Zinc Finger–DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å

Nikola P. Pavletich and Carl O. Pabo

The zinc finger DNA-binding motif occurs in many proteins that regulate eukaryotic gene expression. The crystal structure of a complex containing the three zinc fingers from Zif268 (a mouse immediate early protein) and a consensus DNA-binding site has been determined at 2.1 angstroms resolution and refined to a crystallographic R factor of 18.2 percent. In this complex, the zinc fingers bind in the major groove of B-DNA and wrap partway around the double helix. Each finger has a similar relation to the DNA and makes its primary contacts in a three–base pair subsite. Residues from the amino-terminal portion of an α helix contact the bases, and most of the contacts are made with the guanine-rich strand of the DNA. This structure provides a framework for understanding how zinc fingers recognize DNA and suggests that this motif may provide a useful basis for the design of novel DNA-binding proteins.

The zinc finger is one of the major structural motifs involved in eukaryotic protein–nucleic acid interactions. The fingers that were first discovered in the Xenopus transcription factor IIIA (TFIIIA) (1) contain a sequence motif of the form X3-Cys-X2.4-Cys-X12-His-X3.4-His-X4 (where X is any amino acid), and hundreds of similar finger sequences have been reported (2). Only a few of the proteins that contain such fingers have been studied in detail, but it appears that many of these zinc finger domains are involved in DNA binding. Proteins with zinc finger domains are involved in many aspects of eukaryotic gene regulation. For example, such fingers occur in proteins induced by differentiation and growth signals [EGR1 (3, 4), EGR2 (5, 6)], in proto-oncogenes [GLI (7), Wilms’ tumor gene (8)], in general transcription factors [Sp1 (9)], in Drosophila segmentation genes [Hunchback (10), Kruppel (11)], and in regulatory genes of lower eukaryotic organisms [ADR1 (12), BrA (13)]. The term zinc finger has been used in many different ways, and therefore it is important to realize that the characteristic sequence patterns and the three-dimensional structures of the TFIIIA-like zinc fingers are distinct from those of the cysteine-rich motifs in the steroid receptors (14), the cysteine-rich motif in the yeast transcription factor GAL4 (15), or the Cys-Cys-His-Cys motif in a set of retroviral proteins (16). In our study, we have focused on zinc fingers that are homologous to the fingers in TFIIIA.

Recent NMR (nuclear magnetic resonance) studies (17, 18) have shown that the TFIIIA-like zinc fingers contain an antiparallel β ribbon and an α helix. The two invariant cysteines, which are near the turn in the β-ribbon region, and the two invariant histidines, which are in the COOH-terminal portion of the α helix, coordinate a central zinc ion, and the finger forms a compact globular domain. There has been no structural information about how individual fingers interact with the DNA, or how proteins with tandemly repeated fingers recognize their binding sites. Sequence comparisons and mutational analyses have been used to propose a model for the zinc finger–DNA interactions (19), but no crystal structures or NMR studies of complexes have been reported.

To understand how these zinc finger domains are used in site-specific recognition, we have expressed and purified a peptide that contains the DNA-binding domain from the mouse immediate early protein Zif268 (4) (also known as Krox-24, NGFI-A, and Egr1). We have crystallized this three–zinc finger peptide with a Zif268 consensus binding site (20), solved the structure of this complex at 2.1 Å resolution, and refined it to an R factor of 18.2 percent. Here we report the structure of the complex, discuss the implications for our understanding of zinc finger–DNA interactions and protein-DNA recognition, and consider the prospect of using zinc finger motifs as a basis for designing novel DNA-binding proteins.

Cloning and purification of the Zif268–zinc finger peptide. The portion of the zif268 cDNA that codes for the three zinc fingers, corresponding to residues 349 to 421 of Zif268 (4) (Fig. 1A), was amplified in a polymerase chain reaction (PCR). One of the PCR primers introduced an Nde I restriction enzyme site with an in-frame ATG start codon at the 5' end of the zinc finger domain (a silent mutation that eliminated a second Nde I site was also introduced). The second PCR primer introduced a TAG stop codon, and a Bam HI restriction enzyme site at the 3' end. The amplified DNA was cloned into the Nde I and Bam HI sites of the...
the percent, and the resolution for mM pET3a(21) expression plasmid have been in the cloning. Only residues 3 to 87 are present in the current model. (We presume that the terminal residues are disordered in the crystal.) The three zinc fingers are aligned to show the conserved residues and secondary structures. Helices are boxed, and the β sheets are indicated by zig-zag lines. The approximate positions for these regions of secondary structure could have been predicted from NMR studies of related zinc fingers; the precise positions were determined from our crystal structure. (B) DNA duplex used for co-crystallization. The subunits that the fingers bind to are either in shadowed or in bold letters. These alternate to highlight the 3-bp subunits recognized by the zinc fingers.

