Zif268 protein–DNA complex refined at 1.6 Å: a model system for understanding zinc finger–DNA interactions

Monicia Elrod-Erickson, Mark A Rould, Lena Nekludova and Carl O Pabo*

Background: Zinc fingers of the Cys_2His_2 class recognize a wide variety of different DNA sequences and are one of the most abundant DNA-binding motifs found in eukaryotes. The previously determined 2.1 Å structure of a complex containing the three zinc fingers from Zif268 has served as a basis for many modeling and design studies, and Zif268 has proved to be a very useful model system for studying how TFIIIA-like zinc fingers recognize DNA.

Results: We have refined the structure of the Zif268 protein–DNA complex at 1.6 Å resolution. Our structure confirms all the basic features of the previous model and allows us to focus on some critical details at the protein–DNA interface. In particular, our refined structure helps explain the roles of several acidic residues located in the recognition helices and shows that the zinc fingers make a number of water-mediated contacts with bases and phosphates. Modeling studies suggest that the distinctive DNA conformation observed in the Zif268–DNA complex is correlated with finger–finger interactions and the length of the linkers between adjacent fingers. Circular dichroism studies indicate that at least some of the features of this distinctive DNA conformation are induced upon complex formation.

Conclusions: Our 1.6 Å structure should provide an excellent framework for analyzing the effects of Zif268 mutations, for modeling related zinc finger–DNA complexes, and for designing and selecting Zif268 variants that will recognize other DNA sites.

Address: Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*Corresponding author.
E-mail: pabo@pabo1.mit.edu

Key words: protein-DNA recognition, X-ray crystallography, Zif268, zinc finger

Received: 5 July 1996
Revisions requested: 2 August 1996
Revisions received: 16 August 1996
Accepted: 21 August 1996

Structure 15 October 1996, 4:1171–1180

 ISSN 0969-2126

Introduction

Zinc fingers of the Cys_2His_2 class, first discovered in TFIIIA [1], constitute one of the most abundant and important DNA-binding motifs found in eukaryotes [2,3]. Naturally occurring zinc-finger proteins recognize a wide variety of different DNA sequences. Zif268 has proved to be a very useful model system for the study of zinc finger–DNA interactions, and previous structural studies of this complex have provided a starting point for many modeling, design, and selection studies (e.g. [4–10]). Here we report the structure of the Zif268 zinc finger–DNA complex refined at 1.6 Å. Our structure confirms all the basic features of the model reported by Pavletich and Pabo [11] and allows us to address several important questions about the details of zinc finger–DNA interactions. This detailed information is relevant to continuing discussions about codes or patterns in zinc finger–DNA recognition, and our structure will provide a useful reference point for the high-resolution study of Zif268 variants that recognize novel DNA sites.

Results and discussion

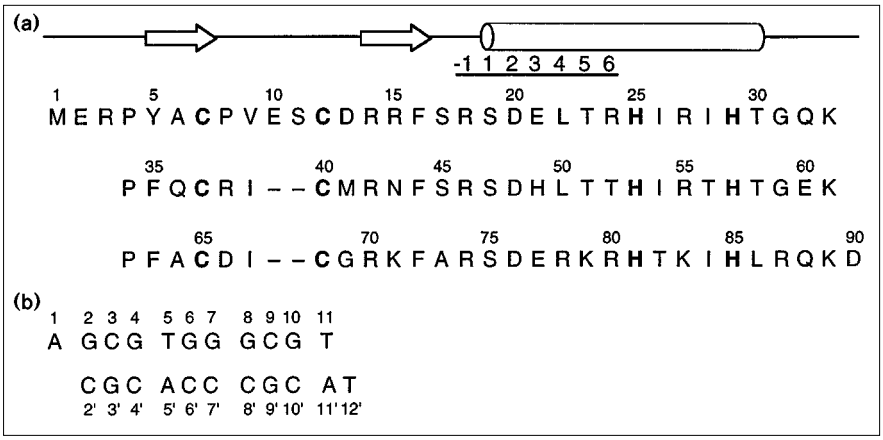
Overall structure of the zinc finger–DNA complex

As expected, the overall structure of the Zif268–DNA complex reported here (Figs 1 and 2) is very similar to the 2.1 Å resolution structure reported by Pavletich and Pabo [11]. (Comparison of these structures shows that the complexes can be superimposed with an rms difference of 0.21 Å for the α carbons, 0.22 Å for the C1′ atoms, and 0.77 Å for all atoms.) Zif268 has three zinc fingers. Each finger contains a short antiparallel β sheet and an α helix that fold to form a compact globular structure, which is held together by a small hydrophobic core and by a zinc ion. The zinc is coordinated by conserved residues, with two cysteines contributed by the β sheet and two histidines by the α helix. The three Zif268 fingers wrap around the DNA, with the α helices fitting into the major groove. Residues from the N-terminal portions of these α helices contact the bases, and each finger makes its primary base contacts within a three-base-pair subsite. For a detailed description of the basic architecture of the complex, the reader is referred to the previous paper [11]. Here we focus on some critical details of the protein–DNA interface that can be seen more clearly and described more confidently in our 1.6 Å structure. We also examine the structure of the DNA-binding site, and describe some modeling and circular dichroism (CD) studies that help us better understand the role that DNA conformation plays in zinc-finger recognition.

