Structural and Biochemical Studies of Zinc Finger-DNA Complexes

by

Monicia Elrod-Erickson
B.S., Molecular Biology
Vanderbilt University, 1991

Submitted to the Department of Biology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy
at the
Massachusetts Institute of Technology
June 1998

©1998 Monicia Elrod-Erickson
All rights reserved

The author hereby grants to MIT permission to reproduce and to
distribute publicly paper and electronic copies of this thesis document
in whole or in part.

Signature of Author.................   ....
                                                          Department of Biology

Certified by ....
                        Carl O. Pabo, Professor of Biophysics and Structural Biology
                                                                    Thesis Supervisor

Accepted by ....
                                                Frank Solomon, Professor of Biology
                                          Chairman, Biology Graduate Committee

Structural and Biochemical Studies of Zinc Finger-DNA Complexes

by

Monicia Elrod-Erickson

Submitted to the Department of Biology on May 14, 1998 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

ABSTRACT

Zinc fingers of the Cys_2His_2 class comprise one of the largest families of eukaryotic DNA-binding motifs and recognize a diverse set of DNA sequences. These proteins have a relatively simple modular structure, and key base contacts are typically made by a few residues from each finger. These features make the zinc finger motif an attractive system for designing novel DNA-binding proteins and for exploring fundamental principles of protein-DNA recognition. This thesis describes the structural and biochemical characterization of the Zif268 zinc finger-DNA complex and of related complexes involving variants of Zif268.

Chapter 1 provides a review of zinc finger-DNA interactions, including the biological roles of zinc finger proteins, use of the zinc finger motif in the design of novel DNA-binding proteins, and conclusions drawn from analysis of those zinc finger-DNA complexes whose structures have been solved.

Chapter 2 describes the structure of the Zif268 zinc finger-DNA complex refined at 1.6 Å resolution, along with related modeling and circular dichroism experiments. This chapter has been published as "Zif268 protein-DNA complex refined at 1.6 Å: A model system for understanding zinc finger-DNA interactions" (Elrod-Erickson, M., Rould, M.A., Nekludova, L. & Pabo, C.O. 1996. *Structure* **4**, 1171-1180).

2

Chapter 3 describes the high resolution structures of seven complexes involving three variants of Zif268: three complexes that include each of the variants bound to its corresponding DNA target site, plus four other complexes containing various combinations of these peptides with alternative binding sites. This chapter has been published as "High resolution structures of variant Zif268-DNA complexes: Implications for understanding zinc finger-DNA recognition" (Elrod-Erickson*, M., Benson*, T.E. & Pabo, C.O. 1998. *Structure* **6**, 451-464. *shared authorship).

Chapter 4 describes how crystal packing forces affect the structure of the Zif268 DNA binding site when it is crystallized by itself, as indicated by circular dichroism experiments and by analysis of the DNA crystal structure (Rould, M.A.*, Chambers, K.A.*, Elrod-Erickson, M.* & Pabo, C.O. *shared authorship).

Chapter 5 describes the results of binding studies involving mutants of Zif268 with changes in the base-contacting residues of finger one.

Chapter 6 summarizes the conclusions drawn from the experiments and structures described in this thesis and discusses prospects for future research regarding DNA recognition by zinc fingers.

Thesis Supervisor: Carl O. Pabo
Title: Professor of Biophysics and Structural Biology

To my parents, Lewis and Anna Elrod,
and my husband, Matthew J. Elrod-Erickson.

Acknowledgements

Many people have helped make my years in graduate school rewarding, and I would like to take this opportunity to offer my appreciation to them.

I am particularly grateful to my advisor, Carl Pabo. Over the years I have benefited greatly from his knowledge, advice, and sincere commitment to the people in his lab.

I also wish to thank the members of my thesis and defense committees - Tania Baker, Bob Sauer, Jamie Williamson, and Steve Harrison - for their guidance, advice, and interest.

The members of the Pabo lab have my gratitude for helping to create a stimulating and pleasant intellectual environment. I owe a special debt to those people with whom I have worked on various projects. Mark Rould offered his time and knowledge of crystallography unstintingly. He, Lena Nekludova, Tim Benson, Kristen Chambers, Michael Robinson, and Sandra Fay-Richard have been excellent collaborators, and I have enjoyed working with them all.

I have also been fortunate to have so many fine scientists and good friends among my classmates. In particular, I thank the Company of the Singing Sword (Matt Elrod-Erickson, Elizabeth Hong, Mark Borowsky, Adam Grancell, Sandy Gilbert, Ann-Marie White, Christopher Cilley, and Grant Wheeler) for many evenings of magic and merriment and, most of all, for their friendship. Arma manumque Ilshara canet.

I am very grateful to Gerald Stubbs for introducing me to protein biochemistry and structural biology and for his encouragement.

Finally, I thank my family - my parents, Lewis and Anna Elrod, and my husband, Matthew Elrod-Erickson - for their love, encouragement, and unfailing support.

Table of Contents

List of Tables

Chapter 5

List of Figures

10

Chapter One

Introduction

Introduction

DNA-binding proteins fill many essential roles in such crucial cellular processes as DNA replication, recombination, and repair, and a large number are devoted to ensuring that an appropriate subset of each cell's many genes are transcribed at the proper times and at adequate levels. The importance of such proteins is reflected by their prevalence (estimates indicate that gene regulatory proteins, for example, comprise up to 10% of the eukaryotic genome [1, 2]) and by the observation that disruption of protein-DNA interactions can lead to a number of different cancers, developmental abnormalities, or other disorders [e.g., 3-9]. Understanding how such proteins recognize their specific DNA target sites is a necessary step toward a complete understanding of these proteins and their roles in diverse cellular processes.

Many DNA-binding proteins are modular: the DNA-binding activity of a given protein is often separable from its other activities, such as activation or repression of transcription [e.g., 10, 11]. Such observations permit protein-DNA recognition to be studied through a reductionist strategy that has proven to be extremely useful. In this approach, the DNA-binding domain of the protein of interest is delimited and the structure of this domain in complex with an appropriate DNA binding site is determined by x-ray crystallographic or nuclear magnetic resonance spectroscopic (NMR) techniques. To date, the structures of more than sixty protein-DNA complexes have been solved. Comparing these structures has led to several generalizations about protein-DNA recognition, reviewed in [12] and briefly summarized below.

Structural and sequence comparisons have revealed that most DNA-binding domains fall into one of a relatively small number of classes,

each of whose members uses a particular structural motif to interact with DNA in a broadly similar manner [12, 13]; a few examples are the helix-turn-helix motif, the homeodomain, the zinc finger, and the helix-loop-helix motif. Many DNA-binding domains, such as the homeodomain and the zinc finger, position an α helix in the major groove. Others use different strategies, for example, positioning β strands (e.g., Arc repressor [14]) or surface loops (e.g., NF-κB [15, 16]) to contact the DNA.

Recognition frequently depends on contacts between the protein and the edges of the base pairs in the major groove of the DNA. Hydrogen bonds between protein side chains and the bases appear to be especially important, but hydrogen bonds made by the protein backbone and van der Waals interactions are also observed. Although the edges of the base pairs present a distinctive pattern of hydrogen bond donors and acceptors only in the major groove [17] and most proteins make the majority of their contacts there, in some complexes additional contacts to the bases are seen in the minor groove [e.g., 18-20] and, in others, recognition relies entirely upon contacts in the minor groove [21, 22]. Recognition also involves a number of contacts between the protein and the phosphate backbone, usually hydrogen bonds or salt bridges between side chains and the phosphodiester oxygens. Such interactions clearly contribute to binding affinity, and they may also make some contribution to specificity (either indirectly, by helping to fix the orientation of the DNA-binding domain with respect to the DNA such that inappropriate contacts with the bases are not made, for example, or directly, to whatever extent the base sequence affects the conformation of the phosphate backbone [23]).

Multiple DNA-binding domains are frequently used to achieve

site-specific recognition. The association of such domains is often noncovalent, as in the formation of homodimers [24], heterodimers [25], or heterotrimers [26]. Alternatively, a single polypeptide can contain multiple DNA-binding domains. These domains can be all of the same type (multiple zinc fingers, for example) or they can be from different families (for instance, a POU-specific domain and a homeodomain [27]).

Zinc fingers - Biological roles

Zinc fingers of the Cys_2His_2 class comprise one of the largest families of DNA-binding motifs found in eukaryotes. The motif was discovered by analysis of the sequence of TFIIIA [28, 29] and is characterized by the consensus sequence $CX_{2-4}CX_3FX_5LX_2HX_{3-5}H$ [30]. Other zinc-binding motifs, some of which also bind DNA (such as the steroid hormone receptor DNA-binding domain), are characterized by different consensus sequences and structures but have also been referred to as zinc fingers (reviewed in [30, 31]). However, in this thesis use of the term will be restricted to the Cys_2His_2-type finger. Since their initial discovery, zinc fingers have been found in a multitude of eukaryotic proteins. The human genome, for example, is estimated to contain from 300 to 700 zinc finger proteins, corresponding to about 1% of all human proteins [32-34]. A similarly large number of zinc finger-containing proteins has been found in other organisms [32], particularly other vertebrates (relatively few zinc finger proteins have been identified in plants to date [35, 36], and there are fewer than two dozen in yeast [37]).

Although many of these proteins remain to be characterized, a large number have been shown to be involved in gene regulation: for instance,

the Drosophila segmentation genes Hunchback and Krüppel [38, 39], the general transcription factor Sp1 [40], and the growth factor-responsive proteins Zif268/Krox24 and Krox20 [41, 42]. Although the most common function of zinc fingers appears to be DNA binding, a few fingers have been demonstrated to have roles other than, or in addition to, DNA binding. For example, TFIIIA, in which the zinc finger motif was first recognized, binds 5S RNA as well as the internal control region of the 5S ribosomal RNA genes (distinct subsets of TFIIIA's nine fingers appear to be primarily responsible for RNA and DNA binding) [43 and references therein]. Another multi-finger protein, p43, also binds 5S RNA but has no detectable DNA binding activity [44]. Zinc fingers have also been shown to mediate protein-protein interactions. For instance, the Ikaros protein contains two C-terminal zinc fingers that do not appear to bind nucleic acid but that instead are capable of mediating both homodimerization [45] and heterodimerization with a second zinc finger protein [46]. In other proteins, the same fingers that bind DNA can also interact with another protein. The three fingers of Sp1, for example, both bind DNA and mediate interactions with the initiator-binding protein YY1 (another zinc finger protein) [47]. Zinc fingers are not limited to contacting other fingers: interactions have been demonstrated between the Sp1 zinc fingers and the DNA-binding domain of GATA-1 [48] and between the YY1 fingers and the basic-leucine zipper region of ATFa2 [49].

The zinc finger fold

Zinc fingers tend to occur in tandem, with proteins typically containing stretches of three to seven fingers (although as few as one [50] and as many as 37 fingers [51] have been reported in a single protein).

16

Each finger consists of about 30 amino acids that fold into a short two-stranded antiparallel β sheet and an α helix (Figure 1). The sheet and the helix are held together by a small hydrophobic core and by a zinc ion that is coordinated by the two conserved cysteines (from the sheet region) and the two conserved histidines (from the α helix). This basic fold has been observed in numerous crystallographic and NMR structures of both single and tandem fingers, alone and in complex with DNA [52-67].

Coordination of zinc is essential for the stability and therefore DNA-binding activity of zinc fingers; in the absence of zinc, fingers are unstructured (reviewed in [68]). The highly conserved phenylalanine and leucine residues, with the first of the two conserved histidines and several other moderately conserved hydrophobic residues [e.g., 59], form a hydrophobic core that helps stabilize the folded finger. The basic finger can tolerate some variation in these hydrophobic core residues. For example, substitution of an isoleucine or leucine for the highly conserved phenylalanine results in the same basic fold [65, 69], as does changing the phenylalanine to tyrosine and/or altering its position from $CX_{2-4}CX_3FX_5LX_2HX_{3-5}H$ to $CX_{2-4}CXF/YX_7LX_2HX_{3-5}H$ [61, 62, 67, 69, 70]. A few fingers are known to require additional elements in order to fold properly. The amino-terminal fingers of the tramtrack and SWI5 proteins each require several residues N-terminal of the consensus finger sequence for their structural stability; the additional residues form a third strand in the β sheet of these fingers [56, 57, 62].

Linker sequences

The sequence connecting adjacent tandem fingers, called the linker

or H-C link, is also quite highly conserved among zinc finger proteins as TGEKPF/YX. Several NMR studies have shown that, unlike the finger domains, the linkers do not adopt a single well defined conformation in the absence of DNA [54, 71-73]. In a peptide containing the first three fingers of TFIIIA, for example, in the absence of DNA the fingers display rigid-body motions and the linkers display considerable flexibility [72, 73]. Once the peptide binds to DNA, however, the linkers lose flexibility and assume a defined conformation [66]. Similarly well defined conformations of canonical linkers have been observed in crystal structures of zinc finger-DNA complexes (see for example [59, 61, 67]).

Roles have been suggested for many of the residues in the consensus linker sequence [59, 60]. As observed in the Zif268 zinc finger-DNA complex, the threonine is the last residue in the α helix of the finger preceding the linker. Its methyl group is involved in hydrophobic core interactions and its hydroxyl group hydrogen bonds to the backbone amide of the third residue in the linker; this interaction presumably helps stabilize the conformation of the linker seen in the complex. The second residue in the linker, glycine, makes the last hydrogen bond in the α helix and may help terminate the helix [74]. This glycine also assumes a $\Phi\Psi$ combination not accessible to larger residues. The role of the conserved glutamate is not clear, since it has not been observed to interact with other side chains or with the DNA. The conserved lysine makes water-mediated contacts with the phosphate backbone. The proline and tyrosine or phenylalanine are in van der Waals contact with each other, and their interaction could help restrict the conformation of the polypeptide at the beginning of the next finger. The linker residues fill similar roles in other complexes containing canonical linker sequences

[e.g., 66], and the importance of these residues is borne out by the observation that mutating single amino acids in the linker can reduce binding affinity by up to 24-fold [75, 76].

The extent to which the linkers help to determine the orientation that adjacent fingers assume with respect to DNA is still unclear. Some zinc finger proteins contain linkers that differ substantially from the consensus in length and amino acid sequence. TFIIIA, for example, contains both consensus and non-consensus linkers. The TFIIIA fingers connected by canonical linkers assume orientations with respect to DNA that are roughly similar to orientations observed in other consensus linker-containing zinc finger-DNA complexes [66, 67], while the fingers connected by non-consensus linkers are positioned very differently with respect to the DNA [67]. This observation seems to support the idea that the sequence of the linker can be a major factor in determining how fingers bind. The two fingers from the tramtrack protein, however, are also connected by a non-canonical linker, but in this complex the fingers nevertheless assume orientations with respect to the DNA like those assumed by consensus linker-containing zinc finger peptides such as Zif268 [62].

Structure of the Zif268-DNA complex

The crystal structure of the three fingers from Zif268 (also known as Krox24, NGFI-A, and Egr1) complexed with a consensus binding site provided the first view of a zinc finger-DNA complex [59]. As described in Chapter 2, this structure was later refined to 1.6 Å resolution, revealing further details of the protein-DNA interface [60]. A brief description of this complex, focusing on direct side chain-base interactions, is given

below. (The fingers also make direct contacts with the phosphate backbone, as well as a number of water-mediated contacts with both bases and phosphates, that are covered in Chapter 2.)

