34. S. L. Buchwald and S. M. King, *ibid.* **113**, 258 (1991).
35. S. L. Buchwald, E. A. Lucas, J. C. Dewan, *ibid.* **109**, 4396 (1987).
36. This value was calculated from standard heats of hydrogenation of an alkyne and the values calculated by L. Radom [*Pure Appl. Chem.* **58**, 75 (1986)].
37. D. P. Hsu, E. A. Lucas, S. L. Buchwald, *Tetrahedron Lett.* **31**, 5563 (1990).
38. S. L. Buchwald and R. B. Nielsen, *J. Am. Chem. Soc.* **110**, 3171 (1988).
39. S. L. Buchwald, B. T. Watson, M. W. Wannamaker, J. C. Dewan, *ibid.* **111**, 4486 (1989).
40. M. Jensen and T. Livinghouse, *ibid.*, p. 4495.
41. N. Coles, R. J. Whitby, J. Blagg, *Synlett.*, 271 (1990).
42. ———, *ibid.*, 143 (1992).
43. R. Halterman, *Chem. Rev.* **92**, 965 (1992).
44. R. B. Grossman, W. M. Davis, S. L. Buchwald, *J. Am. Chem. Soc.* **113**, 2321 (1991).
45. F. R. W. P. Wild, M. Wasiucionek, G. Huttner, H. H. Brintzinger, *J. Organomet. Chem.* **288**, 63 (1985).
46. A. Schafer, E. Karl, L. Zsolani, G. Huttner, H. H. Britzinger, *ibid.* **328**, 87 (1987).
47. This method is based on the method described by Brintzinger for the analogous Ti compounds [F. R. W. P. Wild, L. Zsolnai, G. Hettner, H. H. Brintzinger, *ibid.* **232**, 233 (1982)].
48. R. Noyori, *Science* **248**, 1194 (1990).
49. W. A. Nugent, T. V. RajanBabu, M. J. Burk, *ibid.* **259**, 479 (1993).
50. R. F. Jordan and D. F. Taylor, *J. Am. Chem. Soc.* **111**, 778 (1989).
51. V. M. Dzhemilev *et al.*, *Izv. Akad. Nauk Kaz. SSR Ser. Khim.* **38**, 207 (1989).
52. K. S. Knight and R. M. Waymouth, *J. Am. Chem. Soc.* **113**, 6268 (1991).
53. U. Wischmeyer, K. S. Knight, R. M. Waymouth, *Tetrahedron Lett.* **33**, 7735 (1992).
54. A. H. Hoveyda, J. P. Morken, A. F. Houri, Z. Xu, *J. Am. Chem. Soc.* **114**, 6692 (1992).
55. T. Takahashi *et al.*, *ibid.* **113**, 6266 (1991).
56. D. P. Lewis, P. M. Muller, R. J. Whitby, R. V. H. Jones, *Tetrahedron Lett.* **32**, 6797 (1991).
57. S. C. Berk, R. B. Grossman, S. L. Buchwald, *J. Am. Chem. Soc.* **115**, 4912 (1993).
58. Supported at MIT by the National Institutes of Health (GM 34917), the Petroleum Research Fund (administered by the American Chemical Society), National Cancer Institute training grant #CA 09112 (R.D.B.), and the Office of Naval Research; S.L.B. was supported as an Alfred P. Sloan Fellow and a Camille & Henry Dreyfus Teacher-Scholar. We thank our many co-workers whose efforts provide a partial basis for this article, whose names appear in the references; and J. Stubbe, R. Grossman, J. Tidwell, and B. Warner for useful comments on the manuscript.

RESEARCH ARTICLE

Crystal Structure of a Five-Finger GLI-DNA Complex: New Perspectives on Zinc Fingers

Nikola P. Pavletich* and Carl O. Pabo

Zinc finger proteins, of the type first discovered in transcription factor IIIA (TFIIIA), are one of the largest and most important families of DNA-binding proteins. The crystal structure of a complex containing the five Zn fingers from the human *GLI* oncogene and a high-affinity DNA binding site has been determined at 2.6 Å resolution. Finger one does not contact the DNA. Fingers two through five bind in the major groove and wrap around the DNA, but lack the simple, strictly periodic arrangement observed in the Zif268 complex. Fingers four and five of GLI make extensive base contacts in a conserved nine base-pair region, and this section of the DNA has a conformation intermediate between B-DNA and A-DNA. Analyzing the GLI complex and comparing it with Zif268 offers new perspectives on Zn finger-DNA recognition.

Zinc fingers, of the type found in TFIIIA (*1*), are one of the most common DNA-binding motifs in eukaryotic transcription factors. This family of zinc finger proteins is characterized by the consensus sequence X_3-Cys-X_{2-4}-Cys-X_{12}-His-X_{3-5}-His-X_4 (where X is any amino acid residue); more than a thousand such zinc finger sequences have been reported (*2*). The zinc finger forms a compact globular structure that contains a β sheet and an α helix held together by a central Zn ion (*3*). The two cysteines, which are in the β sheet region, and the two histidines, which are in the α helical region, are tetrahedrally coordinated to the Zn. Crystallographic studies of a complex containing the three Zn fingers from the Zif268 protein (*4*) revealed that the Zif fingers bind in the major groove and wrap partway around the double helix. Residues from the NH_2-terminal portion of each α helix contact the bases, and a conserved pattern of side chain-base interactions is observed in the Zif complex.

Although only a small number of the known Zn finger proteins have been characterized in detail, it is clear that this family of proteins can recognize a diverse set of DNA sequences. For example, the *Drosophila* Hunchback protein recognizes a site that includes the sequence AAAAA (*5*); the human Sp1 protein recognizes a site that includes the sequence GGGGGC (*6*); and the human glioblastoma (GLI) protein recognizes a site that includes the sequence TGGGTGGTC (*7*). Preliminary attempts to model other Zn finger complexes suggested that Zif might provide a reliable basis for modeling complexes formed by closely related proteins [such as Sp1 and WT1 (*8*)], but it did not appear that Zif would provide a satisfactory basis for modeling complexes formed by other, more distantly related Zn finger proteins.

To help understand how Zn fingers can recognize such a diverse set of binding sites, we have studied a complex that contains the five Zn fingers from the human *GLI* oncogene. The *GLI* gene was first discovered because it was amplified in human glioblastomas (*9*), and GLI was later found to be amplified in other tumors (*10*). In vitro studies have shown that the GLI protein, in conjunction with the adenovirus E1A protein, can transform primary rodent cells (*11*). GLI is a sequence-specific DNA-binding protein, and three high-affinity sites have been recovered from a pool of human genomic DNA (*7*). Our crystals contain the five Zn fingers of the human GLI protein (Fig. 1A) bound to a 21–base pair (bp) DNA fragment (Fig. 1B) that includes a high affinity DNA-binding site. We now describe the crystal structure of the GLI complex at 2.6 Å resolution, compare it with Zif, and consider the broader implications for our understanding of Zn finger–DNA interactions.