Base and phosphate contacts: details of the protein–DNA interface and correlation with biochemical studies

Detailed biochemical studies have raised some interesting questions about the sequence specificity of Zif268. The primary contacts in this complex involve arginine–guanine and histidine–guanine interactions along one strand of the

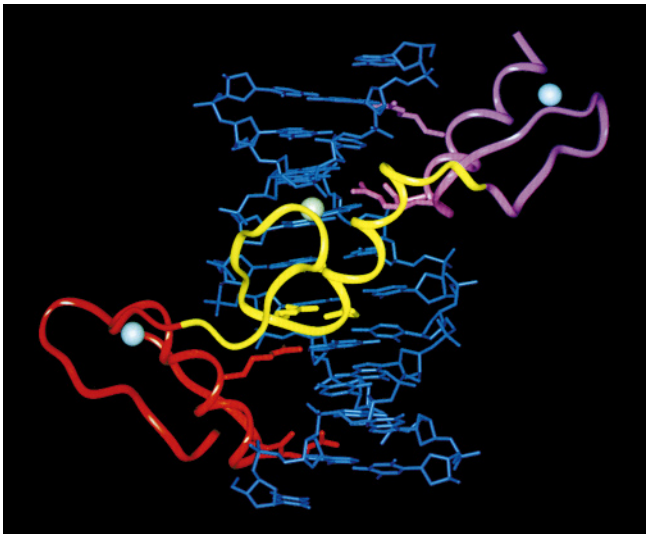
Figure 1



Sequence of the Zif268 zinc finger peptide and of the DNA binding site used in the cocrystallization. (a) Sequence of the zinc finger peptide, aligned by conserved residues and secondary structure elements. Helices are indicated by cylinders, β strands by arrows. Our model includes residues 3 to 87; the terminal residues are disordered in the crystal. The conserved cysteine and histidine residues are highlighted in bold. (b) Sequence of the duplex oligonucleotide used in the cocrystallization. (The figure has been adapted from [11], with permission.)

DNA (these contacts are highlighted in Fig. 3), and the basis for sequence specificity at these positions was clear from the 2.1 Å structure. However, Zif268 does show some sequence specificity at other positions, and the basis for these preferences has not been entirely clear. Does Zif268 make additional, weaker contacts at these positions, or do sequence changes at these other positions have subtle effects on the DNA structure that allow 'indirect readout' via the key contacts reported by Pavletich and Pabo [11]? Our refined structure provides important new information about these issues. In particular, the 1.6 Å structure helps to elucidate the roles of several acidic residues that occur at positions 2 and 3 of the recognition helices, and it reveals a number of water-mediated contacts between the zinc fingers and the DNA. Since residues at equivalent positions in different fingers of Zif268 often play very similar roles in recognition, we have organized our discussion in a way that facilitates the comparison of corresponding residues in the three fingers (residue positions are numbered with respect to the start of the α helix; see top of Fig. 1a).

Figure 2



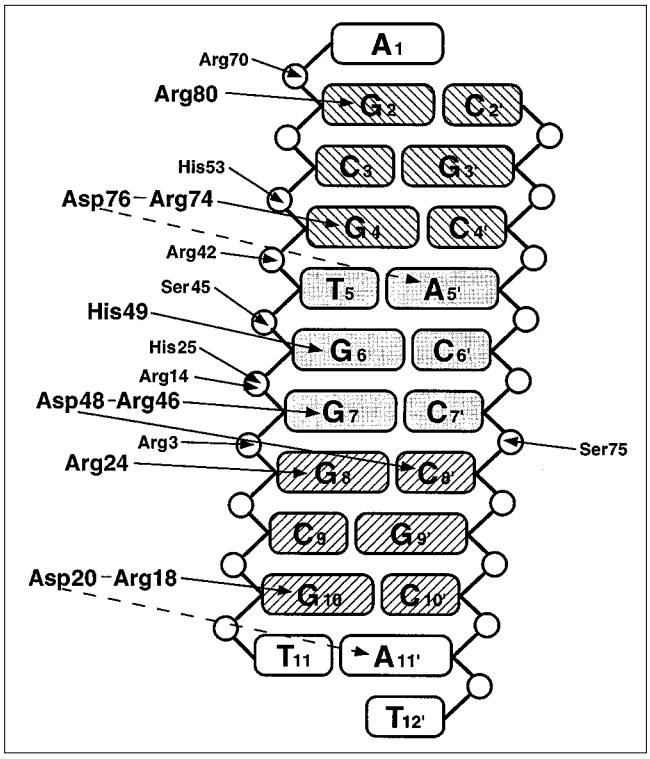
Overview of the Zif268–DNA complex, showing the side chains that make direct base contacts. The peptide is color-coded by finger: finger one is red, finger two is yellow, and finger three is purple. The DNA is shown in dark blue, and the zinc ions in pale blue.

The roles of Arg–1 and Asp2

As emphasized in the previous paper, the aspartic acid at position 2 of each α helix (Asp20, 48, and 76) makes a pair of hydrogen bonds with the arginine immediately preceding the start of the helix (Arg18, 46, and 74), and each of these arginines makes a pair of hydrogen bonds with a guanine. The hydrogen bonds between the aspartic acid at position 2 of each helix and the arginine at position –1 presumably help to orient the arginine side chains and thus increase the specificity of the arginine–guanine interactions. Our structure confirms these critical contacts. However, biochemical studies and our 1.6 Å structure suggest that these residues play some additional roles in sequence-specific recognition.

Our structure shows that these coupled Arg–1/Asp2 residue pairs make water-mediated contacts with the cytosine that is base paired to the guanine contacted by Arg–1 and with the phosphate on the 5′ side of this guanine (Fig. 4a). In all three fingers, we see that the aspartic acids at position 2 of the α helix form water-mediated contacts with the cytosines of these critical G·C base pairs: Asp20 makes a water-mediated contact with the N4 of cytosine 10′, Asp48 with cytosine 7′, and Asp76 with cytosine 4′. In fingers one and three, the arginines at position –1 of the helix make water-mediated contacts to the phosphate backbone. Arg18 makes a water-mediated contact to the phosphate on the 5′ side of guanine 10, and Arg74 has a corresponding interaction with the phosphate on the 5′ side of guanine 4. (In finger two, Arg46 is slightly further

Figure 3



Summary of direct base and phosphate contacts. The DNA bases are shaded to highlight the canonical three-base-pair subsites. Residues that make direct hydrogen bonds to a base or phosphate group are shown in large and small type, respectively. Arrows indicate hydrogen bonds; dotted arrows represent bonds with marginal geometry. All of the direct contacts reported in [11] are observed in our refined structure, but (as shown in later figures) we now have a much more detailed view of the water-mediated contacts.

from the corresponding phosphate, and two ordered water molecules bridge this gap.) We do not have any information on the energetic significance of the water-mediated contacts made by these aspartic-acid and arginine residues. However, it certainly appears that they will help ensure that the coupled Arg–1/Asp2 residues bind very tightly and specifically to the G·C base pair.

