tively, at the surface. Locations of the offshore earthquakes are not as accurate as those onshore because of the network geometry.
9. R. S. Cockerham, *Bull. Seismol. Soc. Am.* **74**, 569 (1984); S. R. Walter, *ibid.* **76**, 583 (1986).
10. T. V. McEvilly, *Nature* **220**, 901 (1968); J. P. Eaton, *Eos* **62**, 959 (1981); T. J. Lay, J. W. Given, H. Kanamori, *Bull. Seism. Soc. Am.* **72**, 439 (1982).
11. D. S. Wilson, *J. Geophys. Res.* **94**, 3065 (1989).
12. A. C. Lawson, *Carnegie Inst. Washington Publ.* **87**, 54 (1908); W. Thatcher and M. Lisowski, *Eos* **68**, 1507 (1987).
13. J. R. Curray and R. D. Nason, *Geol. Soc. Am. Bull.* **78**, 413 (1967); L. Seeber, M. Barazangi, A. Nowroozi, *Bull. Seismol. Soc. Am.* **60**, 1669 (1970).
14. T. H. Heaton and H. Kanamori, *Bull. Seismol. Soc. Am.* **74**, 993 (1984); T. H. Heaton and S. H. Hartzell, *ibid.* **76**, 675 (1986).
15. B. F. Atwater, *Science* **236**, 942 (1987); M. E. Darienz and C. D. Peterson, *Tectonics* **9**, 1 (1990); A. R. Nelson, *Quat. Res.* **38**, 74 (1992); G. S. Vick, thesis, Humboldt State University (1989); D. W. Valentine, thesis, Humboldt State University (1992).
16. D. Merritts and W. B. Bull, *Geology* **17**, 1020 (1989).
17. S. H. Clarke, Jr., and G. A. Carver, *Science* **255**, 188 (1992).
18. We inverted for the main shock source parameters using the formalism of B. Romanowicz and T. Monfret [*Ann. Geophys.* **4**, 271 (1986)] and the spherical-Earth model PREM [A. M. Dziewonski and D. L. Anderson, *Phys. Earth Planet. Inter.* **25**, 297 (1981)] to correct for propagation effects. Rayleigh waves with periods between 140 and 320 s from a global distribution of stations were sampled at 10 frequencies.
19. An analysis of broad-band surface waves indicates that the main shock ruptured to the southwest (azimuth = 230°) with the largest moment release beginning 5 km from and extending to 20 km offshore [C. J. Ammon, A. A. Velasco, T. Lay, *Geophys. Res. Lett.* **20**, 97 (1993)].
20. For the aftershocks, we computed path-dependent amplitude and phase corrections based on analysis of the main shock. The approach of path-dependent corrections is advantageous for use of data at regional distances as well as shorter periods [M. Pasyanos and B. Romanowicz, *Eos* **73**, 372 (1992)].
21. H. Benioff, *Calif. Div. Mines Geol. Bull.* **171**, 199 (1955); A. Ben Menahem, *Bull. Seismol. Soc. Am.* **51**, 401 (1961).
22. A. Shakal *et al.*, *Calif. Div. Mines Geol. Rep. OSMS 92-05* (1992).
23. Where colonies of organisms were completely killed, the vertical range of the organisms represents a minimum measure of the uplift. On some reefs, only the upper parts of some colonies of plants and sessile animals, such as mussels, died. The vertical extent of mortality in these colonies provided the best basis for estimates of the amount of uplift. The mean and standard deviation of the vertical extent of mortality for each colony (Fig. 6, inset) is estimated from multiple measurements made by a laser total station with an instrument precision of 1 mm.
24. We assumed that slip is uniform on a rectangular, planar fault embedded in an elastic half space. We performed a Monte Carlo search of the parameters describing the geometry of the fault plane and, for each geometry, used the horizontal and vertical GPS and coastal-uplift data to estimate the magnitude and rake of the slip vector. The scatter of the residuals for the best fitting models, including our preferred model (Table 2), is 2.2 times the a priori observational errors. An F-ratio test indicates other models with residual scatter less than 2.6 times the a priori errors do not differ significantly from the best fitting models at the 95% confidence level.
25. In the projections, we assumed that faults are planar, the ocean depth is 2.3 km at the seaward edge of the Cascadia subduction zone at the latitude 40°27'N, and the mean elevation of the seismic network is 0.9 km. Main-shock parameters are from Table 1. We assumed the upper boundary of the background Gorda seismicity dips 6° east at a depth of 17.5 km at the coordinates of the main shock hypocenter.
26. G. G. Shor, P. Dehlinger, H. K. Kirk, W. S. French, *J. Geophys. Res.* **73**, 2175 (1968); S. W. Smith, J. S. Knapp, R. C. McPherson, *ibid.* **98**, 8153 (1993).
27. W. P. Chen and P. Molnar, *ibid.* **88**, 4183 (1983); P. Molnar and P. England, *ibid.* **95**, 4833 (1990).
28. J. J. Tabor and S. W. Smith, *Bull. Seismol. Soc. Am.* **75**, 237 (1985).
29. D. Verdonck and G. Zandt, *Lawrence Livermore Natl. Lab. Rep. UCRL-JC-111629* (1992).
30. K. Wang and G. Rogers, *Eos* **73**, 504 (1992).
31. We estimated static stress changes caused by the main shock using a rectangular dislocation surface (Table 2) in an elastic half-space [Y. Okada, *Bull. Seismol. Soc. Am.* **82**, 1018 (1992)] to stimulate the earthquake rupture. The results of the model are insensitive to small variations in the assumed fault geometries or slip distributions, although the maximum values and spatial details of the stress fields are.
32. K. Hudnut *et al.*, *Bull. Seismol. Soc. Am.* **79**, 282 (1989).
33. We thank B. Ellsworth, C. Weaver, D. Wilson, and D. Merritts for reviews of the manuscript. The measurement of coastal uplift was supported by a grant from the Pacific Gas and Electric Company.

RESEARCH ARTICLE

NMR Structure of a Specific DNA Complex of Zn-Containing DNA Binding Domain of GATA-1

James G. Omichinski, G. Marius Clore,* Olivier Schaad, Gary Felsenfeld, Cecelia Trainor, Ettore Appella, Stephen J. Stahl, Angela M. Gronenborn*

The three-dimensional solution structure of a complex between the DNA binding domain of the chicken erythroid transcription factor GATA-1 and its cognate DNA site has been determined with multidimensional heteronuclear magnetic resonance spectroscopy. The DNA binding domain consists of a core which contains a zinc coordinated by four cysteines and a carboxyl-terminal tail. The core is composed of two irregular antiparallel β sheets and an α helix, followed by a long loop that leads into the carboxyl-terminal tail. The amino-terminal part of the core, including the helix, is similar in structure, although not in sequence, to the amino-terminal zinc module of the glucocorticoid receptor DNA binding domain. In the other regions, the structures of these two DNA binding domains are entirely different. The DNA target site in contact with the protein spans eight base pairs. The helix and the loop connecting the two antiparallel β sheets interact with the major groove of the DNA. The carboxyl-terminal tail, which is an essential determinant of specific binding, wraps around into the minor groove. The complex resembles a hand holding a rope with the palm and fingers representing the protein core and the thumb, the carboxyl-terminal tail. The specific interactions between GATA-1 and DNA in the major groove are mainly hydrophobic in nature, which accounts for the preponderance of thymines in the target site. A large number of interactions are observed with the phosphate backbone.

The erythroid-specific transcription factor GATA-1 is responsible for the regulation of transcription of erythroid-expressed genes and is an essential component required for the generation of the erythroid lineage (*1*). GATA-1 binds specifically as a monomer to the asymmetric consensus target sequence (T/A)GATA(A/G) found in the cis-regulatory elements of all globin genes and most other erythroid-specific genes that have been examined (*2*). GATA-1 was the first member of a family of proteins, which now includes regulatory proteins expressed in other cell lineages, characterized by their recognition of the GATA DNA sequence and by the presence of two metal binding regions of the form Cys-X-X-Cys-$(X)_{17}$-Cys-X-X-Cys separated by 29 residues (*2*, *3*). Mutation and deletion studies on

J. G. Omichinski, G. M. Clore, O. Schaad, and A. M. Gronenborn are in the Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892. G. Felsenfeld and C. Trainor are in the Laboratory of Molecular Biology, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892. E. Appella is in the Laboratory of Cell Biology, Building 37, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892. S. J. Stahl is in the Protein Expression Laboratory, Building 6B, Office of the Director, National Institutes of Health, Bethesda, MD 20892.

*To whom correspondence should be addressed.

GATA-1 have indicated that the amino-terminal metal binding region is not required for specific DNA binding (*4*). More recently, studies with synthetic peptides have shown that a 59-residue fragment (residues 158/216 of chicken GATA-1), comprising the carboxyl-terminal metal binding region complexed to zinc and 28 residues carboxyl terminal to the last Cys, constitutes the minimal unit required for specific binding (association constant K_{assoc} $\approx 1.2 \times 10^8$ M^{-1}) (*5*). In this article we present the determination of the solution structure of the specific complex of a 66-residue fragment (residues 158 to 223) comprising the DNA binding domain of chicken GATA-1 (cGATA-1) with a 16–base pair (bp) oligonucleotide containing the target sequence AGATAA by means of multidimensional heteronuclear filtered and separated nuclear magnetic resonance (NMR) spectroscopy.