Overall structure of the GLI complex. The overall structure of the GLI-DNA complex shows that fingers 2 to 5 fit in the major groove and wrap around the DNA for a full helical turn (Fig. 2 and Table 1). Finger 1 surprisingly does not contact the DNA but instead makes extensive protein-protein contacts with finger 2. The overall arrangement of the other fingers is generally similar to that observed for the fingers in the Zif complex. The α helix of each finger fits into the major groove, and the NH_2-terminal portion of each of these α helices

The authors are in the Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

*Present address: Memorial Sloan-Kettering Cancer Center, Cellular Biochemistry and Biophysics Program, 1275 York Avenue, New York, NY 10021.

is closest to the bases. Fingers 4 and 5 of GLI appear to be the most important for recognition. These fingers make extensive base contacts in a 9-bp region that is conserved in all of the known binding sites (*7*). Fingers 2 and 3 make a set of contacts with the DNA backbone, but a comparison of the high-affinity GLI binding sites (*7*) shows that the DNA sequence of this region can vary, and the only base contact in this region is a single hydrogen bond contributed by finger 2.

The DNA is most readily described by dividing it into three regions that have distinct conformations. The conserved region, recognized by fingers 4 and 5, has a structure that is intermediate between that expected for B-DNA and that expected for A-DNA. This region is underwound (relative to B-DNA), and has a deep major groove. The second region involves the

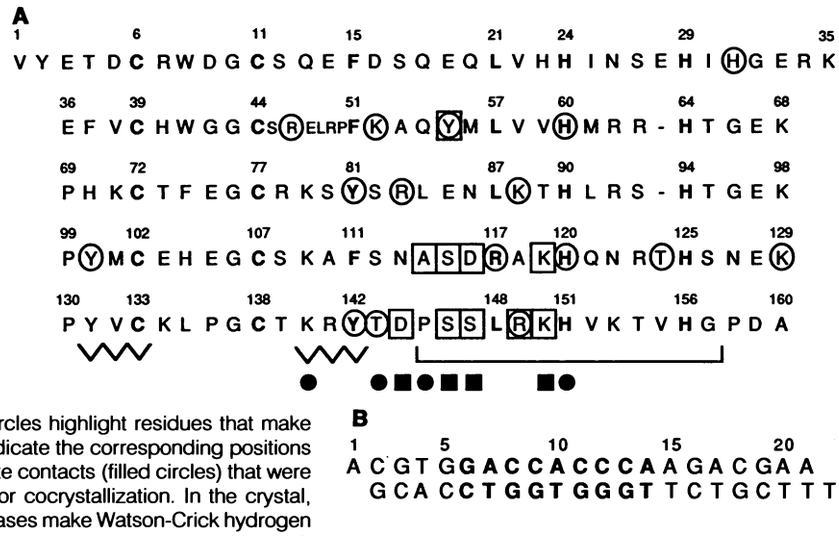


Fig. 1. Sequences of the GLI zinc finger domain and the DNA-binding site used for cocrystallization. (**A**) The peptide contains 160 residues [corresponding to amino acids 232 to 391 of the GLI protein (9)], and the current model includes residues 3 to 157 of this peptide. (The terminal residues are disordered in the crystal.) The GLI peptide (A) was cloned, expressed in *Escherichia coli*, and purified as described for Zif (*4*). The cloned gene was sequenced, and the identity of the purified peptide was confirmed by high-resolution mass spectroscopy (*27*). The five zinc fingers of GLI are aligned to show the conserved residues and secondary structures. The approximate position of the α helix is underlined, and that of the β sheet is indicated by zig-zag lines. In finger 2, the sheet region is perturbed by a three-residue insertion, and the second strand of the β sheet contains residues 49 to 51. Open boxes highlight residues that make base contacts in the crystal structure, and open circles highlight residues that make phosphate contacts. Symbols below the GLI sequence indicate the corresponding positions of side chain–base (filled boxes) and side chain–phosphate contacts (filled circles) that were observed in the Zif complex (4). (**B**) DNA duplex used for cocrystallization. In the crystal, these molecules stack head-to-tail, and the overhanging bases make Watson-Crick hydrogen bonds. Bold letters highlight the 9 base pairs that are conserved in the three high-affinity GLI binding sites (*7*). Binding studies were performed with the purified GLI peptide and a 45-bp DNA fragment that contained the site shown above. This complex has a dissociation constant of approximately 20 nM (in the presence of 200 mM NaCl and nonspecific DNA at 25 μg/ml. 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.

