IMMUNOMETHODS 2, 3–8 (1993)

Structural Studies of Immunoglobulin-Binding Domains of Streptococcal Protein G

Angela M. Gronenborn and G. Marius Clore

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20814

The solution and X-ray structures of the IgG-binding domains of streptococcal protein G are described and compared. Each domain comprises a core of 56 residues and exhibits extreme thermal stability (~90°C), despite the absence of any disulfide bridges. The structure has an unusual fold comprising a four-stranded β-sheet with a −1, +3x, −1 topology on top of which lies an α-helix. The central two strands, comprising the N- and C-termini, are parallel; the outer two strands, which are antiparallel to the inner strands, are connected by the helix in a +3x crossover. This fold is also found in ubiquitin, a protein with no sequence similarity or functional relationship to the IgG-binding domains of protein G. The thermal stability of the domains can be accounted for by the unusual topology, coupled with an extensive hydrogen bonding network and a tightly packed and buried hydrophobic core. Possible sites of interaction with IgG are discussed in the light of the structure. © 1993 Academic Press, Inc.

Protein G is a large multidomain cell surface protein of group G *Streptococcus* which is thought to help the organism evade the host defenses via unique protein-binding properties (1). A repeating 55-residue domain binds to immunoglobulin G and to α_2-macroglobulin, a major protease inhibitor of human plasma (1). There are two such repeats in protein G from strain GX7809 and three for the protein from strain GX7805, and the sequence identity between the various repeats is greater than 90% (1). Microcalorimetry of single domains, known as B1 and B2, reveals extreme thermal stability, with melting temperatures of 87 and 79°C, respectively (2, 3). Further, the unfolding transition on urea gradient gel electrophoresis cannot be observed in full as the protein remains native up to ~8 M urea (2, 4). These features are highly unusual considering the small size of the domain and the *absence* of any disulfide bridges or tightly bound prosthetic group. The potential importance of the IgG-binding domains of protein G as analytical tools in immunology, together with their extreme stability, has led to a variety of structural and physicochemical studies (2–10).

To shed light on the remarkable properties of this protein, as well as the interaction with immunoglobulins, we have determined high-resolution structures of two IgG-binding domains of protein G, using NMR spectroscopy for the B1 domain and X-ray crystallography for the B2 domain (2, 8). The B1 domain comprises 56 residues, while our version of B2 has 13 residues preceding the 56-residue B2 domain and 14 residues following it, resulting in a total of 83 residues. Within the 56 residues, B1 and B2 differ in amino acid sequence at six positions: Ile-6 → Val, Leu-7 → Ile, Glu-19 → Lys, Ala-24 → Glu, Val-29 → Ala, and Glu-42 → Val. A comparison of the primary sequences for the B1 and B2 domains is given in Fig. 1.

SOLUTION STRUCTURE OF THE B1 DOMAIN OF PROTEIN G

The structure of the B1 domain comprises a four-stranded β-sheet made up of two antiparallel β-hairpins connected by an α-helix (Figs. 2 and 4A). The topology of the sheet (−1, +3x, −1) is highly unusual: namely, the two central strands of the sheet, β1 (residues 1–8) and β4 (residues 50–56), are parallel; while two outer strands, β2 (residues 13–20) and β3 (residues 42–47), are antiparallel to β1 and β4, respectively. Strands β1 and β2 are connected by a type I turn (residues 9–12), while strands β3 and β4 are connected by an unusual 6-residue turn from residues 46 to 51 in which Lys-50 has a positive ϕ angle of ~50° in the left-handed helical region of the Ramachandran plot. The two outer strands, β2 and β3, of the sheet are connected via a long helix (residues 22–37) and a short extended structure (residues 38–41). The long axis of the helix lies at ~140° to the axes of β2 and β3. In addition to the characteristic CO(i)–NH(i + 4) hydrogen bonds found in α-helices, the helix also displays a number of bifurcating CO(i)–NH(i + 3) hydrogen bonds. Further, the last 5 residues are tightened into a 3–10 helix characterized exclusively by two CO(i)–NH(i + 3) hydrogen bonds.