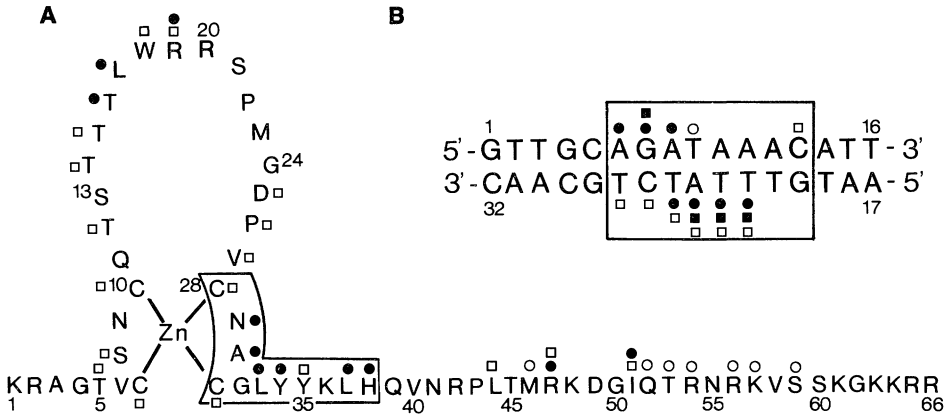

Fig. 1. (**A**) Sequence of the cGATA-1 DNA binding domain (residues 158 to 223 of intact cGATA-1) with structural residues and residues that contact the DNA indicated (*50*). (**B**) Sequence of the 16-bp DNA illustrating sites of contact with the cGATA-1 DNA binding domain. In (A), the region encompassing the single helix is boxed and the symbols are as follows: open squares represent residues involved in maintaining the structural integrity of the zinc binding region; whereas closed and open circles are residues that interact with the DNA in the major and minor grooves, respectively. In (B), the region of the oligonucleotide interacting with the cGATA-1 DNA binding domain is boxed in and the symbols are as follows: closed and open circles are sites of interaction with bases in the major and minor grooves, respectively, and closed and open squares are sites of interaction with the sugar or phosphate or both in the major or minor grooves, respectively.

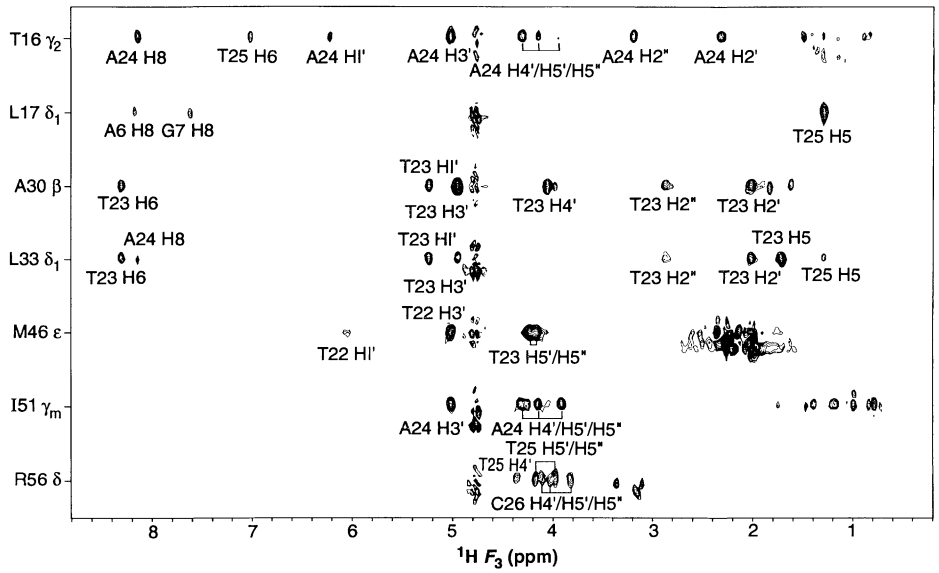

Fig. 2. Composite of ^{13}C-H strips taken from the 120-ms 3D $^{13}C(F_2)$-separated, ^{12}C-filtered NOESY spectrum of the complex of the cGATA-1 DNA binding domain with DNA specifically illustrating NOEs between protons of the DNA (attached to ^{12}C along the F_3 axis) and protons of the peptide (attached to ^{13}C). The $^1H(F_1)/^{13}C(F_2)$ shifts of T16γ_2, L17δ_1, A30β, L33δ_1, M46ε, I51γ_m, and R56δ are 1.40/23.07, 0.75/26.57, 1.74/18.74, 0.79/28.60, 2.25/17.11, 0.88/17.58, and 3.24/42.56 ppm, respectively.

Proteins that contain one or more zinc binding domains and regulate transcription by binding sequence specifically to DNA are widely distributed throughout nature (*6*). Five distinct classes of zinc-dependent DNA binding proteins have been identified (*7*). The prototypical members of classes I through V are TFIIIA (*8*), the glucocorticoid receptor (*9*), GAL-4 (*10*), GATA-1 (*3*), and MyT1 (*11*), respectively. The five classes differ from each other on the basis of the distribution of Cys or His residues involved in chelation, the number of amino acids located in the loop between the chelating residues, the number of zinc binding domains required for high-affinity DNA binding, and the quarternary structure of the protein in the complex (monomer or dimer). Structures of class I, II, and III zinc binding domains free in solution have been determined by NMR (*12*) and in complexes with their specific DNA binding sites by x-ray crystallography (*13, 14*). The crystallographic results demonstrate that classes I to III each interact with DNA in a distinctive manner. No structural information of any kind is available to date on classes IV or V, either in the free state or complexed to DNA. Our results on cGATA-1 (class IV) demonstrate that some structural similarity exists between the zinc binding region of the DNA binding domain of cGATA-1 and the amino-terminal zinc module of the glucocorticoid receptor (class II). The mode of DNA recognition, however, is very different.

Structure determination. All experiments were carried out at pH 6.5 on a 2 mM complex containing 1 equivalent (eq) cGATA-1 DNA binding domain uniformly (>95%) labeled with either ^{15}N or both ^{15}N and ^{13}C, 1.1 eq of zinc, and 1 eq of a double-stranded 16-bp oligonucleotide at natural abundance. The amino acid sequence of the 66-residue polypeptide comprising the DNA binding cGATA-1 peptide and the sequence of the DNA are shown in Fig. 1. Details of the cloning, expression, and purification of the cGATA-1 DNA binding domain, the synthesis and purification of the two strands of the DNA target site, and the preparation of the complex are given in (*15*), (*16*), and (*17*), respectively.

The ^1H, ^{15}N, and ^{13}C assignments of the cGATA-1 DNA binding domain in the complex were obtained with an array of three-dimensional (3D) double and triple resonance NMR experiments (*18, 19*). Sequential assignments of the exchangeable and nonexchangeable protons of the DNA in the complex were obtained by standard procedures (*20*) from a 2D nuclear Overhauser enhancement (NOESY) spectrum (150-ms mixing time) in water with a 1-1 semiselective excitation pulse (*21*) to assign the exchangeable and A(H2) protons (*22*),

and 2D ^{12}C-filtered NOESY (50- and 150-ms mixing time) (*23*) and homonuclear Hartmann-Hahn (HOHAHA) (*24*) spectra in D_2O to observe correlations involving only protons attached to ^{12}C, suppressing the signals of ^{13}C attached protons from the protein.

Structural information in the form of interproton distance restraints was derived from isotope-edited and isotope-filtered 2D and 3D NOESY spectra. NOEs involving only the DNA were detected with the same 2D NOESY and ^{12}C-filtered NOESY spectra, recorded in H_2O and D_2O, respectively, used for assignment purposes. The NOEs involving protons of the protein were obtained from 3D ^{15}N- and ^{13}C-separated NOESY spectra (mixing time, 120 ms) (*25*) recorded in H_2O and D_2O, respectively. Finally NOEs specifically between protein and DNA protons were identified in 3D ^{12}C-filtered/^{13}C-separated (*26*) and ^{12}C, ^{14}N-filtered/^{15}N-separated (*27*) NOESY spectra (120-ms mixing time), recorded in D_2O and H_2O, respectively. An example of the quality of the data is provided by the composite of $^{13}C(F_2)$-$^1H(F_1)$ strips taken from the 3D ^{12}C-filtered/^{13}C-separated NOESY spectrum shown in Fig. 2, illustrating specifically NOE interactions between nonexchangeable DNA and protein protons.

Approximate interproton distance restraints were derived from the above NOESY spectra as follows. The NOEs between protein protons and between protein and DNA protons were grouped into four ranges, 1.8 to 2.7, 1.8 to 3.3 (1.8 to 3.5 for NOEs involving NH protons), 1.8 to 5.0, and 2.2 to 6.0 Å, corresponding to strong, medium, weak, and very weak NOEs, respectively (*28, 29*). NOEs between DNA protons were classified into six ranges: 1.8 to 2.5, 2.3 to 3.0, 2.3 to 3.5, 2.3 to 4.0, 2.5 to 5.0, and 3.0 to 6.0 Å, corresponding to strong, medium-strong, medium, medium-weak, weak, and very weak NOEs, respectively (*29, 30*). Stereospecific assignments and torsion angle (ϕ, ψ, χ_1) restraints (*31*) for the cGATA-1 DNA binding domain were obtained with the conformational grid search program STEREOSEARCH (*32*) on the basis of three-bond coupling constants (*33*) and the intraresidue and sequential interresidue NOEs involving the NH, CαH, and CβH protons (*34*). The χ_2 torsion angle restraints (*31*) for Leu and Ile were obtained from $^3J_{CC}$ coupling constants (*33*) and the pattern of intraresidue NOEs as described in (*35*). Stereospecific assignments were obtained for the β-methylene protons of 20 of the 47 β-methylene groups, for the α-methylene protons of 2 of the 5 Gly residues, and for the methyl groups of 2 of the 4 Val residues and 3 of the 4 Leu residues.

