Two Crystal Structures of the B1 Immunoglobulin-Binding Domain of Streptococcal Protein G and Comparison with NMR[†,‡]

Travis Gallagher,[*] Patrick Alexander, Philip Bryan, and Gary L. Gilliland[*]

Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, and the National Institute of Standards and Technology, 9600 Gudelsky Drive, Rockville, Maryland 20850

Received November 24, 1993; Revised Manuscript Received January 28, 1994[●]

ABSTRACT: The structure of the 56-residue B1 immunoglobulin-binding domain from streptococcal protein G has been determined in two different crystal forms. The crystal structures were deduced by molecular replacement, based on the structure of the B2 domain (Brookhaven accession code 1PGX). Final R values are 0.174 and 0.198 for orthorhombic and trigonal forms, for diffraction data from 6.0 to 2.07 Å and from 6 to 1.92 Å, respectively. The orthorhombic crystals have an unusually high packing density for protein crystals, with V_m = 1.66 and a solvent content of 26%. The protein structure is found to be very similar (rms deviation 0.25 Å for 56 Cα's) in the two crystal forms, with an efficiently packed hydrophobic core between a four-stranded β-sheet and a four-turn α-helix. The B1 domain has the same fold and general structure as the B2 domain (rms deviations 0.36 and 0.39 Å), despite the six residue differences between them. The crystallographic models differ from NMR-derived models in several local regions, primarily in the loop involving residues 46–51; other significant variations are observed in the helix and in the structure of bound water. The primary crystal contact is the same in both crystal forms, involving both sheet edges to form extended β-sheets throughout the crystals.

Many infectious bacteria display surface proteins that bind tightly to immunoglobulins and other host proteins, and some of these bacterial proteins have found useful roles in the laboratory and clinic. The most familiar example is protein A from *Staphylococcus*, whose affinity for IgG confers wide utility in antibody separation procedures. Protein G from *Streptococcus*, Lancefield group G, also binds human IgG in the Fc region (Myhre & Kronvall, 1977). This protein consists of a series of small binding domains separated by linkers, and a cell-wall anchor near the C-terminus. Two (in some strains, three) of the domains bind IgG; another two or three domains bind serum albumin. The binding interactions are believed to help the pathogen evade the host's immune response by mimicking the "self"-markers commonly displayed by host cells. The IgG-binding domains of protein G are of increasing interest due to their broad biomedical utility (e.g., in purifying antibodies) and for studies of folding and stability. They are minimal folding units of high thermostability, without disulfides or cofactors. The domains are also remarkable for their sequence parsimony: they are devoid of serine, cysteine, arginine, proline, histidine, and, in the native form, methionine.

The IgG-binding domains of protein G are identified as B1, B2, etc., numbering from the N-terminus of the native protein G molecule. The subject of the present study, B1, is identical to the B1 domain in strain GX7809 (Fahnestock et al., 1986), wherein there are six residue differences between domains B1 and B2. These six, with the B1 residue given first, are 16V,[1] L7I, E19K, A24E, V29A, and E42V. The melting temperatures of B1 and B2 at pH 5.4 are 87.5 and 79.4 °C, respectively (Alexander et al., 1992). Protein G, proteolytically released from streptococci, binds human IgG with an association constant of 6.7×10^{10} M^{-1} (Akerstrom & Bjork, 1986). The B1 and B2 domains bind to the Fc fragment of IgG with association constants of 3×10^6 and 1×10^7 M^{-1}, respectively, as determined by titration calorimetry (Alexander, unpublished data). The isoelectric points for the two domains are 4.0 and 4.7, respectively (Alexander et al., 1992a). The three-dimensional structure of the B1 domain was first determined by NMR techniques (Gronenborn et al., 1991). The fold was shown to consist of a four-stranded sheet and a helix. Both sheet and helix are strongly amphipathic, with a hydrophobic core between them. The five secondary structure elements are connected by four loops, L1, L2, L3, and L4 (see Table 1 for sequence and secondary structure information). The structures of both domains, from a different bacterial strain in which B1 contains the I6V and L7I substitutions, were determined by NMR (Lian et al., 1992). The crystal structure of the B2 domain has also been reported (Achari et al., 1992), as has a complex of a B3 domain and an Fab fragment (Derrick & Wigley, 1992). While the domains consist of 56 residues, some of these studies involve constructs with extra lengths of polypeptide at either end, with the result that residues are numbered differently. In the present paper, residue numbers are identical to those used by Gronenborn et al. (1991), 5 less than those used by Lian et al. (1992), and 13 less than those of Achari et al. (1992).

Table 1: Amino Acid Sequence and Locations of Loops in the B1 Domain of Protein G

```
          L1          L2   (..helix..)    L3        L4
         ** **       * **                *** *     ****
MTYKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK TFTVTE
         10          20          30          40          50    56
```

[†] This work was supported in part by NSF Grant MCB-92-19309.

[‡] The refined models have been submitted to the Brookhaven Protein Data Bank under filenames 1PGA and 1PGB.

[*] To whom correspondence should be addressed at the Center for Advanced Research in Biotechnology. Phone: (301) 738-6272. FAX: (301) 738-6255.

[●] Abstract published in *Advance ACS Abstracts*, March 15, 1994.

[1] Abbreviations: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile;, K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg: S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr; rms, root mean square; NMR, nuclear magnetic resonance. Single-site mutations are denoted by the sequence number preceded by the original amino acid (using the one-letter abbreviation) and followed by the substituted amio acid.

0006-2960/94/0433-4721$04.50/0

The NMR structure of the B1 domain and the crystal structure of the B2 domain were compared by Achari et al. (1992). The structures were found to be similar (rms coordinate shift for Cα's 1.13 Å), with the main differences in the loops connecting secondary structure elements. Also noted was a difference of 10° in the angle of the helix relative to the sheet, and a variation in the structure of the C-terminal portion of the helix. A subsequent NMR investigation of ordered water in the solution structure has provided evidence for two specific found water sites, one of which is postulated to correlate with a distortion in the helix by involvement with the helix hydrogen bond 29O–33N (Clore & Gronenborn, 1992). Alexander et al. (1992) have investigated the unfolding and refolding of B1 and B2, providing measurements of the relevant thermodynamic and kinetic parameters. Their studies have been applied toward an explanation of the cooperative folding/unfolding transition and the high melting temperatures of IgG-binding domains, which are characteristic of small folding units and do not necessarily imply extreme stability at physiological temperatures. In addition, NMR sequence assignments and preliminary amide exchange studies have also been carried out on the B2 domain (Orban et al., 1992).