1058-6687/93 $5.00

What features of the structure could be responsible for the unusually high thermal stability of this small protein domain? First, the involvement of nearly all residues in regular secondary structure ensures a large number of stabilizing hydrogen bonds (Fig. 3). There are a total of 45 hydrogen bonds; 41 are backbone–backbone, 3 side chain–backbone, and 1 side chain–side chain. Over 60% of the backbone donor and acceptor groups (38 backbone amides and 36 backbone carbonyls) participate in hydrogen bonding. The side chain–backbone hydrogen bonds between the carboxylate of Asp-22 and the NH of Thr-25, the side-chain amide of Asn-37 and the carbonyl of Tyr-33, and the carboxylate of Asp-46 and the NH of Ala-48 stabilize and cap the first and last turns of the helix and the turn between $\beta 3$ and $\beta 4$, respectively (Fig. 3). In addition, there is one side chain–side chain hydrogen bond in which the OηH atom of Tyr-3 donates a hydrogen bond to the Oη atom of Tyr-45.

Second, the interior of the protein is highly hydrophobic, while the exterior is very hydrophilic. This is associated with a large negative solvation free energy of folding (11) of -55 ± 2 kcal $\cdot$ mol^{-1}. This value is comparable to that for another small and very stable protein, bovine pancreatic trypsin inhibitor, and is $\sim$15% more negative than the predicted SFE for a protein of this size (12). Nine hydrophobic residues (Leu-5, Leu-7, Gly-9, Ala-26, Phe-30, Ala-34, Val-39, Phe-52, and Val-54) are completely buried, with surface accessibilities less than 0.5 Å^2, and only five hydrophobic residues (Ile-6, Val-21, Ala-24, Val-29, and Ala-48) are significantly exposed to solvent. The helix itself constitutes a perfect example of an amphiphilic helix. The packing of the internal side chains is illustrated in Fig. 4B. The aromatic rings are packed approximately orthogonally to each other. Leu-5 interacts with Phe-30, Trp-43, and Phe-52, Leu-7 with Trp-43 and Tyr-33, and Leu-12 with Val-39 and Trp-43. The hydrophilic functional groups of the polar aromatics (Tyr-3, Trp-43, and Tyr-45) and threonine residues (Thr-16 and Thr-18) located on the inside surfaces of the sheet and helix are directed toward solvent, while their hydrophobic portion is directed toward the interior and constitutes part of the hydrophobic core. Thus, the methyl groups of Thr-16 and Thr-18 interact with Leu-5 and Tyr-3, respectively, while the aromatic rings of Trp-43 and Tyr-45 interact with the methyl groups of Val-54 and the aromatic ring of Phe-52, respectively. In this regard, it should also be noted that these residues are located either on the outer strands (Thr-16 and Thr-18 in $\beta 2$, and Trp-43 and Tyr-45 in $\beta 3$) or at the beginning of a strand (Tyr-3 in $\beta 1$). The exposed surfaces of the sheet and helix are made up of positively charged Lys residues, negatively charged Glu and Asp residues, and polar Asn, Gln, and Thr residues. Interestingly, the solvent-exposed surface of the sheet has a large preponderance of Thr residues.

```
                              .   .   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
```

FIG. 1. Amino acid sequences of the streptococcal protein G B1 and B2 domains. Differences are highlighted in boldface.

X-RAY STRUCTURE OF THE B2 DOMAIN OF PROTEIN G AND COMPARISON WITH THE SOLUTION STRUCTURE OF B1 DOMAIN

The X-ray structure of the B2 domain was determined at 1.67 Å resolution using a combination of single isomorphous replacement phasing and manual fitting of the coordinates of the solution NMR structure of the B1 domain (8). The overall fold of the central 56 residues is identical to that of the B1 domain, with the electron density at each end of the chain of the protein fading into disorder. The fact that the N- and C-termini are disordered in the crystal is also reflected in the high values for the B factors in these regions which increase as one moves away from the core domain.

A comparison of the structures for residues 1–56 of B1 with residues 14–69 of B2 (Fig. 5) yields an atomic rms

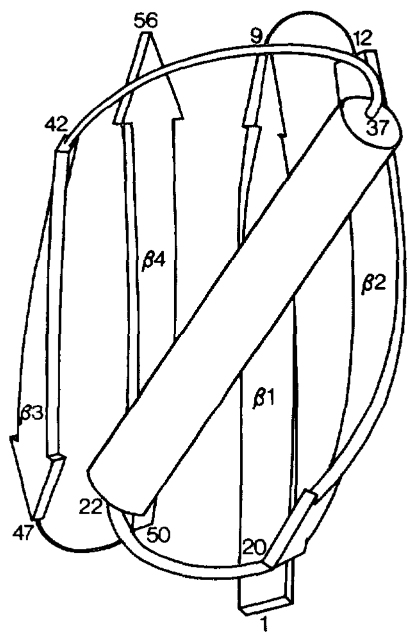

FIG. 2. Schematic ribbon representation of the polypeptide fold of the B1 domain.