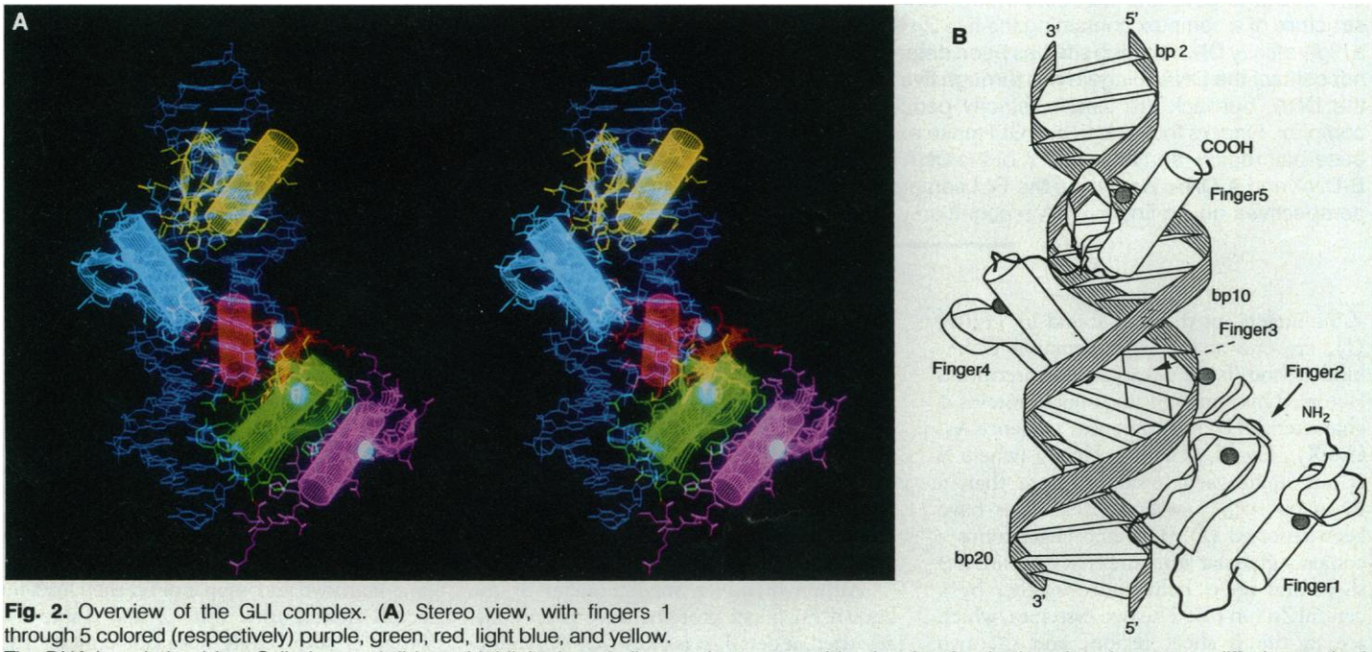


Fig. 2. Overview of the GLI complex. (**A**) Stereo view, with fingers 1 through 5 colored (respectively) purple, green, red, light blue, and yellow. The DNA is a darker blue. Cylinders and ribbons highlight the α helices and β sheets, and the cobalt ions are shown as light blue spheres. Note that finger 3 (in red) is directly behind the DNA in this view. (**B**) Sketch of the complex in a similar orientation. Samples were prepared for crystallization, by the addition of $CoCl_2$ to the purified peptide (2.0 molar equivalents of Co^{2+} per finger), adjusting the pH to 7.0 with 125 mM bis-tris-propane-HCl (BTP-HCl), and adding 1.2 molar equivalents of the DNA fragment (Fig. 1B) in 100 mM BTP-HCl pH 7.0. Crystals were grown in an anaerobic chamber by the hanging drop vapor diffusion method. The best crystals were obtained when hanging drops were prepared by mixing a solution of the complex (containing 0.5 mM protein and 0.6 mM DNA) with an equivalent volume of well buffer containing 60 to 100 mM $MgCl_2$, 20 to 25 percent PEG 400, and 50 mM BTP-HCl (pH 7.0). Crystals grew in 2 to 3 weeks. They form in space group $P2_12_12_1$ with a = 148.4 Å, b = 50.2 Å, and c = 45.4 Å; they have one complex in the asymmetric unit (*28*)].

subsites where fingers 2 and 3 bind, and here the structure is more characteristic of standard B-DNA. Finally, there is a region just beyond finger 5 where the DNA is overwound, with a narrow helical radius and a shallow major groove. The transition to this overwound region occurs just at the edge of the conserved recognition sequence, and thus is next to the bases that finger 5 contacts.

Protein-DNA interactions. Fingers 4 and 5 of GLI make most of the base contacts, and each of these fingers recognizes approximately half of a 9-bp conserved region (Fig. 1B). As observed with the Zif fingers, fingers 4 and 5 of GLI use residues from the NH_2-terminal portion of their α helices to contact the bases. Finger 4 has four residues—Ala^{114}, Ser^{115}, Asp^{116}, and Lys^{119}—that make base contacts (Figs. 3A and 4). These residues correspond to the first, second, third, and sixth residues of the α helix, and they make contacts over a 5-bp region of the GLI binding site. (Detailed contacts are listed in the legend to Fig. 4.) Finger 5 has five residues—Asp^{144}, Ser^{146}, Ser^{147}, Arg^{149}, and Lys^{150}—that make base contacts (Figs. 3A and 4). These residues correspond to the residue immediately preceding the α helix and to the second, third, fifth, and sixth residues of this helix. The contacts made by finger 5 are spread over a 4-bp region of the GLI binding site (Fig. 4). When the finger sequences are aligned (Fig. 1A), there is a clear correlation between the position of the GLI residues that make base contacts and the position of Zif residues that make base contacts. However, there are differences in the precise way that each finger docks against the DNA (Fig. 5), and there are corresponding differences in the detailed pattern of interactions between side chains and bases interactions (Fig. 6). Although these complexes are related (Fig. 5), the arrangements of the GLI fingers are different enough that these base contacts could not readily have been predicted from a knowledge of the Zif contacts.

Fingers 4 and 5 also make a number of contacts with the phosphodiester oxygens. Hydrogen bonds are made by the side chains of Tyr^{100}, Arg^{117}, His^{120}, and Thr^{124} in finger 4; by Lys^{129}, which is in the linker between fingers 4 and 5; and by the side chains of Tyr^{142} and Arg^{149} in finger 5 (Fig. 4). Finger 5 also makes one backbone-to-backbone contact. The peptide-NH of Thr^{143} hydrogen bonds to a phosphodiester oxygen. Comparison of the positions of these residues in the GLI sequence with the positions of Zif residues that make phosphate contacts (Fig. 1A) reveals only a weak correlation. Knowledge of the Zif contacts certainly would not have allowed us to predict the GLI contacts.

The docking arrangements for fingers 2 and 3 of GLI are generally similar to those observed with other fingers (Fig. 5), but these fingers only make a single base contact; Tyr^{55}, which is the second residue in the α helix of finger 2, accepts a hydrogen bond from the N4 of the cytosine at base pair 19 (Fig. 4). Fingers 2 and 3 make a number of contacts with the phosphodiester oxygens. Hydrogen bonds are made by the side chains of Arg^{46}, Lys^{52}, and His^{60} in finger 2, and by the side chains of Tyr^{81}, Arg^{83}, and Lys^{88} in finger 3. The peptide -NH of Tyr^{55} also contacts a phosphodiester oxygen. (The side chain of this residue makes the only base contact in this region.) An additional contact is provided by the side chain of His^{31}, which is in the linker region between fingers 1 and 2. Again, if we align fingers 2 and 3 of GLI with the Zif fingers, we find only a modest correlation in the positions of residues that make backbone contacts. Although the overall docking arrangements are similar in these two complexes, the fingers make significantly different contacts with the DNA backbone.