The insoluble peptide-Co2+-DNA complex was harvested by centrifugation, dissolved in a solution of 6.4 M guanidinium-HCl and 50 mM tris, pH 7.4, and reduced with 150 mM DTT at 75°C for 30 minutes. The peptide was extracted from the solution in batch mode with the use of C4 reversed-phase resin (Vydac), eluted with 40 percent CH3CN and 0.1 percent trifluoroacetic acid (TFA), and was lyophilized. The reduction reaction was repeated, and the peptide was purified by chromatography as follows. (i) The peptide was purified on a C4 reversed-phase (Vydac) high-performance liquid chromatography (HPLC) column in 0.1 percent TFA, with a CH3CN gradient. (ii) The peptide was reconstituted with zinc in a buffer of 50 mM 2-[N-morpholino]ethanesulfonic acid–Na+, pH 6.2, and 2 mM DTT, and purified on a MonoS cation exchange column (Pharmacia) with a NaCl gradient. (iii) The peptide was loaded onto a C4 reversed phase HPLC column in 0.1 percent TFA, and eluted with a shallow CH3CN gradient. The HPLC peak fractions were alternately resuspended in water and lyophilized three times to remove the excess TFA, and the final dried product was stored in an anaerobic chamber (Coy Laboratory Products), where the oxygen content was kept below 1 part per million. The identity of the peptide was confirmed by analysis of amino acid composition. Mobility shift experiments showed that the Zif268–zinc finger peptide binds specifically to the consensus DNA sequence shown in Fig. 1B, with a dissociation constant (Kd) of 6 nM (22).

**Crystallography and structure determination.** The Zif268–zinc finger peptide–DNA complex (Zif complex) was prepared by adding 1.5 molar equivalents of cobalt chloride or zinc chloride to the apoprotein, adjusting the pH to 8.0, adding 1 molar equivalent of the buffered DNA-binding site, and solubilizing the complex by the addition of NaCl. Cocrystals were grown in the anaerobic chamber by the hanging drop vapor diffusion method. The best crystals were obtained by mixing a 1 mM solution of the complex (in 450 to 750 mM NaCl, 125 mM bis-tris propane-HCl, pH 8.0) with an equal volume of the buffer used in the crystallization well (0.0 to 10.0 percent polyethylene glycol (PEG) 400, 350 to 650 mM NaCl, and 25 mM bis-tris propane-HCl, pH 8.0); the crystals grew over the course of 2 weeks. The crystals formed in the space group C2221 with a = 45.4 Å, b = 56.2 Å, and c = 130.8 Å, and have one complex in the asymmetric unit.

Diffraction data were collected with the Siemens area detector (Table 1). Isomorphous derivatives were obtained by preparing duplex DNA in which 5-iodouracil was substituted for thymine at specific positions (at base pair 5 on the upper strand for IdU8, and at the 5' end of the lower strand for IdU12). The structure was first solved and refined with data from crystals of the peptide-Co2+-DNA complex. For our final structure, high-resolution data were collected with crystals of the peptide-Zn2+-DNA complex at the Midwest Area Detector User Facility. The structure has been refined against these data, and we see no significant differences between the Zn2+ and the Co2+ structures.

Heavy-atom parameters were first refined with the use of the program REFINe from the CCP4 (23) package of crystallographic programs. These data were used to determine initial multiple isomorphous replacement (MIR) phases at 2.5 Å resolution, and phasing with the program PHARE (from the CCP4 package) gave a mean figure of merit of 0.61. After solvent flattening [Wang's protocol (24)] this MIR map had very good density for the entire complex. The map was further improved by refining the heavy atom parameters against the solvent-flattened phases (25). Two additional cycles of phasing, solvent flattening, and parameter refinement were used to calculate the final map. Phases for the last MIR map had a mean figure of merit of 0.62 for data from 20.0 to 2.5 Å, and this MIR map was used with the FRODO (26) graphics program to build a model of the complex.

The electron density for each of the three zinc fingers was excellent.

### Table 1. Statistics from the crystallographic analysis.

<table>
<thead>
<tr>
<th>Item</th>
<th>Native</th>
<th>Native</th>
<th>IdU8</th>
<th>IdU12</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metal ion</strong></td>
<td>Zinc</td>
<td>Cobalt</td>
<td>Cobalt</td>
<td>Cobalt</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>2.1</td>
<td>2.4</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Measured reflections</td>
<td>34488</td>
<td>27330</td>
<td>27264</td>
<td>25615</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>9458</td>
<td>5880</td>
<td>5221</td>
<td>5320</td>
</tr>
<tr>
<td>Data coverage (percent)</td>
<td>93.4</td>
<td>86.5</td>
<td>85.6</td>
<td>88.0</td>
</tr>
<tr>
<td>Rsym</td>
<td>4.81</td>
<td>4.44</td>
<td>6.67</td>
<td>5.39</td>
</tr>
<tr>
<td><strong>MIR analysis:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution limits (Å)</td>
<td>20.0-2.5</td>
<td>20.0-2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean isomorphous difference</td>
<td>0.18</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phasing power</td>
<td>1.92</td>
<td>2.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cullis R factor</td>
<td>0.60</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Refinement:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution limits</td>
<td>7.0-2.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R factor</td>
<td>0.182</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflections with F &gt; 2σ</td>
<td>9047</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total number of atoms</td>
<td>1290</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water molecules</td>
<td>129</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rms in B values (Å²)</td>
<td>3.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rms in bond lengths (Å)</td>
<td>0.014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rms in bond angles (deg.)</td>
<td>2.34</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Refinement:**

\[
\text{R}_{\text{sym}} = \frac{\sum_i |I_{i|\text{obs}} - I_{i|\text{calc}}|}{\sum_i I_{i|\text{obs}}} \],

