solution in an ultrasonic bath for 10 and 5 min, respectively, and then rinsed in distilled water in an ultrasonic bath for 5 min. From this material, approximately 250 whole-rock grains were picked for irradiation. Samples for irradiation were encapsulated in aluminum cups and arranged in a known geometry along with mineral standards. The sample package was placed in a cadmium lined, 2.5-cm-diameter aluminum tube. Then the sample set was irradiated 10 min at 8 MW in the hydraulic rabbit of the Los Alamos National Laboratory Omega West reactor by fast neutrons to produce the reaction $^{39}K(n,p)$ ^{39}Ar. After irradiation, the samples were transferred to a copper sample holder and loaded into the Ar-extraction system. Fusion was induced by a 6-W continuous Ar-ion laser beam focused to a 2- to 3-mm spot, applied for 30 to 60 s. The gases released from the grains were then scrubbed for reactive species (CO_2, CO, and N_2) by exposure to a 150°C Zr-Fe-V alloy getter for 3 to 5 min. The remaining inert gases, principally Ar, were then admitted to the mass spectrometer, and the argon-isotopic ratios were determined. The mass spectrometer was operated in static mode with the use of automated data-collection procedures. The age is then calculated from the $^{40}Ar/^{39}Ar$ ratio after all interfering Ar-isotopes from atmospheric contamination and undesirable neutron reactions with Ca and K are corrected [N. R. Brereton, *Earth Planet. Sci. Lett.* **8**, 427 (1971); G. B. Dalrymple and M. A. Lanphere, *ibid.* **12**, 300 (1971)].
15. J. R. Taylor, *An Introduction of Error Analysis, the Study of Uncertainties in Physical Measurements* (Oxford Univ. Press, New York, 1982). Weighted means, in the tables, are calculated by multiplying the individual analyses by 1 over the variance. The uncertainty of the weighted average is obtained by pooling the variances. The preferred age given for any set of data is the weighted average because the more precise analyses are given more weight in the determination of the average age.
16. The $^{40}Ar/^{36}Ar$ ratio of atmospheric Ar is 295.
17. P. L. McFadden and D. L. Jones, *Geophys. J. R. Astron. Soc.* **67**, 53 (1981).
18. The geomagnetic field is not constant in direction or intensity at any given location through time, but instead undergoes a secular variation at a rate of about 4° per century [D. E. Champion and E. M. Shoemaker, *NASA Tech. Mem. 78,436* (1977), p. 7; D. E. Champion, *U.S. Geol. Surv. Open-File Rep. 80-824* (1980)].
19. This value (0.029 million years ago) is the minimum difference in age between the two flows that is detectable at the 95% confidence level. See p. 120 of G. B. Dalrymple and M. A. Lanphere, [*Potassium-Argon Dating* (Freeman, San Francisco, 1969)].
20. B. J. Szabo *et al.*, *U.S. Geol. Surv. Open-File Rep. 81-119* (1981).
21. The principal source of error in U-Th dating lies in the assumption of a closed system [G. H. Curtis, *U.S. Nucl. Reg. Comm. Rep. NUREG/CR-2382* (1981)]. It has been shown that U-Th dates on soil carbonates from soils formed on K-Ar dated basaltic lava flows were consistently younger than the K-Ar dates by up to an order of magnitude [J. L. Slate, thesis, University of Arizona (1985)].
22. W. C. Swadley *et al.*, *U.S. Geol. Surv. Open-File Rep. 84-788* (1984).
23. The manuscript was improved by comments by G. B. Dalrymple, M. A. Lamphere, D. Noble, E. H. McKee, and J. C. Dohrenwend.

22 January 1991; accepted 3 June 1991

A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein G

ANGELA M. GRONENBORN,* DAVID R. FILPULA, NINA Z. ESSIG, ANIRUDDHA ACHARI, MARC WHITLOW, PAUL T. WINGFIELD, G. MARIUS CLORE*

The high-resolution three-dimensional structure of a single immunoglobulin binding domain (B1, which comprises 56 residues including the NH_2-terminal Met) of protein G from group G *Streptococcus* has been determined in solution by nuclear magnetic resonance spectroscopy on the basis of 1058 experimental restraints. The average atomic root-mean-square distribution about the mean coordinate positions is 0.27 angstrom (Å) for the backbone atoms, 0.65 Å for all atoms, and 0.39 Å for atoms excluding disordered surface side chains. The structure has no disulfide bridges and is composed of a four-stranded β sheet, on top of which lies a long helix. The central two strands (β1 and β4), comprising the NH_2- and COOH-termini, are parallel, and the outer two strands (β2 and β3) are connected by the helix in a +3x crossover. This novel topology (−1, +3x, −1), coupled with an extensive hydrogen-bonding network and a tightly packed and buried hydrophobic core, is probably responsible for the extreme thermal stability of this small domain (reversible melting at 87°C).