In order to provide a diffraction-based structural model of the B1 domain, for comparison with the B2 domain and for the assessment of differences in the results of NMR and X-ray methods, we undertook crystallographic studies of protein G, domain B1. Crystals in two different space groups were suitable for diffraction analysis and provided data for the two independent structure determinations described below. These structures are compared with each other and with those reported by Gronenborn et al. (1991), Lian et al. (1992), and Achari et al. (1992).

MATERIALS AND METHODS

The B1 domain was engineered for production as a 56-residue protein with N-terminal methionine (this position was threonine in the wild type). The sequences of B1 and B2 correspond to amino acids 228–282 and 298–352 of the gene sequence given by Fahnestock et al. (1986). Aspects of the cloning, expression, and purification of the domain are detailed by Alexander et al. (1992). Protein for crystallization was kept in 50 mM NaOAc, pH 4.5, and stored at 4 °C. Crystals were grown in 10-μL hanging drops by the vapor diffusion technique, using Linbro 24-well tissue culture plates. Crystal conditions were sought in the pH range 3.8–5.0 due to the domain's low p*I*. The only successful crystallization agent found was 2-methyl-2,4-pentanediol, in the range of 50–80%. Many crystal forms were observed from combinations of 2-methyl-2,4-pentanediol and various salts, although few crystals achieved sufficient size and diffraction quality. Crystals grew best at 18 °C, and required several weeks to grow. See Table 2 for crystal conditions.

Diffraction data were collected using X-rays generated by a Siemens rotating anode, operated at 40 kV and 70 mA. A Siemens focusing mirror system was utilized. A Siemens electronic area detector was positioned 100 mm from the crystal, and at a 2θ angle of 24°. Diffraction data collected with the area detector were recorded as a series of discrete frames or electronic images, each comprising a 0.2° scan in ω counted for 90 s. The temperature of data collection was 20 °C. The determination of crystal orientation and the integration of reflection intensities were performed with the XENGEN program system (Howard et al., 1987). Unweighted R_{sym} values were 0.070 and 0.068 for intensities in the orthorhombic and trigonal reduced data sets, respectively. See Table 2 for diffraction statistics.

Table 2: Crystallographic Data for the B1 IgG-Binding Domain in Two Crystal Forms

	orthorhombic	trigonal
crystal conditions	50% MPD[a] 50 mM NaCl 25 mM NaOAc pH 4.5 20% IPA[a] T = 18 °C	70% MPD 300 mM $MgSO_4$ 25 mM NaOAc pH 4.0 T = 18 °C
protein concentration (mg L^{-1})	10	10
crystal size (mm)	0.1 × 0.1 × 0.8	0.3 × 0.4 × 0.4
growth time (weeks)	4	2
habit	rectangular rod	irregular block
space group	$P2_12_12_1$	$P3_121$
a (Å)	37.04	36.63
b (Å)	25.08[b]	36.63
c (Å)	51.28	79.25
V_m	1.66	2.14
diffraction data		
reflections	2701	5014
resolution (Å)	2.04	1.86
R_{sym}[c]	0.070	0.068
$\langle I/\sigma \rangle$ in high-resolution shell	6.4	4.9

[a] Abbreviations: MPD, 2-methyl-2,4-pentanediol; IPA, 2-propanol. [b] The usual convention of naming orthorhombic cell edges in increasing order was broken to facilitate comparison with a monoclinic crystal form with a = 37.9 Å, b = 26.3 Å, c = 50.1 Å, and β = 100° (refinement in progress). [c] $R_{sym} = \sum_{hkl}|I_{ij} - \langle I \rangle_j G_{ij}|/\sum I_{ij}$.

Table 3: Refinement Data for the B1 IgG-Binding Domain in Two Crystal Forms

crystal	orthorhombic	trigonal
refinement		
resolution (Å)	6–2.07	6–1.92
reflections ($I > \sigma_I$[a])	2567	4493
data completeness (%)	85	93
final model		
protein atoms	436	436
solvent molecules	20	24
refined R[b]	0.174	0.198
deviations, rms[c,d]		
bonds (Å)	0.020	0.020
angles (deg)	1.970	2.045

[a] I = intensity. [b] $R = \sum_{hkl}||F_o| - |F_c||/\sum_{hkl}|F_o|$. [c] rms = root mean square. [d] Although the final refinement used PROLSQ, the rms deviations given are from the ideal parameters in XPLOR 3.1 parameter sets tophcsdx.pro and parhcsdx.pro (see Materials and Methods).

The structures were solved by molecular replacement using a suite of programs implemented in XPLOR 3.1 (Brunger et al., 1987). The search probe was a partially refined model of a mutant of the B2 domain (the V21C mutant), a structure currently in refinement that was solved by molecular replacement using the B2 crystal structure 1PGX (Achari et al., 1992) as a search probe. Subsequently, it was found that the 1PGX structure itself would suffice as a probe to solve the present structures. The B1 NMR structure 2GB1.PDB, although chemically identical, gave a weaker signal in the rotation function and hence was not used. The probe was prepared by removing the N- and C-terminal extensions and converting the following residues to alanine: all Glu, Gln, and Lys, both termini, and the seven residues different between B2:V21C and B1. In both cases the stepwise sequence of the molecular replacement method was rotation search, Patterson correlation refinement of rotation function peaks, translation search, rigid body refinement of the whole probe, and rigid body refinement of the probe divided into three parts (1–20, 21–39, 40–56). The resolutions of diffraction data used in

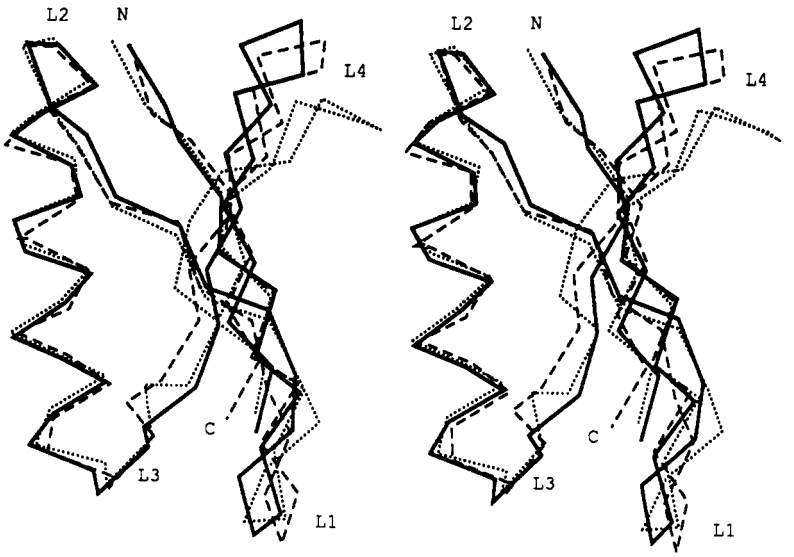

FIGURE 1: Stereo Cα superposition of the trigonal crystal form (solid lines) with the Gronenborn et al. (1991) NMR structure (dashed lines) and the first Lian et al. (1992) NMR structure (dotted lines). The greatest deviations among the models are in the four loops, especially L1 (residues 9–12), L3 (residues 38–41), and L4 (residues 47–50).