The three Zif268 fingers wrap around the DNA, with their α helices fitting into the major groove (Figure 2). Residues from the amino-terminal portion of the helices contact the bases, with each finger making its primary contacts to a three base pair subsite on one strand of the DNA (Figure 3). The Zif268 fingers use residues from four positions in the α helix to make base contacts: the residue immediately preceding the helix (residue -1) and the second, third, and six residues in the helix. Fingers one and three have the same residues at these positions and recognize identical GCG subsites. Finger two has a different set of residues at these four positions and recognizes a TGG subsite.

The residue at position -1 in all three fingers is an arginine, and in all three fingers this arginine makes a pair of hydrogen bonds with the O6 and N7 of the guanine at the 3' end of the finger's subsite (Figure 4a). All three fingers also have an aspartic acid at position 2 of the helix, and in all the fingers this aspartate makes a pair of hydrogen bonds with the arginine from position -1 (Figure 4a). This interaction presumably helps orient the arginine and thus increases the specificity of the arginine-guanine interaction. The aspartic acid from position 2 may also interact with a base on the opposite strand of the DNA that is just outside the finger's subsite (discussed further in Chapter 2; note that, because of this interaction, the fingers could be described as having overlapping four base pair subsites rather than contiguous three base pair subsites).

The residue at position 3 of the helix contacts the middle base in the finger's subsite in all the fingers, but these contacts are of different

types. In finger two, the residue at position 3 is a histidine. This histidine hydrogen bonds to guanine 6 (Figure 4b; although this histidine is depicted in this figure as donating a hydrogen bond to the N7 of the guanine, an arrangement in which the histidine is rotated 180° such that it contacts the O6 of the guanine is also consistent with the crystallographic data). This histidine also makes van der Waals contacts with the methyl group and the C5 and C6 of thymine 5. Fingers one and three have a glutamic acid at position 3 of the helix. As described in Chapter 2, each of these glutamic acids makes hydrophobic contacts with the edge of the base in the middle of the finger's subsite (Glu21 with cytosine 9 and Glu77 with cytosine 3). It is not yet clear how great a role the van der Waals contacts play in determining specificity at these base pairs.

The residue at position 6 of the helix in fingers one and three is an arginine. In both fingers, this arginine makes a pair of hydrogen bonds with the guanine at the 5' end of the finger's subsite (Figure 4c). The residue at the corresponding position of finger two is a threonine that does not make any direct base contacts.

Positions -1, 2, 3, and 6 in recognition

Several observations support the idea that positions -1, 2, 3, and 6 of the α helix play key roles in recognition, both in the Zif268 complex and in other zinc finger-DNA complexes. For example, swapping certain of these residues from an Sp1 finger into a Krox20 finger changes the finger's specificity to that of the Sp1 donor finger [77]. Mutations at several of these same four positions in the fingers from the Wilms' tumor suppressor protein decrease binding affinity and sequence selectivity [3,

78]. Selection experiments in which the residues at positions -1, 2, 3, and 6 were randomized in one of Zif268's three fingers produced variants of Zif268 that recognize novel binding sites [79]. Similarly, alteration of the residues at these positions can produce peptides for which a binding site different from that of the parent peptide can be selected [80, 81]. Finally, residues from these four positions of the α helix have been seen to interact with the DNA bases in numerous other zinc finger-DNA complex structures [59-67].

Structures of other zinc finger-DNA complexes

The structures of several other zinc finger-DNA complexes have been solved since the first Zif268 complex structure. These complexes include the five fingers of the human glioblastoma protein (GLI) [61], the two fingers of the Drosophila tramtrack protein [62], the four fingers of the initiator-binding protein YY1 [63], a three finger designed peptide [64], the single finger from the Drosophila GAGA factor (plus a distinctive amino-terminal extension) [65], and three [66] or six [67] fingers of Xenopus TFIIIA, with their DNA binding sites. In these structures, residues from positions -1, 2, 3, and 6 of the α helix make base contacts in many instances (summarized in Figure 5). Often these contacts are to the same base within the finger's subsite as was observed for the Zif268 complex (i.e., residue -1 to the 3' base of the subsite, residue 2 to a base on the opposite strand just outside the subsite, residue 3 to the middle base of the subsite, and residue 6 to the 5' base of the subsite). In these other structures, however, residues from these four positions are occasionally observed to contact bases in alternative positions of the subsite or to

make simultaneous contacts to more than one base (see Figure 6). In addition, in these other complexes, residues from additional positions in the α helix occasionally make base contacts (Figure 6).

Although most of the fingers observed in these complexes assume roughly the same orientation with respect to DNA as the Zif268 fingers, there are differences in their precise orientations (see Figure 7a for an example). In addition, a few fingers assume orientations with respect to DNA that are quite different from those observed the Zif268 complex. The first finger of GLI, for example, is positioned such that it makes extensive contacts with finger two but no contacts with the DNA (Figure 7b). Fingers four and six of TFIIIA provide another example of fingers with distinctive orientations, and, in this case, the unusual placement of the fingers clearly plays an important role in recognition. Fingers one through three and finger five of TFIIIA bind in the major groove, in much the same manner as the Zif268 fingers. Fingers four and six, however, span the minor groove. These two fingers make a few contacts with the phosphate backbone, but their main role in the TFIIIA-DNA complex seems to be maintaining the proper spacing between fingers one through three and finger five that allows these other fingers to recognize separate regions of the promoter [67].

DNA conformation and recognition by zinc fingers

Structural studies of protein-DNA complexes suggest that sequence-dependent aspects of DNA conformation (such as local changes in the helical parameters) may play a significant role in protein-DNA recognition (reviewed in [82]). Analysis of the structures of the Zif268,

GLI, and tramtrack zinc finger-DNA complexes revealed that the DNA in these complexes is a distinctive form of B-DNA. This conformation, called $B_{enlarged\ groove}$-DNA, is characterized by an unusually wide and deep major groove that results from a slight unwinding of the DNA and an increased displacement of the base pairs from the helical axis [83]. Similar conformations of the DNA have been found in the other multifinger peptide-DNA complexes whose structures have been solved [63, 64, 66, 67, 84].

As described in Chapter 4, the Zif268 binding site was crystallized and its structure was solved in an attempt to determine whether the distinctive $B_{enlarged\ groove}$-DNA conformation observed in the Zif268 complex structure is an inherent characteristic of the DNA that the protein could use to help recognize its binding site or whether this conformation is induced upon protein binding. The DNA in these crystals, however, is nearly canonical A-form. Circular dichroism (CD) studies reveal that the DNA is B-form in solution, indicating that crystal packing forces are responsible for the A-form structure seen in the DNA crystals.

A partial answer as to whether the $B_{enlarged\ groove}$ conformation is an intrinsic characteristic of the binding site or is induced by Zif268 binding is provided by further CD studies (described in Chapter 2), which indicate that the Zif268 DNA site changes conformation as the protein binds. Thus, at least some features of the distinctive $B_{enlarged\ groove}$ conformation observed in the complex are induced upon complex formation. This observation is consistent with results from a plasmid unwinding assay indicating that binding of another zinc finger protein, Sp1, causes a decrease in the helical twist of the DNA [85]. Whether any features of the $B_{enlarged\ groove}$ conformation are characteristic of the free DNA is unclear,

since attempts to obtain B-form crystals of the binding site have been unsuccessful (M.E., unpublished data).

Use of zinc fingers in protein design

Designed proteins with novel DNA-binding specificities have many potential applications in gene therapy and in biological research, and the process of designing such proteins provides an opportunity to test our understanding of the physical-chemical principles underlying protein-DNA recognition. Zinc fingers have proven to be useful starting points for the design of such novel proteins: fingers have been modified to recognize new DNA sites and have been used as components of novel chimeric proteins. In this latter approach, one or more zinc fingers are joined to another type of domain to produce a new, chimeric protein. This approach has been used to produce novel transcription factors capable of activating or repressing transcription in vivo [86, 87], endonucleases with novel specificities [88, 89], and zinc finger-GAL4 dimerization domain fusions capable of homo- and heterodimerizing [90]. Related experiments in which fingers from multiple peptides are joined have produced a six finger peptide with extremely high affinity for DNA [91].

As mentioned above, the specificity of zinc finger peptides can be altered such that the fingers recognize a new DNA binding site. Several features of zinc fingers make them a particularly attractive framework for the generation of DNA-binding proteins with novel specificities. First, recognition is achieved through contacts made by side chains from a relatively limited number of positions within the finger, which can be varied without disrupting the finger's folded structure. Second, the fingers are relatively modular. The contacts each finger makes to its

subsite can be altered without affecting the binding of neighboring fingers, as long as their orientations and any overlapping contacts they make are compatible. Third, zinc finger proteins can recognize asymmetric sites (unlike homodimers). Lastly, although zinc fingers tend to contact purine-rich sites, the sites are otherwise of widely varying base sequence (unlike, for example, basic helix-loop-helix proteins, most of which recognize hexameric CAC/GG/CTG sites [92]).

Attempts at rationally designing zinc finger peptides to recognize desired sites have met with mixed success. Choosing amino acids to occupy positions -1, 2, 3, and 6 on the basis of their frequency of occurrence in a database of zinc fingers or on the basis of a zinc finger-DNA "recognition code" sometimes produces a protein that recognizes the desired site with reasonable affinity and specificity (see [93, 94] for examples), but sometimes produces a protein that binds with low affinity and/or specificity (e.g., [93, 95]). Choosing a set of residues to occupy the four key positions of the α helix, selecting the optimal DNA subsite for that finger, and then combining several such fingers has proven to be a slightly more reliable way of obtaining zinc finger peptides with novel specificities [64, 80, 81]. However, the most successful method for generating new zinc finger peptides has been the affinity-based selection of zinc fingers capable of binding the desired site from among a library of randomized fingers expressed on the surface of phage. This method has produced numerous three finger peptides in which one or more fingers have altered specificities, and many of these selected peptides bind with high affinity and specificity to their target sites [79, 96-100].

Structures of Zif268 variants

The structures of seven complexes containing zinc finger peptides produced from one such selection are described in detail in Chapter 3; a few important features of these complexes are summarized below. The peptides in these complexes resulted from selections in which positions -1, 2, 3, and 6 were randomized in finger one of Zif268 (fingers two and three were unchanged). Variants were then selected for binding to DNA target sites in which the subsite for finger one had been altered [79]. The selection is schematized in Figure 8, and apparent dissociation constants for the resulting variant peptides are listed in Table 1.

As mentioned previously, in structures of complexes containing zinc fingers from several different proteins, the fingers have been seen to assume a range of orientations with respect to DNA. It is not clear how much of this variation is due to the differing base contacts made in these complexes by the residues at positions -1, 2, 3, and 6 and how much is due to other sequence differences between the peptides (resulting in a different set of potential phosphate contacts and thereby influencing the finger's orientation, for example). However, since only the residues at the four key base-contacting positions have been changed, any differences observed when comparing the structures of the complexes containing the variant peptides with the wild type complex can be attributed to the changes in the base-contacting residues. As described in Chapter 3, although the orientations of fingers two and three with respect to the DNA are the same in the variant complexes as in the wild type complex, the orientation of finger one with respect to the DNA varies somewhat (see Figure 7 of Chapter 3 for examples). This change in orientation presumably occurs to facilitate the new side chain-base interactions formed in the

variant complexes. Thus, a change only in the potential base contacts can result in an altered orientation of a finger with respect to DNA. Similarly, changes in the conformation of finger one are observed in two of the variant Zif268 peptides; these changes must also be attributable to the changes in base-contacting residues. The adaptations in finger conformation and orientation with respect to the DNA observed in the variant complexes illustrate how new contacts with the DNA (introduced during protein design or evolution, for example) can be accommodated within a generally conserved structural framework.

Comparing the structures of these variant complexes also reveals that side chain-side chain interactions play a role in recognition. One of the variants, RADR, has an arginine at position -1 of the α helix, an alanine at position 2, and an aspartic acid at position 3. In the complex between this peptide and its targeted site, the arginine at position -1 contacts the phosphate backbone and interacts with the aspartate at position 3. In another complex involving the RADR peptide, the arginine at position -1 is seen to assume two conformations: one interacts with a base, the other with the phosphate backbone and with the aspartate at position 3. As mentioned previously, in the wild type Zif268 complex, the arginine at position -1 of the α helix makes a pair of hydrogen bonds with a guanine as well as a pair of bonds with the aspartic acid at position 2. As described in Chapter 3, when Zif268 is bound to a site containing an adenine instead of a guanine, the arginine assumes two conformations - one like that seen in the wild type complex, and a second conformation analogous to that seen in the RADR complex. Thus, it appears that the most favorable conformation for the arginine at position -1 is that which maximizes contacts with nearby acidic residues as well as with the DNA;

the acidic residues at positions 2 and 3 therefore, by orienting the arginine, help determine what sort of contacts the arginine makes with DNA.

Is there a zinc finger-DNA recognition code?

Since the first few structures of protein-DNA complexes were solved it has been clear that there is no simple, generally applicable code that describes recognition by DNA-binding proteins [12]. The initial structure of the Zif268 zinc finger-DNA complex [59] suggested that there might, however, be a simple code underlying DNA recognition by zinc finger domains, since the direct side chain-base contacts in the Zif268 complex (Figure 3) appear to follow a simple pattern (for example, arginine at position 6 of the α helix used to recognize a guanine at the 5' end of the subsite). Ensuing zinc finger-DNA complex structures [60-67] revealed that this view was oversimplified and that no simple general zinc finger-DNA recognition code exists.

Although, as shown in Figure 6, there is a strong correlation between the position of an amino acid residue in the α helix and the location in the finger's subsite of the base that amino acid contacts, this pattern does not equate to a code. There is clearly no one-to-one correspondence between the identity of the base at a given position of the subsite and the identity of the amino acid used to recognize that base (Figure 6). Nor is there a strict correspondence between the position of a base within the subsite and the position within the α helix of the amino acid residue which contacts that base (i.e., the residues at positions -1, 2, 3, and 6 do not always contact the same position within the subsite; these

residues sometimes contact more than one position within the subsite simultaneously; and occasionally residues from other positions in the helix make base contacts). Varying the orientation of a finger with respect to the DNA changes the optimal set of base-contacting residues [101, 102]; such context-dependent effects can not be incorporated into a simple code. Neither can the observed networks of water-mediated hydrogen bonds between the fingers and the DNA that may contribute to affinity and specificity [e.g., 60], side chain-side chain interactions, or potential sequence-dependent aspects of DNA conformation. All of the above problems apply both to any code for use in designing fingers with desired specificities and to any code meant to predict the optimal binding site for a given zinc finger protein. An additional problem arises in this latter case, which is that no means exists of predicting which fingers contact the DNA bases in the first place and which fingers have other functions (serving as spacer elements as seen in the TFIIIA complex [67], or mediating protein-protein interactions, for instance). Although several groups have proposed that a highly restricted code describing recognition by only those zinc finger proteins most closely related to Zif268 can be developed (e.g., [102, 103]), the high resolution structures of complexes involving variants of Zif268 described in Chapter 3 demonstrate that the above problems can still arise even in this limiting case. Proposed recognition codes can not replace sequential optimization in finding the best finger to use at a given binding site or in finding the optimal binding site for a given zinc finger protein.