PROTEIN G IS A LARGE MULTIDOMAIN cell surface protein of group G *Streptococcus*, which is thought to help the organism evade the host defenses through its protein binding properties (*1*). A repeating 55-residue domain binds to the F_c region of immunoglobulin G (IgG) and to α2-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 one of these domains, known as B1 (*2*), reveals extreme thermal stability with a melting temperature (T_m) of 87°C and completely reversible thermal denaturation (*3*). Further, the unfolding transition on urea gradient gel electrophoresis (*4*) cannot be observed in full as the protein remains native up to ~8 M urea. These features are highly unusual considering the small size of the domain and the absence of any disulfide bridges or tightly bound prosthetic group. For comparison, the average T_m in a recent compilation of the thermodynamic parameters of unfolding for a large number of proteins is ~63°C, and only three proteins in this collection are more stable than the B1 domain, namely, the Ca^{2+} bound form of parvalbumin (90°C), neurotoxin II (96°C), and bovine pancreatic trypsin inhibitor (BPTI) (100°C) (*5*). The potential importance of the B1 domain of protein G as an analytical tool in immunology, together with its extreme physicochemical properties, prompted us to undertake the determination of its three-dimensional (3D) structure in solution by NMR (*6*) spectroscopy.

The ^1H-NMR spectrum of the B1 domain was assigned by using conventional 2D NMR methodology (*7*) on a Bruker AM600 spectrometer. Spin systems were delineated by using PE.COSY (*8*) and HOHAHA (*9*) spectroscopy to demonstrate direct and relayed through-bond connectivities, respectively, while NOESY (*10*) spectroscopy was used to identify through-space (<5 Å) interactions. The assignment was slightly more complex than expected because of the duplication of the resonances for 22 residues arising from the presence of two species in a ~3:7 mixture with and without post-translational processing of the NH_2-terminal Met, respectively. The pattern and relative intensities of the NOEs for the two species, however, were the same within experimental error. Approximate interproton distance restraints were obtained from NOESY spectra recorded at mixing times of 50, 100, and 150 ms, and grouped into three classes, 1.8 to 2.7 Å, 1.8 to 3.3 Å (1.8 to 3.5 Å for distances involving NH protons), and 1.8 to 5.0 Å, which correspond to strong, medium, and weak NOEs, respectively (*11, 12*). To help resolve ambiguities in NOE assignments, spectra were recorded at 25° and 41°C. The $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$ coupling constants were measured from PE.COSY spectra. Stereospecific assignments and φ, ψ, and χ1 torsion angle restraints were obtained

A. M. Gronenborn and G. M. Clore, Laboratory of Chemical Physics, Building 2, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892.
D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, Genex Corporation, 16020 Industrial Drive, Gaithersburg, MD 20877.
P. T. Wingfield, Protein Expression Laboratory, Building 6, Office of the Director, National Institutes of Health, Bethesda, MD 20892.

*To whom correspondence should be addressed.

with the conformational grid search program STEREOSEARCH on the basis of the coupling constants and the intraresidue and sequential interresidue NOEs involving the NH, CαH, and CβH protons (*13, 14*). Structures were calculated with the use of the hybrid distance geometry-simulated annealing method of Nilges *et al.* (*15, 16*) with a few minor modifications (*17*). An iterative strategy was used (*17*). At the end of this procedure, all 30 slowly exchanging NH protons could be accounted for by hydrogen bonds, and stereospecific assignments were obtained for 24 of the 30 β-methylene groups, for the α-methylene groups of all four Gly residues, and for the methyl groups of three of the four Val residues and two of the three Leu residues. Although the chemical shifts of the methyl groups of Val^{21} and Leu^{7} are degenerate, many of the NOEs involving these methyl protons could be assigned to one or other group on the basis of the initial structures (*17*). The same was also true of the aromatic ring protons on either side of the ring, which are degenerate owing to rapid ring flipping (*17*). The χ_1 side chain torsion angles for the six residues whose β-methylene protons were not stereoassigned are disordered on the basis of $^3J_{\alpha\beta}$ coupling constant values of 6 to 7 Hz. The final structure calculations were based on a total of 1058 experimental restraints made up of 854 interproton distance restraints, 144 torsion angle restraints (for 54 ϕ, 51 ψ, and 39 χ_1 angles), and 60 distance restraints for 30 hydrogen bonds associated with slowly exchanging backbone NH protons (*18*).

A total of 60 simulated annealing (SA) structures were calculated (Table 1) (*18*). The structure is exceptionally well defined with an atomic root-mean-square (rms) distribution about the mean coordinate positions of 0.27 Å for the backbone atoms and 0.39 Å for all atoms excluding disordered surface side chains, and an average angular ϕ and ψ rms deviation between the structures of 8.9° ± 6.4°, respectively (Table 2). This accuracy reflects the large number of experimental NMR restraints (*19*) with an average of ~30 restraints affecting the conformation of each residue. The stereochemistry of the structures is good as judged by several criteria: (i) all of the ϕ, ψ angles lie within the allowed regions of a Ramachandran plot; (ii) the nonbonded contacts are characterized by a large negative Lennard-Jones van der Waals energy; (iii) and the deviations from idealized covalent geometry are small. A best fit superposition of the backbone atoms is shown in Fig. 1, and a schematic representation of the structure, together with the hydrogen-bonding pattern, is shown in Fig. 2.