the rotation functions were 8.0–3.3 Å (orthorhombic form) and 9.0–3.4 Å (trigonal form). In both cases the solution was among the top three rotation function peaks, and Patterson correlation refinement amplified the solution to a clear maximum. The translation function results were unambiguous, and in the case of the trigonal space group, clearly distinguished between space group $P3_121$ and its enantiomorph $P3_221$. The R values of the translation function solutions were 0.39 and 0.42 for the orthorhombic and trigonal structures, and the corresponding values after rigid body refinement were 0.35 and 0.39.

The two structures were refined independently, in parallel, without reference to each other. Atomic positions and thermal parameters were initially refined using XPLOR, in a series of steps at increasing resolution. The topology and parameter sets used were tophcsdx.pro and parhcsdx.pro, respectively. Diffraction data with $I > \sigma(I)$ and with d spacings under 6 Å were included. Each step typically involved a round of simulated annealing and/or Powell minimization, followed by examination of electron-density maps to check progress, guide the building of the missing residues, and search for waters. Simulated annealing temperatures ranged from 4500 °C in early rounds to 1500 °C in late rounds. The program FRODO (Jones, 1987) was used on an Evans and Sutherland PS300 graphics system for inspection of electron-density maps and adjustment of the model. When nearly all the reflections had been included, several rounds of additional refinement were carried out by the restrained parameter least-squares procecedure of PROLSQ (Hendrickson & Konnert, 1980), which has been modified by Sheriff (1987) to restrain contacts between different molecules. Thermal parameters, which had been released in the late rounds of XPLOR refinement, were now refined to convergence under the least-squares restraints of PROLSQ.

To reduce bias to the probe structure and intermediate phasing models, annealed omit maps were used throughout refinement to guide manual refitting. About half the side chains were rebuilt or manually adjusted at least once in each refinement. This fact, and the large number of passes of simulated annealing (at least 20 for each structure), helps to ensure the independence of the final models. In a final test, atoms were deliberately mispoisitioned (to incorrect rotamers) and Fourier maps calculated. The R value invariably increased in such experiments, and the maps contained new peaks reflective of the displacements. The refinements of both structures are summarized in Table 3. The refined models are in the Brookhaven Protein Data Bank (Bernstein et al., 1977) under filenames 1PGA (orthorhombic) and 1PGB (trigonal).

Hydration of the protein was determined by examination of difference maps, with new water molecules being added to the model at chemically reasonable sites as indicated by electron-density maxima. Waters whose temperature factors refined to exceed 40 Å^2 or whose omit-map electron density dropped below the 3σ level were removed from the model. Finally waters were ranked by occ$^2/B$ (James & Sielecki, 1983). Crystal packing was analyzed as follows. First, atoms within 3.8 Å of another protein molecule were determined using the SAM CONTACTS routine in FRODO (Jones, 1987). These atoms were grouped into contacts (a "contact" was defined as the set of atoms on one molecule that are close to another single molecule). Contacts were ranked according to the number of atom pairs involved.

RESULTS

The two refined structures of the B1 IgG-binding domain from protein G superimpose with a root mean square deviation over 56 Cα positions of 0.25 Å, indicating that the structure of this protein is very similar in the two different crystals. In the core region, the deviations between the two structures are under 0.15 Å. The only differences in rotamers are in the side chains of the following surface residues: Met1, Lys4, Lys10, Glu15, Glu19, Lys28, Lys31, Gln32, Asn35, Glu42, and Lys50. All of these except Lys31 and Asn35 are involved in symmetry contacts in one or both of the crystal forms. The greatest Cα deviation (0.53 Å) is at residue 38 in the loop L3 (see Table 1 for sequence and secondary structure information). Due to the strong consensus between the two structures, the trigonal form will stand for both in most of the following comparisons. The trigonal form is chosen because its diffraction data extend to higher resolution, and because it has fewer crystal contacts.

The NMR structure of the chemically identical molecule (Gronenborn et al., 1991) and the NMR structure of a molecule differing only slightly (Lian et al., 1992) afford good comparisons of the X-ray and NMR methodologies. The Lian et al. (1992) report describes two domains referred to as II

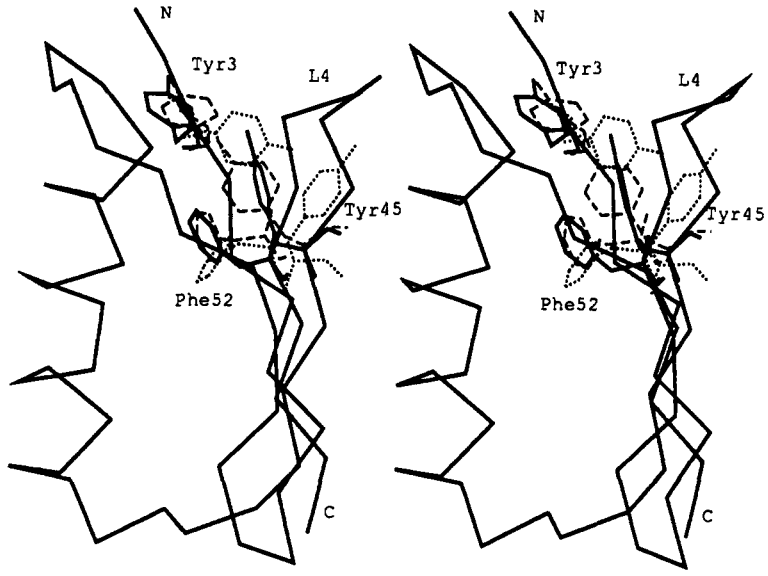

FIGURE 2: Stereogram of the region of the core close to L4, showing differing conformations of three aromatic residues. Solid lines, including the Cα tracing, are from the trigonal crystal structure. Dashed lines are from the NMR model of Gronenborn et al. (1991). Dotted lines are from the first NMR model of Lian et al. (1992).