In fingers one and three, these aspartic acids also make water-mediated contacts with the base on the 5′ side of the critical guanine. Specifically, our structure reveals that a water bridges the carboxylate of Asp20 and the N4 of cytosine 9 (Fig. 4b). An analogous water-mediated contact occurs between Asp76 and cytosine 3. As discussed later in the paper, these contacts may play some role in specifying the identity of the corresponding base pair.

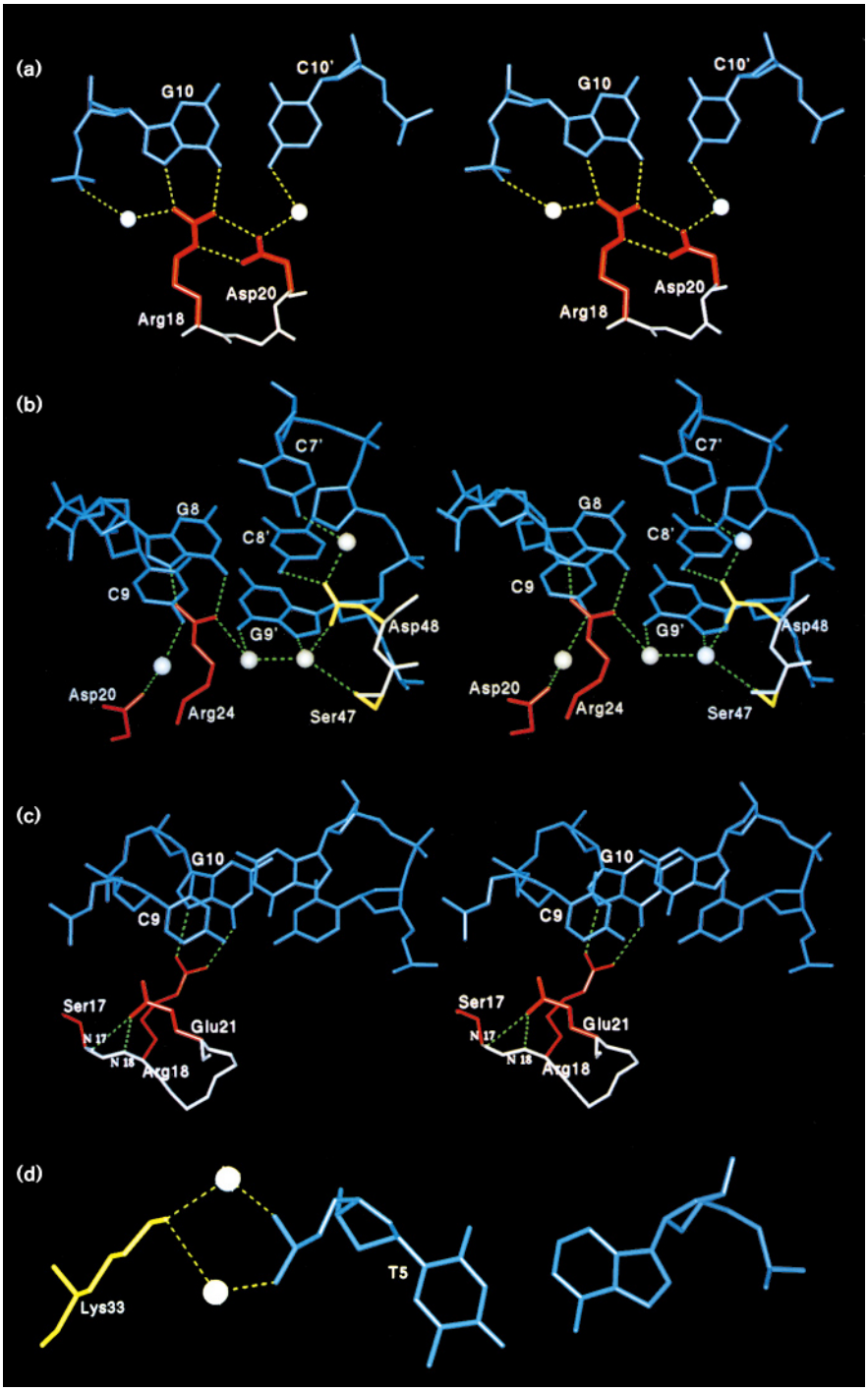
Biochemical studies and our 1.6 Å structure suggest that each of these aspartic acids may also have some weak interactions with a base that is just outside of the canonical three-base-pair subsite and is on the secondary (C-rich) strand of the DNA. These contacts, not discussed in the original report, had been mentioned when discussing comparisons of the Zif268–DNA complex with the GLI–DNA complex [12]. In the following section, we consider the biochemical and structural data for each of these proposed contacts.

Our structure shows that Asp48 clearly contacts cytosine 8′: the carboxylate is 3 Å from the cytosine N4, and the hydrogen-bonding geometry is reasonable (Fig. 4b). The corresponding interactions are less favorable in fingers one and three, where the distance between the aspartic acid and the exocyclic amine is greater (Asp20 is 3.8 Å from the N6 of adenine 11′; Asp76 is 3.5 Å from the N6 of adenine 5′). In these cases, the orientation is also less favorable for hydrogen bonding, but selection and binding experiments suggest that there is a net favorable interaction at these positions. Binding-site selections [13] reveal a strong preference for adenine or cytosine at positions 5′ and 11′, and either of these bases would have a hydrogen-bond donor that could interact with aspartic acid. Competition experiments and K_d determinations also show that Zif268 binds oligos with adenine or cytosine at these positions about 5 to 10-fold more tightly than oligos containing guanine or thymine ([13]; Bryan Wang and COP, unpublished data). The preference for adenine or cytosine at these positions may reflect weak favorable interactions with the corresponding aspartic acid residues. However, aspartic acid could also contribute to the observed sequence preferences by tending to exclude guanine or thymine from these positions (the nearby aspartic acid presumably would have weak unfavorable electrostatic interactions with either of these bases).