where \(I_{i|\text{obs}}\) is the mean intensity of the \(i\) observations of reflection \(h\). Mean isomorphous difference = \(\sum_{h \neq h} F_{h|\text{calc}} - F_{h|\text{obs}}\), where \(F_{h|\text{calc}}\) and \(F_{h|\text{obs}}\) are the derivative and native structure factor amplitudes, respectively. Phasing power = \((|F_{h|\text{calc}}|^2 - |F_{h|\text{obs}}|^2)^{1/2}\), Cullis R factor = \(\sum_{h \neq h} |I_{i|\text{calc}} - I_{i|\text{obs}}| / \sum_{h \neq h} |I_{i|\text{obs}}|\). **R factor** = \(\frac{\sum_{h|\text{calc}} |F_{h|\text{calc}} - F_{h|\text{obs}}|}{\sum_{h|\text{calc}} |F_{h|\text{calc}}|}\). The rms in B values is the rms deviation in temperature factors between bonded atoms. The rms in bond lengths and bond angles is the rms deviation in bond lengths and bond angles from ideal values.
and was fit with the NMR model of the Xin31 zinc finger (18). The linkers that connect neighboring fingers had strong density and were readily added in at this stage. The electron density for the DNA was fit with individual nucleotide bases that were extracted from a model of uniform B-DNA. We built 88 percent of the complex into this MIR map, leaving out residues 1 to 4 and 87 to 90 of the peptide since they did not have clear electron density. Even though all the side chains that made base contacts were well defined in the electron density map, they were omitted in the early stages of refinement in order to minimize model bias.

The model was refined with the program X-PLOR (27, 28), which applies molecular dynamics with crystallographic restraints. The first cycle of simulated annealing gave an R factor of 27.0 percent for data from 10.0 to 2.5 Å. Simulated annealing was repeated after minor rebuilding and after adding three of the missing residues and all of the critical side chains. The second cycle of simulated annealing refinement gave a model with an R factor of 25.0 percent for data from 7.0 to 2.4 Å.

This partially refined structure of the peptide-Co²⁺-DNA complex was used as the starting model for refinement against the high-resolution data from the peptide-Zn²⁺-DNA crystals. We used several additional rounds of simulated annealing, gradually added 129 water molecules, and refined individual temperature factors. In the final stages of refinement, the TNT (29) package was used for least squares refinement. The structure presented here has an R factor of 18.2 percent for data from 7.0 to 2.1 Å. The root-mean-square (rms) deviation from ideality for bond lengths is 0.014 Å, and the rms deviation for bond angles is 2.34°.

To determine whether the high salt concentration used for cocrystallization perturbed the structure in any way, data were also collected from cobalt cocrystals that had been equilibrated in low ionic strength buffer (about 150 mM NaCl). The low salt structure has been refined to an R factor of 23.2 percent for data from 7.0 to 2.2 Å, with individually refined temperature factors. Comparison with the high salt structure does not reveal any significant changes. Thus there is no indication that the high salt in the other crystals affected any key aspects of the structure. The crystals are quite stable at 150 mM NaCl, but higher salt concentrations are required to initially solubilize the complex for crystallization.

Overall structure of the zinc finger–DNA complex. The overall structure of the complex reveals why tandemly repeated zinc fingers are such efficient motifs for protein-DNA recognition: the three zinc fingers are arranged in a semicircular (C-shaped) structure that fits snugly into the major groove of B-DNA (Fig. 2). As expected from NMR studies of individual fingers (17, 18), each zinc finger domain consists of an antiparallel β sheet and an α helix, held together by a zinc ion and by a set of hydrophobic residues. Two cysteines from the β-sheet region and two histidines from the α helix coordinate the zinc ion.

Our cocystal structure shows that the α helix of each zinc finger fits directly into the major groove and that residues from the NH₂-terminal portion of each α helix contact the base pairs in the major groove. Each of the three Zif268 zinc fingers uses its α helix in a similar fashion, and each finger makes its primary contacts with a 3-bp subsite. The overall structure of the complex exhibits periodicity, with neighboring fingers related in a way that reflects the 3-bp periodicity of the subsites. A rotation of approximately 96° (3 × 32°) around the DNA axis, and a translation of approximately 10 Å (3 × 3.3 Å) along the DNA axis, move one finger onto the next. Although the α helix fits into the major groove, its axis is only approximately aligned with the groove, and the α helix is tilted at a somewhat steeper angle (about 45° with respect to the plane of the base pairs) than the angle of the major groove (32°). The β sheet is on the back of the helix away from the base pairs and is shifted toward one side of the major groove. The two strands of the β sheet have very different roles in the complex. The first β strand does not make any contacts with the DNA, whereas the second β strand contacts the sugar phosphate backbone along one strand of the DNA.