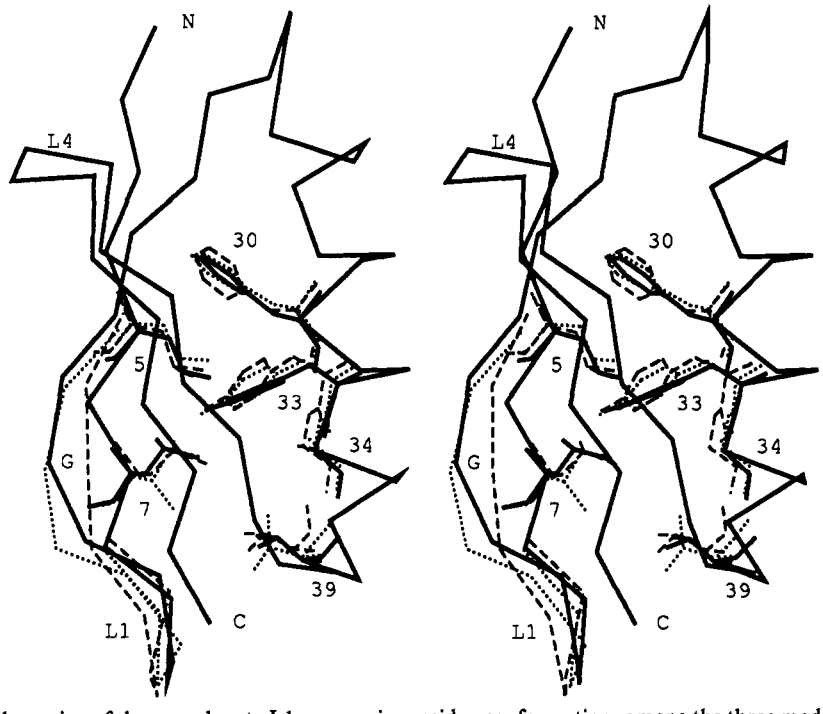

FIGURE 3: Stereogram of the region of the core close to L1, comparing residue conformations among the three models Gronenborn et al. (1991) (dashed lines), Lian et al. (1992) (dotted lines), and the trigonal crystal form (solid lines). Most of the residues are in good agreement. See Table 1 for residue types. Residue 7 is the only nonsolvated mutation site among the reported sequence variants of IgG-binding domains; it is Ile in the Lian et al. (1992) structure and Leu in the others. The letter G indicates that position of Gly14, which has a nearly fully extended conformation in all reported structures, giving the sheet an outward bend.

and III. Except for short extensions at the N- and C-termini, their II domain is very similar to the B1 domain of the present paper. The differences are at positions 6 and 7, where domain II has the same sequence as B2, Val-Ile, instead of Ile-Leu as in B1. Domain II is thus a sequence hybrid of B1 and B2. The rms deviations (Cα) between the present structures and the solution structure of Gronenborn et al. (1991) (Brookhaven accession code 2GB1) are 1.06 and 1.18 Å for the orthorhombic and trigonal forms, respectively. PDB coordinate set 2IGG contains 27 models of domain II resulting from the NMR structure determination by Lian et al. (1992). The average rms deviation (Cα) of these structures from the orthorhombic crystal form of B1 is 1.67 Å; the corresponding figure for the trigonal form is 1.72 Å. In general the Gronenborn structure more closely resembles the X-ray structure than do the Lian structures (see Figure 1). The differences are most significant at L1 and L4. In these regions, the solution structures (and especially the Lian et al. (1992) structures) appear less compact than the X-ray structure. The average rms deviation (Cα) between the Gronenborn structure and the Lian structures is 1.55 Å; thus, the crystal structure agrees with one NMR model better than they agree with each other. The present crystal structures of B1 resemble the B2 crystal structure (Achari et al., 1992) more closely than they resemble either B1 NMR structure. The rms deviations between the B2 structure and the B1 crystal structures are 0.39 and 0.36 Å for the

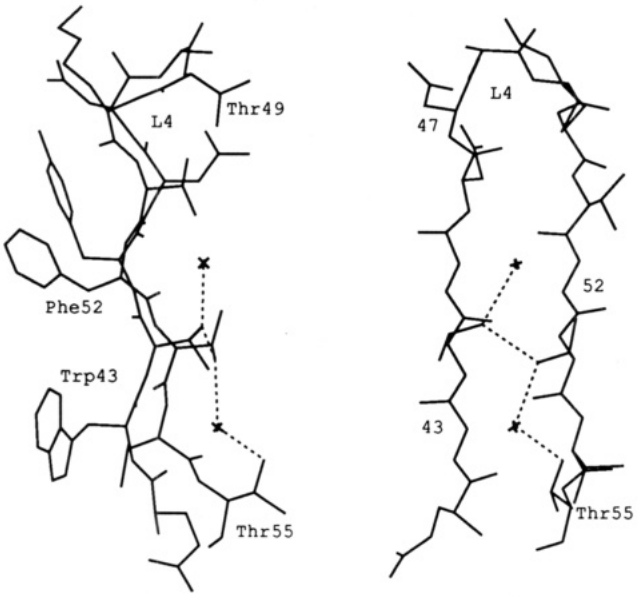

FIGURE 4: Two orthogonal views of the two strongest water sites. These waters are cradled by pleats on the external face of the β-sheet, where most of their neighbors are threonine residues. In addition to the hydrogen bonds indicated by dotted lines, the waters are within H-bonding distance of the peptide carbonyls and amides in the underlying antiparallel β-structure.

orthorhombic and trigonal crystal forms. The only differences in side chain rotamers are for surface residues. The greatest backbone deviations are in the L3 region, especially residues 39–42 and the adjacent residue 56. The mutation at residue 42 may have a role in these structural differences, although the effect is not obvious since the residue projects into the solvent. That this is found to be the region of greatest dissimilarity between the crystal structures of B1 and B2 is consistent with the NMR results of Lian et al. (1992). None of the six residue differences between B1 and B2 has a marked effect on the backbone or on another residue. There are no intraprotein hydrogen bonds involving any of the six side chains in either domain. Thus, there is no immediate explanation for the difference in stability. Most of the residue differences are on the protein surface. Residue 7 contributes to the core; the alternative residues here (Leu/Ile) are found to have similar conformations (i.e., similar χ_1 and χ_2 torsions).