coordinate shift of 1.1 Å for the backbone atoms and 1.7 Å for all atoms. The largest differences between the two structures are seen around residues 11(24), 36(49), and 48(61). Thr-11(24) is in the turn between β-strand 1 and β-strand 2. Asp-36(49) is the one but last residue in the α-helix and is in close proximity to Thr-11(24). Ala-48(61) is located in the turn between β-strands 3 and 4. All these residues are close to five intermolecular hydrogen bonds between β-strand 2 of one molecule and β-strand 3 of a symmetry-related molecule in the crystal lattice, which contains an extended β-sheet running through the entire crystal.

The α-helix is rotated ~10° in the X-ray structure compared to the NMR structure, rendering it slightly more parallel to the β-strands of the sheet. This rotation affects residues 19(32) to 37(50) and it may be worth pointing out that three of the six residues that differ in the B1 and B2 domains are located in this part of the structure. As with the other differences seen between the NMR and X-ray structures, the α-helix rotation can be explained by crystal packing interactions since the side chain of Glu-37 in the B2 domain is involved in a hydrogen bond with the side chain of Thr-10 of a symmetry-related molecule in the crystal lattice. In the B1 domain, the equivalent residue is Ala-24, which clearly cannot participate in such an interaction.

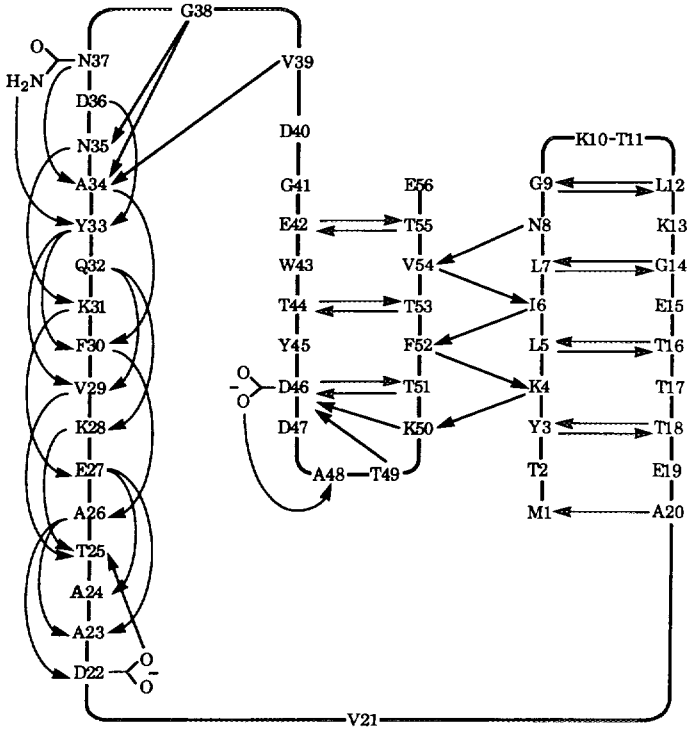

FIG. 3. Diagrammatic representation of hydrogen bonding within the B1 domain.

STRUCTURAL SIMILARITY OF THE IgG-BINDING DOMAINS OF PROTEIN G AND UBIQUITIN

An unexpected structural similarity between the IgG-binding domains of protein G and ubiquitin has been observed (13, 14). The topology of the fold in both proteins is very similar, with ubiquitin containing a short additional β-strand in the sheet and the helix running at a slightly more tilted angle. Indeed, a least-squares best match between the B1 domain of protein G and ubiquitin reveals that 41 residues of the two proteins can be superimposed with a backbone rms value of 2.2 Å (14). This structural similarity is clearly surprising given the lack of discernible amino acid sequence similarity. Like the protein G domains, ubiquitin is an extremely stable molecule, and it may well be that this particular structural motif represents a very robust folding unit which cannot simply unfold from one of the two ends since both ends of the polypeptide chain comprise the central strands of the β-sheet.