The roles of residue 3

His49, which is the third residue in the α helix of finger two, clearly plays an important role in recognition. As in the previous study, our crystallographic model has this histidine donating a hydrogen bond from its Nε to the N7 of guanine 6. However, we note that rotating the ring 180° about the Cβ–Cγ bond would allow the histidine to contact the O6 of the guanine instead. (These arrangements are so similar that they cannot be reliably distinguished even at 1.6 Å resolution.) This histidine also stacks against thymine 5, making van der Waals contacts with the methyl group and with the C5 and C6 atoms [11]. (Contacting the edge of one step in the 'double helical staircase' allows the histidine to rest on top of the preceding step.) As suggested by Swirnoff and Milbrandt [13], this histidine–thymine interaction may be significant for site-specific recognition. Binding-site selections have revealed a preference for thymine over guanine at position 5, despite the fact that either adenine or cytosine should be acceptable at position 5′ and could donate a hydrogen bond to Asp76 [13,14]. Other binding studies confirm that Zif268 binds slightly more tightly to oligos containing a thymine at this position

Figure 4



Details of the protein–DNA interface. Side chains from finger one are shown in red and from finger two in yellow. The peptide backbone is shown in gray and the DNA in blue. Water molecules are depicted as gray spheres. (a) Stereo view showing the network of contacts that Arg18 and Asp20 make with base pair 10. (These are the Arg–1/Asp2 pair from finger one; an analogous set of interactions occurs in fingers two and three.) (b) Stereo view showing the water-mediated interactions that Asp20 and Arg24 make with base pair 9. (These residues occupy positions 2 and 6 in the helix of finger one. As described in the text, a similar set of interactions occurs between finger three and base pair 3, except that there is no water contacting the N7 of guanine 3′.) The contact between Asp48 and cytosine 8′ is also visible in this figure, as is the water-mediated contact between Asp48 and cytosine 7′. (Guanine 7 and Arg46 have been omitted for clarity.) (c) Stereo view showing the conformation of Glu21, with its carboxylate group hydrogen bonding to the backbone amides of Ser17 and Arg18. As described in the text, this side chain makes some van der Waals contacts with the edge of the cytosine. (d) Water-mediated contacts made by Lys33, which is in the linker between fingers one and two, to phosphate 5.

than to oligos containing a guanine [13,15], and it has been shown that substitution of uracil at this position results in reduced binding [13].

Fingers one and three have glutamic acid, instead of histidine, at the third position of the α helix, and there are interesting questions about the role of these acidic residues. Could each of these glutamic acids, in analogy with the histidine that occurs at this position in finger 2, contact the base in the center of its finger's subsite? This idea is appealing since it would be consistent with ideas about a simple recognition code, and since site-selection and binding studies do show a clear preference for cytosine at the center of the GCG triplets recognized by these fingers [8,13,15].

Our 1.6 Å structure clearly defines the conformation of these glutamic acid residues and suggests how they may contribute to specificity.

The carboxylate groups of Glu21 and Glu77 clearly do not make any base contacts: the carboxylate of Glu21 is 5.3 Å away from the N4 of cytosine 9, and that of Glu77 is 5.7 Å away from the N4 of cytosine 3. Instead, these side chains hydrogen bond to the backbone amides of the residues immediately preceding the α helix. Glu21 makes a good hydrogen bond to the backbone -NH of Arg18 (the residue immediately preceding the α helix) and can also hydrogen bond to the backbone amide of Ser17 (Fig. 4c). Corresponding interactions occur in finger three, where Glu77 hydrogen bonds to the backbone -NH groups of Arg74 and Ala73. These interactions may help to stabilize the conformation of the residues immediately preceding the α helix and may thus enhance the specificity of the contacts made by the arginine residues at position –1.

The distinctive and well-ordered conformation observed for these glutamic acid residues (which have their terminal atoms interacting with a neighboring region of the polypeptide backbone) allows each of these side chains to make hydrophobic contacts with the edge of the corresponding cytosine: the Cγ and Cδ atoms of the side chain approach the C5 and C6 positions of the base. (The Cγ and Cδ atoms of Glu21 are, respectively, 4.0 and 4.1 Å from the C5 and 4.8 and 4.5 Å from the C6 of cytosine 9. Analogous contacts, involving Glu77 and cytosine 3, occur in finger 3.) These van der Waals contacts may make some modest contribution to the recognition of cytosine, and the position of the Cγ and Cδ with respect to the base may play a role in discrimination against other bases. (The glutamic acid side chain might interfere with normal hydration of the N7 position of adenine or guanine, and the side chain would have to change conformation to accommodate the methyl group of thymine.)

In addition, as Nardelli *et al.* [4] suggested, an electrostatic phenomenon resulting from the proximity of the glutamic acids to the bases may play some role in discrimination. Our structure indicates that specificity at these positions may also involve water-mediated interactions with these C·G base pairs (Fig. 4b). As described above, Asp20 and Asp76 (at position 2 of the helices) make water-mediated contacts to cytosines 3 and 9. In addition, Arg24 and Arg80 (position 6 of the helices) make water-mediated contacts to the O6 of the corresponding guanines. This is consistent with Swirnoff and Milbrant's suggestion [13] that water-mediated arginine–guanine interactions might contribute to specificity at these positions. There is also a water molecule that interacts with the N7 position of guanine 9′. This water is stabilized by Ser47 and Asp48 from finger two, and it also interacts with the water that bridges Arg24 and the O6 of this guanine. (There is no water contacting the N7 of guanine 3′ in our structure.)

The roles of residue 6

Fingers one and three have an arginine at position 6 of the α helix. As was readily apparent in the 2.1 Å structure, each of these arginines makes a pair of hydrogen bonds to a guanine (Arg24 to guanine 8, and Arg80 to guanine 2). As mentioned above, our 1.6 Å structure reveals that these arginines also make water-mediated contacts with a neighboring base on the opposite strand: Arg24 makes a water-mediated contact with the O6 of guanine 9′, and Arg80 with the corresponding position of guanine 3′. Arg80 also makes a water-mediated contact with the O5′ of the nucleotide at position 1.

