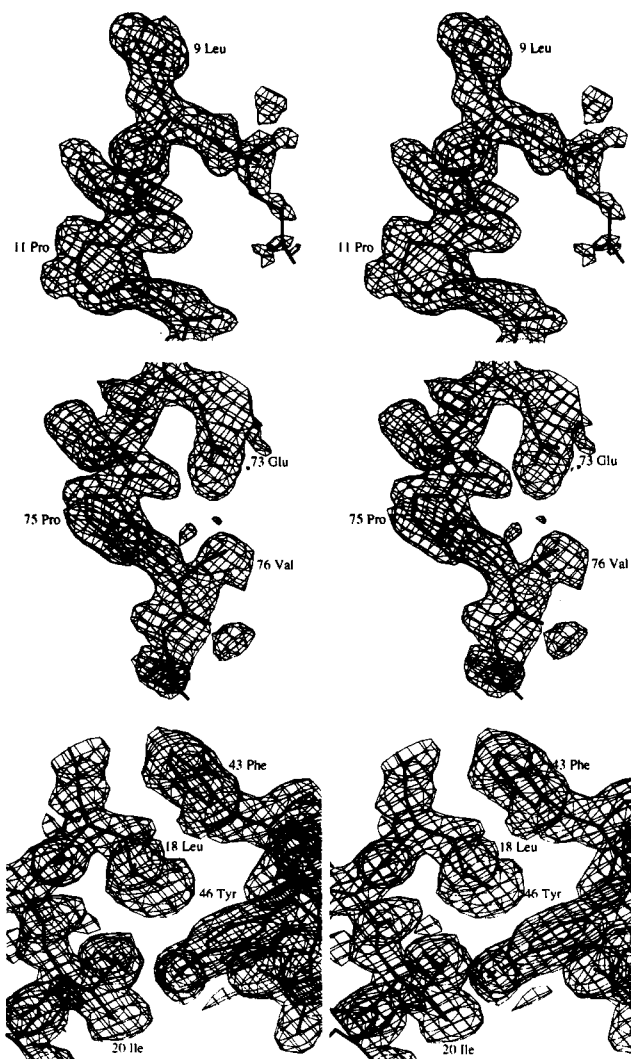

FIGURE 6: Final $2F_o - F_c$ electron density maps of the B2 domain of protein G: (A, top) N-terminal domain, residues 8–13; (B, middle) C-terminal domain, residues 74–77; (C, bottom) electron density around helical residues Phe 43 and Tyr 46.

Alternative folding patterns could also be studied by changing the order of the β-strands, such that one of the terminal strands is no longer central in the β-strand.
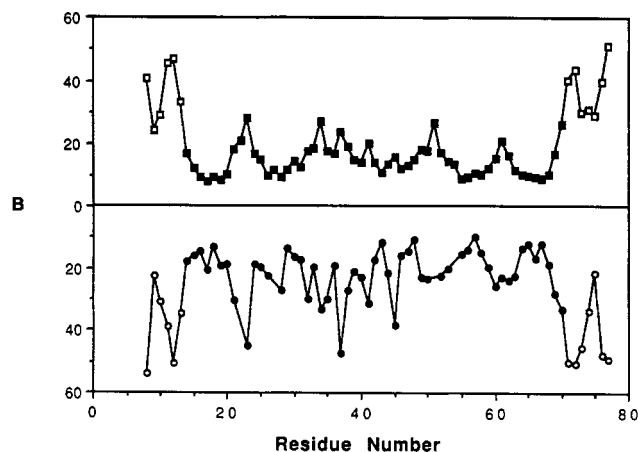
FIGURE 7: Variation in *B* factors along the polypeptide chain of protein G B2 domain crystallographic structure. The main-chain *B* factors are shown as squares, and the side-chain *B* factors are shown as circles. Residues in the core of the B2 domain are shown as solid circles or squares; residues of the N- and C-terminal linking regions are shown as open circles or squares.
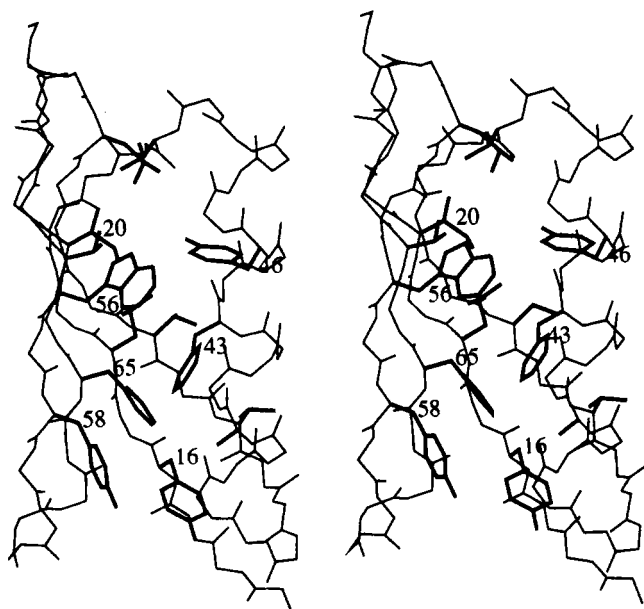