TIGHTLY BOUND WATER IN THE SOLUTION STRUCTURE OF THE B1 DOMAIN OF PROTEIN G

We also have identified two bound water molecules in the solution structure of the B1 domain (15). Both bound water molecules may contribute to the unusual stability of this structure. As the lifetime of the bound water detected in the NMR experiment is greater than about 1 ns, it is likely that the two bound water molecules participate in a bifurcating hydrogen bonding network comprising a CO–NH hydrogen-bonded pair, such that the water molecule accepts a hydrogen bond from the NH proton and donates one to the carbonyl oxygen, with the result that the amide proton is involved in a three-center hydrogen bond. On the basis of the structure, one water molecule participates in such an interaction with the Ala-20(NH)–Met-1(CO) hydrogen-bonded pair at the beginning of an antiparallel β-sheet, and the other with the Tyr-33(NH)–Val-29(CO) hydrogen-bonded pair in the single α-helix. The latter, which is external and solvent accessible, is associated with a distortion in the α-helix centered around Tyr-33 and results in a significant increase in the CO(i-4)–N(i) and CO(i-4)–NH(i) distances relative to those in the rest of the helix, as well as a significant departure in the ϕ, ψ angles of Tyr-33 relative to regular helical geometry. Thus, water molecule W1 could stabilize the β-sheet at the N-terminus of the polypeptide chain, while W2 may stabilize a distortion in the α-helix around Tyr-33 arising as a consequence of the extremely well-packed hydrophobic core. Such solvent-induced dis-