Biochemical studies of the Zif268 complex

Although the Zif268 complex been very well characterized

structurally, little biochemical data is available. Binding site selection data indicate that TGCGTG/AGGCGG/T is the most favored binding site for Zif268 [104], and competition experiments using mutated oligonucleotides confirm this [104, 105]. Methylation interference data [105] are consistent with the contacts seen in the structure of the complex [59, 60].

How much individual contacts observed in the structure contribute to binding affinity and specificity, however, has been unclear. A few mutants with changes in the base-contacting residues have been constructed, but the binding studies that have been reported involving these mutants used peptides expressed on the surface of phage rather than purified peptides (multiple copies of each peptide can be present on any given phage particle, and this multivalency may affect binding) [106, 107]. We have produced peptides containing mutations in the base-contacting residues of finger one of Zif268. The results of preliminary binding studies with these mutant peptides are described in Chapter 5, followed in Chapter 6 by a brief summary of what has been learned about zinc finger-DNA interactions and some possible future directions for further study of zinc finger-DNA recognition.

References

1.   Adams, M.D., Kerlavage, A.R., Fields, C. & Venter, J.C. (1993). 3400 new expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet.* **4**, 256-267.

2.   Johnson, A.D. (1995). The price of repression. *Cell* **81**, 655-658.

3.   Pelletier, J., *et al.* (1991). germline mutations in the Wilms' tumor suppressor gene are associated with abnormal urogenital development in Denys-Drash syndrome. *Cell* **67**, 437-447.

4.   Little, M.H., Prosser, J., Condie, A., Smith, P.J., Heyningen, V.V. & Hastie, N.D. (1992). Zinc finger point mutations within the *WT1* gene in Wilms tumor patients. *Proc. Natl. Acad. Sci. USA* **89**, 4791-4795.

5.   Cho, Y., Gorina, S., Jeffrey, P.D. & Pavletich, N.P. (1994). Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations. *Science* **265**, 346-355.

6.   Xu, W., Rould, M.A., Jun, S., Desplan, C. & Pabo, C.O. (1995). Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for pax developmental mutations. *Cell* **80**, 639-650.

7.   Georgopoulos, K., *et al.* (1994). The Ikaros gene is required for the development of all lymphoid lineages. *Cell* **79**, 143-156.

8.   Winandy, S., Wu, P. & Georgopoulos, K. (1995). A dominant mutation in the *Ikaros* gene leads to rapid development of leukemia and lymphoma. *Cell* **83**, 289-299.

9.   Sakai, T., Ohtani, N., McGee, T.L., Robbins, P.D. & Dryja, T.P. (1991). Oncogenic germ-line mutations in Sp1 and ATF sites in the human retinoblastoma gene. *Nature* **353**, 83-86.

10.   Brent, R. & Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* **43**, 729-736.

11.   Hollenberg, S.M. & Evans, R.M. (1988). Multiple and cooperative *trans*-activation domains of the human glucocorticoid receptor. *Cell* **55**, 899-906.

12.   Pabo, C.O. & Sauer, R.T. (1992). Transcription factors: Structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053-1095.

13.   Harrison, S.C. (1991). A structural taxonomy of DNA-binding domains. *Nature* **353**, 715-719.

14.   Raumann, B.E., Rould, M.A., Pabo, C.O. & Sauer, R.T. (1994). DNA recognition by β-sheets in the Arc repressor-operator crystal structure. *Nature* **367**, 754-757.

15.   Ghosh, G., Duyne, G.V., Ghosh, S. & Sigler, P.B. (1995). Structure of NF-κB p50 homodimer bound to a κB site. *Nature* **373**, 303-310.

16.   Muller, C.W., Rey, F.A., Sodeoka, M., Verdine, G.L. & Harrison, S.C. (1995). Structure of the NF-κB p50 homodimer bound to DNA. *Nature* **373**, 311-317.

17.   Seeman, N.C., Rosenberg, J.M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* **73**, 804-808.

18.   Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. & Pabo, C.O. (1990). Crystal structure of the engrailed homeodomain-DNA complex at 2.8 Å resolution: A framework for understanding homeodomain-DNA interactions. *Cell* **63**, 579-590.

19.   Schumacher, M.A., Choi, K.Y., Zalkin, H. & Brennan, R.G. (1994). Crystal structure of LacI member, PurR, bound to DNA: Minor groove binding by α helices. *Science* **266**, 763-770.

20.   Muller, C.W. & Herrmann, B.G. (1997). Crystallographic structure of the T domain-DNA complex of the *Brachury* transcription factor. *Nature* **389**, 884-888.

21. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-520.

22. Kim, J.L., Nikolov, D.B. & Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520-527.

23. Otwinowski, Z., *et al.* (1988). Crystal structure of *trp* repressor/operator complex at atomic resolution. *Nature* **335**, 321-329.

24. Beamer, L.J. & Pabo, C.O. (1992). Refined 1.8 Å crystal structure of the λ repressor-operator complex. *J. Mol. Biol.* **227**, 177-196.

25. Glover, J.N.M. & Harrison, S.C. (1995). Crystal structure of the heterodimeric bZIP transcription factor cFos-cJun bound to DNA. *Nature* **373**, 257-261.

26. McNabb, D.S., Xing, Y. & Guarente, L. (1995). Cloning of yeast *HAP5*: A novel subunit of a heterotrimeric complex required for CCAAT binding. *Genes Dev.* **9**, 47-58.

27. Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. & Pabo, C.O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* **77**, 21-32.

28. 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.

29. Brown, R.S., Sander, C. & Argos, P. (1985). The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS Lett.* **186**, 271-274.

30. Berg, J.M. & Shi, Y. (1996). The galvanization of biology: A growing appreciation for the roles of zinc. *Science* **271**, 1081-1085.

31. Schwabe, J.W.R. & Klug, A. (1994). Zinc mining for protein domains. *Nat. Struc. Biol.* **1**, 345-349.

32.	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.

33.	Bray, P., Lichter, P., Thiesen, H.-J., Ward, D.C. & Dawid, I.B. (1991). Characterization and mapping of human genes encoding zinc finger proteins. *Proc. Natl. Acad. Sci. USA* **88**, 9563-9567.

34.	Hoovers, J.M.N. & al., e. (1992). High-resolution localization of 69 potential human zinc finger protein genes: A number are clustered. *Genomics* **12**, 254-263.

35.	Meissner, R. & Michael, A.J. (1997). Isolation and characterisation of a diverse family of *Arabidopsis* two and three-fingered C_2H_2 zinc finger protein genes and cDNAs. *Plant Mol. Biol.* **33**, 615-624.

36.	Kubo, K., *et al.* (1998). Cys_2/His_2 zinc-finger protein family of petunia: Evolution and general mechanism of target-sequence recognition. *Nuc. Acids Res.* **26**, 608-615.

37.	Saccharomyces genome database. http://genome-ftp.stanford.edu/ Saccharomyces

38.	Tautz, D., *et al.* (1987). Finger protein of novel structure encoded by *hunchback*, a second member of the gap class of *Drosophila* segmentation genes. *Nature* **327**, 383-389.

39.	Rosenberg, U.B., *et al.* (1986). Structural homology of the product of the *Drosophila Kruppel* gene with *Xenopus* transcription factor IIIA. *Nature* **319**, 336-339.

40.	Briggs, M.R., Kadonaga, J.T., Bell, S.P. & Tijan, R. (1986). Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. *Science* **234**, 47-52.

41.	Christy, B.A., Lau, L.F. & Nathans, D. (1988). A gene activated in mouse 3T3 cells by serum growth factors encodes a protein with "zinc finger" sequences. *Proc. Natl. Acad. Sci. USA* **85**, 7857-7861.

42.   Lemaire, P., Revelant, O., Bravo, R. & Charnay, P. (1988). Two mouse genes encoding potential transcription factors with identical DNA-binding domains are activated by growth factors in cultured cells. *Proc. Natl. Acad. Sci. USA* **85**, 4691-4695.

43.   Clemens, K.R., *et al.* (1993). Molecular basis for specific recognition of both RNA and DNA by a zinc finger protein. *Science* **260**, 530-533.

44.   Joho, K.E., Darby, M.K., Crawford, E.T. & Brown, D.D. (1990). A finger protein structurally similar to TFIIIA that binds exclusively to 5S RNA in Xenopus. *Cell* **61**, 293-300.

45.   Sun, L., Liu, A. & Georgopoulos, K. (1996). Zinc finger-mediated protein interactions modulates Ikaros activity, a molecular control of lymphocyte development. *EMBO J.* **15**, 5358-5369.

46.   Morgan, B., *et al.* (1997). Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. *EMBO J.* **16**, 2004-2013.

47.   Lee, J.-S., Galvin, K.M. & Shi, Y. (1993). Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proc. Natl. Acad. Sci. USA* **90**, 6145-6149.

48.   Merika, M. & Orkin, S.H. (1995). Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol. Cell. Biol.* **15**, 2437-2447.

49.   Zhou, Q., Gedrich, R.W. & Engel, D.A. (1995). Transcriptional repression of the c-*fos* gene by YY1 is mediated by a direct interaction withATF/CREB. *J. Virol.* **69**, 4323-4330.

50.   Soeller, W.C., Oh, C.E. & Kornberg, T.B. (1993). Isolation of cDNAs encoding the *Drosophila* GAGA transcription factor. *Mol. Cell. Biol.* **13**, 7961-7970.

51.   Altaba, A.R.i., Perry-O'Keefe, H. & Melton, D.A. (1987). *Xfin*: An embryonic gene encoding a multifingered protein in *Xenopus*. *EMBO J.* **6**, 3065-3070.

52.   Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. & Wright, P.E. (1989). Three-dimensional solution structure of a single zinc-finger DNA-binding domain. *Science* **245**, 635-637.

53.   Omichinski, J.G., Clore, G.M., Appella, E., Sakaguchi, K. & Gronenborn, A.M. (1990). High-resolution three-dimensional structure of a single zinc finger from a human enhancer binding protein in solution. *Biochemistry* **29**, 9324-9334.

54.   Omichinski, J.G., Clore, G.M., Robien, M., Sakaguchi, K., Appella, E. & Gronenborn, A.M. (1992). High-resolution solution structure of the double Cys_2His_2 zinc finger from the human enhancer binding protein MBP-1. *Biochemistry* **31**, 3907-3917.

55.   Klevit, R.E., Herriott, J.R. & Horvath, S.J. (1990). Solution structure of a zinc finger domain of yeast ADR1. *Proteins* **7**, 215-226.

56.   Neuhaus, D., Nakaseko, Y., Schwabe, J.W.R. & Klug, A. (1992). Solution structures of two zinc-finger domains from SWI5 obtained using two-dimensional ^{1}H nuclear magnetic resonance spectra: A zinc-finger structure with a third strand of β-sheet. *J. Mol. Biol.* **228**, 637-651.

57.   Dutnall, R.N., Neuhaus, D. & Rhodes, D. (1996). The solution structure of the first zinc finger domain of SWI5: A novel structural extension to a common fold. *Structure* **4**, 599-611.

58.   Narayan, V.A., Kriwacki, R.W. & Caradonna, J.P. (1997). Structures of zinc finger domains from transcription factor Sp1. *J. Biol. Chem.* **272**, 7801-7809.

59.   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.

60.   Elrod-Erickson, M., Rould, M.A., Nekludova, L. & Pabo, C.O. (1996). Zif268 protein-DNA complex refined at 1.6 Å: A model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171-1180.

61.   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.

62.   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.

63.   Houbaviy, H.B., Usheva, A., Shenk, T. & Burley, S.K. (1996). Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl. Acad. Sci. USA* **93**, 13577-13582.

64.   Kim, C.A. & Berg, J.M. (1996). A 2.2 Å resolution crystal structure of a designed zinc finger protein bound to DNA. *Nat. Struc. Biol.* **3**, 940-945.

65.   Omichinski, J.G., Pedone, P.V., Felsenfeld, G., Gronenborn, A.M. & Clore, G.M. (1997). The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nat. Struc. Biol.* **4**, 122-132.

66.   Wuttke, D.S., Foster, M.P., Case, D.A., Gottesfeld, J.M. & Wright, P.E. (1997). Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: Determinants of affinity and sequence specificity. *J. Mol. Biol.* **273**, 183-206.

67.   Nolte, R.T., Conlin, R.M., Harrison, S.C. & Brown, R.S. (1998). Differing roles for zinc fingers in DNA recognition: Structure of a six-finger transcription factor IIIA complex. *Proc. Natl. Acad. Sci. USA* **95**, 2938-2943.

68.   Berg, J.M. (1990). Zinc finger domains: Hypotheses and current knowledge. *Annu. Rev. of Biophys. Biophys. Chem.* **19**, 405-421.

69.   Mortishire-Smith, R.J., Lee, M.S., Bolinger, L. & Wright, P.E. (1992). Structural determinants of Cys_2His_2 zinc fingers. *FEBS Lett.* **296**, 11-15.

70.   Jasanoff, A., Kochoyan, M., Fraenkel, E., Lee, J.P. & Weiss, M.A. (1992). Aromatic-aromatic interactions in the zinc finger motif: Analysis of the two-dimensional nuclear magnetic resonance structure of a mutant domain. *J. Mol. Biol.* **225**, 1035-1047.

71.   Nakaseko, Y., Neuhaus, D., Klug, A. & Rhodes, D. (1992). Adjacent zinc-finger motifs in multiple zinc-finger peptides from SWI5 form structurally independent, flexibly linked domains. *J. Mol. Biol.* **228**, 619-636.

72.   Bruschweiler, R., Liao, X. & Wright, P.E. (1995). Long-ranged motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* **268**, 886-889.

73.   Foster, M.P., Wuttke, D.S., Radhakrishnan, I., Case, D.A., Gottesfeld, J.M. & Wright, P.E. (1997). Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. *Nat. Struc. Biol.* **4**, 605-608.

74.   Aurora, R., Srinivasan, R. & Rose, G.D. (1994). Rules for α-helix termination by glycine. *Science* **264**, 1126-1130.

75.   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.

76.   Clemens, K.R., Zhang, P., Liao, X., McBryant, S.J., Wright, P.E. & Gottesfeld, J.M. (1994). Relative contributions of the zinc fingers of transcription factor IIIA to the energetics of DNA binding. *J. Mol. Biol.* **244**, 23-35.

77.   Nardelli, J., Gibson, T.J., Vesque, C. & Charnay, P. (1991). Base sequence discrimination by zinc-finger DNA-binding domains. *Nature* **349**, 175-178.

78.   Borel, F., Barilla, K.C., Hamilton, T.B., Iskandar, M. & Romaniuk, P.J. (1996). Effects of Denys-Drash syndrome point mutations on the DNA binding activity of the Wilms' tumor suppressor protein WT1. *Biochemistry* **35**, 12070-12076.

79.   Rebar, E.J. & Pabo, C.O. (1994). Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**, 671-673.

80. Desjarlais, J.R. & Berg, J.M. (1994). Length-encoded multiplex binding site determination: Application to zinc finger proteins. *Proc. Natl. Acad. Sci. USA* **91**, 11099-11103.