The hydrogen bonding in the present models is basically as described by Gronenborn et al. (1991), with the following elaborations. The first β-ribbon is fully H-bonded, including two bonds between residue 1 and residue 20. The loop L1 is a type I β-turn, but without a hydrogen bond linking the carbonyl of 9 and the amide of 12. The conformational angles of residue 11 incline the 11–12 peptide so as to make the 9O–12N distance about 4.4 Å. This is consistent with amide exchange rates observed by Orban et al. (1992). The present structures concur with the B2 crystal structure of Achari et al. (1992) in this turn. In the helix, hydrogen bonding is normal from 22O through 37N. In particular, the hydrogen bond from 29O to 33N is between 2.9 and 3.0 Å, while the ϕ, ψ angles at Tyr33 are −55°, −52° (orthorhombic form) and −64°, −50° (trigonal form). Although in several instances the O_i–N_{i+3} distances are under 3.3. Å (suggesting a 3–10 helix), the helix is best described as α in its entire length. In its final turn, 34O receives hydrogen bonds from amides 37 and 39; the distance to 38N is suitable for a long H bond, but the angle is poor (the ON vector is over 60° from the peptide plane). The lack of a hydrogen bond 34O–38N is consistent with fast amide exchange observed by Orban et al. (1992). The second β-ribbon has three pairs of H bonds with standard

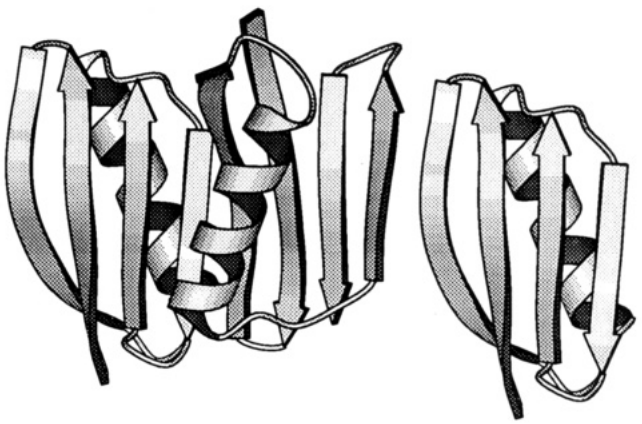

FIGURE 5: Three adjacent G:B1 molecules in the crystal lattice. The sheet–edge contact shown here is the primary crystal contact in both the orthorhombic and trigonal forms. Interacting molecules are related by 2-fold screw symmetry, forming molecular chains with extended β-sheets throughout the crystal. This diagram was produced with the program MOLSCRIPT (Kraulis, 1991).

antiparallel geometry, linking residues 42, 44, and 46 with residues 55, 53, 51, respectively. The central, parallel portion of the β–sheet has six regular hydrogen bonds, from 4N–50O to 8O–56N. In addition to these main chain hydrogen bonds, the following four well-ordered H bonds involving side chains could be important in stabilizing the protein structure. Two pairs of threonines form hydrogen bonds between their hydroxyl groups, 49–51 and 44–53, and the side chain carboxylate of Glu 56 tucks into the space between L1 and L3, uniting them by hydrogen bonding to the amides of residue 10 and residue 40.

The two Phe side chains are completely buried in the core, in van der Waals contact and with the planes of their rings nearly perpendicular in the manner described by Burley and Petsko (1985). The three tyrosine rings and the single Trp side chain are clustered loosely around the phenyl rings, generally with perpendicular or oblique contact geometry of their aromatic systems. The crystal structures [both the present B1 structures and the B2 structure of Achari et al. (1992)] have the same rotamers for all the aromatic side chains. The Gronenborn et al. (1991) structure agrees in every rotamer except χ_2 of Tyr45. The Lian et al. (1992) first model disagrees with the crystal and Gronenborn et al. (1991) models in the χ_1 rotamers for Tyr3, Trp43, and Phe52. While it has the same rotamers as the crystal structures for Tyr45, this residue is in a completely different position in the Lian et al. (1992) model due to the unique open structure of L4 (see Figure 1). The situation of the side chain of Tyr45 is different among the various models. Gronenborn et al. (1991) reported a hydrogen bond between the hydroxyl groups of Tyr3 and Tyr45. In the crystal structures this bond is absent, and Tyr45 is instead H-bonded to the side chain of Asp47. This situation holds also in the first model reported by Lian et al. (1992), despite the overall different structure of L4. In this region, as throughout the core, the B1 crystal structures are in close agreement with the B2 crystal structure of Achari et al. (1992). Figure 2 compares the conformations of the aromatic residues Tyr3, Tyr45, and Phe52 in the present crystal structure and the two NMR structures. The main chain at Lys50 is in the left-handed helical conformation, with ϕ, ψ angles of about +60°, +40° in the present crystal structure. In all reported structures of protein G domains, especially the X-ray structures, Gly14 has an extended, nearly flat conformation with ϕ, ψ angles near 180°, 180°. The effect is to remove a β-pleat, producing a bend in the strand that serves to accommodate the core residues Leu(or Ile)7 and Tyr33 (see Figure 3).

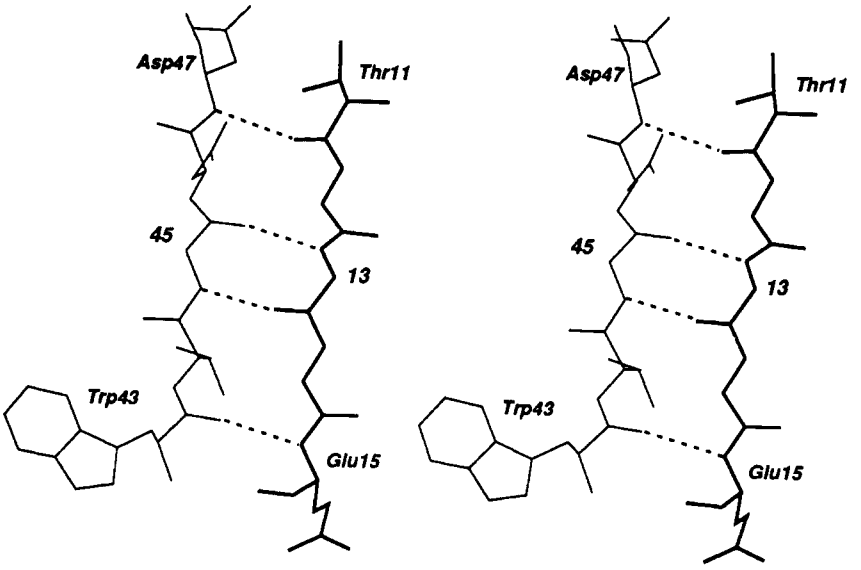

FIGURE 6: Stereogram close-up of the primary crystal contact in both structures. The four hydrogen bonds have standard antiparallel geometry.