When considering the GLI complex, it is important to recognize that the binding site used in our study was obtained by in vitro selection from genomic DNA (*7*), and some caution must be exercised because the biologically relevant binding site is unknown. These concerns are relevant to fingers 1, 2, and 3 because the dearth of base contacts in this region raises the possibility that these fingers may represent a type of "nonspecific complex". We cannot

Table 1. Data collection and structure determination. Data were collected with a RAXIS-IIC area detector. At room temperature, the crystals deteriorated after x-ray exposure for 1 to 2 hours. Diffraction data were collected at −175°C to minimize radiation damage. Crystals were prepared for freezing by transferring them to a solution containing 15 percent isopropanol, 28 percent PEG 400, 50 mM $MgCl_2$, 50 mM BTP-HCl (pH 7.0), and 15 μM $CoCl_2$. Crystals were then placed in a thin wire loop and frozen in a stream of nitrogen at minus 175°C. Derivatives were prepared by substituting 5-iodouracil for thymine at bp 15, at bp 20, and at bp 21 (giving derivatives IdU^{15}, IdU^{20}, and IdU^{21}). Not all of the frozen crystals were isomorphous, and it was necessary to collect several native and several derivative data sets before we could solve the structure. Heavy-atom parameters for the three derivatives were refined (with the program REFINE from the CCP4 package) (*29*), and multiple isomorphous replacement (MIR) phases were calculated to 3.0 Å resolution with the program PHARE. Initial phases were improved by solvent flattening (*30*), and a model containing the DNA and fingers 2 through 4 was built into the MIR map by means of the program FRODO (*31*). This partial model was refined by simulated annealing with the program X-PLOR (*32*), and fingers 1 and 5 were located in $2|F_o| - |F_c|$ maps. After several cycles of refinement with the full model, X-PLOR omit maps were used to systematically check and rebuild every part of the complex. About 2 to 5 percent of the structure was deleted in each calculation, and simulated annealing was used to reduce model bias in the omit maps. The only region of the complex that consistently had poor electron density was the linker connecting fingers 1 and 2 (residues 32 to 35), and we presume that this region is disordered in the crystal. The TNT package (*33*) was used in the final stages of refinement. To correct for anisotropic diffraction and for absorption problems during data collection, a local scaling program was used (*34*); this program scales the observed and calculated structure factor amplitudes. After several additional cycles of refinement, 44 water molecules were gradually added and individual (correlated) temperature factors were included.

	Native 2	Native 1	IdU^{15}	IdU^{20}	IdU^{21}
Resolution (Å)	2.6	2.9	3.0	2.9	3.0
Measured reflections	28039	18499	17690	13389	7517
Unique reflections	10856	6738	5879	6811	4451
Data coverage (%)	97.9	88.3	85.2	89.3	64.5
R_{sym}*	6.72	9.35	8.16	5.90	8.70
MIR analysis (Native 1)					
Resolution limits (Å)			20.0–3.0	20.0–3.0	20.0–3.5
Mean isomorphous difference			0.15	0.14	0.12
Phasing power			1.60	1.48	1.26
Cullis *R* factor			0.74	0.73	0.89
Refinement: (Native 2)					
Resolution limits	7.0–2.6				
R factor	0.228				
Reflections with $F > 2\sigma$	9187				
Total number of atoms	2130				
Water molecules	44				
Rms in B values (Å^2)	2.50†				
Rms in bond lengths (Å)	0.015‡				
Rms in bond angles (deg.)	3.07‡				

*$R_{sym} = \Sigma_h\Sigma_i|I_{h,i} - I_h| / \Sigma_h\Sigma_i I_{h,i}$, where I_h is the mean intensity of the i observations of reflection h. Mean isomorphous difference $= \Sigma |F_{PH} - F_P| / \Sigma F_{PH}$, where F_{PH} and F_P are the derivative and native structure factor amplitudes, respectively. Phasing power $= [(F_{H(calc)})^2/|(F_{PH(obs)} - F_{PH(calc)})^2|]^{1/2}$. Cullis *R* factor $= \Sigma ||F_{der} \pm F_{nat}| - F_{H(calc)}| / \Sigma |F_{der} - F_{nat}|$ for centric reflections, where $F_{H(calc)}$ is the calculated heavy atom structure factor. *R* factor $= \Sigma ||F_{obs}| - |F_{calc}|| / \Sigma F_{obs}$. †The rms in B values is the rms deviation in temperature factors between bonded atoms. ‡Rms in bond lengths and bond angles is the rms deviation of bond lengths and bond angles from ideal values.

exclude the possibility that these fingers might bind in a different way or might make additional contacts at some other binding site. Such concerns appear to be less relevant in regard to fingers 4 and 5, since these fingers clearly bind in a sequence-specific manner and since the sequence of this region was the same in each of the high-affinity sites that was isolated.

DNA structure. Examination of the DNA revealed three regions with distinct conformations, and structural parameters were calculated with a separate helical axis for each of these segments. The region containing base pairs 15 to 21, which corresponds to the nonconserved portion of the GLI binding site, has structural parameters characteristic of standard B-DNA. This section of the DNA (region III in Table 2) has an average helical twist of 34.7°, which corresponds to 10.4 bp per turn. There are small displacements of the bases from the helical axis, but the major groove is only slightly deeper than in canonical B-DNA.

Region II, which consists of base pairs 7 to 15 and thus is essentially coincident with the 9-bp consensus sequence, has a structure that is intermediate between that expected for B-DNA and that expected for A-DNA. This section of the DNA has an average helical twist of 30.7° (11.7 residues per turn) and a rise of 3.08 Å per base pair. Region II has an unusually deep major groove (the base pairs are displaced by about 3 Å) and a rather wide minor groove. The base pairs have a significant inclination (about 9°), and the backbone δ angles show a 50/50 mixture of values expected for B-DNA and values expected for A-DNA (*12, 13*). It is interesting that this region contains the conserved part of the GLI binding site and includes almost all of the base contacts.

Region I, which consists of base pairs 2 to 7, has several features that are characteristic of C-DNA. This region is overwound, with an average helical twist of 37.5° (corresponding to 9.6 bp per turn). This section of the DNA also has a shallow major groove (the helical axis passes through the minor groove) and an unusually small helical radius. Although region I is not the primary binding site for any of the fingers, the transition between Regions I and II occurs near one end of the consensus sequence and seems to affect the position of a few bases that are contacted by finger 5.

Finger structure. As expected, the structures of the GLI fingers are very similar to those observed in other TFIIIA-like zinc fingers: Each finger contains a β sheet and an α helix that are held together by a central Zn ion. Superimposing fingers also shows that their tertiary structures are very similar. For example, residues 70 to 95 from finger 3 of GLI can be superimposed on finger 1 of Zif with a root-mean-square deviation of 0.79 Å for the Cα coordinates (*14*).