81. Kim, C.A. & Berg, J.M. (1995). Serine at position 2 in the DNA recognition helix of a Cys_2-His_2 zinc finger peptide is not, in general, responsible for base recognition. *J. Mol. Biol.* **252**, 1-5.

82. Travers, A.A. (1992). DNA conformation and configuration in protein-DNA complexes. *Current Opinion in Structural Biology* **2**, 71-77.

83. 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.

84. Elrod-Erickson, M., Benson, T.E. & Pabo, C.O. (1998). High resolution structures of variant Zif268-DNA complexes: Implications for understanding zinc finger-DNA recognition. *Structure* **6**, 451-464.

85. Shi, Y. & Berg, J.M. (1996). DNA unwinding induced by zinc finger protein binding. *Biochemistry* **35**, 3845-3848.

86. Pomerantz, J.L., Sharp, P.A. & Pabo, C.O. (1995). Structure-based design of transcription factors. *Science* **267**, 93-96.

87. Kim, J.-S., Kim, J., Cepek, K.L., Sharp, P.A. & Pabo, C.O. (1997). Design of TATA box-binding protein/zinc finger fusion for targeted regulation of gene expression. *Proc. Natl. Acad. Sci. USA* **94**, 3616-3620.

88. Kim, Y.G., Shi, Y., Berg, J.M. & Chandrasegaran, S. (1997). Site-specific cleavage of DNA-RNA hybrids by zinc finger/FokI cleavage domain fusions. *Gene* **203**, 43-49.

89. Nahon, E. & Raveh, D. (1998). Targeting a truncated Ho-endonuclease of yeast to novel DNA sites with foreign fingers. *Nuc. Acids Res.* **26**, 1233-1240.

90. Pomerantz, J.L., Wolfe, S.A. & Pabo, C.O. (1998). Structure-based design of a dimeric zinc finger protein. *Biochemistry* **37**, 965-970.
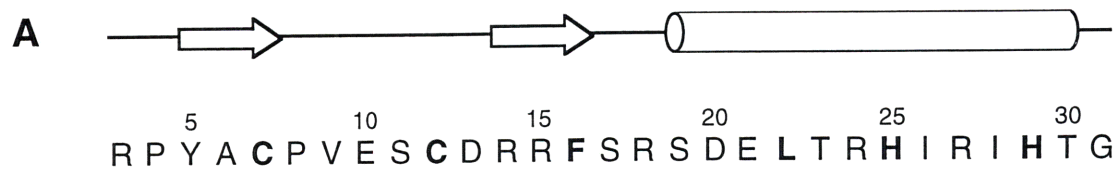
91. Kim, J.-S. & Pabo, C.O. (1998). Getting a handhold on DNA: Design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl. Acad. Sci. USA* **95**, 2812-2817.

92. Ellenberger, T. (1994). Getting a grip on DNA recognition: Structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains. *Current Opinion in Structural Biology* **4**, 12-21.

93. 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.

94. Corbi, N., Perez, M., Maione, R. & Passananti, C. (1997). Synthesis of a new zinc finger peptide; Comparison of its 'code' deduced and 'CASTing' derived binding sites. *FEBS Lett.* **417**, 71-74.

95. Desjarlais, J.R. & Berg, J.M. (1992). Redesigning the DNA-binding specificity of a zinc finger protein: A data base-guided approach. *Proteins* **12**, 101-104.

96. 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.

97. 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.

98. 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.

99. Jamieson, A.C., Wong, H. & Kim, S.-H. (1996). A zinc finger directory for high-affinity DNA recognition. *Proc. Natl. Acad. Sci. USA* **93**, 12834-12839.

100. Greisman, H.A. & Pabo, C.O. (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657-661.

101. Suzuki, M., Gerstein, M. & Yagi, N. (1994). Stereochemical basis of DNA recognition by Zn fingers. *Nuc. Acids Res.* **22**, 3397-3405.

102. Desjarlais, J.R. & Berg, J.M. (1992). Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc. Natl. Acad. Sci. USA* **89**, 7345-7349.

103. Choo, Y. & Klug, A. (1997). Physical basis of a protein-DNA recognition code. *Current Opinion in Structural Biology* **7**, 117-125.

104. Swirnoff, A.H. & Milbrandt, J. (1995). DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Mol. Cell. Biol.* **15**, 2275-2287.

105. Christy, B. & Nathans, D. (1989). DNA binding site of the growth factor-inducible protein Zif268. *Proc. Natl. Acad. Sci. USA* **86**, 8737-8741.

106. Choo, Y. (1998). End effects in DNA recognition by zinc finger arrays. *Nuc. Acids Res.* **26**, 554-557.

107. Isalan, M., Choo, Y. & Klug, A. (1997). Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc. Natl. Acad. Sci. USA* **94**, 5617-5621.

108. Rebar, E.J. (1997). Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Figure 1. (a) Amino acid sequence of the first zinc finger from Zif268. Conserved zinc ligands and hydrophobic residues are highlighted in bold. The secondary structure of the finger is sketched at the top of the panel: β strands are represented by arrows and the α helix by a cylinder.

(b) Backbone trace of the first finger from Zif268, which exhibits the typical zinc finger fold. The two conserved cysteine and two conserved histidine side chains that coordinate the zinc ion (gray sphere) are shown, as are the conserved phenylalanine and leucine residues.

Figure 1
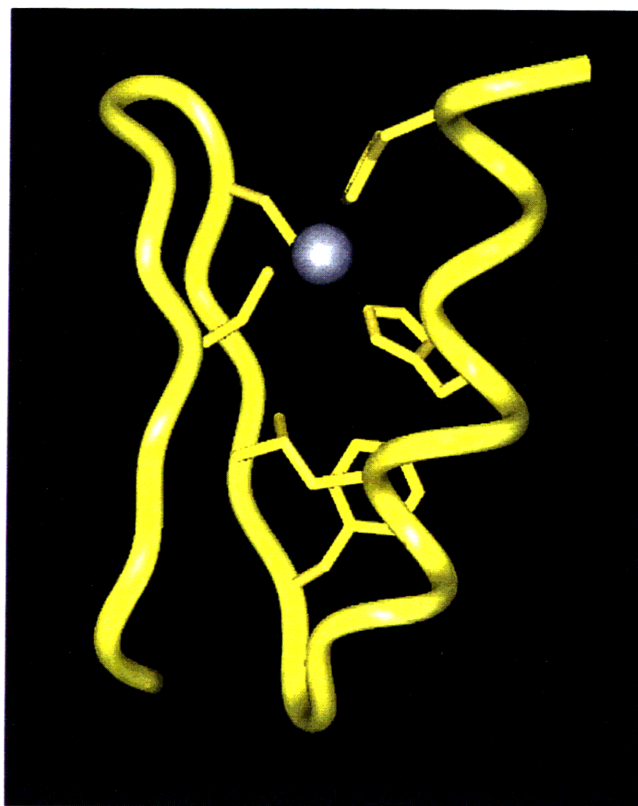
A


```
         5              10             15             20             25             30
R  P  Y  A  C  P  V  E  S  C  D  R  R  F  S  R  S  D  E  L  T  R  H  I  R  I  H  T  G
```

B

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 (residues 3 to 32) is red, finger two (residues 33 to 60) is yellow, and finger three (residues 61 to 87) is pink. The DNA is shown in blue, and the zinc ions in gray.
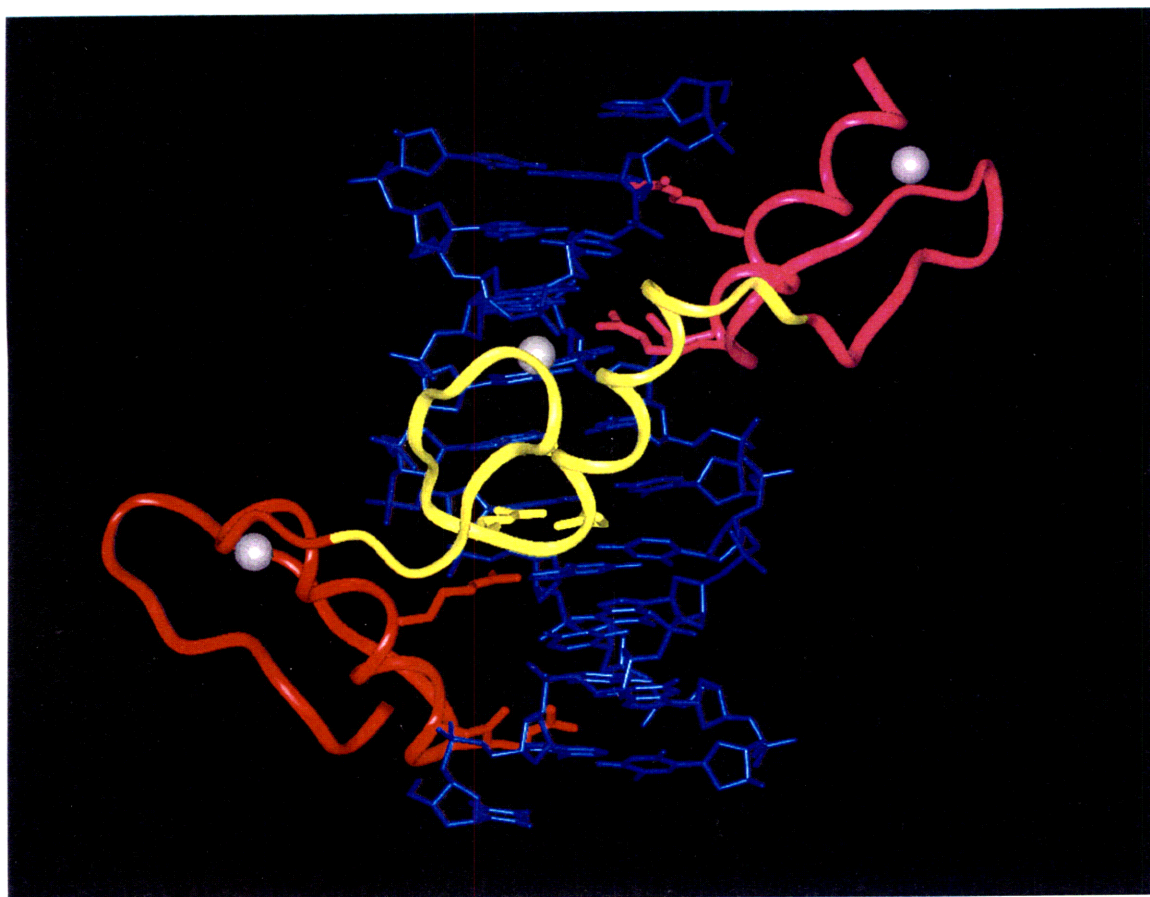
Figure 2

Figure 3. Schematic diagram of the base contacts made by Zif268. Arrows represent hydrogen bonds; dotted arrows represent hydrogen bonds with marginal geometry. Figure adapted from [59].

Figure 3

Finger 3

Finger 2

Finger 1

Arg 80

Asp 76

Arg 74

His 49

Asp 48

Arg 46

Arg 24

Asp 21

Arg 18

A

G   C

C   G

G   C

T   A

G   C

G   C

G   C

C   G

G   C

T   A

T
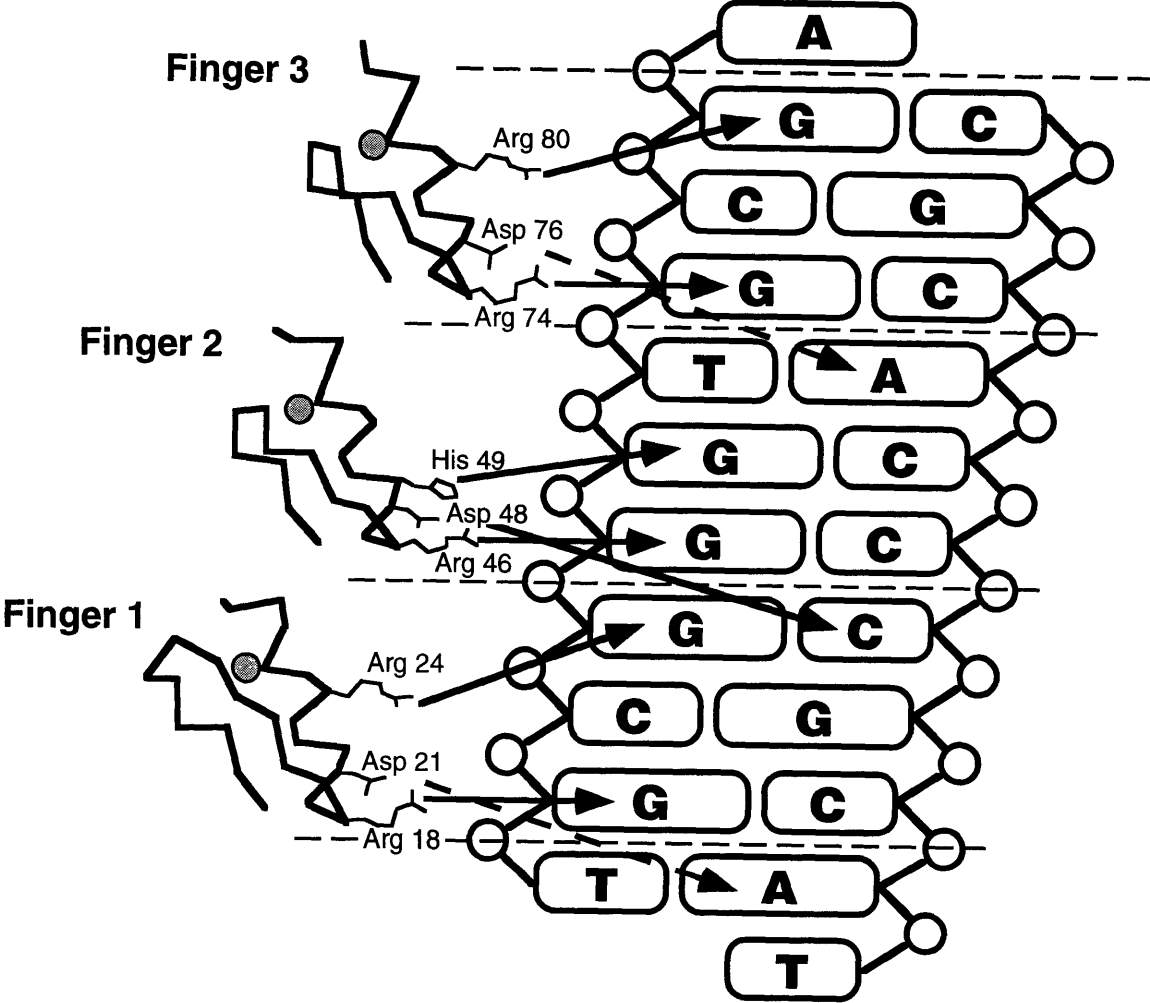
Figure 4. Views of the Zif268-DNA interface, illustrating the types of base contacts made by the Zif268 fingers. The electron density is from the final 2Fo-Fc map (Chapter 2), contoured at either 1.5 σ (panels a and c) or 2.0 σ (panel b). **(a)** Arg 18 (from position -1 of the α helix in finger one) makes a pair of hydrogen bonds with guanine 10, and Asp 20 (from position 2 of the helix) makes a pair of hydrogen bonds with the Arg. Analogous interactions occur in fingers two and three between corresponding residues and bases. **(b)** His 49 (from position 3 of the α helix in finger two) contacts guanine 6. **(c)** Arg 24 (from position 6 of the α helix in finger one) makes a pair of hydrogen bonds with guanine 8. An analogous interaction occurs in finger three, between Arg 80 and guanine 2.