The Zif268–zinc finger peptide makes 11 critical hydrogen bonds with the bases in the major groove. The important side chains include an arginine that immediately precedes the α helix in each of the three fingers and also include the second, the third, and the sixth residues in the α helices. All of these hydrogen bonds involve bases on the G-rich strand of the consensus binding site (5' GIFCTGGCGCGC-3'). Using the 5' to 3' convention for the direction of a DNA strand and the N to C convention for the direction of a polypeptide strand, we might say that the overall arrangement of the peptide is "antiparallel" to the DNA strand that has most of the contacts. The peptide is arranged so that finger 1 binds near the 3' end of the primary strand (GCG TGG GCG); finger 2 binds near the center (GCG TGG GCG); and finger 3 binds near the 5' end (GCG TGG GCG). (Chemical analysis of the TFIIIA-DNA complex also indicated that most of the contacts involve the guanine-rich strand of the DNA and that the TFIIIA fingers also are "antiparallel" to the G-rich strand (30, 31).)
Fig. 3. The zinc fingers make extensive contacts with the bases, primarily along
the guanine-rich DNA strand. (A) Stereo diagram of the complex in an orientation similar to
that of Fig. 2A. To make the base contacts easier to see, the DNA is slightly tilted toward
the observer. Backbone atoms are shown for residues 3 to 87, and side chains are shown
for residues that contact the base pairs: R18, D20, R24 of finger 1; R46, D48, H49 of
finger 2; and R74, D76, R80 of finger 3. The zinc ions are shown as circles. (B) Sketch
summarizing the critical base contacts. The contacts of finger 3 are identical to those of
finger 1, but are not shown since they are obstructed by the DNA in this view. (C) Stereo
view of the calculated electron density from a 2F_o − F_c map in the vicinity of fingers
1 and 2. The peptide is shown in yellow, and the side chains of R24, R46, D48, and H49
are labeled. The DNA is shown in red, and the electron density is shown in blue. The
“stars” indicate water molecules. The electron density is contoured at a level of 1 rms
deviation above the average density. (D) Drawing of the Asp-Arg-guanine interaction
that is present in all three fingers. (E) Drawing of the Arg-guanine interaction that is
present in fingers 1 and 3. (F) Drawing of the His-guanine interaction that occurs in
finger 2.

Most of the contacts to the backbone of the DNA are made with
the primary, G-rich strand. In each finger an arginine on the second
β strand (two residues after the second cysteine) makes a contact to
a phosphodiester oxygen, and the first metal binding histidine on
the α helix makes a contact to another phosphodiester oxygen.

Contacts with the bases in the major groove. The zinc finger
motif uses the residue that immediately precedes the α helix, as well
as the second, third, and sixth residues of the α helix to contact the
base pairs (Fig. 3). Fingers 1 and 3 have exactly the same residues at
these critical positions, and they recognize identical sub sites (GCG).
Finger 2 has different residues at the third and sixth positions of the
α helix, and it recognizes a distinct sub site (TGG). The orientation of
each α helix, like that of the peptide as a whole, is “antiparallel”
to the primary (G-rich) strand of the DNA, with the NH_2-terminus
of the α helix near the 3' end of the sub site.

An arginine residue immediately precedes each of the three α
helices, and in each finger this arginine hydrogen bonds with the N7
and O6 of the guanine at the 3’ end of the subsite. Arg18 hydrogen bonds with the G at base pair 10, Arg49 hydrogen bonds with the G at base pair 7, and Arg74 hydrogen bonds with the G at base pair 4 (Fig. 3, A and B, and Fig. 5). Each of these arginine-guanine contacts is stabilized by a conserved aspartic acid that occurs as the second residue in each of the α helices. Both oxygens of the carboxylate group of the aspartic acid are in a hydrogen bond–salt bridge interaction with the Ne and Nq of the guanidinium group of the arginine (Fig. 3D). We presume that this side chain–side chain interaction helps position and stabilize the long side chain of the arginine and enhances the specificity of the arginine-guanine contacts. In addition to stabilizing the arginine, the aspartic acid at the second position in each of the α helices has one of its carboxylate oxygens within hydrogen bonding distance of a neighboring base on the secondary strand (C-rich strand) of the DNA. However, the geometry of these hydrogen bonds is not favorable, and we presume they do not play a major role in recognition.

Residue 3 is the next critical recognition residue on the α helix. In fingers 1 and 3, this position is occupied by a glutamic acid (Glu21 and Glu78, respectively) that does not contact the DNA. However, in finger 2 this third position is occupied by a histidine (His49) which forms a hydrogen bond with the guanine at the middle of the subsite (Fig. 3, A, B, and F). (This histidine is not a zinc ligand—those histidines occur later in the α helix.) In our model, the Ne of His49 donates a hydrogen bond to the N7 of the G at base pair 6, but a structure that has the His rotated 180° about the Cβ-Cγ bond and allows hydrogen bonding to the O6 of the G is equally consistent with the crystallographic data. The imidazole ring of His49 is coplanar with the G and stacks against the T of base pair 5. The stacking seems to limit the conformational flexibility of His49, thus enhancing the specificity of this hydrogen bonding interaction.

Residue 6 of the α helix is the last residue involved in base contacts. In fingers 1 and 3, this position is occupied by an arginine that donates a pair of charged hydrogen bonds to the N7 and O6 of the guanine at the 5’ end of the subsite (Fig. 3, A, B, and E). Arg24 hydrogen bonds with the G at base pair 8, and Arg80 hydrogen bonds with the G at base pair 2 (Fig. 3, A and B). Finger 2 has a threonine at this position, but Thr52 does not seem to participate in recognition.