Approximately 20 ordered water molecules have been refined with each protein. Their loci are usually not conserved between the two models, often because of adjacency to crystal contacts. However, several strongly occupied waters that interact with only one protein molecule are conserved. Waters have been ranked by "quality", defined as the square of occupancy divided by the thermal parameter (James & Sielecki, 1983). The highest quality water in both structures occupies a distinct location in the threonine-rich outer surface of the β-sheet. This water site is cradled by a pleat of the sheet and surrounded by residues 42, 43, 54, and 55. This location gives the water several polar atoms within hydrogen-bonding distance. The water is close to two hydrogen bonds in an antiparallel β-arrangement, as well as the Oγ atoms of three proximal threonines. A similar site in the adjacent pleat of the sheet holds the second-ranked water in the trigonal structure, and the first water in the B2 structure reported by Achari et al. (1992) (see Figure 4).

In both crystals, the strongest intermolecular contact involves the edges of the four-stranded sheet, effectively forming a continuous sheet, or "β-chain", that runs throughout the crystal. Four hydrogen bonds with typical antiparallel β-geometry give adjacent molecules a 2-fold screw relation, with the screw axis running along the molecular chain. This crystal contact (see Figures 5 and 6) is virtually identical in the orthorhombic and trigonal systems. The contact relates molecules along the **a** lattice direction in the orthorhombic crystals, and along both the **a** and **b** directions in the trigonal crystals. Thus, the **a** lattice constant in both crystals corresponds approximately to twice the width of the β-sheet. [The same contact also occurred in the B2 structure of Achari et al. (1992); see Discussion.] Weaker contacts, involving fewer atoms, interrelate different β-chains to give rise to the crystal symmetry. In the orthorhombic crystals, the chains are all (anti)parallel, producing an unusually high packing density for protein crystals, with a solvent content of only about 26%.

DISCUSSION

The fold, core structure, and overall H-bonding pattern of the IgG-binding domains from protein G are now well established. Among the various structure descriptions from X-ray and NMR, and across several sequence variants, there

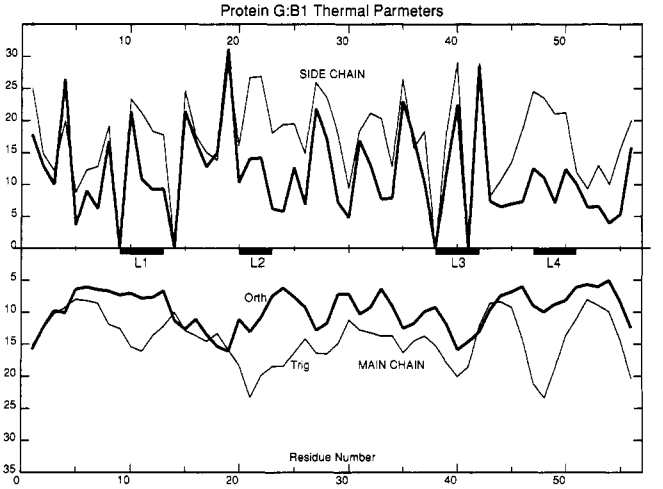

FIGURE 7: Residue-average thermal factors for the two structures for side chains (top half) and the main chain (bottom half) plotted versus sequence: thin lines, trigonal form; thick lines, orthorhombic form. In the side chain (upper) plot, a value of 0 indicates glycine.

is general agreement in the core and backbone hydrogen-bonding pattern, but there remains some question as to the conformation of several regions including the loops. In part, this irresolution results from the greater flexibility of the loops, and it may not be reasonable to expect a precise, unique description of the more mobile components of any protein. While NMR methods may be faithful to the solution "structure" in leaving such regions ambiguous, crystal-based models may give a precise model that is only one of many normal conformations, rigidified by the crystalline environment.

Protein molecules in a crystal must all have a common structure to enable diffraction. Diffraction data report the location of all ordered parts of the protein, without regard for their neighbors. NMR data report interatomic distances. In both methodologies, more observations make a better result, but incomplete data have different effects in the two techniques. While a poorly defined X-ray structure is likely to misconnect its parts and have the correct overall shape, a poorly defined NMR structure is likely to have good local connectivity and the wrong overall shape. In this way, the two methods are complementary. One reason for determining the crystal structure in two different crystal forms is to control for possible