The structure calculations were based on a total of 1407 experimental interproton distance restraints, comprising 919 and 371 restraints for the cGATA-1 DNA binding domain and DNA, respectively, and 117 intermolecular restraints between the protein and the DNA. In addition, 37 hydrogen-bonding restraints to maintain base pairing of the DNA were used (see Table 1). The distance restraints were supplemented by 328 torsion angle restraints comprising 144 (58 ϕ, 56 ψ, 26 $\chi1$, and 4 $\chi2$) angles for the protein and 152 for the DNA. The latter comprised the α, β, γ, ε, and ζ torsion angles of the DNA backbone that were restricted to broad ranges of 180° ± 50°, −85° ± 50°, −70° ± 50°, 180° ± 50°, and 60° ± 35°, respectively, describing the values characteristic of both right-handed B- and A-form DNA (*36*), in order to avoid problems associated with local mirror images commonly found in DNA structures determined from NMR distance restraints (*30, 37*).

A total of 30 simulated annealing (SA) structures (*38*) were calculated with the hybrid distance geometry–SA method (*39, 40*). A summary of the structural statistics is provided in Table 1, and a stereoview of a best-fit superposition of the 30 SA structure (residues 2 to 59 of the protein and base pairs 6 to 13 of the DNA) is shown in Fig. 3. All of the structures satisfy the experimental restraints (no violations greater than 0.5 Å and 5° in the interproton distance and torsion angle restraints, respectively), display very small deviations from idealized geometry and exhibit good nonbonded contacts. The amino (residue 1) and carboxyl (residues 60 to 66) termini of the protein are disordered in solution based on the absence of any nonsequential or intermolecular NOEs, and values of $^3J_{HN\alpha}$ couplings (6 to 7 Hz) characteristic of a random coil conformation. Base pairs 6 to 13 of the DNA are in contact with the cGATA-1 DNA binding domain and are well defined both locally and globally. The orientation, however, of the first five and last three base pairs of the DNA, which are not in contact with the protein, is poorly

Table 1. The notation of the NMR structures is as follows: <SA> are the final 30 SA structures; $\overline{SA}$ is the mean structure obtained by averaging the coordinates of the individual SA structures best fitted to each other (with respect to residues 2 to 59 of the protein and base pairs 6 to 13 of the DNA); ($\overline{SA}$)r is the restrained minimized mean structure obtained by restrained regularization of the mean structure $\overline{SA}$. Only residues 2 to 59 of the protein and base pairs 6 to 13 of the DNA were used to calculate ($\overline{SA}$)r. The number of terms for the various restraints is given in parentheses: the first number indicates the values for the whole system, while the second gives the numbers for residues 2 to 59 of the protein and base pairs 6 to 13 of the DNA; rms, root-mean-square, and expt, experimental.

Structural statistics	<SA>	($\overline{SA}$)r
Rms deviations from expt distance restraints (Å)*		
All (1444/1207)	0.034 ± 0.001	0.030
Protein		
Interresidue sequential ($\|i - j\| = 1$) (242/237)	0.011 ± 0.003	0.011
Interresidue short range ($1 < \|i - j\| \leq 5$) (161/161)	0.018 ± 0.003	0.017
Interresidue long range ($\|i - j\| > 5$ (182/182)	0.016 ± 0.002	0.043
Intraresidue (334/307)	0.024 ± 0.002	0.024
DNA		
Intraresidue (157/78)	0.083 ± 0.001	0.061
Sequential intrastrand (180/87)	0.032 ± 0.002	0.026
Interstrand (34/20)	0.020 ± 0.005	0.026
Hydrogen bonds (37/18)†	0.034 ± 0.005	0.049
Protein-DNA (117)	0.024 ± 0.004	0.021
Rms deviations from expt dihedral restraints (degree) (296/217)*‡	0.20 ± 0.04	0.52
Deviations from idealized covalent geometry§		
Bonds (Å) (3257/1507)	0.007 ± 0.001	0.006
Angles (degree) (5895/2723)	1.55 ± 0.003	1.40
Impropers (degree) (1640/759)	0.37 ± 0.02	0.51
$E_{L\text{-}J}$ (kcal mol^{-1})¶	−610 ± 12	−360

*None of the structures exhibited distance violations greater than 0.5 Å or dihedral angle violations greater than 5°. †The hydrogen-bonding restraints for the DNA were as follows: for GC base pairs, $r_{G(N1)\text{-}C(N3)} = 2.95 \pm 0.2$ Å, $r_{G(O6)\text{-}C(N4)} = 2.91 \pm 0.2$ Å, and $r_{G(N2)\text{-}C(O2)} = 2.86 \pm 0.2$ Å; for the AT base pairs, $r_{A(N1)\text{-}T(N3)} = 2.82 \pm 0.2$ Å; $r_{A(N6)\text{-}T(O4)} = 2.95 \pm 0.2$ Å. These values are taken from the x-ray structure analyses of GpC and ApU (*55*). ‡The torsion angle restraints for the whole complex comprise 144 angles for the protein and 152 angles for the DNA. The torsion angles for the protein consist of 58 ϕ, 56ψ, 26 χ_1, and 4 χ_2 restraints. The torsion angles for the DNA comprise loose restraints on the backbone torsion angles α, β, γ, ε, and ζ to prevent problems associated with local mirror images (*30*). For residues 2 to 59 of the protein and base pairs 6 to 13 of the DNA, there are 143 torsion angle restraints for the protein (with the number of ϕ angle restraints being reduced to 57) and 74 torsion angle restraints for the DNA. §The improper torsion restraints serve to maintain planarity and chirality. The bond and angle restraints include the following restraints for the tetrahedral coordination of the zinc atom to the Sγ of the Cys^7, Cys^{10}, Cys^{28}, and Cys^{31}: $r_{Zn\text{-}S} = 2.30$ Å, 109° for the S(i)–Zn–S(j) angle, and 107.94° for the Cβ(i)–S(i)–Zn angle. ¶$E_{L\text{-}J}$ is the Lennard-Jones van der Waals energy calculated with the CHARMM PARAM19/20 protein and PARNAH1ER1 DNA parameters (*56*) and is *not* included in the target function for SA or restrained minimization.

defined with respect to the core of the complex, although the conformation of each of these bases at a local level is reasonably well defined. This is due to the fact that, in addition to their approximate nature, the interproton distance restraints within the DNA are solely sequential. Hence, they are inadequate to ascertain the relative orientation of base pairs separated by more than five to six steps with any great degree of precision and accuracy. The global conformation of the central eight base pairs, on the other hand, is determined not only by the restraints within the DNA, but more importantly by the large number of intermolecular interproton distance restraints between the protein and DNA. Hence, all subsequent discussion is restricted to the complex proper comprising residues 2 to 59 of the protein and base pairs 6 to 13 of the DNA (that is, the ordered region of the protein and the region of the DNA in contact with the protein). The atomic root-mean-square (rms) distribution of the 30 SA structures about the mean coordinate positions for the complex proper is 0.70 ± 0.13 and 1.13 ± 0.08 Å for protein backbone plus DNA and all protein atoms plus DNA, respectively (Table 2).

Tertiary structure of the cGATA-1 DNA binding domain. The protein can be divided into two modules: the protein core, which consists of residues 2 to 51 and contains the zinc coordination site, and an extended carboxyl-terminal tail (residues 52 to 59).

A schematic ribbon drawing of the core is presented in Fig. 4A. The core starts out with a turn (residues 2 to 5), followed by two short irregular antiparallel β sheets, a helix (residues 28 to 38), and a long loop (residues 39 to 51), which includes a helical turn (residues 44 to 47) as well as an Ω-like loop (residues 47 to 51). β strands 1 (residues 5 to 7) and 2 (residues 11 to 14) form the first β sheet, and β strands 3 (residues 18 to 21) and 4 (residues 24 to 27) form the second β sheet. In both cases, the strands of each sheet are connected by a hairpin turn. β strands 2 and 3 are connected by a short loop (residues 15 to 17), and there is a single backbone hydrogen bond between Ser^8N and Pro^{26}O bridging the carboxyl terminus of β strand 1 to the carboxyl terminus of β strand 4.

The core is dominated by the zinc which is coordinated in a tetrahedral manner to the Sγ atoms of Cys^7, Cys^{10}, Cys^{28}, and Cys^{31} in an S configuration as defined by Berg (*41*). Cys^7 is located in β strand 1, Cys^{10} in the turn between β strands 1 and 2, and Cys^{28} and Cys^{31} in the helix. Hence, the zinc is responsible for determining the orientation of the helix with respect to the first antiparallel β sheet. The Sγ of Cys^7 accepts three hydrogen bonds from the backbone amides of Asn^9, Cys^{10}, and Cys^{28}, while the Sγ atoms of Cys^{10} and Cys^{28} accept a single hydrogen bond each from the backbone amides of Thr^{12} and Cys^{31}, respectively.