Thr52, which is at position 6 in the helix of finger two, does not make any direct contacts with the DNA, but our structure shows that it does make a water-mediated contact with phosphate 4 (Arg74 is another ligand of this water molecule).

The roles of other residues in the α helices

As discussed in the previous paper, the Zif268 complex includes a number of side chain–phosphate contacts (summarized in Fig. 3). Our 1.6 Å structure allows us to see additional, water-mediated phosphate contacts. In two cases (positions 1 and –2 of the helices), we find that where one finger makes a direct side chain–phosphate contact, the other fingers make related, water-mediated contacts.

The pattern is quite striking for the serines that are at position 1 of the α helices (residues 19, 47, and 75). Ser75 (finger 3) hydrogen bonds to phosphate 7′, while Ser47 (finger 2) makes a water-mediated contact with phosphate 9′. (As mentioned earlier, Ser47 also makes a water-mediated contact to the N7 of guanine 9′.) Likewise, Ser19 (finger 1) makes a water-mediated contact to the O5′ of position 12′ (which lacks a phosphate since it is at the 5′ terminus of our synthetic oligonucleotide).

There also are some similarities in the roles of the residues at position –2. Ser45 (finger 2) hydrogen bonds to phosphate 6, while Ser17 (finger 1) makes a water-mediated contact to the 5′ phosphate of base 9. (Finger three has an alanine at this position and cannot make an analogous contact.)

Our structure also reveals two water-mediated contacts from residues that occur later in the α helices. Arg78 (residue 4 in the helix of finger 3) makes a water-mediated contact with phosphate 7′, and one conformation of Thr23 (residue 5 in the helix of finger one) contacts the O5′ of base 12′. (Ser19 is another ligand of this water.)

The role of the lysines in the linkers between fingers

The linker sequence TGEKPF/Y occurs in a large number of zinc finger proteins [16]. The 2.1 Å structure suggested

Table 1

Local helical parameters for the DNA site.*

Base pair		Displacement (Å)	Helical twist (°)	Rise (Å)	Tilt (°)	Roll (°)
2	G·C	–2.60	23.6	3.69	0.33	4.85
3	C·G	–1.91	40.5	2.91	–4.14	6.62
4	G·C	–2.20	27.3	3.40	0.74	5.08
5	T·A	–1.08	36.2	3.23	–4.49	3.45
6	G·C	–1.59	31.0	2.97	–1.66	3.48
7	G·C	–2.01	35.5	3.69	–1.45	9.23
8	G·C	–1.74	28.1	2.88	1.78	–0.96
9	C·G	–0.66	36.1	3.38	–2.61	3.21
10	G·C	–1.63	30.8	3.30	0.74	1.36
11	T·A	–0.70				
Mean		–1.61	32.1	3.27	-1.20	4.04

* Helical parameters were calculated for the duplex portion of the DNA-binding site with NEWHELIX93 [28]. Note that the values for tilt and roll reported in [11] were calculated using older definitions and thus cannot be directly compared to the values reported here. (Using NEWHELIX93 with the coordinates from [11] gives values similar to those shown above.)

roles for most of these conserved residues [11], but it did not explain why glutamic acid and lysine tend to be conserved. After finding that mutation of the corresponding lysine in a peptide derived from TFIIIA reduces its affinity for DNA about sevenfold, Choo and Klug proposed that this lysine might make a phosphate contact [17]. Our 1.6 Å structure provides detailed new information about this region: we find that Lys33, which is located in the linker between fingers one and two, makes a pair of water-mediated contacts to the 5′ phosphate of base 5 (Fig. 4d). Lys61 (in the linker between fingers 2 and 3) makes a similar water-mediated contact with the 5′ phosphate of base 2. These contacts help explain why lysine tends to be conserved in the linker sequence.

Structure of the Zif268-binding site

The DNA in the Zif268 complex is a variant form of B-DNA. The Zif268 site has 11.2 base pairs per turn and also has an unusually deep major groove, with the base pairs displaced about 1.6 Å from their positions in canonical B-DNA. This conformation has been described as $B_{enlarged\ groove}$-DNA, and related structures have been found in the tramtrack [18] zinc finger–DNA complex, the GLI zinc finger–DNA complex, and several other protein–DNA complexes [19]. We find that our coordinates for the DNA site are very similar to the coordinates in the 2.1 Å structure [11], and analysis of the DNA parameters (Table 1) shows that these are also quite similar for the two models. However (as discussed below) our refined structure, modeling studies, and a circular dichroism study have yielded interesting new information about the Zif268–DNA structure.

Table 2

Superimposing the Zif268 DNA site on itself in various registers reveals a subtle three-base-pair periodicity in the structure.*

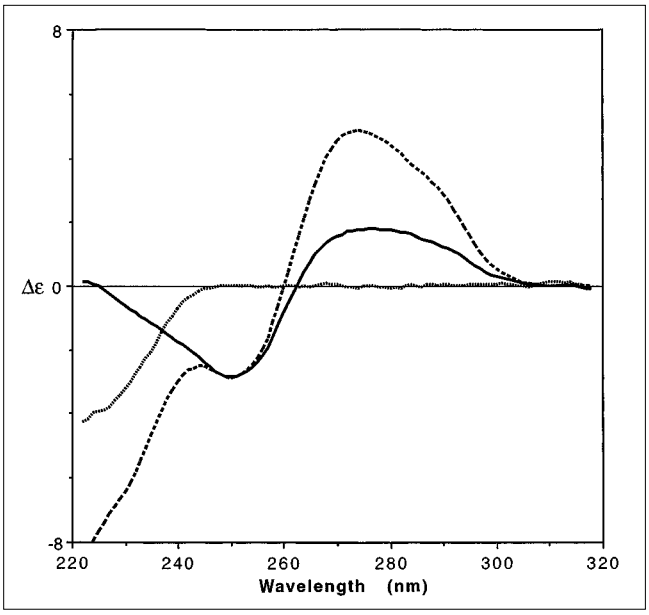
Base pairs superimposed	Offset (bp)	Rms deviation (P and C1′ atoms) (Å)
2–10 and 3–11	1	1.163
2–9 and 4–11	2	1.123
2–8 and 5–11	3	0.963
2–7 and 6–11	4	1.099
2–6 and 7–11	5	1.297
2–5 and 8–11	6	0.686
2–4 and 9–11	7	1.161
2–3 and 10–11	8	1.091

*The closest matches occur when the site is shifted by three or six base pairs.