In summary, whenever a residue is conserved at one of the recognition positions, it makes a conserved base contact. None of the Zif268 zinc fingers contacts all three bases, but there is a relatively simple pattern to recognition: the residue immediately preceding the α helix contacts the third base on the primary strand of the subsite (5’ - G), the third residue on the α helix can contact the second base on the primary strand (5’ - G -), and the sixth residue on the α helix can contact the first base (5’ G -) of the 3-bp subsite.

Contacts with the DNA backbone. As mentioned previously, the first histidine that coordinates the zinc ion also hydrogen bonds to a phosphate on the primary strand of the DNA. This histidine, which is the seventh residue in each of the α helices, coordinates the Zn2+ through its Ne and contacts the phosphodiester oxygens with its N8. This interaction is analogous to the zinc-histidine-carboxylate interaction observed in carboxypeptidase A (32, 33). As shown in Fig. 4, His25 contacts the 5’ phosphate of base pair 7, and His83 contacts the 5’ phosphate of base pair 4. These contacts are to a “neighboring subsite” and overlap a region where the next finger is contacting the bases. Because of this overlap, the phosphate that His81 (finger 3) would contact is outside the consensus binding site, and thus absent from our DNA duplex. Nevertheless, this finger makes an analogous interaction: A water molecule bridges the N8 of His81 to the 5’-OH of base 1. These histidine-phosphate contacts seem remarkable for several reasons. First, the zinc makes an unexpected and direct contribution to the overall binding energy. Second, since this contact is made by an invariant histidine we
expect that it is widely conserved among zinc finger–DNA complexes. Finally, this interaction uses the most central structural feature of the zinc finger—the tetrahedral geometry around the zinc ion—to orient the finger for site-specific recognition.

A conserved arginine on the second β strand also contacts phosphodiester oxygens on the primary DNA strand. In finger 1, Arg₁₄ contacts the 5′ phosphate of base 7; in finger 2, Arg₁₄ contacts the 5′ phosphate of base 5; and, in finger 3, Arg₇₀ contacts the 5′ phosphate of base 2. Fingers 2 and 3 contact equivalent phosphates (with respect to the 3-bp subsites), whereas the finger 1 contact is shifted by one nucleotide.

In addition to the conserved histidine and arginine contacts made by each finger, the Zif268 peptide makes four other phosphate contacts. Arg₉, which precedes finger 1, contacts the 5′ phosphate of base 8 on the primary DNA strand, and Arg₈⁷ contacts a phosphate on the primary strand of a symmetry-related DNA molecule. Ser⁴⁵ (finger 2) hydrogen bonds to the 5′ phosphate of base 6 on the primary strand, and Ser⁷⁵ (finger 3) hydrogen bonds to the phosphate between base pairs 7 and 8 on the secondary strand. The Ser⁷⁵ contact is the only backbone contact that the zinc finger makes with the secondary strand of the DNA.

Structure of the zinc finger. The overall structures of the three fingers are similar, and superimposing them gives excellent alignment of the corresponding residues (Fig. 6A). The α carbons of fingers 2 and 3 (residues 34 to 57 and 62 to 85) can be superimposed with a rms deviation of 0.45 Å. Finger 1 (excluding the two extra residues in the loop between the two cysteines of this finger) can be superimposed on finger 2, or on finger 3, with rms deviations of 0.83 Å and 0.87 Å, respectively. The NMR structure of Xfin31 (18) aligns best with finger 2 (residues 35–57) with an rms deviation of 0.74 Å. Since all three fingers have essentially the same secondary structures, we describe only the structure of finger 1 (Fig. 6B) and note significant differences as they occur.

Pro⁴ is the first zinc finger consensus residue of finger 1, and it has van der Waals contacts with the side chain of Tyr⁶, which is a highly conserved aromatic residue. This proline-aromatic interaction is conserved in fingers 2 and 3, and it may play an important role in restricting the conformation of the polypeptide chain at the start of the zinc finger.

The polypeptide chain then forms a hairpin with an antiparallel β-sheet stem and a turn near the conserved cysteines. This region encompasses residues 5 to 16 of finger 1. The antiparallel β-sheet has three backbone hydrogen bonds: the first one is between Tyr⁵-NH and Phe¹⁶-CO, the second one is between Tyr²-CO and Phe¹⁶-NH, and the third one is between Cys²-NH and Arg¹⁴-CO. The φ-ψ angles in this region are typical of an antiparallel β-sheet. In finger 1, the turn between these two β strands contains residues 8 to 13. In this turn, residues 9, 10, and 11 appear to be flexible and are poorly defined in our electron density maps. Fingers 2 and 3 have a shorter loop (two residues between the cysteines) and the turns in these fingers have well-defined conformations. Each of these fingers has a pair of conserved hydrogen bonds from backbone amides to the sulfhydryl of the first cysteine. In finger 2, Ile³⁹-NH and Cys⁴⁰-NH

Table 2. Local helical parameters for the DNA site.