In addition to the tight packing around the zinc ion, the core of the cGATA-1 DNA binding domain is stabilized by a number of hydrophobic interactions, which are illustrated in Fig. 4B. The hydrophobic cluster formed by Thr^5, Thr^{15}, Trp^{18}, and Pro^{26} serves to stabilize the relative orientation of the two antiparallel β sheets. The cluster formed by Val^{27}, Gly^{32}, and the aliphatic portion of the side chain of Arg^{19} serves to position the helix with respect to the second antiparallel β sheet. There are three hydrophobic clusters that stabilize the position of the extended loop (residues 39 to 51) with respect to the helix and the first antiparallel β sheet. Thus Tyr^{35} of the helix packs against the backbone of Val^{40}, thereby establishing the directionality of the loop. The cluster formed by Thr^{12}, Leu^{44}, and the aliphatic portion of the side chain of Arg^{47} positions the helical turn (residues 44 to 47) in proximity to the hairpin loop connecting β strands 1 and 2. Finally, the cluster formed by Thr^{14}, Gly^{50}, and Ile^{51} bring the carboxyl terminus of the core in proximity to β strand 2.

There are also two side chain hydrogen-bonding interactions within the core: namely a hydrogen bond between the Oγ and OγH of Thr^{12} and Thr^{14} and between the OγH of Ser^8 and the side chain carboxylate of Asp^{25}. The former stabilizes β strand 2 and the latter stabilizes the orientation of β strand 4 with respect to β strand 1.

Part of the core of the cGATA-1 DNA binding domain is structurally similar to that of the amino-terminal zinc-containing module of the DNA binding domain of the glucocorticoid receptor [GR (*14*)]. Thus the Cα atoms of 30 residues of these two proteins can be superimposed with an

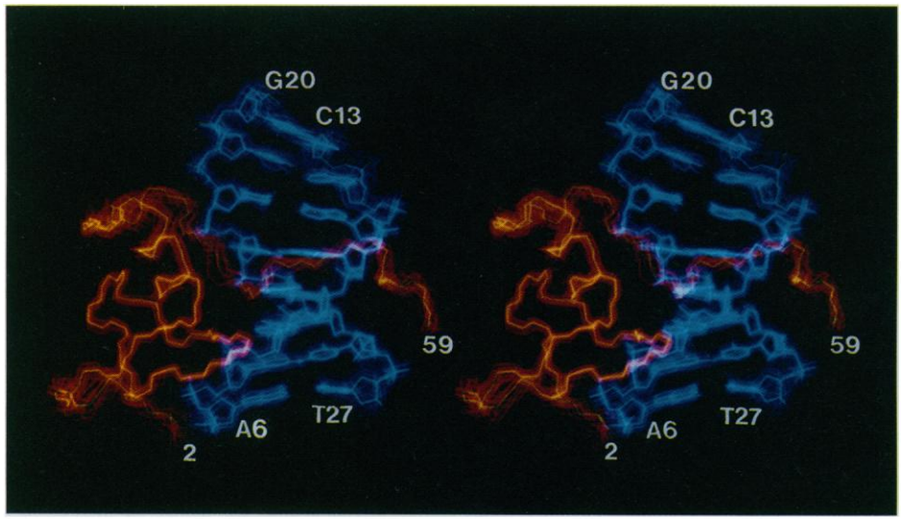

Fig. 3. Stereoview showing a best-fit superposition of the 30 SA structures of the complex of the cGATA-1 DNA binding domain with DNA. The backbone (N, Cα, and C) atoms of the cGATA-1 DNA binding domain are shown in red, and all non-hydrogen atoms of the DNA are shown in blue; the restrained minimized mean structure of the complex is also included in the superposition and is highlighted. The residues shown are 2 to 59 of the protein and base pairs 6 to 13 of the oligonucleotide. The amino (residue 1) and carboxyl (residues 60 to 66) termini of the protein are disordered, and the orientation of base pairs 1 to 5 and 14 to 16 of the DNA with respect to the complex proper are poorly defined because their conformation is only restrained by sequential NOEs. In contrast, the structure of base pairs 6 to 13, which are in contact with the protein, are determined not only by the NOEs within the DNA but equally importantly by the NOEs between the DNA and the protein.

Table 2. Atomic rms differences (in angstroms).

Structures	Protein (2 to 59) plus DNA (base pairs 6 to 13)		Protein (2 to 59)		DNA (base pairs 6 to 13)
	Backbone	All atoms	Backbone	All atoms	
<SA> vs $\overline{SA}$	0.70 ± 0.13	1.13 ± 0.08	0.70 ± 0.11	1.30 ± 0.11	0.55 ± 0.10
<SA> vs $\overline{(SA)r}$	0.18	0.40	0.24	0.53	0.12
$\overline{(SA)r}$ vs $\overline{SA}$	0.74 ± 0.09	1.20 ± 0.09	0.74 ± 0.11	1.40 ± 0.13	0.56 ± 0.10

rms difference of only 1.4 Å (Fig. 4C). The resulting sequence alignment based on the structural superposition is shown in Fig. 4D. Apart from the four Cys residues that coordinate the zinc atom, only one residue (Lys^{36} in the cGATA-1 DNA binding domain and Lys^{465} in GR) is conserved between the two proteins. The structural similarity extends from the amino terminus up to the end of the helix (residues 3 to 39 of the cGATA-1 DNA binding domain and residues 436 to 468 of GR), and the Zn-Sγ geometry, as well as the side chain conformations of the four coordinating cysteines, are identical. The loop between strands β2 and β3 has three deletions, and the turn between strands β3 and β4 has one deletion in GR with respect to cGATA-1. The topology and polypeptide trace following the carboxyl end of the helix, however, are entirely different in the two proteins. Thus, in the DNA binding domain of GR there is a second compact zinc-containing module (residues 470 to 514) made up of two strands and two helices, whereas in the cGATA-1 DNA binding domain there is a long loop (residues 38 to 51) and extended strand (residues 52 to 59). Interestingly, however, within this region, Met^{46} of the GATA-1 DNA binding domain aligns with Lys^{490} of GR. Both of these residues are involved in sugar-phosphate contacts with the DNA, although the polypeptide chains run in opposite directions at this point.

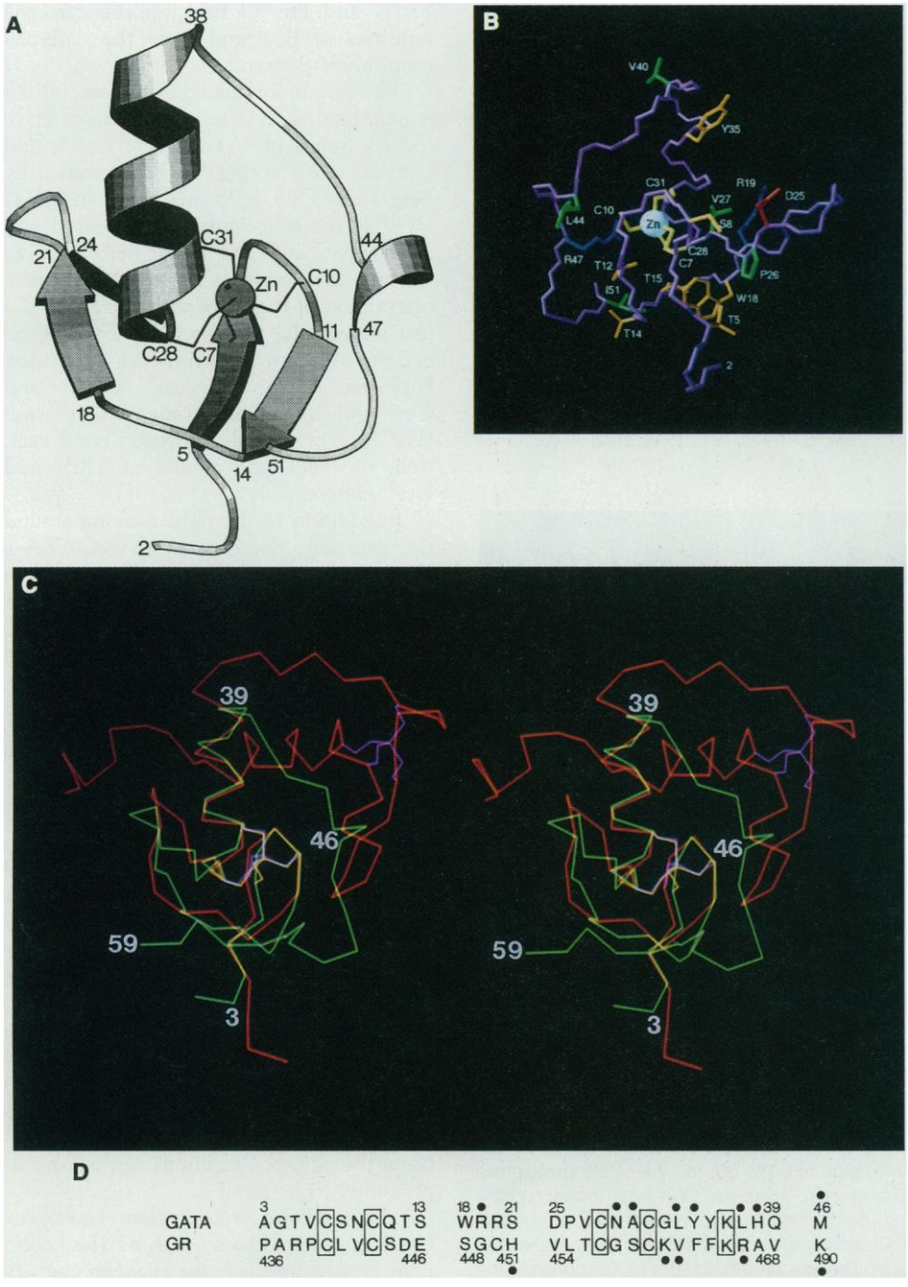

Fig. 4. (**A**) Schematic ribbon drawing of the core of the cGATA-1 DNA binding domain (residues 2 to 51). (**B**) Backbone (N, Cα, and C) trace (violet) of the core of the cGATA-1 DNA binding domain together with the zinc atom and side chains that are important for maintaining its structural stability (zinc, white; Cys, yellow; hydrophilic residues, orange; hydrophobic residues, green; and positively and negatively charged polar residues, blue and red, respectively). (**C**) Superposition of the Cα atoms of the cGATA-1 (green) and GR (red) DNA binding domains (the zinc and coordinating cysteine side chains are shown in yellow for cGATA-1 and in purple for the glucocorticoid receptor). (**D**) Sequence alignment of the cGATA-1 (*2*) and GR (*14*) DNA binding domains based on the structural alignment shown in (C). (The closed circles indicate residues that interact with DNA.) The GR sequence shown in (C) is that of the x-ray structure (*14*); residues 436 to 439 of the natural sequence, however, are PPKL. The programs Molscript (*51*) and VISP (*52*) were used to generate (A) and (B), respectively. The superposition and structural alignment in (C) was carried out with the program O (*53*).