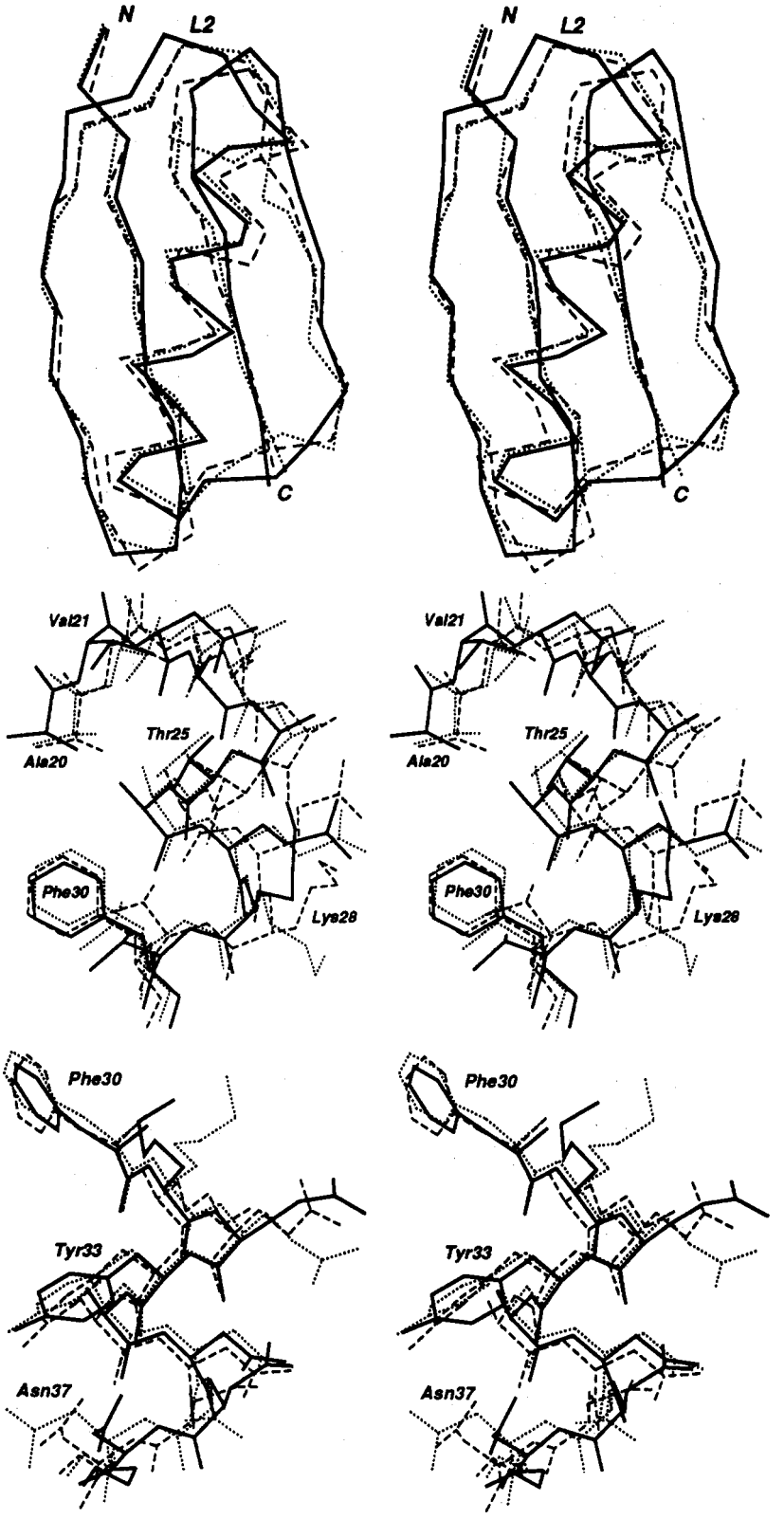

FIGURE 8: Three stereograms of the superposed Cα models in the region of the α-helix: NMR model of Gronenborn et al. (1991), dashed lines; NMR model of Lian et al. (1992), dotted lines; trigonal crystal form, solid lines; (a, top) emphasizing the angular relation of the helix to the sheet; (b, middle) superposed models in the N-terminal half of the helix; (c, bottom) superposed models in the C-terminal half of the helix.

structural perturbations due to crystal packing forces. In the present study, several surface side chains (and water molecules) are affected by the crystal environment, but there is no evidence for a significant effect on the protein backbone. The sheet–edge contact common to both crystal forms does not make the sheet's width or general structure different from those in the NMR models.

In general, the NMR models appear less compact than the X-ray models, as illustrated for loops L1 and L4 in Figure 1. NMR methods require long-range distance restraints to define the relation of loop L4 to the rest of the molecule. The Lian et al. (1992) structure is based on 478 distance restraints, while the Gronenborn et al. (1991) structure utilizes 854 interproton distance restraints. Lian et al. (1992) describe

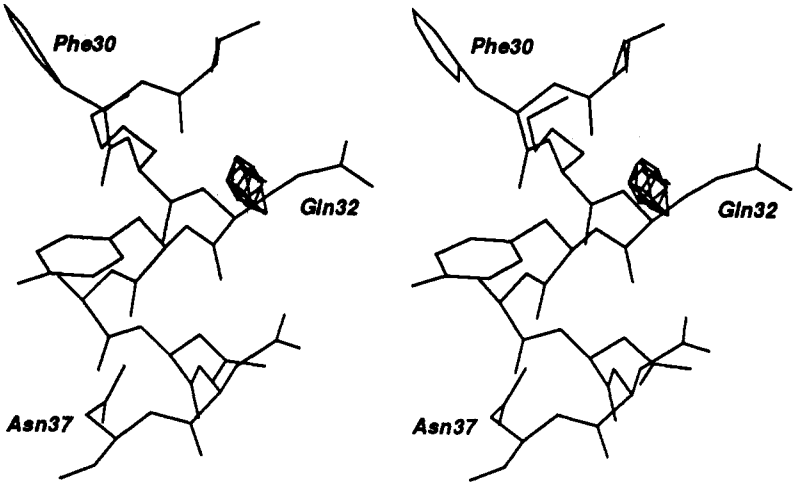

FIGURE 9: Final difference electron density near the helix hydrogen bond 29O–33N. The map is contoured at 1.7σ. This site coincides closely with a water site in the Achari et al. (1992) B2 domain model.

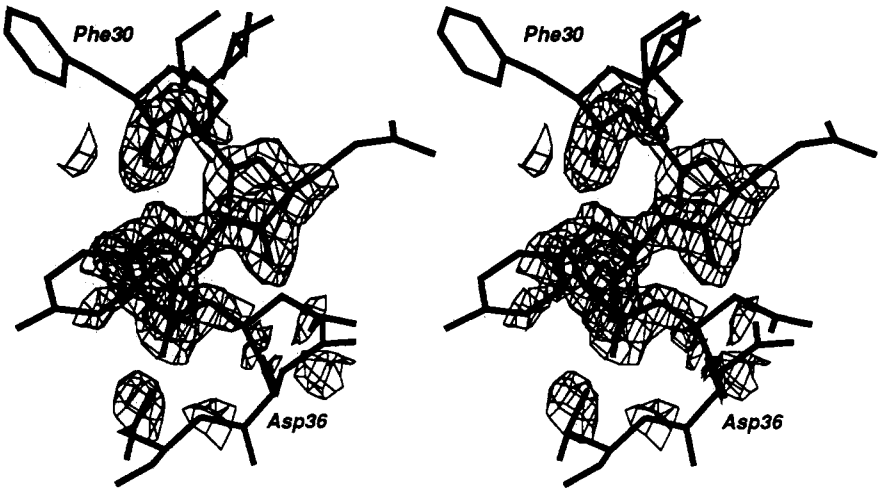

FIGURE 10: Final $2F_o - F_c$ electron density for the helix backbone near Tyr33. The map is contoured at 2σ.

two distinct classes of conformations for L4 among their models, with large differences in the ϕ, ψ angles of residues 47–50, and large differences in local interatomic distances. While the possibility of these alternative conformations and the implied motion between them are interesting, this model is more a result of molecular dynamics simulation than of structure determination. The predominant conformation of L4 described by Lian et al. (1992) is essentially the same as the one in all other reported structures. L4 has the highest main chain thermal parameters in the trigonal crystal structure, and a significant peak in the orthohombic form, indicating that even in the crystal this region is somewhat mobile (see Figure 7).