<table>
<thead>
<tr>
<th>BASE PAIRS</th>
<th>HELICAL TWIST</th>
<th>RISE/_bp</th>
<th>PROPELLER Twist</th>
<th>TILT</th>
<th>ROLL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 C : G</td>
<td>39.392</td>
<td>2.938</td>
<td></td>
<td>-12.175</td>
<td>4.099</td>
</tr>
<tr>
<td>5 T : A</td>
<td>36.214</td>
<td>3.369</td>
<td></td>
<td>-10.329</td>
<td>2.054</td>
</tr>
<tr>
<td>6 G : C</td>
<td>31.203</td>
<td>3.266</td>
<td></td>
<td>-7.233</td>
<td>1.537</td>
</tr>
<tr>
<td>7 C : G</td>
<td>35.854</td>
<td>3.191</td>
<td></td>
<td>-11.769</td>
<td>5.805</td>
</tr>
<tr>
<td>9 C : G</td>
<td>39.706</td>
<td>3.451</td>
<td></td>
<td>-7.413</td>
<td>13.613</td>
</tr>
<tr>
<td>10 G : C</td>
<td>29.210</td>
<td>3.177</td>
<td></td>
<td>-5.057</td>
<td>12.863</td>
</tr>
</tbody>
</table>

MEAN: 31.967 3.298 -6.366 7.813 4.784
hydrogen bond with the Cys$^{57}$-S$_Y$, and in finger 3, Ile$^{67}$-NH and Cys$^{68}$-NH hydrogen bond with the Cys$^{65}$-S$_Y$

Only two residues are required to make the transition from the β sheet to the α helix. In finger 1, Ser$^{17}$ and Arg$^{18}$ connect these secondary structures. Although the backbone carbonyl of Arg$^{18}$ makes the first α-helical hydrogen bond with the backbone amide of Leu$^{22}$, the backbone of Arg$^{18}$ is not in an α-helical conformation. The regular α helix begins with Ser$^{19}$ and continues for several residues. A kink occurs near the center of the helix, and the kink is followed by several hydrogen bonds characteristic of a 3$\_10$ helical arrangement. Specifically, the kink occurs because Thr$^{23}$-CO and Arg$^{24}$-CO are too far away from their normal hydrogen bond partners (4.0 Å and 4.7 Å, respectively). The backbone carbonyl of His$^{25}$ hydrogen bonds with the Ile$^{28}$-NH in a 3$\_10$ helical arrangement, and Ile$^{28}$-CO makes a bifurcated hydrogen bond to both His$^{29}$-NH (as expected for a 3$\_10$ helix) and to Thr$^{29}$-NH (as expected for an α helix). This allows for a smooth transition back to a regular α helix. The next, and last, α-helical bond is between Arg$^{37}$-CO and Gly$^{38}$-NH. Gly$^{38}$ has nonhelical ϕ-ψ angles and essentially terminates the α helix. Overall, the kink bends the second half of the α helix toward the zinc binding site and thus puts the second histidine in a position where it can coordinate the metal. The same kink (involving an identical hydrogen bonding pattern) is also observed in fingers 2 and 3.

In addition to the zinc ion, each finger also is stabilized by a hydrophobic core involving the highly conserved Phe (which is residue 16 in finger 1), Leu (residue 22), and His (residue 25). This core also involves a number of moderately conserved hydrophobic residues (Val$^9$, Ile$^{26}$, Ile$^{28}$, and Thr$^{30}$ in finger 1) which form hydrophobic patches on both sides of the finger and help shield the zinc binding site from the solvent.

The sequence Thr-Gly-Glu-Lys has been called the “linker” [or H-C link (34)] since it occurs between fingers and since this sequence is conserved in a large number of zinc finger proteins. In the Zif268 peptide, the linker between fingers 1 and 2 (Thr-Gly-Glu-Lys) deviates slightly from the consensus, whereas the linker between fingers 2 and 3 exactly matches the consensus sequence. Our cocrystal structure shows that the first linker residue, Thr$^{30}$, is actually in the α helix of the zinc finger. As mentioned earlier, its methyl group is involved in hydrophobic interactions on one side of the finger. In addition, the -OH of this threonine hydrogen bonds to the backbone amide of the third linker residue (Gln$^{32}$). The second linker residue, Gly$^{31}$, makes the last hydrogen bond of the α helix and seems to play an important role in terminating the α helix. In our maps, the side chains for the third and fourth linker residues (Gln$^{32}$ and Lys$^{33}$) have weaker electron density, and they do not seem to have any important contacts with the rest of the protein or with the DNA. Both of the linkers in our cocrystal structure have the same well-defined backbone conformations, and they probably play an important role in controlling the orientation and spacing of adjacent fingers.

There are not many contacts between the fingers, but the side chain of the conserved Arg$^{27}$ hydrogen bonds to the backbone carbonyl of Ser$^{45}$, which is in the turn between the β sheet and the α helix of the next finger. The same interaction is also observed between fingers 2 and 3, where Arg$^{65}$ hydrogen bonds to the backbone carbonyl of Ala$^{73}$.