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. 1 and 2A). The two central strands of the sheet, β1 (residues 1 to 8) and β4 (residues 50 to 56), are parallel. The two outer strands, β2 (residues 13 to 20) and β3 (residues 42 to 47), are antiparallel to β1 and β4, respectively. Strands β1 and β2 are connected by a type I turn (residues 9 to 12), whereas strands β3 and β4 are connected by an unusual six 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. Interestingly, residues 46 to 51 superimpose

Table 1. Structural statistics. The notation of the structures is as follows: ⟨SA⟩ are the 60 simulated annealing structures; $\overline{SA}$ is the mean structure obtained by averaging the coordinates of the individual SA structures best fitted to each other; and $(\overline{SA})r$ is the restrained minimized mean structure obtained from $\overline{SA}$. The number of terms for the various restraints is given in parentheses.

Structural statistics	⟨SA⟩	$(\overline{SA})r$
Rms distance deviations (Å)*†		
All (914)	0.026 ± 0.007	0.028
Interproton distances		
Interresidue short range ($\lvert i-j \rvert \le 5$) (254)	0.020 ± 0.001	0.021
Interresidue long range ($\lvert i-j \rvert > 5$) (291)	0.035 ± 0.001	0.040
Intraresidue (309)	0.021 ± 0.001	0.020
Hydrogen bonds (60)‡	0.019 ± 0.001	0.021
Rms dihedral deviations (degrees) (144)*§	0.108 ± 0.026	0.082
F_{NOE} (kcal mol^{-1})‖	18.7 ± 0.8	21.9
F_{tor} (kcal mol^{-1})‖	0.121 ± 0.048	0.064
F_{repel} (kcal mol^{-1})‖	11.4 ± 0.6	15.9
$E_{L\text{-}J}$ (kcal mol^{-1})¶	−210 ± 7	−186
Deviations from idealized covalent geometry		
Bonds (Å) (861)	0.003 ± 0.0002	0.004
Angles (degrees) (1549)	1.930 ± 0.001	2.013
Impropers (degrees) (347)#	0.561 ± 0.007	0.585

*Root-mean-square deviations from experimental restraints. †None of the structures exhibit distance violations greater than 0.3 Å or dihedral angle violations greater than 1°. ‡Each hydrogen bond is characterized by two distance restraints: $r_{NH\text{-}O} \le 2.3$ Å and $r_{N\text{-}O} = 2.5$ to 3.3 Å. All hydrogen-bonding restraints involve slowly exchanging backbone amide protons. §The torsion angle restraints comprise 54 ϕ, 51 ψ, and 39 χ_1 angles. ‖The values of the square-well NOE (F_{NOE}) and torsion angle (F_{tor}) potentials [compare with equations 2 and 3 in (*11*)] are calculated with force constants of 50 kcal mol^{-1} $Å^{-2}$ and 200 kcal mol^{-1} rad^{-2}, respectively. The value of the quartic van der Waals repulsion term F_{rep} [compare with equation 5 in (*15*)] is calculated with a force constant of 4 kcal mol^{-1} $Å^{-4}$ with the hard-sphere van der Waals radii set to 0.8 times their standard values. ¶$E_{L\text{-}J}$ is the Lennard-Jones van der Waals energy calculated with the CHARMM (*28*) empirical energy function and is *not* included in the target function for simulated annealing or restrained minimization. #The improper torsion restraints serve to maintain planarity and chirality.