Structure of the DNA target site. The overall conformation of a DNA oligonucleotide can be readily deduced from a qualitative interpretation of the NOE data (*20*). In the case of the cGATA-1 DNA target site, the NOE data indicate that the conformation is that of right-handed B-form DNA. Thus, the observation of H8/H6(i)-H5/CH_3(i+1) NOEs establishes that the DNA helix is right-handed, and the B-type structure is ascertained from the observation of a $NOE_{H2'(i)\text{-}H8/H6(i)} >> NOE_{H2''(i-1)\text{-}H8/H6(i)} > NOE_{H2'(i-1)}$-H8/H6($i$) pattern of NOE intensities throughout the oligonucleotide. In addition, one can readily deduce that the sugar pucker conformation ranges from O1′-*endo* to C2′-*endo* and that the glycosidic bond is anti on account of the observation of a $NOE_{H2'(i)\text{-}H8/H6(i)} >> NOE_{H1'(i)\text{-}H8/H6(i)} > NOE_{H3'(i)\text{-}H8/H6(i)}$ pattern of intraresidue sugar-base NOE intensities.

This qualitative interpretation is confirmed by the structure calculations. The central eight base pairs (base pairs 6 to 13) of the 16-bp oligonucleotide that comprise the target site for the cGATA-1 DNA binding domain are well defined both globally and locally. The atomic rms distribution of the 30 SA structures about their mean coordinate positions for base pairs 6 to 13 of the DNA is 0.55 ± 0.10 Å (Table 2). The backbone α, β, δ, ε, ζ, and the χ glycosidic torsion angles have average values of −66° ± 14°, −176° ± 11°, 55° ± 8°, 117° ± 6°, 172° ± 12°, −84° ± 6°, and −118° ± 7°, respectively, which are typical of right-handed B DNA (36). Likewise the average values of 35.0° ± 3.2° and 3.36 ± 0.44 Å for the helical twist and rise,

respectively, are characteristic of B DNA (36, 42). The variation in helical twist, however, is quite substantial, with a low value of 31.0° between base pairs 9 and 10 and a high value of 39.6° between base pairs 7 and 8. In addition, there is a kink in the direction of the minor groove of ~15° in the DNA between base pairs 10 and 11. This kink is caused by the large roll angle of −16°, which results in an opening between these two base pairs toward the major groove (that is, there is a widening of the major groove and a narrowing of the minor groove at this location). The presence of this kink or bend is consistent with the results of gel mobility-shift assays (43).

The overall structure of the complex. The overall topology and structural organization of the complex is shown in Fig. 5. The helix and the loop connecting strands β2 and β3 (which is located directly beneath the helix) are located in the major groove, and the carboxyl-terminal tail wraps around the DNA and lies in the minor groove, directly opposite the helix. The overall appearance is analogous to that of a right hand holding a rope, with the rope representing the DNA, the palm and fingers of the hand the core of the protein, and the thumb the carboxyl-terminal tail. This view is most evident in Fig. 5D, which displays the solvent-accessible surface of the DNA and the protein. It is this pincer-like configuration of the protein that causes the kink in the DNA. The long axis of the helix lies at an angle of ~40° to the base planes of the DNA (Fig. 5A), whereas the carboxyl-terminal tail is approximately parallel to the base planes (Fig. 5B).

Despite the structural similarity of residues 3 to 39 of the core of the cGATA-1 DNA binding domain with the amino-terminal zinc module of GR, the mode of binding to DNA is entirely different. Thus, the helix of the first module in GR lies approximately parallel to the base planes, no contact is made with the minor groove, and, unlike cGATA-1, GR only binds as a symmetric dimer. It is therefore not surprising that only three locations contacting the DNA are shared between the two proteins within the 30 residues that are structurally superimposable (Fig. 4D). Two are located within the helix (Leu^{33} and Val^{462}, as well as Leu^{37} and Arg^{466}), and the third (Met^{46} and Lys^{490}) is located in the loop following the helix in cGATA-1 and in the first of the two helices of the carboxyl-terminal zinc module in GR.

The binding mode, however, shares some features with that of the homeodomain, a member of the helix-turn-helix class of DNA binding proteins, which also binds as a monomer and in which a helix is positioned in the major groove, and the amino-terminal tail, as opposed to the carboxyl-terminal tail in cGATA-1, lies in the minor groove (44). In contrast to the cGATA-1 complex, however, the amino-terminal tail in the homeodomain complex does not lie directly opposite the helix but is located several base pairs above it.

Contacts between the cGATA-1 DNA binding domain and the DNA. Views of side chain contacts with the DNA in the major and minor grooves are shown in Figs. 6A and 6B, respectively, and a schematic representation of all of the contacts is provided in Fig. 6C.

The cGATA-1 DNA binding domain makes specific contacts with eight bases, seven in the major groove (A6, G7, A8, T25, A24, T23, and T22) and one in the minor groove (T9). All of the base contacts in the major groove involve the helix and the loop connecting β strands 2 and 3. In contrast to other DNA binding proteins, the majority of base contacts involve hydrophobic interactions. Thus, Leu^{17} interacts with A6, G7, and T25, Thr^{16} interacts with A24 and T25, Leu^{33} interacts with A24 and T23, and Leu^{37} interacts with T23 and T22. This finding accounts for the predominance of thymidines in the DNA target site. Indeed, there are only three hydrogen-bonding interactions: namely, between the side chain of Asn^{29} and the N6 atoms of A24 and A8 in the major groove, which may involve a bridging water molecule; and between the $N\zeta H_3^+$ of Lys^{57} and the O2 atom of T9 in the minor groove. In this regard, it is interesting to note that there is a reduction of 1127 $Å^2$ in the surface accessible area of the cGATA-1 DNA binding domain in the presence of DNA (corresponding to a 20% decrease in the accessible surface), and a decrease in the calculated solvation free energy of folding (45) of 13 kcal mol^{-1}. This latter effect can make a sizeable contribution to K_{assoc} [$\sim 1.2 \times 10^8$ M^{-1}; (5)].

The remaining contacts involve the sugar-phosphate backbone, the majority of which are located on the second strand (that is G20 to T27). Salt bridges or hydrogen bonds or both are made by Arg^{19}, Arg^{47}, and His^{38} with the phosphates of G7, A24, and T22, respectively, in the major groove, and with the phosphates of C13, T25, C26, and T27 by Arg^{54}, Thr^{53},

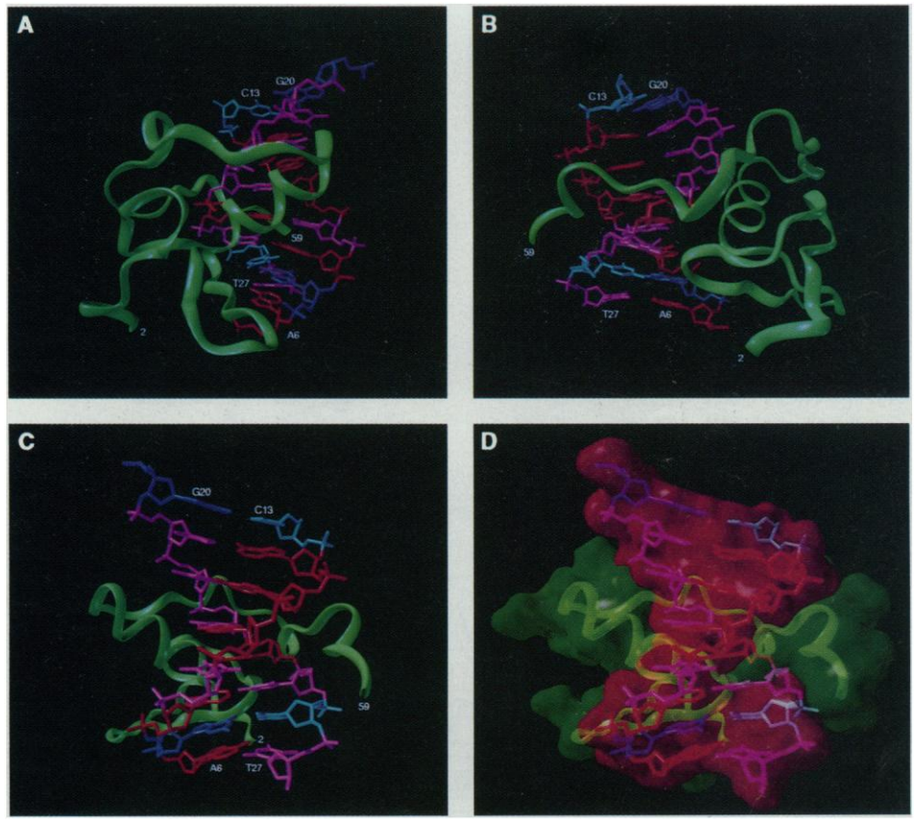

Fig. 5. (**A** to **C**) Three views illustrating the interaction of the cGATA-1 DNA binding domain with DNA. The protein is shown as a schematic ribbon drawing in green, and the color coding used for the DNA bases is red for A, lilac for T, dark blue for G, and light blue for C. (**D**) Same view as in (C) but the solvent-accessible surface (*54*) (green for cGATA-1 and red for the DNA), which illustrates the intimate contact between the peptide and DNA, is also included. Residues 2 to 59 of the cGATA-1 peptide and base pairs 6 to 13 of the DNA are shown. These models were generated with the program VISP (*52*).