Figure 4

A
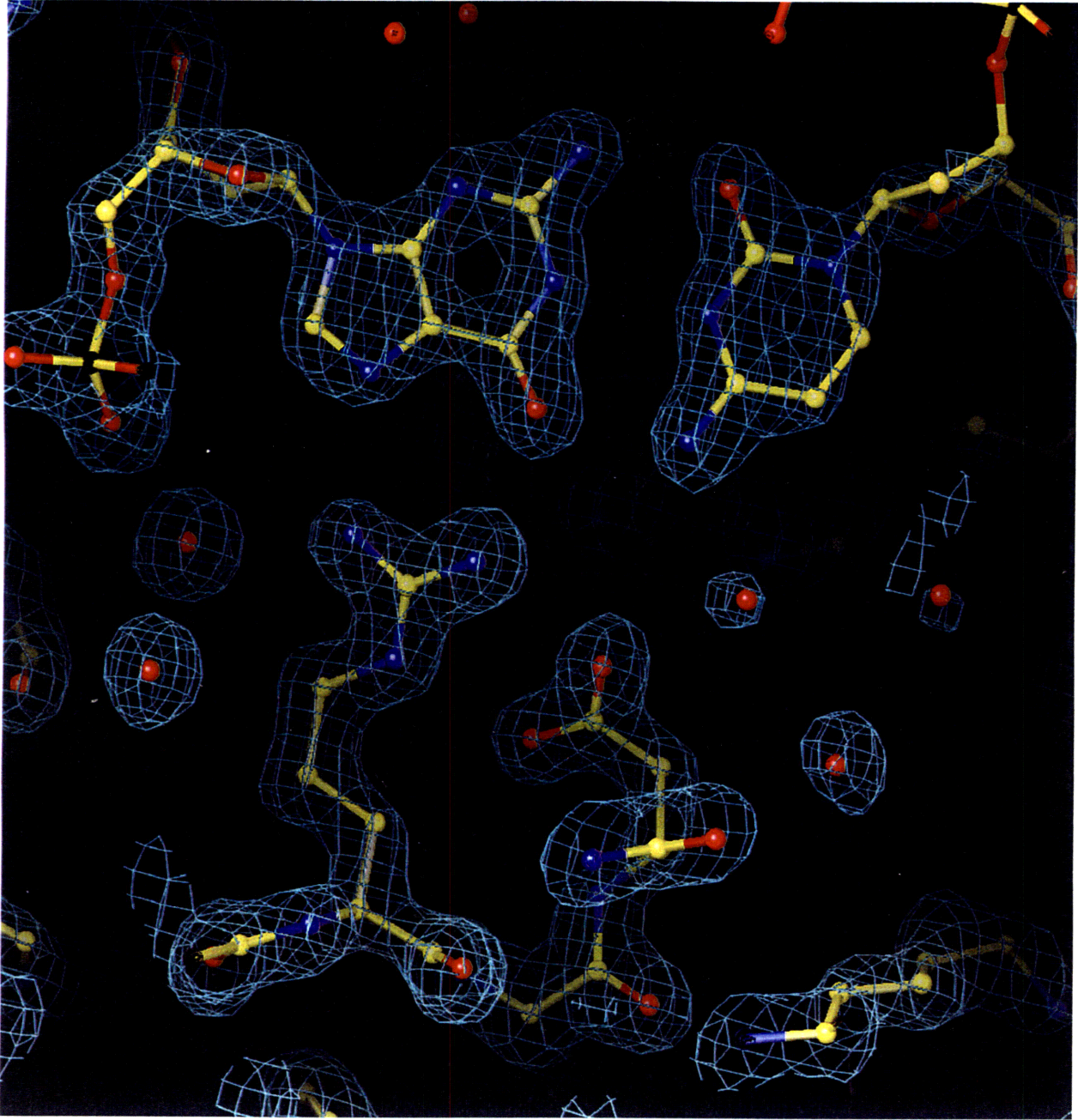
B
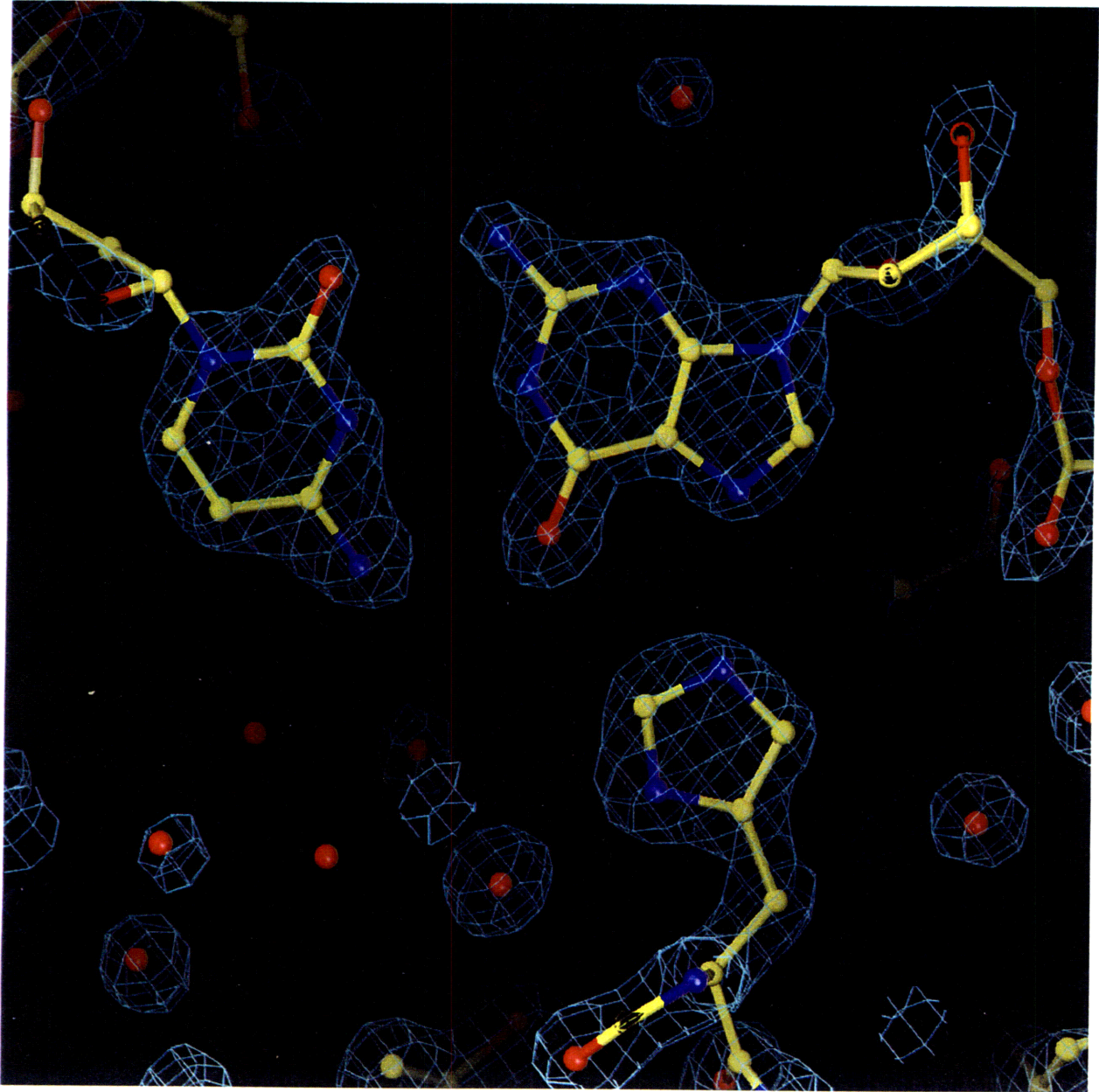
C
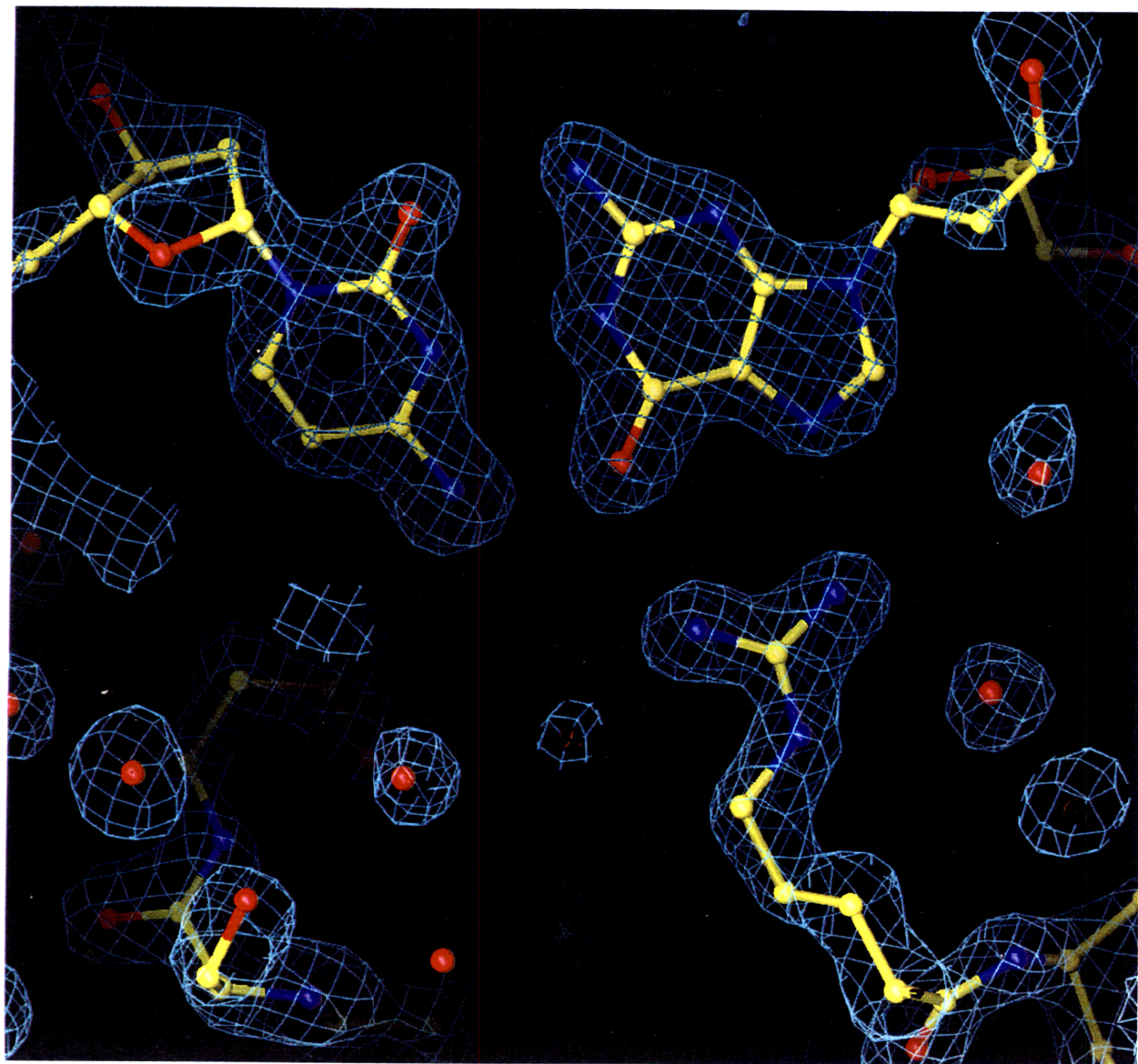
Figure 5. Residues seen to contact DNA in the structures of natural zinc finger-DNA complexes (summarized from references [59-63, 65-67]). Squares indicate residues that make base contacts; circles indicate residues that make phosphate contacts (gray circles indicate residues thought to make phosphate contacts in the NMR structure of the first three fingers of TFIIIA with DNA [66] but not seen to make contacts in the crystal structure of the first six fingers bound to DNA [67]). Complete amino acid sequences of the peptides used in structural studies of the Zif268, GLI, tramtrack, and YY1 complexes are given, but several amino-terminal residues have been omitted from the sequences of the GAGA and TFIIIA peptides. Zif 1 indicates the first finger of Zif268, TTK 1 the first finger of tramtrack, and so on. The sequences were aligned by conserved residues (bold type) and secondary structure elements. Numbers indicate the location of the four key base-contacting positions with respect to the start of the α helix.

Figure 5

```
                                                    -1    2 3        6

Zif 1    M E Ⓡ P Y A C P V E S C D Ⓡ R F S [R] S [D] E L T [R]Ⓗ I R - I H T
Zif 2    G Q K P F Q C R - - I C M Ⓡ N F Ⓢ [R] S [D]Ⓗ L T T Ⓗ I R - T H T
Zif 3    G E K P F A C D - - I C G Ⓡ K F A [R]Ⓢ [D] E R K [R] H T K - I H L R Q K D

GLI 1      V Y E T D C R W D G C S Q E F D S Q E Q L V H H I N S E H IⒽ
GLI 2    G E R K E F V C H W G G C SⓇE L R P F Ⓚ A Q Ⓨ M L V V Ⓗ T K - I H T
GLI 3    G E K P H K C T F E G C R K S Ⓨ S Ⓡ L E N L Ⓚ T H L R S - H T
GLI 4    G E K P Ⓨ M C E H E G C S K A F S N [A][S][D]Ⓡ A K Ⓗ Q N R Ⓣ H S
GLI 5    N E Ⓚ P Y V C K L P G C T K R Ⓨ Ⓣ[D] P SS L [R]Ⓚ H V K T V H G P D A

TTK 1  M E F T K E G E H T Y R C K - - V C S Ⓡ V Ⓨ ⓉⒽ I [S][N] F C Ⓡ H Y V T S H K
TTK 2    R N V K V Y P C P - - F C F Ⓚ E F T [R]Ⓚ[D][N] M T A Ⓗ V K I I H K I

YY1 1    M E P R T I A C P H K G C T Ⓚ M F Ⓡ D N S A M Ⓡ[K]Ⓗ L H - T H G
YY1 2    - P Ⓡ V H V C A - - E C G K A F V E S S Ⓚ L K [R]Ⓗ Q L - Ⓥ H T
YY1 3    G E Ⓚ P F Q C T F E G C G Ⓚ R F Ⓢ L D F N L Ⓡ T Ⓗ V R - I H T
YY1 4    G D R P Y V C P F D G C N K K F A [Q] S [T][N] L K S H I L - Ⓣ H A K A K N N Q

GAGA    ···E Q P A T C P - - I C Y A V ⒾⓇⓆ S [R][N] L ⓇⓇⒽ L E Ⓛ R H F A K P G V

TFIIIA 1 ···V Y K Ⓡ Ⓨ I C S F A D C G A A Ⓨ N [K] N [W]Ⓚ L Q A H L C - Ⓚ H T
TFIIIA 2   G E Ⓚ P F P C K E E G C E Ⓚ G F Ⓣ S L [H]Ⓗ L T [R]Ⓗ S L - Ⓣ H T
TFIIIA 3   G E Ⓚ N F T C D S D G C D L R F Ⓣ T K [A][N] M Ⓚ Ⓗ F N [R] Ⓕ H N
TFIIIA 4 I K I C V Y C H F E N C G K A F K K H N [Q] L K V H Q F - S H T
TFIIIA 5 Q Q L P Ⓨ E C P H E G C D Ⓚ R F S [L] P [S][R] L Ⓚ[R]Ⓗ E K - Ⓥ H A
TFIIIA 6 - - - G Y P C K K D D S C S F V G Ⓚ T W T L Y L K H V A E C H Q D
```

Figure 6. Summary of the base contacts observed in structures of natural zinc finger-DNA complexes [59-63, 65-67]. The contacts are organized to emphasize the tendency of each base-contacting position in the finger to interact with a preferred base in the finger's subsite. Figure adapted from [108].