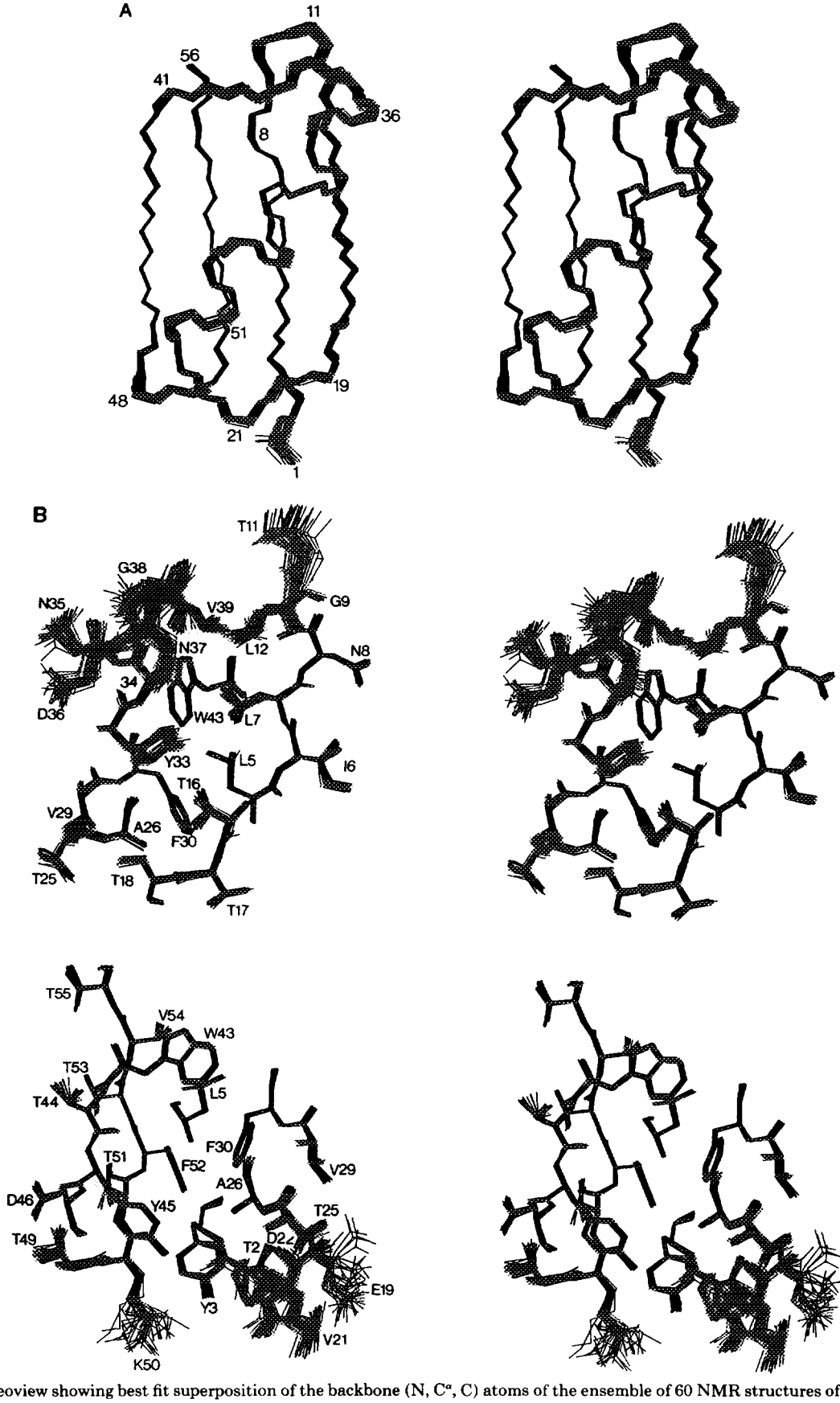

FIG. 4. (A) Stereoview showing best fit superposition of the backbone (N, C^{α}, C) atoms of the ensemble of 60 NMR structures of the B1 domain; (B) stereoviews showing best fit superpositions of all atoms (excluding protons).

tortions in α-helices have been previously noticed in crystal structures and were postulated as possible folding intermediates for helical structures. The present observation of this phenomenon in solution indicates, however, that these water molecules are tightly bound and represent an integral part of the protein framework. Interestingly, the W2 water molecule is also observed in the crystal structure of the B2 domain (8).

IDENTIFICATION OF SITES OF INTERACTION WITH IgG

Having structures for both the B1 and the B2 domains in hand, we examined both structures for clues as to where and how IgG binding occurs. The association constants for the binding of IgG to the B1 and B2 domains of protein G are 0.3 and 2.1 $\times$ 10^8 M^{-1}, respectively (16). In an attempt to deduce which residues might be important for binding, we compared the location and local conformation around the six amino acid substitutions between the two domains: Ile-6 → Val, Leu-7 → Ile, Glu-19 → Lys, Ala-24 → Glu, Val-29 → Ala, and Glu-42 → Val. Leu-7 is completely buried and therefore cannot be involved in binding. The other five residues, on the other hand, are solvent accessible and hence potential contact residues: in particular, Glu-19, Ala-24, and Val-29 are clustered around the end of strand β2 and the beginning of the α-helix, while Ile-6 and Glu-42 lie on the solvent-exposed surface of the β-sheet. Comparison of the structure of the IgG-binding domain of staphylococcal protein A (17) with that of protein G reveals no structural or sequence similarity. In protein A two α-helices are involved in IgG binding, and little similarity between these helices and that of the protein G domains is observed apart from a similar pattern of surface residue for helix 2 of protein A and the protein G helix.

Competition binding studies using peptide fragments derived from an IgG-binding domain of protein G have implicated an 11-amino-acid region centered around the loop connecting the α-helix to β-strand 3 (Asn-35 to Tyr-45) in IgGFc binding (18). One of the nonconservative changes between the B1 and B2 domains of protein G is located in this region, namely, Glu-42 → Val. It may therefore be tempting to invoke this amino acid change to account for the difference in binding affinities between the B1 and B2 domains. In contrast, a recent crystal structure of an IgG-binding domain of protein G complexed with an Fab fragment reveals that β-strand 2 and to a lesser degree the end of the α-helix are the primary regions of interaction with the constant heavy-chain (C_H1) domain of the Fab (10; see also Derrick and Wigley (20), this issue). In this complex, the β-sheet of the immunoglobulin is aligned in an antiparallel fashion with that of the protein G domain, resulting in a contiguous antiparallel β-sheet. The interacting residues of the α-helix in the protein G domain are Tyr-33 and Asn-37, with the tyrosine side-chain hydroxyl donating a hydrogen bond to a backbone carbonyl on the last β-strand of the