The variance in the angular relation of the helix to the sheet was at first suspected to be due to the B1/B2 residue substitutions, but it now appears that this is not the case. The model of Gronenborn et al. (1991) is unique in this parameter; in particular it disagrees with other NMR models (see Figure 8a) and with the present X-ray models. Dissecting the correlates of the 8–10° difference at the level of individual amino acids reveals that, because the ends of the helix contribute no large side chains to the core, the rotation does not demand any striking changes in side chain conformation. Thus, the observed main chain displacements on the order of 1 Å at the ends of the helix may involve little energetic cost. The side chains of Asn37 and Leu12 are in van der Waals contact, implying that the apparent rotational mobility of the helix may be dynamically coupled with observed displacements in L1. In addition to its unique rotation with respect to the sheet, the Gronenborn et al. (1991) helix is somewhat distorted, as described in Clore and Gronenborn (1992). Careful consideration must be given to the sample conditions and techniques used in deducing the several models, in an effort to understand the observed differences in helix structure. In comparing the structures of the helix in various models, the following features are noted. Several of the charged residues adopt varying rotamers, as may be expected for solvent-interacting Glu and Lys side chains. Two possibly significant conformational differences are in the Thr25 rotamer (see Figure 8b) and in the torsion for the 36–37 peptide (see Figure 8c). Regarding Thr25, the Gronenborn et al. (1991) rotamer is rare for threonine in α-helices, and unique among the reported models of either domain.

The two water sites deduced from solution NMR data and described by Clore and Gronenborn (1992) were examined in the present structures. Both sites in both structures are sufficiently distant from crystal contacts that normal hydration may be expected. The crystallographically refined water sites do not include waters at these positions, indicating there are no well-ordered waters at these locations. However, some electron density can be observed at these sites at the 2.4σ level (and below) in final difference maps. It thus appears that water does bind at both these positions, but not with sufficient occupancy to warrant inclusion in the refined model. Figure

9 shows the final difference map for the low-density water site near the 29O–33N hydrogen bond, while Figure 10 presents the final $2F_o - F_c$ map in the backbone region of the adjacent helix. Two other hydrogen bonds of the helix, 24O–28N and 32O–36N, do bind water molecules in the crystallographic models. Both models also feature a water site bound to the amide of Thr11 (as does the Achari et al. (1992) B2 model). The waters that appear to bind the tightest are at the two sites described in the previous section, and depicted in Figure 4. These waters refine with full occupancy and thermal parameters comparable to those of the best-ordered protein atoms. Figure 1 in Clore and Gronenborn (1992) shows several resonances attributed to waters near threonines 44, 51, and 55 (with Thr53 conspicuously absent) and likely arising from these crystallographically observed waters.

The higher packing density in the orthorhombic crystals places many more atoms in crystal contacts (118 as opposed to 88, using a 3.8-Å distance cutoff), providing a succinct explanation for the generally lower temperature factors in that structure (the average temperature factor also depends on the resolution and completeness of diffraction data). It is noteworthy that both the present structures, as well as the B2 structure of Achari et al. (1992), feature the same predominant crystal contact, namely, the sheet–edge association. This type of interaction has also been reported in the binding of a protein G domain to Fab (Derrick & Wigley, 1992). The binding of protein G to Fc could also involve a sheet–edge interaction, but this is unlikely, since the known binding site for protein A (Diesenhofer, 1983), with which protein G competes for binding, is not close to a sheer–edge in the Fc. More likely, protein G interacts with Fc by its helix in a manner similar to that of the all-helix protein A. Recent evidence from NMR solution studies supports this hypothesis (Gronenborn & Clore, 1993). The structures of two crystal forms of the B2 domain (consisting of 56 residues, without the N- and C-terminal extensions of the Achari et al. (1992) structure) are currently being refined and lack the sheet–edge contact, proving that it is not universal in protein G crystals.

ACKNOWLEDGMENT

The authors gratefully acknowledge helpful discussions with John Orban and Joseph Nachmann in the development of this paper. Certain commercial equipment, instruments, and materials are identified in this paper in order to specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment are necessarily the best available for the purpose.

REFERENCES

Achari, A., Hale, S. P., Howard, A. J., Clore, G. M., Gronenborn, A. M., Hardman, K. D., & Whitlow, M. (1992) *Biochemistry 31*, 10449–10457.
Agarwal, R. C. (1978) *Acta Crystallogr. A34*, 791–809.
Akerstron, B., & Bjork, L. (1986) *J. Biol. Chem. 261*, 10240–10247.
Alexander, P., Fahnestock, S. R., Lee, T., Orban, J., & Bryan, P. N. (1992) *Biochemistry 31*, 3597–3603.
Alexander, P., Orban, J., & Bryan, P. N. (1992) *Biochemistry 31*, 7243–7248.
Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.
Brunger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458–460.
Burley, S. K., & Petsko, G. A. (1985) *Science 229*, 23–28.
Clore, G. M., & Gronenborn, A. M. (1992) *J. Mol. Biol. 223*, 853–856.
Cohen, G. H. (1986) *J. Mol. Biol. 190*, 593–604.
Derrick, J. P., & Wigley, D. B. (1992) *Nature 359*, 752–754.
Fahnestock, S. R., Alexander, P., Nagle, J., & Filpula, D. (1986) *J. Bacteriol. 167*, 870–880.
Finzel, B. C. (1987) *J. Appl. Crystallogr. 20*, 53–55.
Gronenborn, A. M., & Clore, G. M. (1993) *J. Mol. Biol. 233*, 331–335.
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., & Konnert, J. H. (1980) in *Computing in Crystallography* (Diamond, R., Ramaseshan, S., & Venkatesan, K., Eds.) pp 13.01–13.23, Indian Institute of Science, Bangalore.
Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H., & Salemme, F. R. (1987) *J. Appl. Crystallogr. 20*, 383–387.
James, M. N. G., & Sielecki, A. R. (1983) *J. Mol. Biol. 163*, 299–361.
Jones, T. A. (1987) *J. Appl. Crystallogr. 11*, 268–272.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.
Lian, L.-Y., Derrick, J. P., Sutcliffe, M. J., Yang, J. C., & Roberts, G. C. K. (1992) *J. Mol. Biol. 228*, 1219–1234.
Myhre, E. B., & Kronvall, G. (1977) *Infect. Immun. 17*, 475–482.
Orban, J., Alexander, P., & Bryan, P. (1992) *Biochemistry 31*, 3604–3611.
Sheriff, S. (1987) *J. Appl. Crystallogr. 20*, 55–57.
Ten Eyck, L. F. (1973) *Acta Crystallogr. A29*, 183–191.