Figure 6
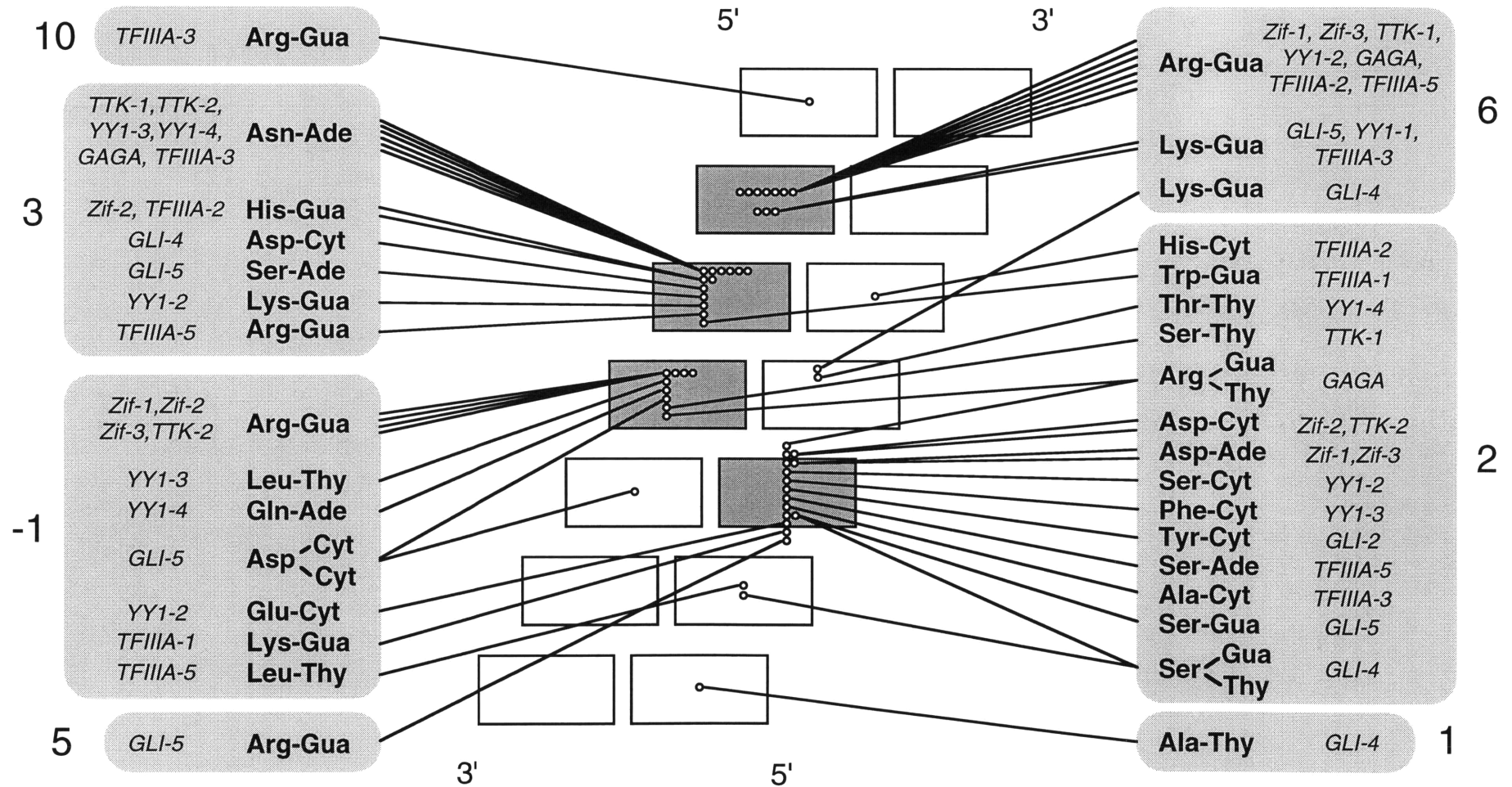
Figure 7. Zinc fingers can assume different orientations with respect to the DNA. **(a)** The base-contacting fingers of GLI dock against the DNA in roughly the same manner as do the Zif268 fingers, but there are differences in their precise orientations. This figure shows the Cα trace of Zif268 finger one (red), Zif268 finger two (purple), Zif268 finger three (orange), GLI finger four (yellow), and GLI finger five (green), along with their respective three base pair DNA subsites (all in blue). The fingers and their subsites were aligned by superimposing the P atoms of the DNA backbone. **(b)** Sketch of the GLI zinc finger-DNA complex, showing that finger one (which makes only a single contact with the DNA) has a very different orientation with respect to the DNA than do the other fingers. Panel reproduced from [61].

Figure 7

A
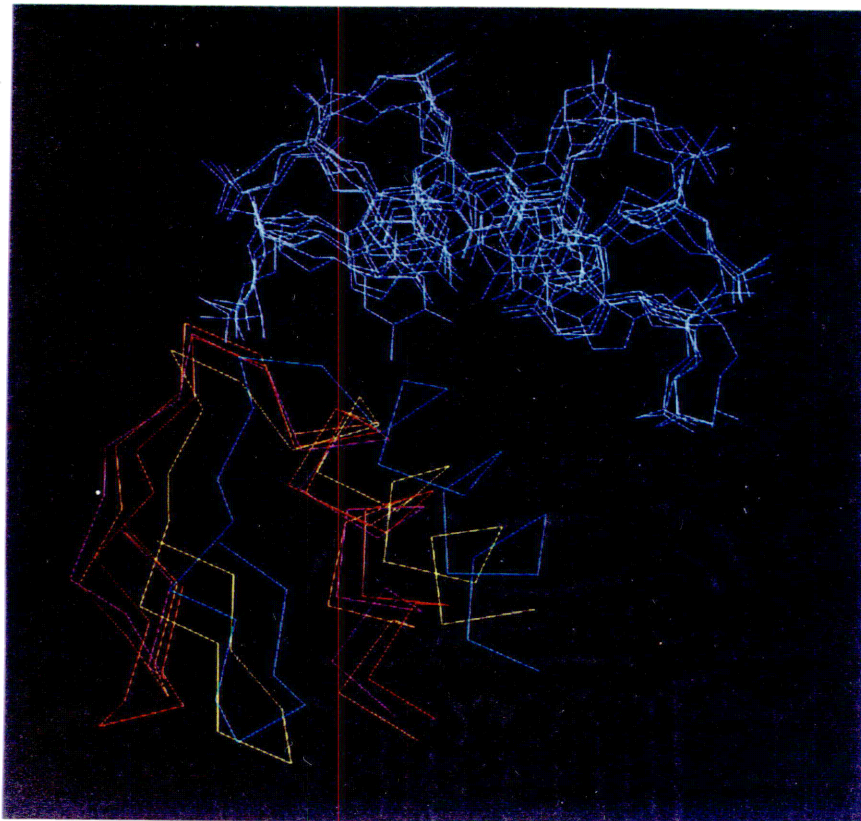
B
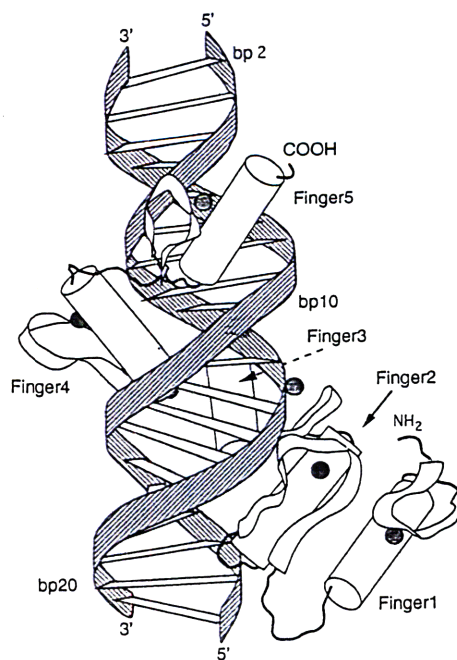
Figure 8. Schematic drawing illustrating the selection for variants of Zif268 with altered DNA binding specificities performed by Rebar and Pabo [79]. The residues at positions -1, 2, 3, and 6 of the α helix in finger one of Zif268 were randomized; fingers two and three were unchanged. Phage display methods were used to select for peptides capable of binding to DNA target sites in which the subsite for finger one had been altered (the subsites for fingers two and three were unchanged). A consensus sequence of Asp, Ser, Asn, and Arg was obtained at positions -1, 2, 3, and 6 when GACC was the targeted subsite for finger one. Two distinct consensus sequences were obtained when GCAC was the targeted subsite for finger one: Arg, Ala, Asp, and Arg when nonspecific DNA was used as a competitor in the selections, and Gln, Gly, Ser, and Arg when the wild type Zif268 binding site was used as a competitor.

Figure 8
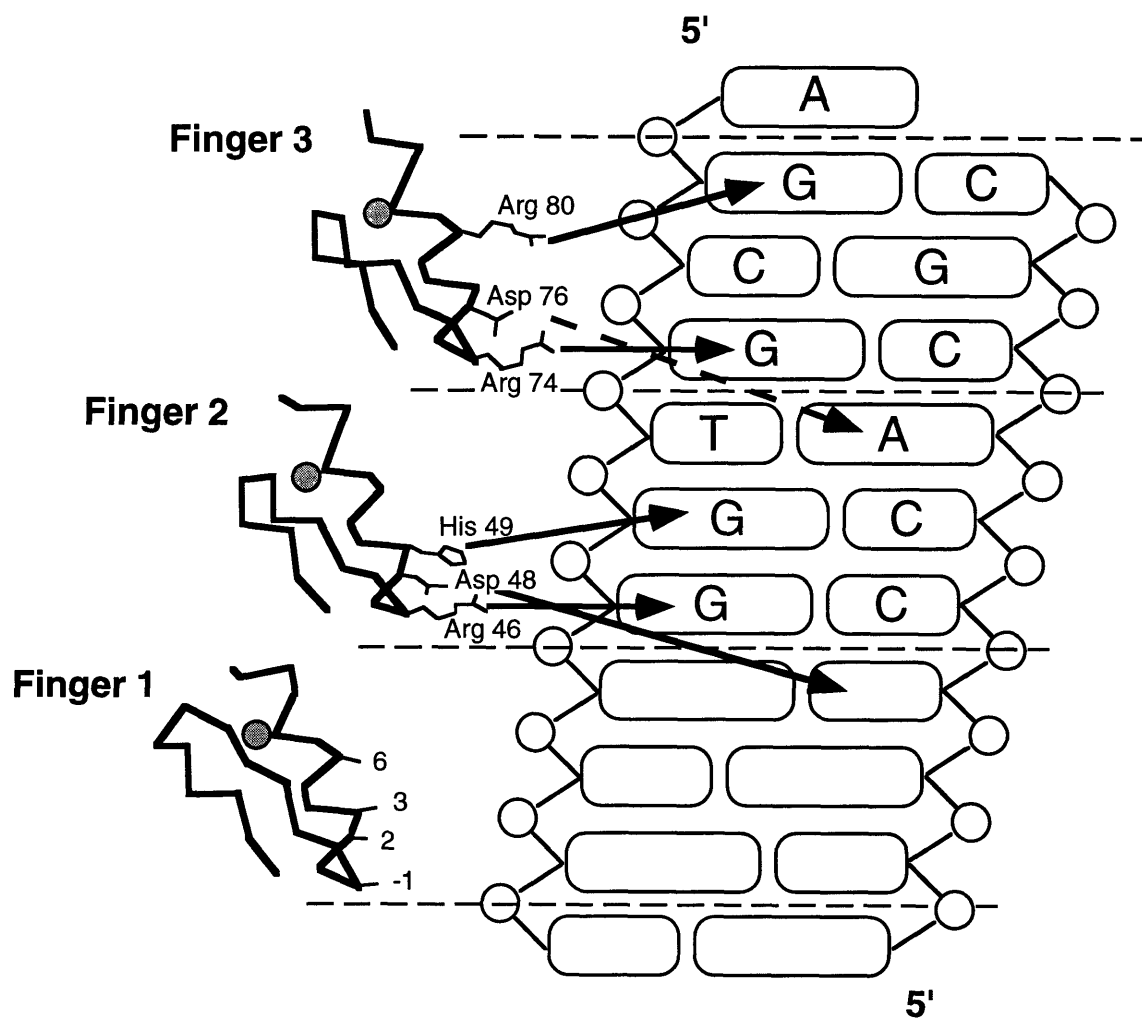
Table 1. Apparent dissociation constants measured for the peptides selected by Rebar and Pabo with various binding sites [79]. Outlined boxes indicate complexes between the variant peptides and the sites against which they were selected. Gray boxes indicate complexes whose structures are described in Chapter 3 (the wild type complex is described in Chapter 2). Table adapted from [79].

Table 1

Finger one sequence	Apparent K_d (nM)		
-1 2 3 6	GACC	GCAC	GCGC (w.t.)
DSNR	0.019	2.5	1.8
RADR	9.3	0.068	0.035
QGSR	1.8	0.055	0.54
RDER (w.t.)	33.	5.6	2.7

Chapter Two

Zif268 Protein-DNA Complex Refined At 1.6 Å: A Model System For Understanding Zinc Finger-DNA Interactions

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.