**Structure of the DNA.** There has been considerable speculation that zinc fingers may bind to A-DNA or to some very distinct form of B-DNA. However, we find that the 11-bp DNA in the cocrystal is essentially a B-type helix. The average rise per base pair is 3.3 Å (Table 2), which is very close to the 3.4 Å rise expected for B-DNA. The average helical twist of 32.0° (11.3 bp per turn) is a few degrees smaller than that expected for B-DNA (35) (34.3°, which corresponds to 10.5 bp per turn). This small difference could result from crystal packing forces since the 11-bp DNA duplexes stack end-to-end to form a pseudo-continuous helix, and thus the DNA is constrained to have an 11-bp repeat. Although the overall structure clearly is characteristic of B-DNA, the subsites for fingers 1 and 3 show considerable internal variations in their base pair twists. The helical twist between base pairs 2 and 3 is 24.3°, and the helical twist between base pairs 3 and 4 is 39.2° (Table 2). A corresponding arrangement is found in the other subsite. The helical twist between base pairs 8 and 9 is 26.6°, and the twist between base pairs 9 and 10 is 39.7°. These deviations in helical twists among neighboring base pairs tend to cancel each other, resulting in overall twists for subsites 1 and 3 that are very close to the average. Within each subsite however, these twists cause the positions of the bases C3 and C9 to be substantially different from what they would have been in an idealized B-DNA helix.

**Comparison of the orientations of the three fingers.** Each of the three fingers binds in a similar orientation and makes similar contacts with 3 bp of the DNA. This conserved spatial relation implies that each finger is related to the next by a simple helical motion. Formally speaking, this combination of a rotation and a translation is a “screw” motion. Rotating by about 96° (three times the average helical twist per base pair) and translating by about 10 Å (three times the average rise per base pair) will superimpose one finger on the next. This structure is consistent with the model for zinc finger binding in the major groove, which was favored by Berg (36, 37) after modeling the zinc finger domain. It is inconsistent with the model that has alternate fingers in rather different orientations (38).

The conserved spatial relation can be illustrated by dissecting the complex into three subsites (with a single finger bound to each) and then superimposing the DNA backbone for each subsite. The three fingers superimpose extremely well, even though the protein structure was not used when aligning the subsites (Fig. 7). This demonstrates that the three fingers of the Zif complex have very similar orientations with respect to the DNA. This conserved spatial arrangement is consistent with our observation that the three fingers contact the DNA with a similar set of residues and that they make a closely related set of side chain–base interactions. It appears that this orientation is ideal for the arginine-guanine contacts that are so critical for the Zif268 zinc fingers. However, it is possible that a somewhat different arrangement would be needed with shorter side chains, and it will be interesting to see how many different zinc finger–DNA relationships are found in other complexes.
Implications for understanding protein-DNA interactions. This structure reemphasizes the central role that α helices have in site-specific recognition. Previous crystallographic studies of protein-DNA complexes had shown how helices are used by the prokaryotic helix-turn-helix proteins (39–41) and by the eukaryotic homeodomain (42); our work now proves that helices are used by another major family of DNA-binding proteins—zinc fingers. NMR studies of the steroid receptor DNA-binding domain (14) and circular dichroism studies of the basic region of leucine zipper peptides (43) suggest that these proteins also use helices for site-specific recognition. Although extended secondary structures, such as the NH₂-terminal arms of λ repressor and Engrailed, and β sheets, as in the prokaryotic MetJ repressor (44), can also have a major role in protein-DNA recognition, it now appears that the major families of DNA-binding proteins use α-helical regions to make the critical contacts with the bases.

Although Zif268 also uses α helices for recognition, the Zif complex is significantly different from previously characterized complexes. Several features of the Zif complex are distinctive.

1) Unlike the helix-turn-helix proteins, this zinc finger complex is based on modular units (individual fingers) which are repeated in a way that allows each finger to contact 3 bp. This appears to be the first clear instance where the periodicity of the protein structure is a simple function of the periodicity of the double-helical DNA. Recognition is based on a modular system that can be used to recognize extended, asymmetric sites.

2) The majority of the contacts are made with a single strand of the DNA.

3) Recognition appears to rely heavily on base contacts in the zinc finger–DNA complex. There are fewer hydrogen bonds with the DNA backbone and they generally appear to play a less critical role in orienting the protein (although the phosphate contact made by the histidine ligand may be quite important). Base contacts appear to play a greater role in orienting the fingers. Since these may change from one complex to the next, it is possible that this makes the zinc finger more “adaptable” than other motifs.

4) Although studies of other protein-DNA complexes have suggested that there is no recognition code, arginine-guanine contacts seem to be very important for the Zif complex.

In spite of these important differences, the Zif complex shows some broad similarities with the prokaryotic helix-turn-helix proteins and with the eukaryotic homeodomain. In some sense, it appears that the finger structure just provides another mechanism for getting an α helix into the major groove, and the critical role of the α-helical regions tends to unite the major families of DNA-binding proteins. As observed in the prokaryotic helix-turn-helix proteins and the eukaryotic homeodomain, other parts of the conserved structural motif contact the DNA backbone and help to precisely position the helix within the major groove. These contacts may have similar functional roles even though Zif268 uses a β strand to contact the DNA backbone while the λ repressor and the Engrailed homeodomain use the first helix of the helix-turn-helix unit to make the corresponding contacts. In each case these neighboring regions may serve as an “outrigger” that keeps the helix from rolling in the major groove. Although the Zif complex has fewer contacts with the DNA backbone, the total number of hydrogen bonds between the protein and the DNA is comparable to that reported with other protein-DNA complexes. As noted in the λ repressor-operator complex, side chain–side chain interactions are important for site-specific recognition (39). Finally, most of the major groove contacts in other complexes also involve purines. This may occur simply because the purines occupy a greater portion of the major groove and offer more hydrogen bonding sites than the pyrimidines. However, it is possible that having a pair of hydrogen bonds (instead of a single hydrogen bond with a pyrimidine) with a rigid planar structure may give enhanced specificity in recognition.