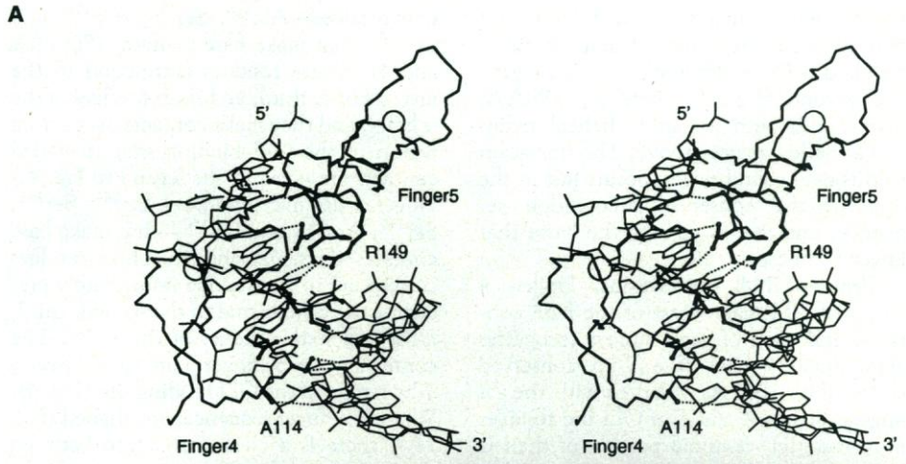


Fig. 3. Fingers 4 and 5 of GLI make extensive base contacts. (**A**) Fingers 4 and 5 of GLI and bp 6 to 15 of the DNA. The orientation is similar to that in Fig. 2, but the DNA has been tilted slightly to make the base contacts easier to see. Backbone atoms are shown for residues 99 to 157, and side chains are shown for residues that contact the bases: A^{114}, S^{115}, D^{116}, and K^{119} from finger 4; D^{144}, S^{146}, S^{147}, R^{149}, and K^{150} from finger 5. The cobalt ions are shown as circles. (To prevent the diagram from becoming too crowded, only residues A^{114} and R^{149} have been labeled.) (**B**) Electron density (blue) from a $2|F_o| - |F_c|$ map in the vicinity of finger 4 (contoured at 1.5 rms above the average density). The peptide is shown in yellow, and the side chains of Ser^{115}, Asp^{116}, and Lys^{119} are labeled. The DNA is shown in red. (The cytosine that is contacted by Asp^{116} is just beyond the block of density that is displayed in this figure.)

Some of the fingers in GLI have a His-X_3-His spacing in the α helical region while others have a His-X_4-His spacing. Thus fingers 2 and 3 of GLI have three residues between the histidines, and their α helices are similar to the Zif helices. Fingers 1, 4, and 5 of GLI have four residues between the conserved histidines. In these fingers, the regular α helix typically ends one residue before the second histidine, and the backbone continues with a wider turn. However, the second histidine is still constrained by the zinc ion, and the additional residue does not change the overall arrangement of the polypeptide chain as it comes out of the helical region. It is possible that structural differences in this COOH-terminal portion of the α helix affect the arrangement of the subsequent linker and finger (*15*) but the differences are complex enough that it is not possible to describe the His-X_3-His to His-X_4-His change in terms of a simple structural switch.

Finger 1, which does not contact the DNA, makes extensive protein-protein interactions with finger 2. Surprisingly, these fingers are related by a local twofold rotation axis, and they have a pair of hydrogen bonds connecting their polypeptide backbones. (The CO of residue 7 and the -NH of residue 9 are involved in hydrogen bonds with the -NH of residue 40 and the CO of residue 42.) These backbone hydrogen bonds flank tryptophans (Trp^8 and Trp^{41}) that make critical hydrophobic contacts at the interface of the two fingers. Many other residues contribute hydrophobic contacts or hydrogen bonds that stabilize this interaction (*16*).

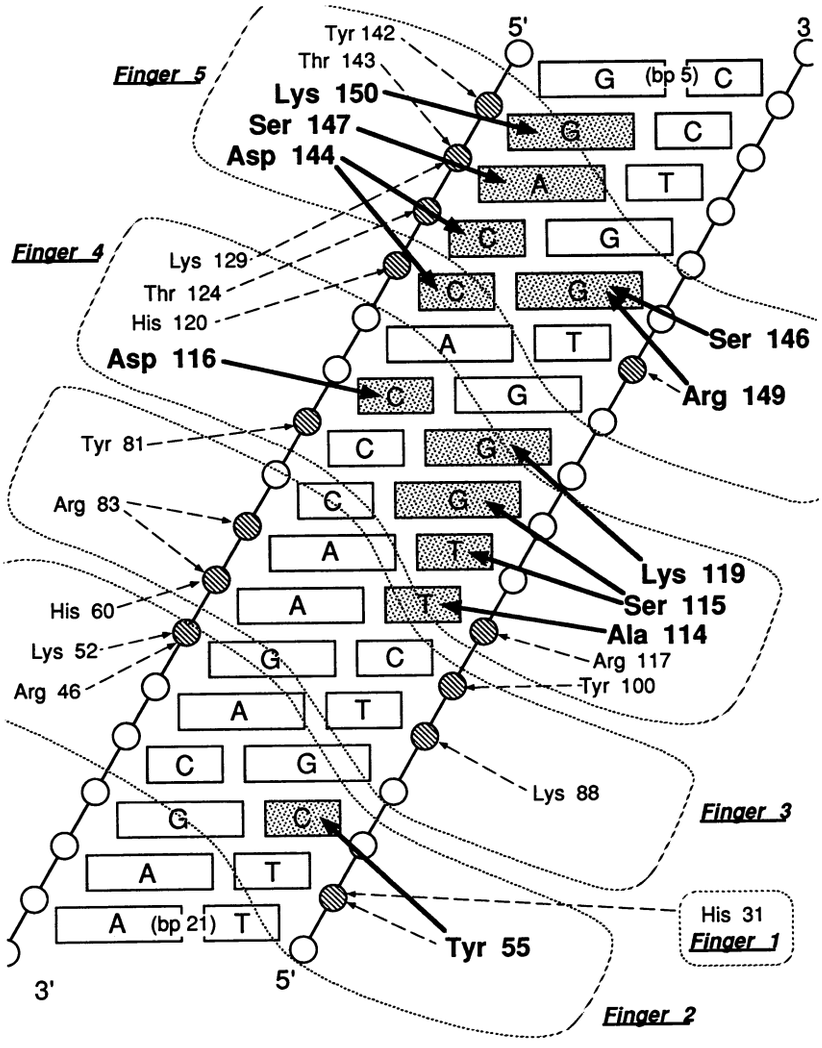
There only are a few contacts between the other fingers, and these contacts are similar to ones seen in Zif (*17*). The conformations of the linker regions between these other fingers are generally similar to those seen in Zif. However, there are differences in the precise arrangements, and these modest differences may help to deter-

Fig. 4 (right). Sketch summarizing base and phosphate contacts made by the GLI peptide. The DNA is represented as a cylindrical projection that has been "unwrapped" to provide a continuous view of the major groove over a region containing about 1.5 helical turns. Solid arrows indicate base contacts and dotted arrows indicate phosphate contacts. (The contacted bases and phosphates are shaded for clarity.) Base contacts made by fingers 4 and 5 are as follows: Ala 114 makes van der Waals contact (3.5 Å) with the methyl group of the T; Ser^{115} donates a hydrogen bond to the O6 of the G while the C_β of this residue makes van der Waals contact (3.1 Å) with the methyl group of the T; Asp^{116} accepts a hydrogen bond from the N4 of the C; Lys^{119} hydrogen bonds to the O6 of the G at bp 12. (The amino group of Lys^{119} also comes within 3.4 Å of the O6 of the G at bp 13.) Asp^{144} accepts a hydrogen bond from the N4 of each C; Ser^{146} hydrogen bonds to the O6 of the G; Ser^{147} hydrogen bonds to the N7 of the A; Arg^{149} donates hydrogen bonds to (i) the N7 of the G, (ii) the hydroxyl of Ser^{146}, and (iii) a phosphodiester oxygen from the T at bp 10; Lys^{150} hydrogen bonds to the N7 and O6 positions of the G.

mine the precise spacing and orientation of the neighboring fingers.