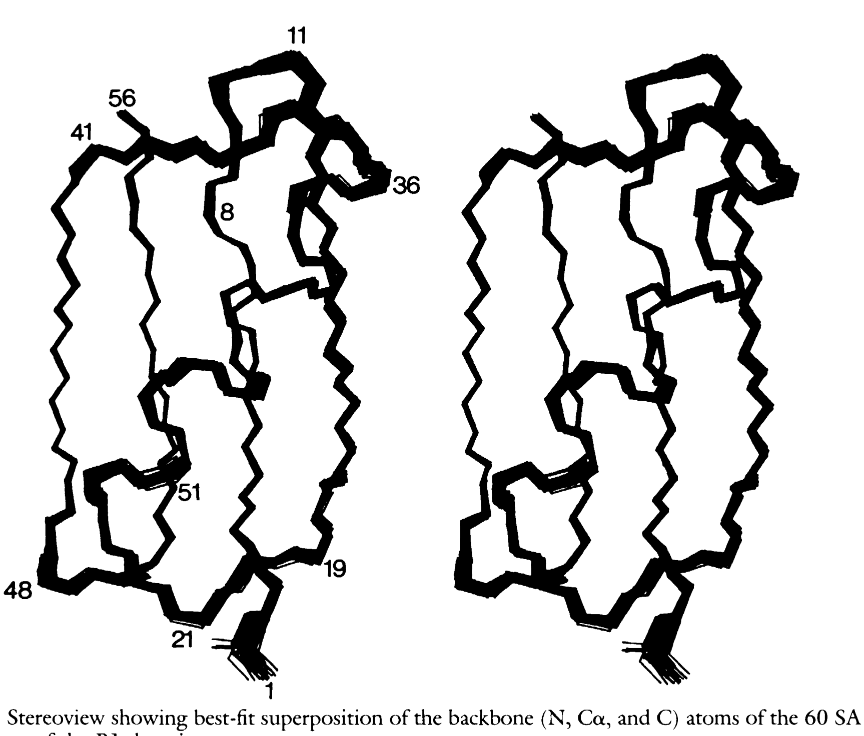

Fig. 1. Stereoview showing best-fit superposition of the backbone (N, Cα, and C) atoms of the 60 SA structures of the B1 domain.

Table 2. Atomic rms differences.

Atomic rms differences (Å)	Backbone atoms	All atoms	All atoms excluding disordered side chains*
⟨SA⟩ versus $\overline{SA}$	0.27 ± 0.03	0.65 ± 0.05	0.39 ± 0.03
⟨SA⟩ versus ($\overline{SA}$)r	0.29 ± 0.03	0.74 ± 0.06	0.45 ± 0.041
($\overline{SA}$)r versus $\overline{SA}$	0.12	0.36	0.22

*The disordered surface side chains that were excluded are: Met[1] from the Cβ position onward, Lys[4] from Cδ, Lys[10] from Cγ, Lys[13] from Cδ, Glu[15] from Cγ, Glu[19] from Cγ, Lys[28] from Cγ, Gln[32] from Cγ, Lys[31] from Cδ, Glu[42] from Cδ, and Glu[56] from Cδ.

within <0.4 Å on residues 235/235A-239 of *Streptomyces griseus* proteinase A, residues 27 to 32 of crambin, and residues 95 to 100 of trypsin (*20*). The latter has a Leu with a φ angle of ~60° at the position equivalent to Lys[50], whereas the other two proteins have the more usual Gly in this position. The two outer strands, β2 and β3, of the sheet are connected through a long helix (residues 22 to 37) and a short extended structure (residues 38 to 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 five residues are tightened into a 3_{10} helix characterized exclusively by two CO(i)-NH(i+3) hydrogen bonds.

The overall structure of the B1 domain exhibits two highly unusual features. The first is the four-stranded sheet with a parallel pair of strands in the middle; the second is the +3x crossover in which the outer two β strands are connected by the helix. Thus, in the notation of Richardson (*21*), the sheet displays a −1, +3x, −1 topology. A search of the Brookhaven Protein Data Bank reveals that this topology is novel and has not been observed before. In contrast, +2x crossover connections in a triple-stranded sheet are common (*21, 22*).

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 of the residues in regular secondary structure ensures a large number of stabilizing hydrogen bonds (Fig. 2B). The average secondary content of proteins is typically ~75% (*21*) compared to ~95% in the case of the B1 domain of protein G. There are a total of 45 hydrogen bonds, of which 41 are backbone-backbone, 3 are side chain–backbone, and 1 is side chain–side chain. More than 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 turn of the helix and the turn between β3 and β4, respectively. 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] (Fig. 3B).

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 (SFE) of folding (*23*) of −55 ± 2 kcal mol^{-1}. This value is comparable to that for another small, stable protein, BPTI, and is ~15% more negative than the predicted SFE for a protein of this size (*24*). 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. 3. 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 Thr 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, we note that these residues are located either on the outer strands (Thr[16] and Thr[18] in β2 and Trp[43] and Tyr[45] in β3) or at the beginning of a strand (Tyr[3] in β1). 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.

The above features, while clearly important, are probably not sufficient in their own right to fully account for the stability of this protein. It is quite possible that the novel topology of the polypeptide fold itself may play a key role in this regard. In particular, as the NH_2- and COOH-terminal strands, β1 and β4, constitute the two internal strands of the four-stranded β sheet, unfolding can presumably only be accomplished by disruption of the hydrogen bonds between these strands. Further, the interaction and relative orientations of strands β1 and β4 are not only determined by backbone hydrogen bonding but also by an extensive network of hydrophobic interactions with the overlaying helix (Fig. 3). Thus, it seems likely that unfolding cannot proceed in a sequential manner, but rather occurs in a highly cooperative fashion involving all parts of the structure simultaneously. Further, while most proteins may unfold in a cooperative manner, the topology of the B1 domain prevents fraying of the NH_2- and COOH-termini that could potentially initiate the unfolding process.