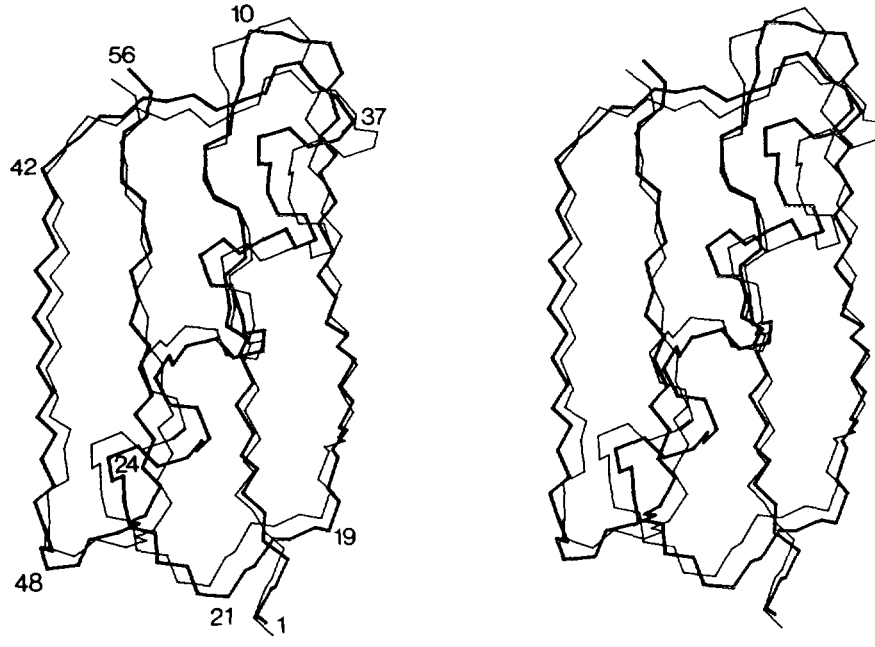

FIG. 5. Superposition of residues 14–69 of the X-ray structure of the B2 domain (thick line) and residues 1–56 of the NMR structure of the B1 domain (thin line). Residue numbering follows that of the B1 domain.

C_H1 domain, and the asparagine side chain accepting a hydrogen bond from a backbone amide. Both of these residues are located in a region close to the one identified by the peptide mapping studies (18). It is interesting to note that the recognition of the Fab by the protein G domain is mediated mainly by backbone interactions. This may account for the observation that protein G binding to Fab is rather weak compared to that with Fc (19). Thus, the main site of IgG binding to protein G involves the Fc portion of the molecule, and it is by no means clear that the type of interaction found in the Fab complex is similar to that in the Fc complex. In this regard it is worth pointing out that the same type of intermolecular contact, namely, the formation of a contiguous β-sheet, was also found in the crystal structure of the B2 domain of protein G alone. Thus, it may solely constitute a preferred nonspecific way of packing IgG-binding domains of protein G with other β-sheet domains in a crystal lattice. Further characterization of the recognition between Fc and protein G will necessitate structural studies of a protein G IgG-binding domain–Fc complex.

ACKNOWLEDGMENT

This work was supported by the AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health.

REFERENCES

1. Fahnestock, S. R., Alexander, P., Filpula, D., and Nagle, J. (1990) *in* Bacterial Immunoglobulin Binding Proteins (Boyle, M. D. P., Ed.), Vol. 1, pp. 133–148, Academic Press, San Diego.
2. Gronenborn, A. M., Filupa, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., and Clore, G. M. (1991) *Science* **253,** 657–661.
3. Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (1992) *Biochemistry* **31,** 3597–3603.
4. Gronenborn, A. M. Unpublished data.
5. Orban, J., Alexander, P., and Bryan, P. (1992) *Biochemistry* **31,** 3604–3611.
6. 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) *Biochemistry* **30,** 5335–5340.
7. Alexander, P., Orban, J., and Bryan, P. (1992) *Biochemistry* **31,** 7243–7248.
8. Achari, A., Hale, S. P., Howard, A. J., Clore, G. M., Gronenborn, A. M., Hardman, K. D., and Whitlow, M. (1992) *Biochemistry* **31,** 10449–10457.
9. Lian, L.-Y., Yang, J. C., Derrick, J. P., Sutcliffe, M. J., Yang, J. C., and Roberts, G. C. K. (1992) *J. Mol. Biol.* **228,** 1219–1234.
10. Derrick, J. P., and Wigley, D. B. (1992) *Nature* **359,** 752–754.
11. Eisenberg, D., and McLachlan, A. D. (1986) *Nature* **319,** 199–203.
12. Cliche, L., Gregoret, L. M., Cohen, F. E., and Kollman, P. A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 3240–3243.
13. Kraulis, P. J. (1991) *Science* **254,** 581.
14. Gronenborn, A. M., and Clore, G. M. (1991) *Science* **254,** 581–582.
15. Clore, G. M., and Gronenborn, A. M. (1992) *J. Mol. Biol.* **223,** 853–856.
16. Fahnestock, S. Private communication.
17. Deisenhofer, J. (1981) *Biochemistry* **20,** 2361–2370.
18. Frick, I.-M., Wikström, M., Forsén, S., Drakenberg, T., Gomi, H., Sjöbring, U., and Björck, L. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8532–8536.
19. Erntell, M., Myhre, E. B., Sjöbring, U., and Björck, L. (1988) *Mol. Immunol.* **25,** 121–126.
20. Derrick, J. P., and Wigley, D. B. (1993) *Immuno Methods* **2.**