FIGURE 8: Hydrophobic core of the B2 domain of protein G. The following side chains involved in the hydrophobic core are shown in bold: Tyr 16, Leu 18, Ile 20, Leu 25, Thr 29, Thr 31, Ala 39, Phe 43, Tyr 46, Ala 47, Val 52, Trp 56, Tyr 58, Phe 65, and Val 67. The other side chains have been omitted for clarity.

FIGURE 9: Plot of the atomic rms differences and ϕ–ψ rms differences between the X-ray structure of the B2 domain and the NMR structures of the B1 domain. The backbone (A) and all-atom (B) rms differences are the average rms deviations of the 60 individual NMR structures from the X-ray structure. The bars in both (A) and (B) are the standard deviations of these values. The ϕ (C) and ψ (D) differences are between the X-ray structure and the restrained minimized mean NMR structure. The residue numbering is based on the B1 domain.

Table II: Summary of the Root Mean Square Differences between the NMR Structure of the B1 Domain and the X-ray Structure of the B2 Domain

	rms difference (Å)		
	backbone	all atoms	all atoms excl disordered side chains in the NMR struct[a]
⟨SA⟩ vs X-ray	1.17 ± 0.06	1.73 ● 0.07	1.43 ± 0.06
(SA)r vs X-ray	1.14	1.69	1.39

[a] The disordered side chains in the NMR structure are Met 1 from C_β position outward; Lys 10, Glu 15, Glu 19, Lys 28, and Gln 32 from C_γ outward; and Lys 4, Lys 13, Lys 31, Glu 42, and Glu 56 from C_δ outward.

We examined this structure for clues as to where an immunoglobin G would bind. An IgG has an isoelectric point near pH 8, whereas the isoelectric point of protein G is about pH 4.5. An Ig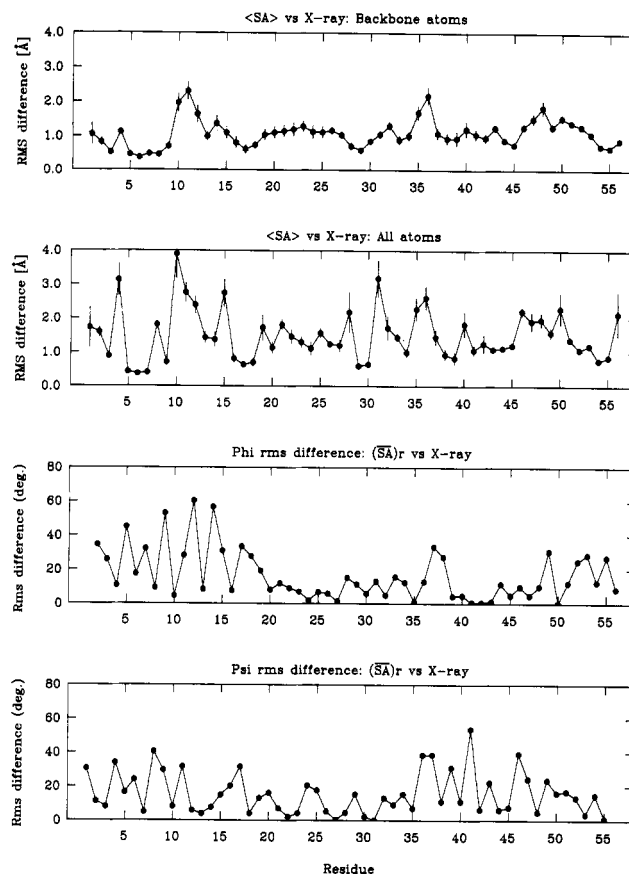G elutes from a protein G affinity column around pH 4, which suggests that a carboxylic acid on protein G is involved in its binding to an IgG. There are 10 carboxylic acids in the B1 domain of protein G and 9 in the B2 domain. Seven of the nine residues containing carboxylic acids in the B2 domain are found in the α-helix and the third β-strand. The α-helix and the third β-strand form a surface that could act as the IgG-binding interface. The B1 and B2 domains of protein G bind to IgGs with affinities of 0.3 X 10[8] and 2.1 X 10[8] M⁻¹, respectively (Fahnestock, personal communication). The lower affinity of the B1 domain may be due to the replacement of Glu 37 (24) on the α-helix in the B2 domain with Ala in the B1 domain.