The association constants for the binding of IgG to the B1 and B2 domains of protein G are 0.3 and 2.1 × 10^8 M^{-1}, respectively (*25*). In an attempt to deduce which residues might be important in IgG binding, we compared the

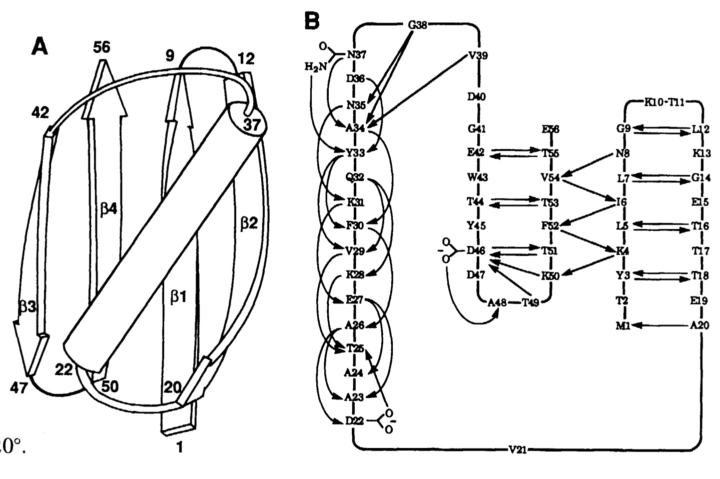

Fig. 2. (**A**) Schematic ribbon drawing representation of the polypeptide fold of the B1 domain; (**B**) diagrammatic representation of hydrogen bonding within the B1 domain. The criteria used for defining a hydrogen bond are that the distance between acceptor and donor heavy atoms is ≤3.3 Å and that the angle between the acceptor heavy atom, the hydrogen, and the donor heavy atom is ≥120°.

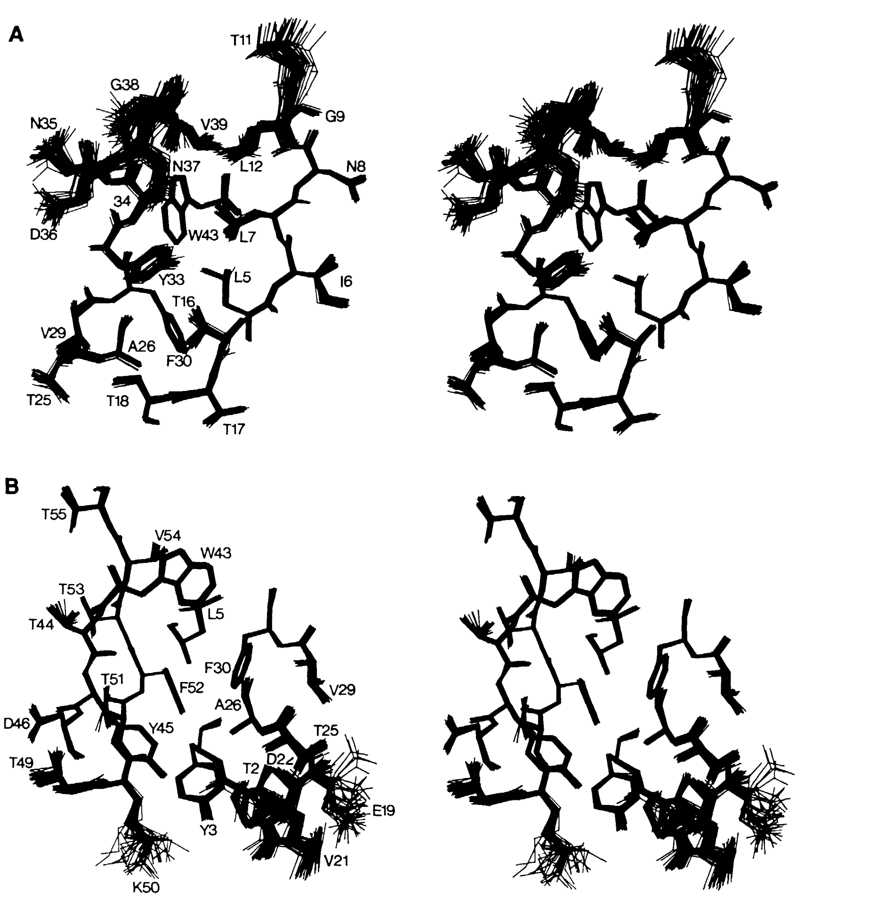

Fig. 3. Stereoviews showing best-fit superpositions of all atoms (excluding protons) of the 60 SA structures for two regions of the protein, illustrating the dense packing of hydrophobic residues within the protein interior. Interactions of residues from strands β1, β2, and β3 and the helix are shown in (**A**), and from strands β1, β3, and β4 and the helix in (**B**).

sequences of B1 and B2. There are six amino acid substitutions between the two domains: I6V, L7I, E19K, A24E, V29A, and E42V (*26*). Leu[7] is completely buried and cannot therefore 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. It should, however, be borne in mind that the difference in affinity for the B1 and B2 domains is small and could also arise from subtle conformational changes rather than direct effects associated with substitution of residues that contact the IgG. We note that there is *no* sequence or structural homology between the B1 domain and the 58-residue IgG binding domain of staphylococcal protein A (*27*). This result, together with the observed differences in specificity (*27*), suggests that the mode of IgG binding for these two proteins is quite distinct. Further characterization would require structural studies of the B1 domain–F_c complex.