Arg[56], and Ser[59], respectively, in the minor groove. The interactions of Arg[54] and Arg[56] above and below the polypeptide chain span the full length of the target site and are probably responsible for the bending of the DNA in the direction of the minor groove. Likewise, all of the sugar contacts involve the second strand. In the major groove, they are hydrophobic in nature and involve contacts between the sugars of T22, T23, and A24 with Tyr[34], Leu[33] and Ala[30], and Ile[51] and Thr[16], respectively. In the minor groove, hydrophobic sugar DNA-protein interactions are made by C13 with the aliphatic portion of the side chain of Arg[54], T23 and T24 with Gln[52], T25 and C26 with the aliphatic portion of the side chain of Arg[56], and C26 with Ser[59]. In addition, there is a hydrogen bond between the side-chain amide of Gln[52] and the sugar O3′ atom of T23.

Comparison with biochemical and mutational data. There is relatively little biochemical and mutational data available on the interaction of cGATA-1 with DNA. Deletion of residues 204 to 304 in the intact protein results in loss of binding (*4*). This deletion starts at the position of Arg[47] in the cGATA-1 DNA binding domain and hence results in the removal of the entire carboxyl-terminal tail, thereby abolishing all interactions in the minor groove. Studies in which synthetic peptides have been used show that the minimum peptide required for specific binding comprises residues 1 to 59, and that a smaller peptide containing residues 1 to 53 (that is the protein core) fails to bind specifically (*5*). This result agrees well with the structural results which indicate that binding of both the core to the major groove and the basic carboxyl-terminal tail to the minor groove are important determinants governing specificity. In addition, the binding of a few single-site mutants has been examined. The Cys[10] → Pro mutation abolishes binding (*4*), which can be attributed to the disruption of the zinc coordination site and the accompanying loss of the structural integrity of the protein core. The Leu[33] → Phe mutation reduces specific binding 10-fold (*4*), probably as a result of a change in the optimal hydrophobic packing of this residue with the bases of T23 and A24.

On the DNA side, chemical interference experiments have implicated every position between base pairs 5 and 11, indicating that methylation in either the major or minor grooves affects binding (*4, 5*). This result, as well as the deduced length of the DNA target site involved, is in accord with our structural data. Scanning mutagenesis of the chicken β-globin enhancer has indicated that the GATA sequence is a crucial determinant of activity (*46*), which is completely consistent with the structural observation that four bases in this region (G7, A8, A24, and T23) are involved in specific contacts. Binding studies with short synthetic oligonucleotides have shown that the cGATA-1 DNA binding domain has a higher affinity for the A[(6)]GATAAAC[(13)] sequence than for the T[(6)]GATAGCA[(13)] sequence (*5*). The AC to CA change at positions 12 and 13, would be expected to have no affect on binding, as these two sites are not involved in specific base contacts (Fig. 6). The A → T change at position 6 would be expected to have little if any effect, as the hydrophobic interaction between Leu[17] and the base would be preserved, if not enhanced. The A → G change at position 11, on the other hand, would be expected to reduce the binding affinity significantly as a result of the removal of the specific interaction between the methyl groups of Leu[37] and T22.

Implications for other members of the GATA family of proteins. The GATA family comprises GATA-1,2,3, and 4, each of which is expressed in different cell lineages (*1, 3, 47*). In addition, GATA binding proteins have also been identified in fungi (*1, 48*). Although these proteins differ in sequence and length, there is remarkable conservation of the amino acid sequence of the DNA binding domain consisting of residues 2 to 59 of the cGATA-1 DNA binding domain used in the present study. Thus, there is 80% sequence identity between the sequences of the DNA binding domains of GATA-1, GATA-2, and GATA-3 from species as diverse as chicken, man, and xenopus, and 65% identity between these DNA binding domains and that of xenopus GATA-4 [xGATA-4 (*47*)]. Among these sets of DNA binding domains, there is conservation of all of the residues involved in contacts with the DNA, with the exception of Arg[47], which is on occasion substituted conservatively by a Lys. Moreover, of the 17 residues that

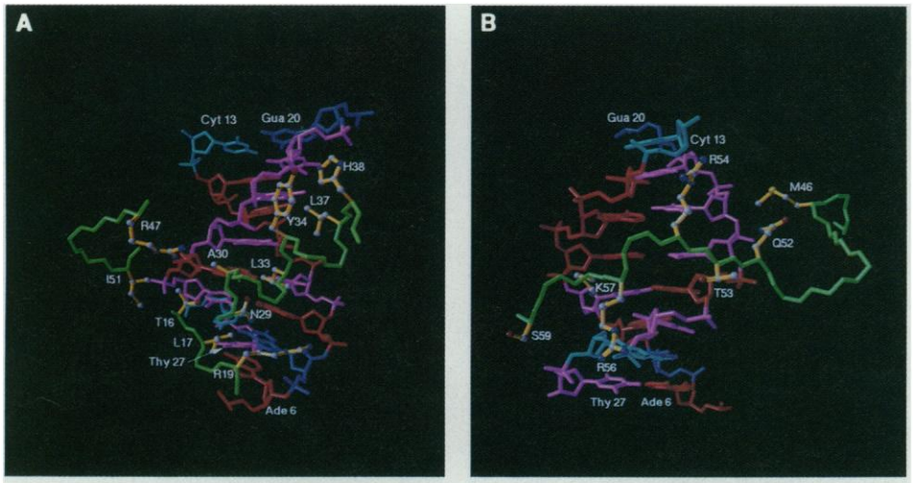

Fig. 6. Views of side chain interactions between the cGATA-1 DNA binding domain and the DNA in the major (**A**) and minor (**B**) grooves. The polypeptide backbone is shown in green, the side chain bonds in yellow, and the side-chain carbon, oxygen, and nitrogen atoms in gray, red, and blue, respectively. The color coding for the DNA is the same as in Fig. 5. (**C**) Sketch summarizing the contacts between the cGATA-1 DNA binding domain and DNA with the DNA represented as a cylindrical projection. Bases interacting with the protein are shaded; phosphates are represented as circles; circles with hatches directed toward the right and left indicate sites of interactions with sugar or phosphate or both in the major and minor grooves, respectively. The models in (A) and (B) were generated with the program VISP (*52*).

play a structural role in stabilizing the polypeptide fold of the core of the cGATA-1 DNA binding domain (compare Fig. 1A with Fig. 4B), only three residues display any variability, namely, Ser^{8} for Thr or Ala, Asp^{25} for Glu in xGATA-4, and Tyr^{35} for Phe in mGATA-1 or Met in xGATA-4. These substitutions are all conservative in nature. One can therefore conclude that the 3D structures and mode of specific DNA binding of this class of DNA binding domains will be similar to that described here for the cGATA-1 DNA binding domain. In the case of the fungal GATA-like proteins, there is a high degree of conservation of the residues that make contact with the DNA in the major groove. Indeed, changes only occur at 2 of the 11 positions: namely, Thr^{16} can be replaced by a Pro and Tyr^{34} by a Phe. As the interactions of these residues with the bases are hydrophobic in nature (Fig. 6, A and C), one would not expect these changes to have a significant effect on DNA binding. On the other hand, for six of the seven residues that make contact with the DNA in the minor groove, substitutions occur. The majority are conservative, and hence will not affect DNA binding to any large degree. However, some of the substitutions at the position of Lys^{57} are not conservative in nature (such as Gly, Ile, and Leu). The only residue that makes a specific base contact in the minor groove is Lys^{57}, so any change at this location would be expected to have a significant impact on specificity.

The degree of sequence identity between the DNA binding domains among the members of the GATA family of proteins from different species is much larger than that between the DNA binding zinc domain and the amino-terminal zinc module present in the mammalian GATA proteins. The separation between the last Cys of the amino-terminal metal-binding module and the first Cys of the DNA binding domain is only 29 residues. As a result, the amino-terminal metal-binding module (residues 105 to 158 of intact cGATA-1) can only be aligned with residues 2 to 55 of the DNA binding domain (residues 159 to 223 of intact cGATA-1) with a sequence identity of 38%. This degree of identity, however, together with the preservation of the spacing between the four Cys residues involved in zinc coordination, are sufficient to predict that the structure of the core of the amino-terminal metal binding module is likely to be similar to that of the DNA binding domain. The absence of the basic carboxyl-terminal tail, which removes a key determinant of specific DNA binding in the minor groove, together with the presence of several substitutions at positions that contact the DNA in the major groove, explains the weak influence of the amino-terminal metal binding module on the DNA binding of cGATA-1 (*4*).