The most dramatic change during refinement of the Zif268 complex involved the overhanging adenine and thymine bases at the ends of the DNA duplex (Fig. 1b). These bases, which had been added to facilitate crystallization [11], form critical crystal-packing contacts. In the 2.1 Å structure, the adenine and thymine had been modeled as a Watson-Crick base pair that would link adjacent duplex segments to form a pseudo-continuous double helix through the crystal. Surprisingly, our refinement at 1.6 Å indicated that the overhanging adenine and thymine actually form a Hoogsteen base pair. This revised base pairing arrangement has interesting implications for understanding the Zif268 crystal packing contacts and the interactions of neighboring complexes in the crystal. However, as these terminal bases are not part of the Zif268-binding site, this change does not affect any of the previous conclusions about the zinc finger–DNA interactions.

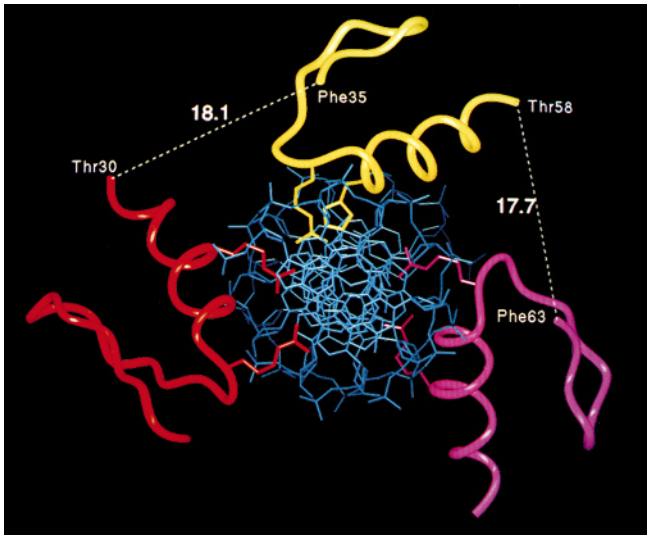
Although it is not obvious from visual inspection of the coordinates or of the helical parameters, there appears to be some subtle three base-pair periodicity in the structure of the DNA. Superimposing the Zif268 site on itself in various registers (by matching corresponding phosphate and C1′ atoms) suggests there is a subtle three-base-pair

Figure 5



Circular dichroism spectra of the Zif268 DNA-binding site (solid line), peptide (short dashes), and complex (long dashes), plotted as the molar extinction coefficient per nucleotide (Δε, in $M^{-1}cm^{-1}$).

Figure 6



Model showing the individual Zif268 fingers docked against ideal B-DNA with ten bp per turn. Each finger has been docked in a way that preserves the local DNA contacts. A dashed line indicates the distance each of the linkers would have to span (measured from α carbon to α carbon of the indicated residues); the omitted residues could span at most 17.5 Å in extended conformation. Finger one is colored red; finger two, yellow; finger three, purple; and the DNA, blue.

periodicity in the DNA structure (Table 2). The significance of this feature is not yet clear, but it is intriguing because it matches the periodicity of the fingers and indicates that subtle structural variations in one subsite tend to be repeated in neighboring subsites.

In trying to understand the significance of the distinctive DNA conformation seen in the Zif268 complex, we would like to know whether structural features observed in the complex represent intrinsic sequence-dependent aspects characteristic of the DNA site or whether these distinctive structural features are induced upon Zif268 binding. Circular dichroism studies (summarized below) suggest that the Zif268 DNA changes conformation during complex formation. As shown in Figure 5, we find that there is a striking difference in the CD spectrum of the Zif268-binding site in the presence and absence of the three-finger peptide. The CD spectrum of the free DNA is similar to that observed for canonical B-DNA, but the height of the maximum observed near 275 nm increases dramatically upon complex formation. (We assume this represents a change in DNA conformation since the peptide has no significant signal in 245–320 nm range.) The observed change is consistent with a decrease in helical twist upon complex formation and/or an increased displacement of the base pairs from the helical axis [20–22]. Both of these features are characteristic of the

Table 3

Linker lengths and inter-finger distances for isolated Zif268 fingers docked against different DNA conformations.

DNA	Parameters for DNA model		Separation of fingers after docking on DNA model	
	bp/turn	displacement (Å)	linker lengths† (Å)	interfinger distances‡ (Å)
Model 1*	10	0	18.1,17.7	8.0,7.4
Model 2	10.5	0	15.9,17.5	6.9,7.5
Model 3	10.5	–1.6	14.8,16.7	4.1,5.1
Model 4	11.2	–1.6	13.4,15.4	3.4,4.2
Zif268§	11.2 (avg)	–1.6 (avg)	14.5,14.4	3.5,2.9

*As described in the text, zinc fingers were docked against idealized DNA sites with the specified displacement and number of bp/turn. †Distance the linker would have to span from fingers 1 to 2 and fingers 2 to 3 (respectively), measured from the threonine Cα to the phenylalanine Cα. ‡Separation of fingers 1 and 2 and of fingers 2 and 3 (respectively), measured from the arginine to the backbone carbonyl. §Values measured for the Zif268 complex are shown for comparison; values for the bp/turn and displacement are averages.

$B_{enlarged\ groove}$-DNA observed in the Zif268 complex (Table 1). Although our CD data do not allow us to determine the precise nature of the conformational change that occurs, our spectra clearly indicate that some features of the distinctive Zif268 DNA conformation are induced by peptide binding. (Our data about structural changes that occur upon protein binding are consistent with a recent report by Shi and Berg [23], who used a plasmid unwinding assay and showed that zinc finger binding causes a slight decrease in the helical twist of the DNA.)