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], and 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-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 (Figures 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 amino-terminal portion 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 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 DNA (these contacts are highlighted in Figure 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 Figure 1A).

i. Roles of Arg -1 and Asp 2

As emphasized in the previous paper, the aspartic acid at position 2 of each α helix (Asp 20, 48, and 76) makes a pair of hydrogen bonds with the arginine immediately preceding the start of the helix (Arg 18, 46, and 74), and each of these arginines makes a pair of hydrogen bonds with a guanine. The hydrogen bonds between the Asp at position 2 of each helix and the Arg 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/Asp 2 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 (Figure 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: Asp 20 makes a water-mediated contact with the N4 of cytosine 10', Asp 48 with cytosine 7', and Asp 76 with cytosine 4'. In fingers one and three, the arginines at position -1 of the helix make water-mediated contacts to the phosphate backbone. Arg 18 makes a water-mediated contact to the phosphate on the 5' side of guanine 10, and Arg 74 has a corresponding interaction with the phosphate on the 5' side of guanine 4. (In finger two, Arg 46 is slightly further from the corresponding phosphate, and two ordered water molecules are seen that 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/Asp 2 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 Asp 20 and the N4 of cytosine 9 (Figure 4B). An analogous water-mediated contact occurs between Asp 76 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 Asp 48 clearly contacts cytosine 8': the carboxylate is 3 Å from the cytosine N4, and the hydrogen bonding geometry is reasonable (Figure 4B). The corresponding interactions are less favorable in fingers one and three, where the distance between the Asp and the exocyclic amine is greater (Asp 20 is 3.8 Å from the N6 of adenine 11', Asp 76 is 3.5 Å from the N6 of adenine 5'). In these cases, the orientation also is 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 Asp. Competition experiments and K_d determinations also show that Zif268 binds oligos with adenine or cytosine at these positions about 5-10 fold more tightly than oligos containing guanine or thymine [13], (Bryan Wang and C.O.P., unpublished data). The preference for adenine or cytosine at these positions may reflect weak favorable interactions with the corresponding Asp residues. However, Asp also could contribute to the observed sequence preferences by tending to exclude guanine or thymine from these positions (the nearby Asp presumably would have weak unfavorable electrostatic interactions with either of these bases).

ii. Roles of residue 3

Histidine 49, 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

69

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 Asp [13, 14]. Other binding studies confirm that Zif268 binds slightly more tightly to oligos containing a thymine at this position 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 simple ideas about a 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 Glu 21 and Glu 77 clearly do not make any base contacts: the carboxylate of Glu 21 is 5.3 Å away from the N4 of cytosine 9, and that of Glu 77 is 5.7 Å away from the N4 cytosine 3. Instead, these side chains hydrogen bond to the backbone amides of the residues immediately preceding the α helix. Glu 21 makes a good hydrogen bond to the backbone -NH of Arg 18 (the residue immediately preceding the α helix) and can also hydrogen bond to the backbone amide of Ser 17 (Figure 4C). Corresponding interactions occur in finger three, where Glu 77 hydrogen bonds to the backbone -NH groups of Arg 74 and Ala 73. 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 Glu 21 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 occur in finger three, between Glu 77 and cytosine 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 Glu 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

71

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 (Figure 4B). As described above, Asp 20 and Asp 76 (at position 2 of the helices) make water-mediated contacts to cytosines 3 and 9. In addition, Arg 24 and Arg 80 (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 also is a water molecule that interacts with the N7 position of guanine 9'. This water is stabilized by Ser 47 and Asp 48 from finger two, and it also interacts with the water that bridges Arg 24 and the O6 of this guanine. (There is no water off the N7 of guanine 3' in our structure.)

iii. 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 (Arg 24 to guanine 8, and Arg 80 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: Arg 24 makes a water-mediated contact with the O6 of guanine 9', and Arg 80 with the corresponding position of guanine 3'. Arg 80 also makes a water-mediated contact with the O5' of the nucleotide at position 1.

Thr 52, 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 (Arg 74 is another ligand of this water molecule).

iv. 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 Figure 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 similar, 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). Ser 75 (finger 3) hydrogen bonds to phosphate 7', while Ser 47 (finger 2) makes a water-mediated contact with phosphate 9'. (As mentioned earlier, Ser 47 also makes a water-mediated contact to the N7 of guanine 9'.) Likewise, Ser 19 (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. Ser 45 (finger 2) hydrogen bonds to phosphate 6, while Ser 17 (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. Arg 78 (residue 4 in the helix of finger 3) makes a water-mediated contact with phosphate 7', and one conformation of Thr 23 (residue 5 in the helix of finger one) contacts the

O5' of base 12'. (Ser 19 is another ligand of this water.)

v. 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 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 seven fold, 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 Lys 33, 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 (Figure 4D). Lys 61 (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 bp 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 zinc finger-DNA complex, the GLI zinc finger-DNA complex, and several other protein-DNA complexes [18]. 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.

The most dramatic change during refinement of the Zif268 complex involved the overhanging adenine and thymine bases at the ends of the DNA duplex (Figure 1B). These bases, which had been added to facilitate crystallization [11, 19], 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 arrangement has interesting implications for understanding the Zif268 crystal packing contacts and the interactions of neighboring complexes in the crystal. However, since 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 3 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) reveals an "autocorrelation" with a three base pair periodicity (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 of the GCGTGGGCG sequence 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 the 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 $B_{enlarged\ groove}$-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 [18]. 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 (Figure 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 Arg 27 and the backbone carbonyl of residue 45; fingers 2 and 3 are connected by an analogous hydrogen bond between Arg 55 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 studies 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 (Figure 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 XPLOR [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 (M.A.R., unpublished) were performed. As refinement continued, 148 water molecules were added, and alternate conformations for five side chains (Pro 4, Arg 15, Thr 23, Gln 36, and Leu 50) 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>2σ (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 sec averaging time. Spectra were taken at a DNA concentration of 0.05 mg/ml and a peptide concentration of 0.075 mg/ml 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.

Acknowledgements

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. *Nuc. 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. & Jäckle, 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. 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.

19. Jordan, S.R., Whitcombe, T.V., Berg, J.M. & Pabo, C.O. (1985). Systematic variation in DNA length yields highly ordered repressor-operator cocrystals. *Science* **230**, 1383-1385.

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.

25. Brünger, A.T. (1992). *XPLOR Manual version 3.1.* Yale University Press, New Haven, CT.

26. Israel, M. & Chirino, A.J. (1994). *TOM/FRODO version 3.0.* University of Alberta, Alberta, Canada, and California Institute of Technology, Pasadena, CA.

27. Engh, R.A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.* **A47**, 392-400.

28. Dickerson, R. (1993). *NEWHELIX93.* University of California, Los Angeles, CA.

Table 1. Local helical parameters for the DNA site*.

base pair		displacement	helical twist	rise	tilt	roll
2	GC	-2.60				
			23.6	3.69	0.33	4.85
3	CG	-1.91				
			40.5	2.91	-4.14	6.62
4	GC	-2.20				
			27.3	3.40	0.74	5.08
5	T A	-1.08				
			36.2	3.23	-4.49	3.45
6	GC	-1.59				
			31.0	2.97	-1.66	3.48
7	GC	-2.01				
			35.5	3.69	-1.45	9.23
8	GC	-1.74				
			28.1	2.88	1.78	-0.96
9	CG	-0.66				
			36.1	3.38	-2.61	3.21
10	GC	-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 can not be directly compared to the values reported here. (Using NEWHELIX93 with the coordinates from reference 11 gives values similar to those shown above.)

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	rms (P and C1' atoms)
2-10 and 3-11	1 bp	1.163 Å
2-9 and 4-11	2 bp	1.123 Å
2-8 and 5-11	**3 bp**	**0.963 Å**
2-7 and 6-11	4 bp	1.099 Å
2-6 and 7-11	5 bp	1.297 Å
2-5 and 8-11	**6 bp**	**0.686 Å**
2-4 and 9-11	7 bp	1.161 Å
2-3 and 10-11	8 bp	1.091 Å

* As highlighted by the bold lettering, the closest matches occur when the site is shifted by 3 or 6 base pairs.

Table 3. Linker lengths and interfinger 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	linker lengths*		interfinger distances†	
			1-2	2-3	1-2	2-3
model 1	1 0	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

* measured from the Thr $C\alpha$ to the Phe $C\alpha$

† distance from the Arg to the backbone carbonyl

§ values measured for the Zif268 complex are shown for comparison; values for the bp/turn and displacement are averages

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
free R-factor (%)	24.2
reflections with F>2σ	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 (deg) protein	1.3
DNA	3.0

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 sheets by arrows. Our model includes residues 3 through 87; the terminal residues are disordered in the crystal. (**B**) Sequence of the duplex oligonucleotide used in the cocrystallization. Figure adapted from [11].

A

-1 1 2 3 4 5 6

```
 1        5        10        15        20        25        30
 M E R P Y A C P V E S C D R R F S R S D E L T R H I R I H T G Q K
```

```
        35        40        45        50        55        60
     P F Q C R I - - C M R N F S R S D H L T T H I R T H T G E K
```

```
        65        70        75        80        85        90
     P F A C D I - - C G R K F A R S D E R K R H T K I H L R Q K D
```
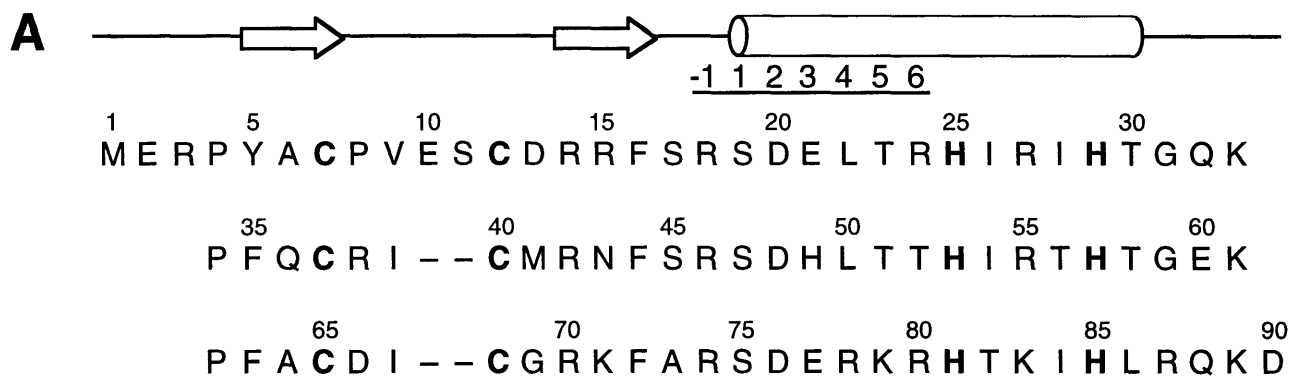
B

```
1   2 3 4   5 6 7   8 9 10  11
A   G C G   T G G   G C G   T

    C G C   A C C   C G C   A T
    2' 3' 4'  5' 6' 7'  8' 9' 10'  11' 12'
```

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 (residues 3 to 32) is red, finger two (residues 33 to 60) is yellow, and finger three (residues 61 to 87) is purple. The DNA is shown in blue, and the zinc ions in gray.
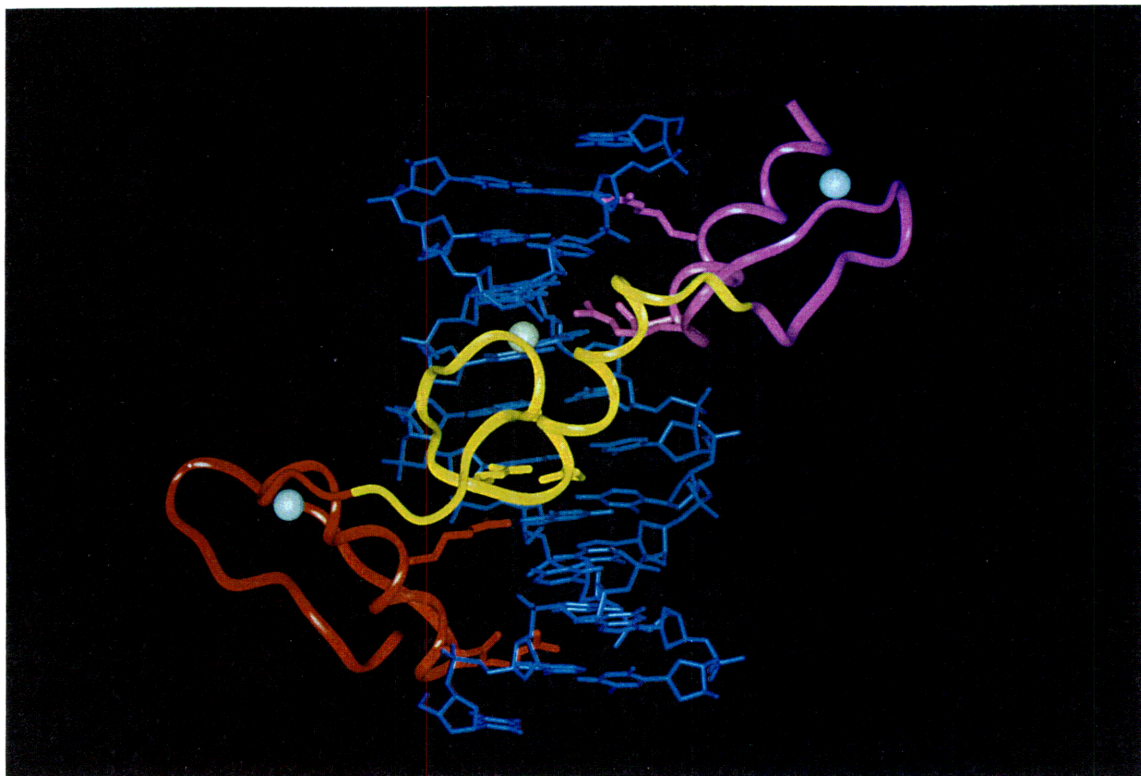
Figure 2