Relation to other zinc-containing DNA binding domains. The mode of specific DNA binding that is revealed in this structure is distinct from that observed for the other three classes of zinc-containing DNA binding domains whose structures have previously been solved (*13, 14*). Features specific to the complex with the DNA binding domain of cGATA-1 include the relatively small size of the DNA target site (eight base pairs, of which only a contiguous stretch of six is involved in specific contacts), the monomeric nature of the complex in which only a single zinc binding module is required for specific binding, the predominance of hydrophobic interactions involved in specific base contacts in the major groove, the presence of a basic carboxyl-terminal tail that interacts with the DNA in the minor groove and constitutes a key component of specificity, and finally the pincer-like nature of the complex in which the core and tail subdomains are opposed and surround the DNA just like a hand gripping a rope. The structure of the cGATA-1 DNA binding domain reveals a modular design. The fold of residues 3 to 39 is similar to that of the amino-terminal zinc binding module of the DNA binding domain of the glucocorticoid receptor, although, with the exception of the four Cys residues that coordinate zinc, there is no significant sequence identity between these regions of the two proteins. Residues 40 to 66 are part of a separate structural motif. In this regard, it is interesting to note that, in addition to both zinc binding modules being encoded on separate exons in the cGATA-1 gene (exons 4 and 5), the next intron-exon boundary lies between amino acids 39 and 40 (current numbering scheme) of the DNA binding domain, thereby separating the carboxyl-terminal zinc binding domain from the basic tail (*49*).

REFERENCES AND NOTES

1. S. H. Orkin, *Blood* **80**, 575 (1992); M. Minie *et al.*, *J. Cell Sci.* (suppl. 16), 15 (1992).
2. T. Evans and G. Felsenfeld, *Cell* **58**, 877 (1989); S. F. Tsai *et al.*, *Nature* **339**, 446 (1989); C. D. Trainor, T. Evans, G. Felsenfeld, M. S. Boguski, *ibid.* **343**, 92 (1990); L. I. Zon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6668 (1990).
3. M. Yamamoto *et al.*, *Genes Dev.* **4**, 1650 (1990); C. I. Ho *et al.*, *EMBO J.* **10**, 1187 (1991).
4. D. I. K. Martin and S. H. Orkin, *Genes Dev.* **4**, 1886 (1990); H-Y. Yang and T. Evans, *Mol. Cell. Biol.* **12**, 4562 (1992).
5. J. G. Omichinski *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1676 (1993).
6. A. Klug and D. Rhodes, *Trends Biochem. Sci.* **12**, 464 (1987); R. M. Evans and S. H. Hollenberg, *Cell* **52**, 1 (1988).
7. S. C. Harrison, *Nature* **353**, 715 (1991).
8. J. Miller, A. D. McLachlan, A. Klug, *EMBO J.* **4**, 1609 (1985).
9. L. P. Freedman *et al.*, *Nature* **334**, 543 (1988).
10. L. Keegan, G. Gill, M. Ptashne, *Science* **231**, 699 (1986).
11. J. G. Kim and L. D. Hudson, *Mol. Cell. Biol.* **12**, 5632 (1992).
12. M. S. Lee, G. P. Gippert, K. Y. Soman, D. A. Case, P. E. Wright, *Science* **245**, 635 (1989); J. G. Omichinski *et al.*, *Biochemistry* **29**, 9324 (1990); J. G. Omichinski *et al.*, *ibid.* **31**, 3907 (1991); R. E. Klevitt, J. R. Herriot, S. J. Horvath, *Proteins* **7**, 215 (1990); M. Kochoyan *et al.*, *Biochemistry* **30**, 3371 (1991); T. Hård *et al.*, *Science* **249**, 157 (1990); J. W. R. Schwabe, D. Neuhaus, D. Rhodes, *Nature* **348**, 458 (1990); P. J. Kraulis, A. R. C. Raine, P. L. Gadhavi, E. D. Laue, *ibid.* **356**, 448 (1992); J. D. Baleja, R. Marmorstein, S. C. Harrison, G. Wagner, *ibid.*, p. 450.
13. N. P. Pavletich and C. O. Pabo, *Science* **252**, 809 (1991); R. Marmorstein, M. Carey, M. Ptashne, S. C. Harrison, *Nature* **356**, 408 (1992).
14. B. F. Luisi *et al.*, *Nature* **352**, 497 (1991).
15. The DNA coding sequence for amino acids 158 to 223 of cGATA-1 (*2*) comprising the DNA binding domain was generated as an Nde I–Bam HI DNA fragment by using the polymerase chain reaction as previously described [S. J. Scharf, G. T. Horn, H. A. Erlich, *Science* **233**, 1076 (1986)], cloned into the *Escherichia coli* vector pET11A, and expressed in the host bacterial strain BL21(DE3) [F. W. Studier, A. H. Rosenberg, J. J. Dunn, J. W. Dubendorf, *Methods Enzymol.* **185**, 60 (1990)]. The cGATA-1 peptide, uniformly labeled (>95%) with ^{15}N or ^{13}C or both, was obtained by growing the cells in minimal medium with $[^{15}N]NH_4Cl$ or $[^{13}C_6]$glucose or both as the sole nitrogen and carbon sources, respectively. Cells were grown overnight at 37°C, after which protein expression was induced at 30°C for 4 hours with 0.5 mM isopropyl-β-D-thiogalactoside. The cells were harvested, resuspended in 50 mM tris buffer, pH 8.0, 5 mM EDTA, 5 mM benzamidine, and 5 mM dithiothreitol (DTT). The cells were lysed by passage through a French press and centrifuged at 100,000*g* for 1 hour. The supernatant was applied to a DEAE–Sepharose Fast Flow (Pharmacia) column (200-ml bed volume) equilibrated with buffer A (50 mM tris, pH 8.0, 5 mM DTT, and 5 mM EDTA). The flow-through from the DEAE column was collected, concentrated by ultrafiltration, and applied to an S–Sepharose Fast Flow (Pharmacia) column. The cGATA-1 DNA binding domain was eluted with a 0 to 1 M NaCl gradient in buffer A, and the fractions containing the product were pooled. The latter were subjected to further purification on a C-8 reversed-phase (Vydac) high-performance liquid chromatography (HPLC) column with a 1 to 100% acetonitrile gradient in 0.05% (v/v) aqueous trifluroacetic acid. The DNA binding domain was characterized by amino acid analysis and amino-terminal sequencing. The lyophilized protein was reconstituted with 1.1 eq of zinc, and the final pH slowly adjusted to 6.5 with NaOH.
16. The DNA binding site was prepared on an Applied Biosystem DNA synthesizer, purified by HPLC, and annealed as described previously [K. Sakaguchi *et al.*, *J. Biol. Chem.* **266**, 7306 (1991)].
17. The DNA binding domain of cGATA-1 was slowly added to a DNA solution (~200 μM DNA, 10 mM NaCl) until a 1:1 ratio of DNA to protein was attained. (Note that if an excess of protein was added, a precipitate quickly formed). The solution was then concentrated with a Centricon-3 (Amicon) concentrator to give a final complex concentration of 2 mM at pH 6.5 with 10 mM NaCl and 2.2 mM $ZnCl_2$. Three samples of the complex were prepared containing: (i) uniformly ^{15}N-labeled cGATA-1 DNA binding domain plus DNA in 90% H_2O/10% D_2O; (ii) uniformly $^{15}N/^{13}C$-labeled cGATA-1 DNA binding domain plus DNA in 90% H_2O/10% D_2O; and (iii) uniformly $^{15}N/^{13}C$-labeled cGATA-1 DNA binding domain plus DNA in 100% D_2O.
18. M. Ikura, L. E. Kay, A. Bax, *Biochemistry* **29**, 4659 (1990); G. M. Clore and A. M. Gronenborn, *Prog. Nucl. Magn. Reson. Spectrosc.* **23**, 43 (1991); A. Bax and S. Grzesiek, *Acc. Chem. Res.* **26**, 131