Further insight into the how protein G binds to IgG's may be found in examining other immunoglobin-binding protein complexes. Deisenhofer (1981) determined the structure of the B fragment of *Staphylococcus* protein A complexed to an immunoglobin Fc fragment. The B fragment of protein A contains two α-helices, both involved in immunoglobin binding. We aligned the α-helix of protein G with each of these α-helices in the Protein Data Bank entry 1FC2. The protein A α-helix from residues 127–138 (helix 1) shows little similarity to the α-helix in protein G, but the α-helix from residues 142–156 (helix 2) had a similar pattern of surface residues. Three of the seven surface residues are the same: Glu 37, Gln 45, and Asp 49 (residues 143, 151, and 155 in protein A) and Asn 50 to Asp 156 is a conservative change. Gln 151, Asp 155, and Asp 156 in protein A are directly involved in immunoglobin binding. In addition, when protein A's helix 2 is aligned with
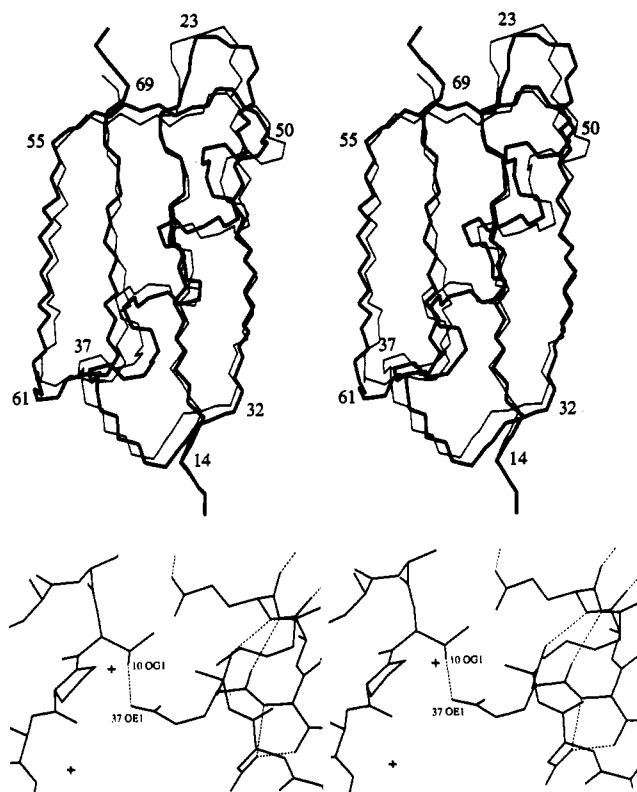
FIGURE 10: (A, top) Superposition of residues 14–69 of the X-ray structure of the B2 domain (thick line) and residues 1–56 of the restrained minimized mean NMR structure of the B1 domain (thin line), main-chain atoms only. For clarity, only the X-ray residue numbers of the B1 domain are shown. (B, bottom) Hydrogen bond between O$_\epsilon$1 of Glu 37 and O$_\gamma$1 of Thr 10 on a symmetry-related molecule.

protein G's α-helix, the third strand of protein G's β-sheet occupies the same position as protein A's helix 1. These similarities suggest one way in which protein G may bind to immunoglobins.

CONCLUSIONS

We have determined the X-ray structure of protein G type 7, which contains the B2 IgG-binding domain, at 1.67 Å with a final R value of 0.191. The NMR structure of the B1 domain was helpful in correcting the errors in the chain tracing of the B2 domain that arose from the weak phasing of the single thallium derivative we used. Without the NMR structure this structure would have taken more work to complete. The 13-residue peptides on either end of the structure faded away into disorder, but the 57-residue core of the protein is in strong density. The B2 domain extends from Thr 14 to Met 70, for a total of 57 residues, and has the same basic fold as the B1 domain solved by NMR.

Comparison of the NMR structure of the B1 domain and the X-ray structure of the B2 domain reveals only minor differences, all of which appear to result from intermolecular hydrogen bonds formed in the X-ray structure. The largest difference between the two structures, a 10° rotation of the α-helix such that it lies more nearly parallel to the strands of the β-sheet in the X-ray structure than in the NMR structure, can be accounted for by one of these intermolecular hydrogen bonds.