Perspectives on Zn finger–DNA recognition. The crystal structure of the Zif268 complex gave an initial view of Zn finger–DNA interactions, but with only one structure available it was impossible to determine which features were peculiar to Zif (and its closest relatives) and which features were characteristic of Zn fingers in general. Comparing the Zif and GLI structures provides a broader perspective on many of the fundamental questions about Zn finger–DNA interactions.

The overall arrangement of fingers 2 through 5 in GLI is similar to the arrangement in Zif. The fingers bind in the major groove with the NH_2-terminal portion of their α helices closest to the bases, and successive fingers wrap around the double helix (Figs. 2 and 5). The Zif and GLI complexes also have related patterns of side chain–base interactions (Fig. 6). Specifically: (i) There are four bases in each subsite that are contacted most frequently. (ii) There are four positions on the α helix that usually provide the critical contacts. (iii) There is a clear correlation between the position of a residue on the α helix and the position (within the finger's subsite) of the base that it contacts. (The correlation of residue positions and base positions is most striking for Zif, but more than half of the GLI contacts involve alignment of analogous residues and bases.) However, there also are differences in the precise docking arrangements, such as: (i) The subsites for fingers 4 and 5 of GLI are somewhat farther apart than any of the Zif subsites. (ii) A majority of the Zif contacts involve one



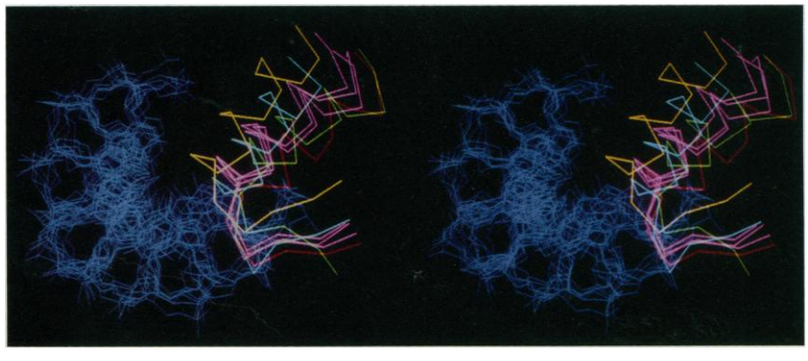


Fig. 5 (above). Comparison of docking arrangements in the GLI and Zif complexes. The complexes have been dissected to show the individual fingers and their subsites, and these "mini-complexes" have been aligned by superimposing the P atoms of the DNA backbone. Residues in the second β strand and residues in the α helix of each finger are shown as Cα traces. The three Zif fingers are purple; the GLI fingers are green (finger 2), red (finger 3), light blue (finger 4), and yellow (finger 5). The DNA subsites are blue, and the complexes are viewed looking down the DNA axes. The DNA subsites used in making these GLI mini-complexes have a simple 3-bp periodicity, and thus the subsites are centered on base pairs 7 to 9 (finger 5), 10 to 12 (finger 4), 13 to 15 (finger 3), and 16 to 18 (finger 2). All of the fingers have basically similar docking arrangements, but differences can involve translations as large as 5 Å and rotations as large as 25°, and these clearly are critical for recognition. (The Zif fingers are most similar to each other; docking arrangements in GLI are more variable.)

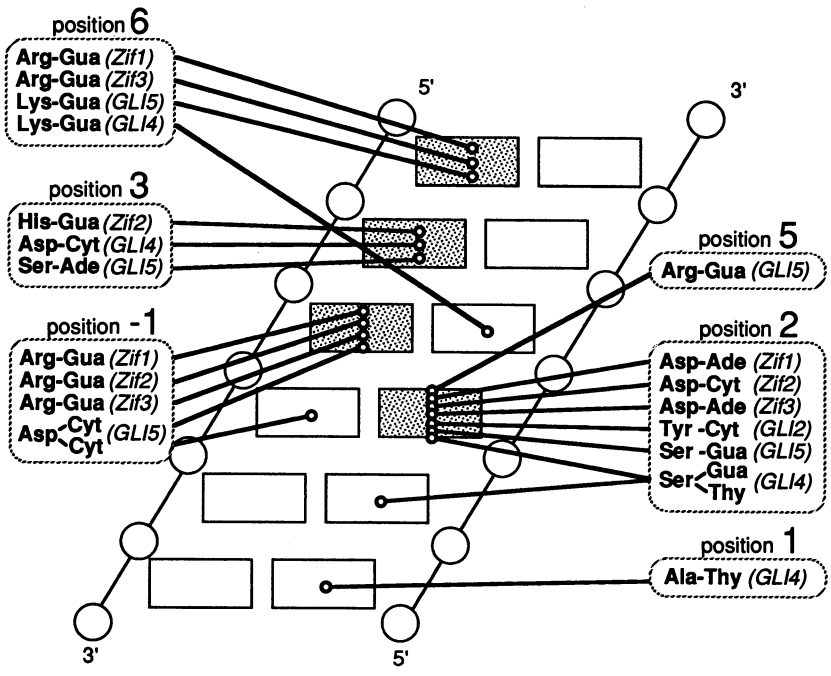


Fig. 6. Comparison of base contacts in the GLI and Zif complexes. Mapping base contacts with the subsites for individual fingers reveals similar patterns relating the position of a residue in the α helix and the position of the base that it contacts. The subsites used in Fig. 5 are used for the Zif fingers and for fingers 2, 3, and 4 of GLI. Finger 5 of GLI is aligned with the use of the side chain–base contacts to provide a frame of reference, and this gives a subsite centered on bp 6 to 8. (In this alignment, the subsites for fingers 4 and 5 are separated by an additional base pair.) Residue positions are numbered with respect to the start of the α helix (where −1 denotes the residue immediately preceding the helix). Rectangles represent the bases in a canonical subsite (viewing the major groove as in Fig. 4), and the three bases corresponding to the primary strand of the Zif subsites (*4*) have been highlighted to facilitate comparisons. This chart includes three hydrogen bonds—one made by each of the Asp residues at position 2 in the Zif helices—that were not emphasized in our original discussion of the Zif complex (*4*). Although their hydrogen bonding geometry is not ideal (*36*), these residues may contribute to specificity and are included here for completeness.

strand of the DNA, but the GLI contacts are more evenly distributed between the two strands. (iii) GLI makes a number of interactions between side chains and bases that were not seen with Zif. Comparing Zif and GLI makes it appear quite unlikely that there will be any simple, general "code" describing Zn finger–DNA interactions.