The high-resolution solution structure of the small B1 domain of protein G presented in this report reveals a highly compact globular protein with a novel topology of secondary structural elements. The unusual stability of this small protein domain, which has no disulfide bridges, suggests that it may serve as a model for protein folding studies and as a template for ab initio protein design and protein engineering of novel functional properties.

REFERENCES AND NOTES

1. L. Björck and G. Kronvall, *J. Immunol.* **133**, 969 (1984); S. R. Fahnestock, P. Alexander, J. Nagle, D. Filpula, *J. Bacteriol.* **167**, 870 (1986); B. Guss *et al.*, *EMBO J.* **5**, 1567 (1986); B. Åkerström and L. Björck, *J. Biol. Chem.* **261**, 10240 (1986); S. R. Fahnestock, *Trends Biotechnol.* **5**, 79 (1987); U. Sjöbring, J. Tronar, A. Grubb, B. Åkerström, L. Björck, *J. Immunol.* **143**, 2948 (1989); S. R. Fahnestock, P. Alexander, D. Filpula, J. Nagle, in *Bacterial Immunoglobulin Binding Proteins* (Academic Press, New York, 1990), vol. 1, pp. 133–148; U. Sjöbring, L. Björck, W. Kastern, *J. Biol. Chem.* **266**, 399 (1991).
2. The B1 domain of protein G was expressed as a soluble protein at over 20% of total cell protein and purified from an *Escherichia coli* strain containing a temperature-sensitive repressor and a plasmid with the synthetic B1 gene under control of the hybrid λ phage promoter O_L/P_R. The B1 protein was affinity-purified on an IgG–Sepharose 6B Fast Flow column from a frozen cell paste. Two species of protein are present, namely the 55-residue B1 domain of protein G and a 56-residue version with an unprocessed NH_2-terminal Met in a ratio of ~3:7.
3. Thermal stability was measured by differential scanning calorimetry with a MC-2 (Microcal, Inc. Northampton, MA) calorimeter. Data were analyzed by a least squares curve fitting procedure assuming a two-state model [J. M. Sturtevant, *Annu. Rev. Phys. Chem.* **38**, 463 (1987)]. The melting curve exhibits a single well-defined transition at 87°C, and no differences in stability could be detected between the two species (with and without the NH_2-terminal Met). In addition, thermal denaturation was fully reversible as judged by ultraviolet-visible and NMR spectroscopy, as well as by repeated calorimetric scans.
4. W. Pfeil, in *Thermodynamic Data for Biochemistry and Biotechnology*, H.-J. Hinz, Ed. (Springer, Berlin, 1986), pp. 349–376.
5. T. E. Creighton, *Methods Enzymnol.* **131**, 157 (1986).
6. Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy; PE.COSY, primitive exclusive correlated spectroscopy; and HOHAHA, homonuclear Hartmann-Hahn spectroscopy.
7. K. Wüthrich, *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986); G. M. Clore and A. M. Gronenborn, *Protein Eng.* **1**, 275 (1987); *CRC Crit. Rev. Biochem. Mol. Biol.* **24**, 479 (1989); K. Wüthrich, *Acc. Chem. Res.* **22**, 36 (1989).
8. L. Mueller, *J. Magn. Reson.* **72**, 191 (1987); D. Marion and A. Bax, *ibid.* **80**, 528 (1988).
9. A. Bax, *Methods Enzymol.* **176**, 151 (1989).
10. J. Jeener, B. H. Meier, P. Bachmann, R. R. Ernst, *J. Chem. Phys.* **71**, 4546 (1979).
11. G. M. Clore *et al.*, *EMBO J.* **5**, 2729 (1986).
12. Upper distance limits for distances involving methyl protons and nonstereospecifically assigned methylene protons were corrected appropriately for center averaging [K. Wüthrich, M. Billetter, W. Braun, *J. Mol. Biol.* **169**, 949 (1983)].
13. M. Nilges, G. M. Clore, A. M. Gronenborn, *Biopolymers* **29**, 813 (1990).
14. The minimum ranges used for φ, ψ, and χ_1 torsion angle restraints were ±30°, ±50°, and ±20°, respectively, as described in (*16*).
15. M. Nilges, G. M. Clore, A. M. Gronenborn, *FEBS Lett.* **229**, 317 (1988).
16. The hybrid distance geometry-simulated annealing protocol for (*15*) makes use of the program DISGEO [T. F. Havel and K. Wüthrich, *Bull. Math. Biol.* **46**, 673 (1984); T. F. Havel, *Quantum Chemistry Program Exchange No. 507* (Indiana University, Bloomington, IN, 1986)] to calculate an initial set of substructures comprising about one third of the atoms by projection from *n*-dimensional distance space into cartesian coordinate space, followed by simulated annealing with all atoms with the program XPLOR [A. T. Brünger, G. M. Clore, A. M. Gronenborn, M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3801 (1986); A. T. Brünger, *XPLOR Manual* (Yale University, New Haven, CT, 1988)]. The target function that is minimized during simulated annealing (as well as conventional Powell minimization) comprises only quartic harmonic potential terms for covalent geometry (that is, bonds, angles, planes, and chirality), square-well quadratic potentials for the experimental distance and torsion angle restraints, and a quartic van der Waals repulsion term for the nonbonded contacts (*15*). All peptide bonds were restrained to be trans. There are *no* hydrogen bonding, electrostatic, or 6-12 Lennard-Jones potential terms in the target function.
17. P. J. Kraulis *et al.*, *Biochemistry* **28**, 7241 (1989); G. M. Clore, E. Appella, M. Yamada, K. Matsushima, A. M. Gronenborn, *ibid.* **29**, 1689 (1990); G. M. Clore, P. T. Wingfield, A. M. Gronenborn, *ibid.* **30**, 2315 (1991); J. D. Forman-Kay, G. M. Clore, P. T. Wingfield, A. M. Gronenborn, *ibid.*, p. 2685.
18. The coordinates of the 60 SA structures and of the restrained minimized mean structure ($\overline{SA}$)r, together with the experimental NMR restraints and ^{1}H-NMR resonance assignments, have been deposited in the Brookhaven Protein Data Bank (PDB accession numbers 1GB1, 2GB1, and R1GB1MR).
19. G. M. Clore and A. M. Gronenborn, *Annu. Rev. Biophys. Chem.* **20**, 29 (1991); *Science* **252**, 1390 (1991).
20. M. N. G. James, A. R. Sielecki, G. D. Brayer, L. T.