The various types of protein G B domain are very thermally stable, such that they can be released from the cytoplasm of *E. coli* by heating the cells to 80 °C. This stability arises from the following: (i) the compact fold of the protein, where

95% of the residues are involved in regular secondary structure elements, and the N and C termini of the B domains are in central strands in the β-sheet; (ii) the tight packing of the hydrophobic core of the B domains; and (iii) the large number of hydrogen bonds in the B domain, with over 70% of the backbone amides and carbonyls being hydrogen-bonded. Because of their small size and high thermal stability, the B domains of protein G could make an ideal system for studying protein folding.

The NMR structure of the B1 domain and this X-ray structure of the B2 domain provide limited insight on how protein G binds to IgG's. The distribution of negative charges around the α-helix and the third strand of the β-sheet forms a surface that could act as the IgG-binding interface. In addition, the α-helix of protein G and one of the two α-helices involved in immunoglobin binding in protein A have similar patterns of surface residues. This question of protein G–IgG interaction will not be definitively answered until a structural study of protein G–IgG complex is complete.

REFERENCES

Alexander, P., Fahnestock, S., Lee, T., Orban, J., & Bryan, P. (1992) *Biochemistry, 31*, 3597–3603.

Agarwal, R. C. (1978) *Acta Crystallogr. A34*, 791–809.

Agarwal, R. C. (1980) in *Refinement of Protein Structures, Proceedings of the Daresbury Conference* (Machin, P. A., Campbell, J. W., Elder, M., Eds.) pp 24–28, Science and Engineering Research Council, Daresbury Laboratory, Warrington, England.

Ben-Bassat, A., Bauer, K., Chang, S.-Y., Myambo, K., Boosman, A., & Chang S. (1987) *J. Bacteriol. 169*, 751–757.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E. J., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tsumi, M. (1977) *J. Mol. Biol. 112*, 535–542.

Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Boyle, M. D. P. (1984) *Biotechniques 2*, 334–340.

Brünger, A. T. (1988) *J. Mol. Biol. 203*, 803–816.

Brünger, A. T., Clore G. M., Gronenborn, A. M., & Karplus, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3801–3805.

Clore G. M., & Gronenborn, A. M. (1992) *J. Mol. Biol. 224*, 853–856.

Deisenhofer, J. (1981) *Biochemistry 20*, 2361–2370.

Eisenberg, D., & McLachlan, A. D. (1986) *Nature 319*, 199–203.

Fahnestock, S. R. (1990) U. S. Patent Number 4,956,296.

Fahnestock, S. R., Alexander, P., Nagle, J., & Filpula, D. (1986) *J. Bacteriol. 157*, 870–880.

Fahnestock, S. R., Alexander, P., Filpula, D., & Nagle, J. (1990) in *Bacterial Immunoglobulin Binding Proteins* (Boyle, M. D. P., Ed.) Vol. 1, pp 133–148, Academic Press, Orlando, FL.

Finzel, B. C. (1987) *J. Appl. Crystallogr. 20*, 53–55.

Gilliland, G. L., & Davies, D. R. (1984) *Methods Enzymol. 104*, 370–381.

Gottesmann, M. E., Adhya, S., & Das, A. (1980) *J. Mol. Biol. 140*, 57–75.

Gronenborn, A. M., & Clore G. M. (1991) *Science 254*, 581–582.

Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., & Clore G. M. (1991) *Science 253*, 657–661.

Hendrickson, W. A. (1985) *Methods Enzymol. 115*, 252–270.

Howard, A. J., Gilliland, G. L., Finzel, B. C., & Poulos, T. L. (1987) *J. Appl. Crystallogr. 20*, 383–387.

Jones, T. A. (1985) *Methods Enzymol. 115*, 157–169.

Kraulis, P. J. (1991) *Science 254*, 581.

Messing, J. (1983) *Methods Enzymol. 101C*, 20–78.

Otwinowski, Z. (1990) *American Crystallographic Association Annual Meeting*, Abstract CO4.

Reis, K. J., Ayoub, E. M., & Boyle, M. D. P. (1984) *J. Immunol., 132*, 3098–3102.

Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.

Scandella, D., Arthur, P., Mattingly, M., & Neuhold, L. (1985) *J. Cell Biochem. 9B*, 203.

Tenan, D. S. (1981) *N. Eng. J. Med. 305*, 1195–1200.

Terwilliger, T. C., & Eisenberg, D. (1987) *Acta Crystallogr. A43*, 6–13.

Tsunasawa, S., Stewart, J. W., & Sherman, F. (1985) *J. Biol. Chem. 260*, 5382–5391.

Wlodawer, A., & Hodgson, K. O. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 398–399.