Comparison of the Zif and GLI complexes also reveals distinctive features of the DNA conformation. Although our original description of the Zif DNA (*4*) focused on features that were similar to those expected for B-DNA, further analysis shows that the Zif DNA is underwound and has a deeper major groove than would be expected for canonical B-DNA (*13, 18*). These features are much more pronounced in certain regions of the GLI complex. The conserved region recognized by fingers 4 and 5 of GLI has a DNA conformation that is intermediate between A-DNA and B-DNA, and this observation is consistent with reports indicating that the binding sites for other Zn fingers may have some features characteristic of A-DNA (*19*).

Table 2. DNA conformational parameters calculated with the program NEWHELIX (*35*). Twist and rise were calculated with the use of the C1′ atoms; slide, inclination, tip, and X dsp (X displacement) were calculated from the C_6 (pyrimidines) or C_8 (purines) atoms. The twist, rise, and slide values are measured with reference to the base pair shown in the next line of the table.

Base pairs	Twist (deg)	Rise (Å)	Slide (Å)	Inclination (deg)	Tip (deg)	X Dsp (Å)
			Region I			
2 C·G	45.18	3.26	2.17	−0.60	0.08	3.76
3 G·C	32.26	3.42	1.23	−0.25	3.24	3.08
4 T·A	46.89	2.89	1.98	−3.40	1.12	2.79
5 G·C	27.45	4.04	1.07	−4.06	−0.37	2.31
6 G·C	35.92	3.82	0.78	−4.67	7.62	3.03
7 A·T				−1.20	9.19	2.13
Mean	37.54	3.48	1.45	−2.37	3.48	2.85
			Region II			
7 A·T	35.59	3.10	−0.05	4.68	3.71	−2.49
8 C·G	26.87	3.06	−1.07	11.51	3.45	−1.91
9 C·G	38.34	3.11	−0.36	9.61	2.41	−2.62
10 A·T	31.22	3.43	−1.17	8.96	−3.38	−3.32
11 C·G	25.53	2.77	−2.40	5.98	3.23	−3.14
12 C·G	28.43	3.29	−0.57	1.63	4.12	−3.65
13 C·G	33.16	2.87	−1.14	11.67	14.13	−2.78
14 A·T	26.30	3.03	−1.09	13.48	12.72	−3.72
15 A·T				14.74	10.14	−4.06
Mean	30.68	3.08	−0.98	9.14	5.62	−3.08
			Region III			
15 A·T	40.33	2.50	−0.41	11.00	7.96	−0.63
16 G·C	35.02	3.51	−0.16	9.06	1.15	−0.77
17 A·T	22.46	3.82	0.01	5.65	1.38	−0.86
18 C·G	37.19	3.30	0.38	8.16	2.08	−1.01
19 G·C	42.47	2.66	−0.17	6.94	2.90	−0.26
20 A·T	30.85	2.69	−0.23	5.80	−6.27	−0.18
21 A·T				1.83	−14.42	−0.84
Mean	34.72	3.08	−0.10	6.92	−0.75	−0.65

In the Zif complex, each of the fingers makes generally similar contributions to DNA-binding, but the GLI complex shows that different fingers can have very different roles in recognition. Some fingers will make base contacts; some will make phosphate contacts; others will not contact the DNA at all. Recent chemical and biochemical studies of TFIIIA (*20*) have emphasized the distinct roles played by different sets of fingers, and the GLI complex provides a new perspective for thinking about such polyfinger complexes. The critical roles that fingers 4 and 5 play in the GLI complex may have a parallel in TFIIIA, where it has been shown that a peptide containing fingers 1 through 3 binds almost as tightly as the intact nine-finger protein (*21*). It should not be surprising that adequate specificity can be provided by a few critical fingers and that different fingers can have very different roles in a polydactyl complex.

The differences between Zif and GLI also suggest that it is necessary to keep track of relationships within families and subfamilies when modeling other Zn finger-DNA

interactions. Some proteins are siblings, while others are distant relatives. For example, Krox-20 (*22*), WT-1, and Sp1 have fingers homologous to the Zif fingers and appear to recognize related DNA sequences. GLI is a more distant relative of Zif, but has its own subfamily. Thus, GLI3 (*23, 24*), tra-1 (*25*), and ciD (*26*) have fingers homologous to those of GLI, and it has been shown that the GLI3 protein binds specifically to the GLI consensus binding site (*24*). Thus, modeling efforts should be far more reliable when the structure of a closely related complex is known.

Finally, the GLI structure has implications for attempts to design novel zinc finger proteins. It still appears that the Zn finger motif will provide an excellent framework for designing and selecting DNA-binding proteins with novel specificities. (The amazing diversity of TFIIIA-like fingers that are found in nature provides the clearest proof of principle for these design strategies.) However, the GLI structure shows that design strategies need not be constrained by rigid rules which assume that all fingers will dock in the same way. Seeing the variety of structural arrangements that occur in Zif and GLI helps us understand the versatility of this zinc finger motif. Screening and selection methods that can test millions of different arrangements should be inherently more powerful than rule-based design strategies. The possibilities for Zn finger–DNA interactions are richer and more complex than one may have assumed.