(1993); G. M. Clore and A. M. Gronenborn, in *Methods Enzymol.*, in press; G. W. Vuister *et al.*, *ibid.*, in press.
19. The sequential assignment of the ^{1}H, ^{13}C, and ^{15}N chemical shifts of the cGATA-1 DNA binding domain in the complex was achieved by means of through-bond heteronuclear correlations along the backbone and side chains. The following 3D experiments were used: ^{15}N-separated HOHAHA, HNHA, CBCANH, CBCA(CO)NH, HBHA(CO)NH, C(CO)NH, HCCH-COSY, and HCCH-TOCSY to demonstrate CαH(i)/CβH(i)-^{15}N(i)-NH(i), CαH(i)-^{15}N(i)-NH(i), ^{13}Cβ/Cα(i-1)-^{15}N(i)-NH(i), CβH/CαH(i-1)-^{15}N(i)-NH(i), ^{13}C$_j$(i-1)-^{15}N(i)-NH(i), H$_j$-^{13}C$_{j\pm1}$-H$_{j\pm1}$, and H$_j$-^{13}C$_{j\pm n}$-H$_{j\pm n}$ correlations, respectively (note that i refers to the residue number, and j to the carbon position along a side chain). Details of these experiments and original references are provided in the reviews cited in (*18*).
20. A. M. Gronenborn and G. M. Clore, *Prog. Nucl. Magn. Reson. Spectrosc.* **17**, 1 (1985); G. M. Clore and A. M. Gronenborn, *EMBO J.* **2**, 2109 (1983); R. M. Scheek *et al.*, *J. Am. Chem. Soc.* **105**, 2914 (1983); D. R. Hare *et al.*, *J. Mol. Biol.* **171**, 319 (1983); M. A. Weiss *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 130 (1984); D. J. Patel, L. Shapiro, D. R. Hare, *Annu. Rev. Biophys. Biophys. Chem.* **16**, 423 (1987); B. R. Reid, *Q. Rev. Biophys.* **20**, 1 (1987).
21. J. Jeener, B. H. Meier, P. Bachmann, R. R. Ernst, *J. Chem. Phys.* **71**, 4546 (1979); P. Plateau and M. Guéron, *J. Am. Chem. Soc.* **104**, 7310 (1982).
22. No isotope filtering is required for the assignment of the base NH protons as their ^{1}H resonances are at very low field (12 to 14 ppm) and do not overlap at all with those of the backbone amide protons of the protein.
23. M. Ikura and A. Bax, *J. Am. Chem. Soc.* **114**, 2433 (1992); G. Wider, C. Weber, R. Traber, H. Widmer, K. Wuthrich, *ibid.* **112**, 9015 (1990).
24. G. M. Clore and A. Bax, unpublished data.
25. D. Marion *et al.*, *J. Am. Chem. Soc.* **111**, 1515 (1989); D. Marion *et al.*, *Biochemistry* **29**, 6150 (1989); S. W. Fesik and E. R. P. Zuiderweg, *J. Magn. Reson.* **78**, 588 (1988); M. Ikura, L. E. Kay, R. Tschudin, A. Bax, *ibid.* **86**, 204 (1990); E. R. P. Zuiderweg, L. P. McIntosh, F. W. Dahlquist, S. W. Fesik, *ibid.*, p. 210.
26. M. Ikura *et al.*, *Science* **256**, 632 (1992).
27. The spectrum was obtained with a ^{1}H-^{15}N NOESY-HMQC sequence (*25*) with a ^{12}C, ^{14}N filter (*23*) prior to the ^{1}H evolution period in t_1 (G. M. Clore, unpublished results).
28. G. M. Clore *et al.*, *EMBO J.* **5**, 2729 (1986).
29. 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)]. In addition, 0.5 Å was added to the upper limit of distances involving methyl protons to account for the higher apparent intensity of methyl resonances [G. M. Clore, A. M. Gronenborn, M. Nilges, C. A. Ryan, *Biochemistry* **26**, 8012 (1987); G. Wagner *et al.*, *J. Mol. Biol.* **196**, 611 (1987)].
30. A. M. Gronenborn and G. M. Clore, *Biochemistry* **28**, 5978 (1989).
31. The minimum ranges used for the Φ, ψ, χ_1, and χ_2 torsion angle restraints were ±30°, ±50°, ±20°, and ±30°, respectively [P. J. Kraulis *et al.*, *Biochemistry* **28**, 7241 (1989)].
32. M. Nilges, G. M. Clore, A. M. Gronenborn, *Biopolymers* **29**, 813 (1990).
33. Quantitative $^{3}J_{HN\alpha}$, $^{3}J_{CC}$, $^{3}J_{CN}$ couplings were measured from a 3D HNHA spectrum (G. Vuister and A. Bax, *J. Am. Chem. Soc.*, in press), a 2D long-range carbon-carbon correlation spectrum [A. Bax, D. Max, D. Zax, *ibid.* **114**, 6924 (1992)] and a 2D ^{13}C-{^{15}N}-spin-echo difference constant time heteronuclear single quantum coherence spectrum (G. Vuister, A. Wang, A. Bax, *ibid.*, in press, respectively. Qualitative $^{3}J_{\alpha\beta}$ and $^{3}J_{N\beta}$ couplings were obtained from 3D ^{15}N-separated HOHAHA [G. M. Clore, A. Bax, A. M. Gronenborn, *J. Biomol. NMR* **1**, 13 (1991)] and HNHB [S. J. Archer *et al.*, *J. Magn. Reson.* **95**, 636 (1991)] experiments, respectively.
34. Approximate intraresidue and sequential interresidue interproton distance restraints involving the NH, CαH, and CβH protons were obtained from 3D ^{15}N- and ^{13}C-separated ROESY experiments (35- and 30-ms mixing times, respectively) [G. M. Clore, A. Bax, P. T. Wingfield, A. M. Gronenborn, *Biochemistry* **29**, 5671 (1990); G. M. Clore, A. Bax, A. M. Gronenborn, *J. Biomol. NMR* **1**, 13 (1991)] to avoid complications arising from spin-diffusion [A. A. Bothner-By *et al.*, *J. Am. Chem. Soc.* **106**, 811 (1984)].
35. R. Powers *et al.*, *Biochemistry*, in press.
36. S. Arnott *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 53 (1983); R. E. Dickerson *et al.*, *ibid.*, p. 13; O. Kennard and W. N. Hunter, *Angew. Chem. Int. Ed. Engl.* **30**, 1254 (1991).
37. Note that we did not assume a priori that the DNA was right-handed. Rather, this was unambiguously established from the pattern of NOEs (*20*).
38. The coordinates of the 30 final SA structures of the complex of the DNA binding domain of cGATA-1 with DNA, together with the coordinates of the restrained minimized mean structure, $\overline{(SA)}$r, and the complete list of experimental NMR restraints and of the ^{1}H, ^{15}N, and ^{13}C assignments have been deposited in the Brookhaven Protein Data Bank.
39. M. Nilges, G. M. Clore, A. M. Gronenborn, *FEBS Lett.* **229**, 317 (1988).
40. The hybrid distance geometry-simulated annealing protocol for (*39*) makes use of the program XPLOR [A. T. Brunger, G. M. Clore, A. M. Gronenborn, M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3801 (1986); A. T. Brunger, *XPLOR Version 3.1: A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, 1993)]. The target function that is minimized during simulated annealing comprises only quadratic 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 (*28*), and a quartic van der Waals repulsion term for the nonbonded contacts (*38*). The initial calculations were carried out in the absence of zinc and indicated that the coordination of the zinc to the Sγ atoms of Cys^{7}, Cys^{10}, Cys^{28}, and Cys^{31} was tetrahedral. In the subsequent calculations, the zinc atom was incorporated by introducing appropriate covalent restraints to ensure that the coordination geometry was tetrahedral with Cys–Zn bond lengths of 2.3 Å [G. P. Diakun, L. Fairall, A. Klug, *Nature* **324**, 698 (1986)]. All of the peptide bonds were restrained to be trans. There were *no* hydrogen-bonding, electrostatic, or 6-12 Lennard-Jones empirical potential energy terms in the target function.
41. J. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 99 (1988).
42. The helical parameters were calculated by using the convention of S. R. Diekmann *et al.* [*J. Mol. Biol.* **205**, 787 (1989)] with the program BASE-PAIR, version 2.0 [E. von Kitzing and S. Diekmann, *Eur. Biophys. J.* **15**, 13 (1987)].
43. G. Schwatzbauer, K. Schlesinger, T. Evans, *Nucleic Acids. Res.* **20**, 4429 (1992).
44. C. R. Kissinger *et al.*, *Cell* **63**, 579 (1990); C. Wolberger *et al.*, *ibid.* **67**, 517 (1991).
45. D. Eisenberg and A. D. McLaghlan, *Nature* **319**, 199 (1986).
46. M. Reitman and G. Felsenfeld, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6267 (1988).
47. C. Kelley, H. Blumberg, L. I. Zon, T. Evans, *Development*, in press.
48. Y. H. Fu and G. A. Marzluf, *Mol. Cell. Biol.* **10**, 1056 (1990); T. S. Cunningham and T. G. Cooper, *ibid.* **11**, 6205 (1991); P. Minehart and B. Magasanik, *ibid.*, p. 6216; B. Kudla *et al.*, *EMBO J.* **9**, 1355 (1990).
49. R. Hannon, T. Evans, G. Felsenfeld, H. Gould, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3004 (1991).
50. 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.
51. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
52. E. de Castro and S. Edelstein, *Visp 1.0 User's Guide* (University of Geneva, Geneva, 1992).
53. T. A. Jones and M. Kjeldgaard, *O Version 5 Manual* (University of Uppsala, Uppsala, Sweden, 1990).
54. M. L. Connoly, *J. Mol. Graphics* **3**, 19 (1985).
55. N. C. Seeman *et al.*, *J. Mol. Biol.* **104**, 109 (1976); J. M. Rosenberg, N. C. Seeman, R. O. Day, A. Rich, *ibid.*, p. 145.
56. B. R. Brooks *et al.*, *J. Comput. Chem.* **4**, 187 (1983); L. Nilsson and M. Karplus, *ibid.* **7**, 691 (1986); W. E. Reiher, thesis, Harvard University, Cambridge, MA (1985).
57. We thank D. Garrett, R. Powers, A. Bax, G. Vuister, F. Delaglio, and S. Grzesiek for useful discussions; R. Tschudin, K. Ikura, and J. Kauffman for technical support; E. de Castro and S. Edelstein for the program VISP; and T. Evans for providing and permitting us to use the sequence of xGATA-4 prior to publication. This work was supported by the AIDS-Directed Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C. and A.M.G.).

7 May 1993; accepted 24 June 1993