Conclusions. The structure of the Zif complex reveals a remarkably simple and efficient mechanism for recognizing specific sites on double-stranded DNA. The α helix of each zinc finger fits directly into the major groove of B-DNA, and side chains from the NH₂-terminal portion of this helix contact the edges of the base pairs. The main contacts from each finger involve a 3-bp subsite. The Zif268 zinc fingers are tandemly arranged in the major groove, and thus the three finger peptide contacts a 9-bp site. Most of the contacts are along one strand of the DNA (the G-rich strand) and the peptide is “antiparallel” to the DNA strand that has the primary contacts. Arginine-guanine contacts, similar to those discussed by Seeman, Rosenberg, and Rich (45), appear to be responsible for much of the specificity in the Zif complex. There are relatively few contacts with the backbone of the DNA, but one of these contacts is made by a histidine that also serves as a zinc ligand. The second β strand of each finger is near the sugar phosphate backbone of the DNA, and an arginine from this strand also makes a conserved contact with a phosphodiester oxygen.

The structure of the Zif complex should provide a useful guide for modeling complexes with closely related fingers, such as Sp1 (9), and it also provides an attractive framework for attempts to design DNA-binding proteins with novel specificities. Since each finger makes its primary contacts along a 3-bp region, it might be possible to design (or find) fingers that would recognize each of the 64 possible base pair triplets. Then one could “mix and match” these fingers to design proteins with any desired sequence specificity. The main issues involved in modeling homologous fingers and in designing new fingers are similar. We need to find out how fingers are arranged in other complexes and how fingers can be used to recognize other sequences. It is possible that zinc fingers (even fingers of the TFIIIA subclass) may be used in different ways in different complexes. Comparing the homeodomain-DNA complex with the λ repressor–operator complex showed that the helix-turn-helix unit can be used in rather different ways (42), and fingers are so common that evolution will have had a chance to explore all possible ways of using them. It will be necessary to solve the structures of several other zinc finger–DNA complexes to determine whether this overall arrangement is conserved and to see the precise contacts made by fingers that recognize A-T-rich sites.

REFERENCES AND NOTES

5. L. J. Joseph et al., ibid., p. 7164.
23. S.E.R.C. [U.K.] Collaborative Computing Project No. 4 (Daresbury Laboratory,
A New Cofactor in a Prokaryotic Enzyme: Tryptophan Tryptophylquinone as the Redox Prosthetic Group in Methylamine Dehydrogenase

WILLIAM S. MCINTIRE,* DAVID E. WEMMER, ANDREI CHISTOSERDOV, MARY E. LIDSTROM

Methylamine dehydrogenase (MADH), an \( \alpha_{2}\beta_{2} \) enzyme from numerous methylotrophic soil bacteria, contains a novel quinonoid redox prosthetic group that is covalently bound to its small \( \beta \) subunit through two amino acyl residues. A comparison of the amino acid sequence deduced from the gene sequence indicates that this small subunit for the enzyme from *Methyllobacterium extorquens* AM1 with the published amino acid sequence obtained by the Edman degradation method, allowed the identification of the amino acyl constituents of the cofactor as two tryptophyl residues. This information was crucial for interpreting \( ^{1}H \) and \( ^{13}C \) nuclear magnetic resonance, and mass spectral data collected for the semicarbazide- and carboxymethyl-derivatized bis(tripeptidyl)-cofactor of MADH from bacterium W3A1. The cofactor is composed of two cross-linked tryptophyl residues. Although there are many possible isomers, only one is consistent with all the data: The first tryptophyl residue in the peptide sequence exists as an indole-6,7-dione, and is attached at its 4 position to the 2 position of the second, otherwise unmodified, indole side group. Contrary to earlier reports, the cofactor of MADH is not 2,7,9-tricarboxyypyrroloquinoline quinone (PQQ), a derivative thereof, or pro-PQQ. This appears to be the only example of two cross-linked, modified amino acyl residues having a functional role in the active site of an enzyme, in the absence of other cofactors or metal ions.

SINCE THE ELUCIDATION OF THE STRUCTURE OF THE REDOX cofactor of methanol dehydrogenase from *Pseudomonas* TP1 as 2,7,9-tricarboxy-1H-pyrrrolo[2,3-f]quinoline-4,5-dione (Fig. 1, 1) (1), this quinone has been shown to be the noncovalently bound redox cofactor of several other bacterial enzymes (2). This prosthetic group was originally given the common name methoxa-tin, but the more descriptive name pyrrolquinoline quinone (PQQ) has come into favor. More properly, this form should be called 2,7,9-tricarboxy-PQQ to distinguish it from other derivatives that may exist in nature. A number of enzymes have been proposed to contain covalently bound PQQ or a PQQ derivative. In this group are the copper-containing amine oxidases: plasma amine oxidase, kidney, and placental diamine oxidases, l-lysyl oxidase, plant diamine oxidase, fungal amine oxidase, and methylamine oxidase from the soil.
Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å
NP Pavletich and CO Pabo

Science 252 (5007), 809-817. DOI: 10.1126/science.2028256