Modeling studies to determine whether the Zif268 zinc fingers could bind to B-DNA

As discussed above, the binding sites in the Zif268, tramtrack, and GLI zinc finger-DNA complexes all have a related $B_{enlarged\ groove}$ conformation [19]. However, it is not immediately obvious why zinc finger peptides cannot recognize canonical B-DNA, so we have undertaken modeling studies to address this question.

These modeling studies began by docking individual Zif268 fingers against various DNA structures. Modeling revealed that single fingers could be docked against B-DNA and still make a relatively normal set of contacts with the appropriate subsites (all of the base contacts and most of the phosphate contacts were preserved). However, when isolated fingers were docked against B-DNA, it was immediately apparent (Fig. 6) that the distance between neighboring fingers was too large to be spanned by the linker and also was too large to allow the normal hydrogen bond between adjacent fingers. (Fingers 1 and 2 of Zif268 are connected by a hydrogen bond between Arg27 and the backbone carbonyl of residue 45; fingers 2 and 3 are connected by an analogous hydrogen bond between Arg55 and the backbone carbonyl of residue 73 [11].) Using a similar modeling strategy but gradually altering the B-DNA structure showed that reducing the helical twist or increasing the groove depth (i.e., making the DNA more like that observed in the Zif268 complex) reduced the distance between fingers (Table 3). The overall implication seems quite clear: the canonical linker length and the observed finger–finger contacts would not allow binding to standard B-DNA.

Biological implications

Zinc fingers of the Cys_2His_2 class constitute one of the most abundant and versatile DNA-binding motifs found in eukaryotes [2,3], and Zif268 has provided a key model system for studying how this family of fingers interacts with DNA. Naturally occurring zinc finger proteins recognize a wide variety of different DNA sequences. Structure-based design and phage display methods have produced fingers capable of recognizing other, novel DNA sites (e.g. [5–10]). Many of these studies were based on the previously reported 2.1 Å structure of the Zif268 protein–DNA complex [11], and many used Zif268 as a starting point for mutation or randomization.

We have refined the structure of the Zif268–DNA complex at 1.6 Å resolution. Our structure confirms all of the main features reported at 2.1 Å [11] and provides important new information about recognition. It reveals auxiliary contacts involving the arginines that make critical guanine contacts, helps explain the role of the acidic residues at positions 2 and 3 of the recognition helices, and reveals water-mediated phosphate contacts that are made by the conserved lysines in the linkers between fingers. The complex networks of interactions that we see highlight the difficulties inherent in trying to develop a simple 'code' that might explain zinc finger–DNA recognition.

Other studies reported in this paper help us understand the role that the distinctive Zif268 DNA conformation plays in recognition. Circular dichroism studies show that the DNA conformation changes as the complex forms, and modeling studies help us rationalize the basis for these changes. In particular, modeling indicates that the fingers would be too far apart if docked against canonical B-DNA and illustrates how the $B_{enlarged\ groove}$-DNA conformation allows the canonical linker sequence to span the gap between neighboring fingers. Our 1.6 Å structure should provide an excellent framework for continued analysis of zinc finger–DNA interactions.

Materials and methods

Crystallization and data collection

The complex we have analyzed contains a peptide corresponding to the three zinc fingers of Zif268 (folded with Zn^{2+}) and a duplex oligonucleotide containing a consensus binding site (Fig. 1). Procedures for purification of the protein and DNA and for crystallization of the complex are described in Pavletich and Pabo [11]. As in the previous study, the complex crystallized in space group $C222_1$, with unit cell dimensions a = 45.4 Å, b = 56.2 Å, and c = 130.8 Å. The current cocrystals diffract beyond 1.6 Å. Data were collected at room temperature from three crystals, using a Rigaku RU-200 generator equipped with mirrors (Molecular Structure Corporation) and an R-Axis IIC image plate system, and were processed with DENZO and SCALEPACK [24]. An R_{merge} of 6.2 % was obtained (149 720 observations of 27 503 reflections); statistics are summarized in Table 4.

Refinement

The 2.1 Å model of Pavletich and Pabo [11], with water molecules deleted, provided the starting point for our refinement. Positional refinement with X-PLOR [25] was initially performed at 2.1 Å with our new data; as refinement continued, data were added in 0.1 Å shells to extend the resolution to 1.6 Å. This process, followed by restrained individual B-factor refinement, produced a model with R_{work} = 29.5 % and R_{free} = 34.4 % for data from 6–1.6 Å.

The $2F_o–F_c$ map calculated from this model was very clear (for example, it indicated unambiguously that adenine 1 adopts a *syn* conformation rather than the *anti* conformation that had been modeled at 2.1 Å). Manual rebuilding using TOM/FRODO [26], further positional refinement, restrained B-factor refinement [25], and local scaling with MAXSCALE (MAR, unpublished) were performed. As refinement continued,

Table 4

Data collection and refinement statistics.

Data collection	
Measured reflections	149 720
Unique reflections	27 503
Completeness to 1.6 Å (%)	94.8
In highest resolution shell (%)	94.9
R_{merge} (%)	6.2
Refinement	
Resolution limits (Å)	6.0–1.6
R factor (%)	19.5
R_{free} (%)	24.2
Reflections with $F > 2\sigma$	19 207
Nonhydrogen atoms of complex	1182
Water molecules	148
Rms ΔB between bonded atoms (Å^2)	1.7
Rms bond lengths (Å)	
Protein	0.007
DNA	0.009
Rms bond angles (°)	
Protein	1.3
DNA	3.0

148 water molecules were added, and alternate conformations for five side chains (Pro4, Arg15, Thr23, Gln36, and Leu50) were incorporated in the model. The last several cycles of positional refinement and B-factor refinement used a data set from which we had temporarily omitted the 3% of the working reflections with the largest $||F_{obs}|-|F_{calc}||$ values. This improved the model (i.e., the new model had a lower R factor for the entire data set), and the complete data set was used in calculating the final R factors.