J. Delbaere, L. A. Bauer, *J. Mol. Biol.* **144**, 43 (1980); W. A. Hendrickson and M. M. Teeter, *Nature* **290**, 107 (1981); J. Walter *et al.*, *Acta Crystallogr.* **B38**, 1462 (1982); T. A. Jones and S. Thirup, *EMBO J.* **5**, 819 (1986).
21. J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981); G. E. Schulz and R. H. Schirmer, *Principles of Protein Structure* (Springer-Verlag, New York, 1979).
22. C. Chothia and A. V. Finkelstein, *Annu. Rev. Biochem.* **59**, 1007 (1990).
23. D. Eisenberg and A. D. McLachlan, *Nature* **319**, 199 (1986).
24. L. Cliche, L. M. Gregoret, F. E. Cohen, P. A. Kollman, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3240 (1990).
25. S. Fahnestock, personal communication.
26. Abbreviations for the amino acid residues are: 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; and Y, Tyr.
27. J. J. Langone, *Adv. Immunol.* **32**, 157 (1982); J. Deisenhofer, *Biochemistry* **20**, 2361 (1981); B. Nilsson *et al.*, *Protein Eng.* **1**, 107 (1987).
28. B. R. Brooks *et al.*, *J. Comput. Chem.* **4**, 187 (1983).
29. Supported by the AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health (G.M.C., A.M.G., and P.T.W.). We thank J. Richardson and R. Feldman for helpful discussions.

22 March 1991; accepted 10 May 1991

Identification of FAP Locus Genes from Chromosome 5q21

KENNETH W. KINZLER, MEF C. NILBERT, LI-KUO SU, BERT VOGELSTEIN,* TRACY M. BRYAN, DANIEL B. LEVY, KELLY J. SMITH, ANTONETTE C. PREISINGER, PHILIP HEDGE, DOUGLAS MCKECHNIE, RACHEL FINNIEAR, ALEX MARKHAM, JOHN GROFFEN, MARK S. BOGUSKI, STEPHEN F. ALTSCHUL, AKIRA HORII, HIROSHI ANDO, YASUO MIYOSHI, YOSHIO MIKI, ISAMU NISHISHO, YUSUKE NAKAMURA

Recent studies suggest that one or more genes on chromosome 5q21 are important for the development of colorectal cancers, particularly those associated with familial adenomatous polyposis (FAP). To facilitate the identification of genes from this locus, a portion of the region that is tightly linked to FAP was cloned. Six contiguous stretches of sequence (contigs) containing approximately 5.5 Mb of DNA were isolated. Subclones from these contigs were used to identify and position six genes, all of which were expressed in normal colonic mucosa. Two of these genes (APC and MCC) are likely to contribute to colorectal tumorigenesis. The MCC gene had previously been identified by virtue of its mutation in human colorectal tumors. The APC gene was identified in a contig initiated from the MCC gene and was found to encode an unusually large protein. These two closely spaced genes encode proteins predicted to contain coiled-coil regions. Both genes were also expressed in a wide variety of tissues. Further studies of MCC and APC and their potential interaction should prove useful for understanding colorectal neoplasia.