REFERENCES AND NOTES

1. J. Miller, A. D. McLachlan, A. Klug, *EMBO J* **4**, 1609 (1985).
2. A. Klug and D. Rhodes, *Trends Biochem. Sci.* **12**, 464 (1987); G. Jacobs and G. Michaels, *New Biol.* **2**, 583 (1990); G. H. Jacobs, *EMBO J.* **11**, 4507 (1992).
3. G. Párraga *et al.*, *Science* **241**, 1489 (1988); M. S. Lee, G. P. Gippert, K. V. Soman, D. A. Case, P. E. Wright, *Science* **245**, 635 (1989); R. E. Klevit, J. R. Herriott, S. J. Horvath, *Proteins* **7**, 215 (1990); J. G. Omichinski, G. M. Clore, E. Appella, K. Sakaguchi, A. M. Gronenborn, *Biochemistry* **29**, 9324 (1990).
4. N. P. Pavletich and C. O. Pabo, *Science* **252**, 809 (1991).
5. D. Stanojevic, T. Hoey, M. Levine, *Nature* **341**, 331 (1989); J. Treisman and C. Desplan, *ibid.*, p. 335.
6. M. R. Briggs, J. T. Kadonaga, S. P. Bell, R. Tjian, *Science* **234**, 47 (1986); J. T. Kadonaga, K. R. Carner, F. R. Masiarz, R. Tjian, *Cell* **51**, 1079 (1987); R. W. Kriwacki, S. C. Schultz, T. A. Steitz, J. P. Caradonna, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9759 (1992).
7. K. W. Kinzler and B. Vogelstein, *Mol. Cell Biol.* **10**, 634 (1990).
8. K. M. Call *et al.*, *Cell* **60**, 509 (1990); M. Gessler *et al.*, *Nature* **343**, 774 (1990); F. J. Rauscher III, J. F. Morris, O. E. Tournay, D. M. Cook, T. Curran, *Science* **250**, 1259 (1990).
9. K. W. Kinzler, J. M. Ruppert, S. H. Bigner, B. Vogelstein, *Nature* **332**, 371 (1988).
10. W. M. Roberts, E. C. Douglass, S. C. Peiper, P. J. Houghton, A. T. Look, *Cancer Res.* **49**, 5407 (1989).
11. J. M. Ruppert, B. Vogelstein, K. W. Kinzler, *Mol. Cell Biol.* **11**, 1724 (1991).
12. W. Saenger, *Principles of Nucleic Acid Structure*, (Springer-Verlag, New York, 1984).
13. L. Nekludova and C. O. Pabo, unpublished data.
14. Finger 2 of GLI is distinctive in that the sheet region is perturbed by an "insertion" just before the second β strand (Fig. 1A), and there are two consecutive reverse turns in this region.
15. M. Kochoyan, H. T. Keutmann, M. A. Weiss, *Biochemistry* **30**, 9396 (1991).
16. It is interesting that the linker between fingers 1 and 2 has two unusual features. (i) This linker is two residues longer than the consensus linker. (ii) Residues 33 to 36 may be relatively flexible, since they do not have clear electron density in our maps. (Other linkers have clear density and thus appear to have a fixed structure in the GLI complex.)
17. The side chain of Arg^{62} (in finger 2) hydrogen bonds to the backbone carbonyl of Ser^{82} (in finger 3), and there is an equivalent interaction between Arg^{92} (finger 3) and the carbonyl of Ser^{112} (finger 4). (Finger 4, which has an extra residue just before this arginine, does not make a corresponding contact with finger 5.)
18. A. A. Travers, *Curr. Opin. Struct. Biol.* **2**, 71 (1992).
19. M. McCall, T. Brown, W. N. Hunter, O. Kennard, *Nature* **322**, 661 (1986); L. Fairall, S. Martin, D. Rhodes, *EMBO J.* **8**, 1809 (1989); P. W. Huber, G. C. Blobe, K. M. Hartmann, *J. Biol. Chem.* **266**, 3278 (1991); P. W. Huber, T. Morii, H-Y. Mei, J. K. Barton, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10801 (1991).
20. M. E. A. Churchill, T. D. Tullius, A. Klug, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5528 (1990); J. J. Hayes and T. D. Tullius, *J. Mol. Biol.* **227**, 407 (1992); J. J. Hayes and K. R. Clemens, *Biochemistry* **31**, 11600 (1992); K. R. Clemens, X. Liao, V. Wolf, P. E. Wright, J. M. Gottesfeld, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10822 (1992).
21. X. Liao, K. R. Clemens, L. Tennant, P. E. Wright, J. M. Gottesfeld, *J. Mol. Biol.* **223**, 857 (1992).
22. P. Chavrier *et al.*, *Mol. Cell Biol.* **9**, 787 (1989).
23. J. M. Ruppert *et al.*, *ibid.* **8**, 3104 (1988); A. Vortkamp, M. Gessler, K.-H. Grzeschik, *Nature* **352**, 539 (1991).
24. J. M. Ruppert, B. Vogelstein, K. Arheden, K. W. Kinzler, *Mol. Cell Biol.* **10**, 5408 (1990).
25. D. Zarkower and J. Hodgkin, *Cell* **70**, 237 (1992).
26. T. V. Orenic, D. C. Slusarski, K. L. Kroll, R. A. Holmgren, *Genes Devel.* **4**, 1053 (1990).
27. Mass was determined by the electrospray ionization method at the Harvard Microchemistry Facility, 16 Divinity Avenue, Cambridge, MA 02138.
28. Attempts to obtain GLI crystals with peptide reconstituted with $ZnCl_2$ have been unsuccessful. Comparative studies with Zif had shown that the Zn complex and the Co complex were virtually identical (*4*). It is possible that Zn interferes with the crystallization of GLI by interacting—at these high peptide concentrations—with the additional (nonconsensus) histidines that appear in fingers 1, 2, and 3 of GLI.
29. S.E.R.C. [U.K.] Collaborative Computing Project No. 4 (Daresbury Laboratory, Warrington, U.K., 1979).
30. B. C. Wang, *Methods Enzymol.* **115**, 90 (1985).
31. T. A. Jones, *J. Appl. Cryst.* **11**, 268 (1978).
32. A. T. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987); A. T. Brünger, *X-PLOR v2.1 Manual* (Yale Univ. Press, New Haven, CT, 1990).
33. D. E. Tonrud, L. F. Ten Eyck, B. W. Matthews, *Acta Crystallogr.* **A43**, 489 (1987).
34. Program DSCALEAD, M. R. Rould (1989).
35. Program NEWHELIX (May 1989 Version), R. E. Dickerson, Molecular Biology Institute, UCLA, Los Angeles, CA 90024.
36. Binding experiments with a variant operator site also suggested that these contacts are not critical for recognition (N. P. Pavletich and C. O. Pabo, unpublished data).

37. Supported by the Howard Hughes Medical Institute, by NIH grant GM-31471 (C.O.P.), and by equipment purchased with support from the PEW Charitable Trusts. We thank K. Kinzler and B. Vogelstein for providing us with the GLI cDNA clone; M. Rould, C. Kissinger, and L. Nekludova for help in structure determination and interpretation; H. Drew and A. Klug for comments on the Zif DNA structure; and the National Cancer Institute for use of the Cray during molecular dynamics refinement. Coordinates are being deposited with the Brookhaven Data Bank. While these are being processed and prepared for distribution, interested scientists may obtain a set of coordinates by sending an appropriate e-mail address to PABO@PABO1.MIT.EDU.

24 February 1993; accepted 17 August 1993