During refinement, the conformations of all side chains and bases were checked in refined omit maps, and all protein–DNA contacts were also checked with simulated annealing omit maps. Our final model has $R_{work} = 19.5\%$ and $R_{free} = 24.2\%$ for data from 6–1.6 Å with $F > 2\sigma$ ($R_{work} = 20.3\%$ and $R_{free} = 25.0\%$ for all data from 6–1.6 Å). The rms deviations in bond lengths and angles are 0.007 Å and 1.3° for the protein (using the dictionary of Engh and Huber [27]) and 0.009 Å and 3.0° for the DNA (using the standard XPLOR dictionary PARAM11X. DNA [25]). The rms ΔB for bonded atoms is 1.7 Å^2. (Statistics are summarized in Table 4.)

Circular dichroism

CD spectra were recorded from 320 to 220 nm (in 1 nm intervals) at 25° C, using an Aviv 60DS spectropolarimeter with a 1.5 nm bandwidth and a 1 s averaging time. Spectra were taken at a DNA concentration of 0.05 mg ml^{-1} and a peptide concentration of 0.075 mg ml^{-1} in 25 mM bis-tris propane, pH 7, in a 1 cm path-length cuvette. Each spectrum shown is the average of two baseline corrected scans, smoothed in 5 nm windows.

Accession numbers

Coordinates are being deposited with the Brookhaven Data Bank. While they are being processed, interested scientists may obtain a set of coordinates by sending an e-mail message to pabo@pabo1.mit.edu.

Acknowledgements

This research was supported by the Howard Hughes Medical Institute, and equipment used had been purchased with support from the Pew Charitable Trusts. We thank Amy Dunn for assistance in preparation of this manuscript, Robert Sauer for use of the spectropolarimeter, Nikola Pavletich for support and advice in the early stages of this work, and Matthew J Elrod-Erickson, Harvey Greisman, and Edward Rebar for critical reading of the manuscript. We also thank the reviewers for helpful comments on this manuscript. ME–E is a Howard Hughes Medical Institute predoctoral fellow; MAR, LN, and COP are in the Howard Hughes Medical Institute.

References

1. Miller, J., McLachlan, A.D. & Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* 4, 1609–1614.
2. Jacobs, G.H. (1992). Determination of the base recognition positions of zinc fingers from sequence analysis. *EMBO J.* 11, 4507–4517.
3. Pellegrino, G.R. & Berg, J.M. (1991). Identification and characterization of 'zinc-finger' domains by the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* 88, 671–675.
4. Nardelli, J., Gibson, T. & Charnay, P. (1992). Zinc finger–DNA recognition: analysis of base specificity by site-directed mutagenesis. *Nucleic. Acids Res.* 20, 4137–4144.
5. Desjarlais, J.R. & Berg, J.M. (1993). Use of a zinc finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA* 90, 2256–2260.
6. Pomerantz, J.L., Sharp, P.A. & Pabo, C.O. (1995). Structure-based design of transcription factors. *Science* 267, 93–96.
7. Rebar, E.J. & Pabo, C.O. (1994). Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* 263, 671–673.
8. Jamieson, A.C., Kim, S.-H. & Wells, J.A. (1994). *In vitro* selection of zinc fingers with altered DNA-binding specificity. *Biochemistry* 33, 5689–5695.
9. Choo, Y. & Klug, A. (1994). Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc. Natl. Acad. Sci. USA* 91, 11163–11167.
10. Wu, H., Yang, W.-P. & Barbas III, C.F. (1995). Building zinc fingers by selection: toward a therapeutic application. *Proc. Natl. Acad. Sci. USA* 92, 344–348.
11. Pavletich, N.P. & Pabo, C.O. (1991). Zinc finger–DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252, 809–817.
12. Pavletich, N.P. & Pabo, C.O. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* 261, 1701–1707.
13. Swirnoff, A.H. & Milbrandt, J. (1995). DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Mol. Cell. Biol.* 15, 2275–2287.
14. Choo, Y. & Klug, A. (1994). Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc. Natl. Acad. Sci. USA* 91, 11168–11172.
15. Christy, B. & Nathans, D. (1989). DNA binding site of the growth factor-inducible protein Zif268. *Proc. Natl. Acad. Sci. USA* 86, 8737–8741.
16. Schuh, R., *et al.*, & Jàckle, H. (1986). A conserved family of nuclear proteins containing structural elements of the finger protein encoded by Krüppel, a *Drosophila* segmentation gene. *Cell* 47, 1025–1032.
17. Choo, Y. & Klug, A. (1993). A role in DNA binding for the linker sequences of the first three fingers of TFIIIA. *Nuc. Acids Res.* 21, 3341–3346.
18. Fairall, L., Schwabe, J.W.R., Chapman, L., Finch, J.T. & Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* 366, 483–487.
19. Nekludova, L. & Pabo, C.O. (1994). Distinctive DNA conformation with enlarged major groove is found in Zn-finger–DNA and other protein–DNA complexes. *Proc. Natl. Acad. Sci. USA* 91, 6948–6952.
20. Baase, W.A. & Johnson, W.C., Jr. (1979). Circular dichroism and DNA secondary structure. *Nuc. Acids Res.* 6, 797–814.
21. Chan, A., Kilkuskie, R. & Hanlon, S. (1979). Correlations between the duplex winding angle and the circular dichroism spectrum of calf thymus DNA. *Biochemistry* 18, 84–91.
22. Johnson, B.B., Dahl, K.S., Tinoco Jr., I., Ivanov, V.I. & Zhurkin, V.B. (1981). Correlations between deoxyribonucleic acid structural parameters and calculated circular dichroism spectra. *Biochemistry* 20, 73–78.
23. Shi, Y. & Berg, J.M. (1996). DNA unwinding induced by zinc finger protein binding. *Biochemistry* 35, 3845–3848.
24. Otwinowski, Z. (1993). *DENZO: an oscillation data processing program for macromolecular crystallography*. Yale University, New Haven, CT, USA.

25. Brünger, A.T. (1992). *X-PLOR Manual version 3.1*. Yale University Press, New Haven, CT, USA.
26. Israel, M. & Chirino, A.J. (1994). *TOM/FRODO version 3.0*. University of Alberta, Alberta, Canada, and California Institute of Technology, CA, USA.
27. Engh, R.A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* 47, 392–400.
28. Dickerson, R. (1993). *NEWHELIX93*. University of California, Los Angeles, CA, USA.