FAMILIAL ADENOMATOUS POLYPOSIS (FAP) is one of the most common autosomal dominant diseases leading to cancer predisposition, affecting nearly 0.01% of the American, British, and Japanese populations (*1*). Affected patients usually develop numerous benign colorectal tumors (polyps) that can progress to malignant forms. The first clue to the location of the gene responsible for FAP was provided by Herrera and colleagues, who demonstrated a constitutional deletion of chromosomal band 5q21 in an FAP patient (*2*). This cytogenetic observation stimulated linkage analyses that demonstrated that 5q21 chromosome markers were tightly linked to the development of polyps in numerous FAP kindreds (*3–5*). Other studies suggest that genes from the same region may be involved in tumorigenesis in kindreds with unusual forms of FAP (*1, 6*), as well as in patients with "sporadic" colorectal cancer (*7–9*). To facilitate the identification of the 5q21 gene or genes responsible for FAP and related disorders, we have cloned a relatively large region from 5q21 and identified several genes from within this region that are expressed in colorectal epithelium.

The cosmid markers YN5.64 and YN5.48 have previously been shown to delimit an 8-cM region containing the locus for FAP (*5*). Further linkage and pulse-field gel electrophoresis (PFGE) analysis with additional markers has shown that the FAP locus is contained within a 4-cM region bordered by cosmids EF5.44 and L5.99 (*10*). To isolate clones representing a significant portion of this locus, a yeast artificial chromosome (YAC) library was screened with 5q21 markers. Twenty-one YAC clones, distributed within six contigs and including 5.5 Mb from the region between YN5.64 and YN5.48, were obtained (Fig. 1A).

Three contigs encompassing approximately 4 Mb were contained within the central portion of this region (Fig. 1B). To initiate construction of each contig, the sequence of a genomic marker cloned from chromosome 5q21 was determined and used to design primers for amplification by the polymerase chain reaction (PCR) (*11*). PCR was then carried out on pools of YAC clones distributed in microtiter trays as described (*12*). Individual YAC clones from the positive pools were identified by further PCR or hybridization-based assays, and the YAC sizes were determined by PFGE. To extend the areas covered by the original YAC clones, "chromosomal walking" was done. YAC termini were isolated by a PCR-based method and sequenced (*13*). PCR primers based on these sequences were then used to rescreen the YAC library. Multipoint linkage analysis with the various markers used to define the contigs, combined with PFGE analysis, showed that contigs 1 and 2 were centromeric to contig 3.

Contig 1 contained the FER gene, which had previously been identified on the basis of its sequence similarity to the oncogene ABL (*14*). Linkage analysis and physical mapping with the YAC clones indicated that FER was tightly linked to previously defined polymorphic markers for the FAP locus (*15*). However, further analysis (*16*) did not indicate any FAP-specific mutations in this gene.

A cross-hybridization approach was used to identify TB1 in contig 2; in this procedure, potential exon sequences are identified by cross-hybridization between human and rodent DNAs (*17–19*). Subclones of all the cosmids shown in Fig. 1 were used to screen Southern blots containing rodent DNA samples. A subclone of cosmid N5.66 was shown to strongly hybridize to rodent DNA, and this clone was used to screen cDNA libraries derived from normal adult colon and fetal liver. The ends of the initial cDNA clones obtained in this screen were then used to extend the cDNA sequence. Sequence analysis of 11 overlapping cDNA clones revealed an open reading frame (ORF) that extended for 1303 bp starting from the most 5′ sequence data obtained (GenBank accession number M74089). The predicted product of this gene (TB1) contained two significant local similarities to a

K. W. Kinzler, M. C. Nilbert, L.-K. Su, B. Vogelstein, T. M. Bryan, D. B. Levy, K. J. Smith, A. C. Preisinger, Molecular Genetics Laboratory, The Johns Hopkins University School of Medicine, Baltimore, MD 21231.
P. Hedge, D. McKechnie, R. Finniear, A. Markham, ICI Pharmaceuticals, Cheshire, United Kingdom SK10 4TG.
J. Groffen, Department of Pathology, Children's Hospital of Los Angeles, Los Angeles, CA 90027.
M. S. Boguski and S. F. Altschul, National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD 20894.
A. Horii, H. Ando, Y. Miyoshi, Y. Miki, I. Nishisho, Y. Nakamura, Department of Biochemistry, Cancer Institute, Tokyo 170, Japan.

*To whom correspondence should be addressed.