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.
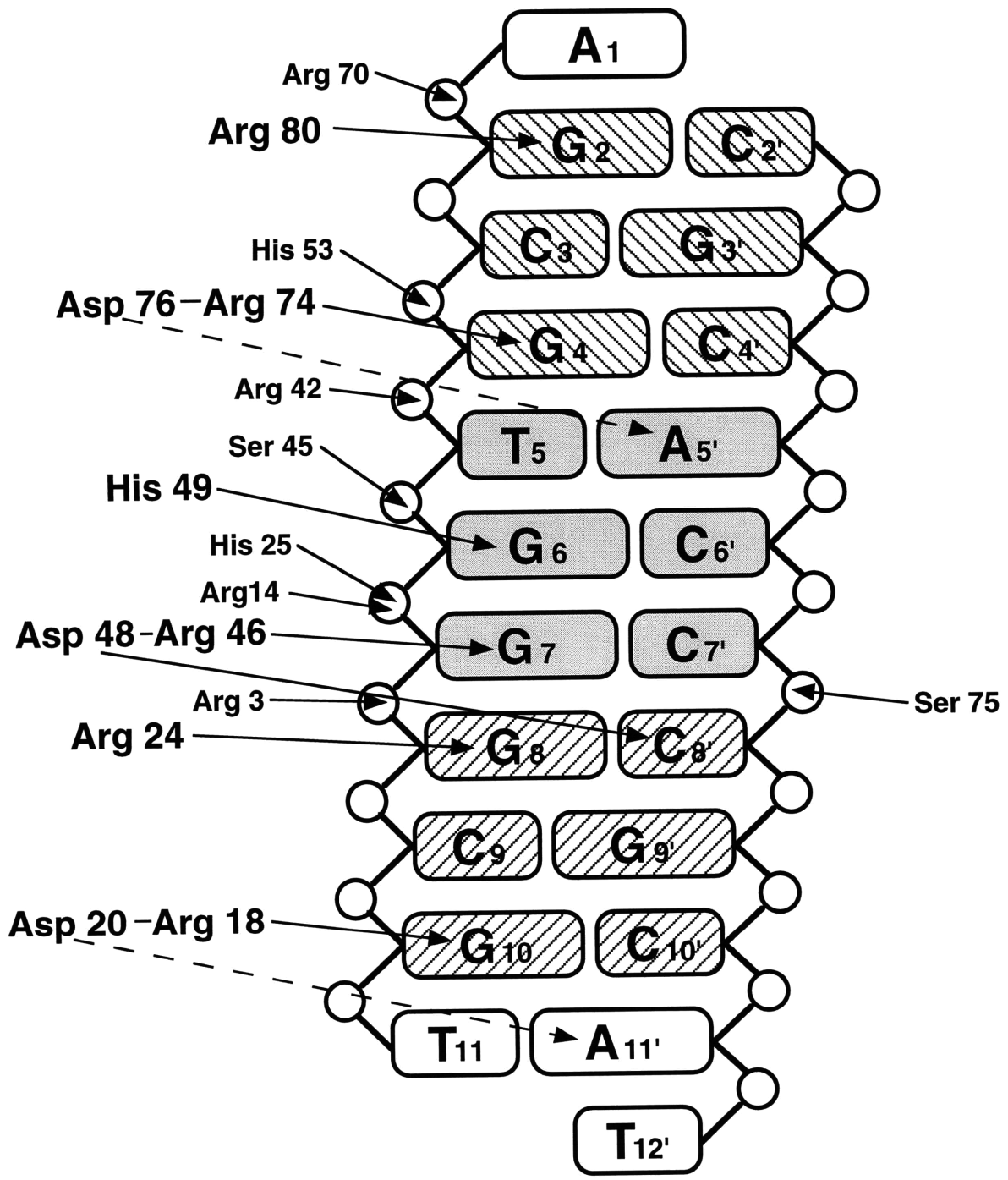
Figure 4. Details of the protein-DNA interface. Side chains from finger one are shown in red, from finger two in yellow, and from finger three in purple. The peptide backbone is shown in gray and the DNA in blue. Water molecules are depicted as gray spheres. (**A**) Stereo view showing the network of contacts that Arg 18 and Asp 20 make with base pair 10. (These are the Arg -1/Asp 2 pair from finger one; an analogous set of interactions occurs in fingers two and three.) (**B**) Stereo view showing the water-mediated interactions that Asp 20 and Arg 24 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 off the N7 of guanine 3'.) The contact between Asp 48 and cytosine 8' is also visible in this figure, as is the water-mediated contact between Asp 48 and cytosine 7'. (Guanine 7 and Arg 46 have been omitted for clarity.) (**C**) Stereo view showing the conformation of Glu 21, with its carboxylate group hydrogen bonding to the backbone amides of Ser 17 and Arg 18. 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 Lys 33, from the linker between fingers one and two, to phosphate 5.

Figure 4

A

B

C

D
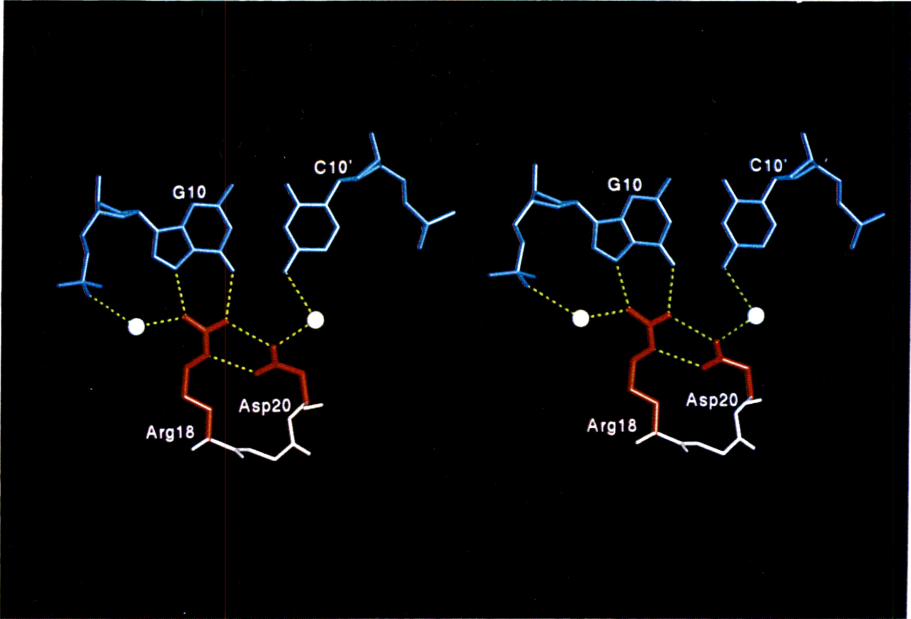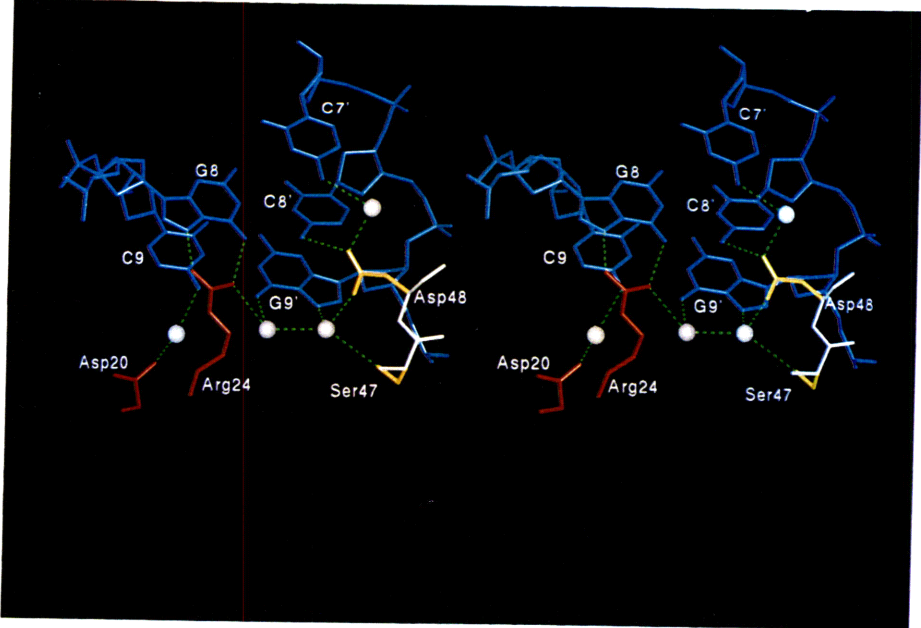
Figure 5. CD spectra of the Zif268 DNA binding site (——), peptide

(··········), and complex (-----), plotted as the molar extinction

coefficient per nucleotide ($\Delta\varepsilon$, in $M^{-1}cm^{-1}$).
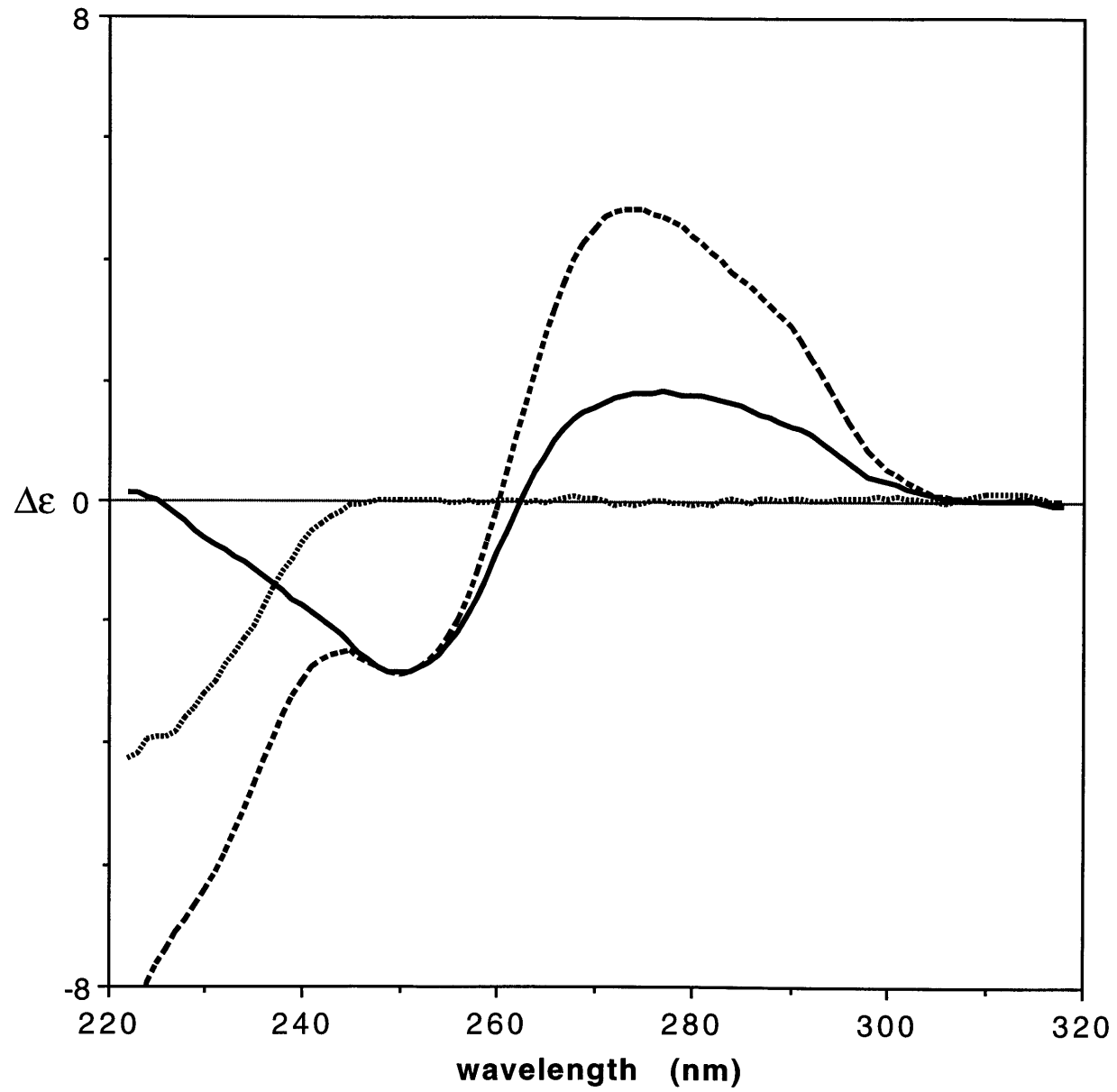
Figure 6. Model showing the individual Zif268 fingers docked against ideal B-DNA with 10 bp/turn. Each finger has been docked in a way that preserves the local DNA contacts. A dashed line indicates the distance each of the linkers would have to span (measured from α carbon to α carbon of the indicated 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.

Figure 6
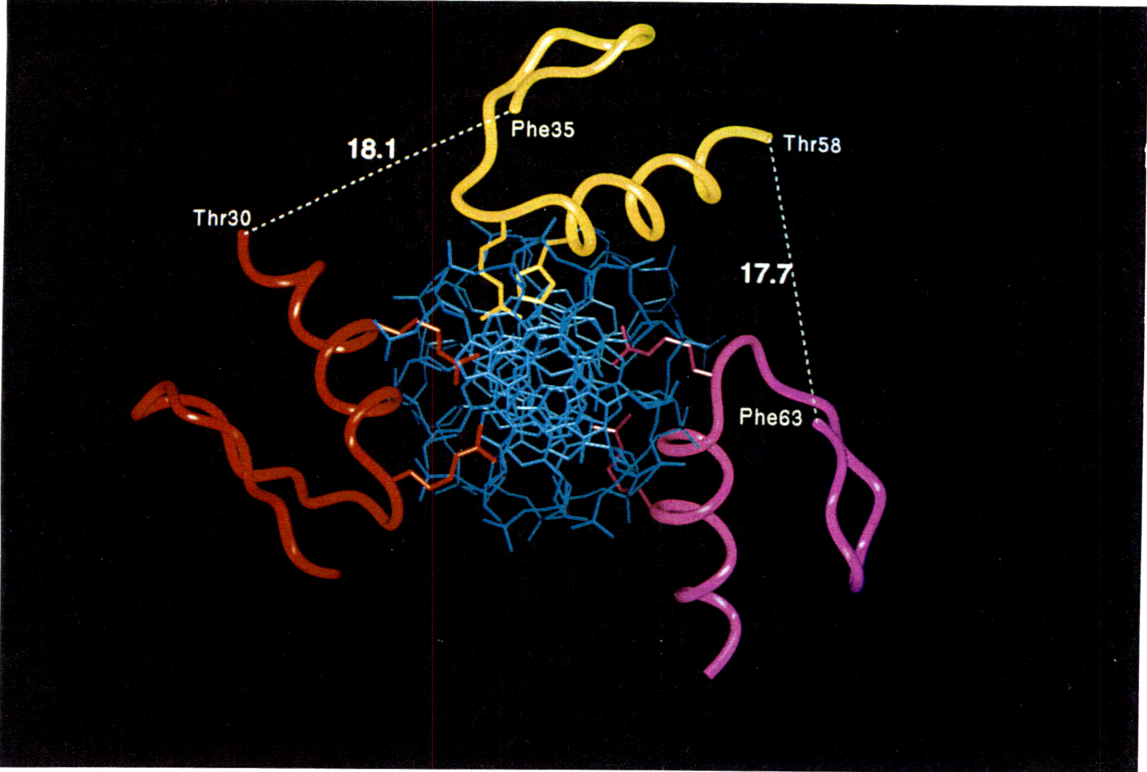
Chapter Three

High Resolution Structures of Variant Zif268-DNA Complexes: Implications for Understanding Zinc Finger-DNA Recognition

Background: Zinc fingers of the Cys_2His_2 class comprise one of the largest families of eukaryotic DNA-binding motifs and recognize a diverse set of DNA sequences. These proteins have a relatively simple modular structure, and key base contacts are typically made by a few residues from each finger. These features make the zinc finger motif an attractive system for designing novel DNA-binding proteins and for exploring fundamental principles of protein-DNA recognition.

Results: Here we report the X-ray crystal structures of zinc finger-DNA complexes involving three variants of Zif268 (with multiple changes in the recognition helix of finger one) that were selected by Rebar and Pabo [*Science* **263**, 671-673 (1994)]. We describe the structure of each of these three-finger peptides bound to its corresponding target site. To help elucidate the differential basis for site-specific recognition, we have also determined the structures of four other complexes containing various combinations of these peptides with alternative binding sites.

Conclusions: The protein-DNA contacts observed in these complexes reveal the basis for the specificity demonstrated by these Zif268 variants. Many, but not all, of the contacts can be rationalized in terms of a recognition code, but the predictive value of such a code is limited. Our structures illustrate how modest changes in the docking arrangement accommodate the new side chain-base and side chain-phosphate interactions. Such adaptations help explain the versatility of naturally occurring zinc finger proteins and their utility in design.

Introduction

Designing and selecting novel zinc finger proteins provides an exciting opportunity to explore the principles of protein-DNA recognition, and structural analysis of the new complexes is critical for careful interpretation of the results. Many zinc fingers with modified specificities have been produced via design and selection efforts [e.g., 1-8]. However, only one of the resulting zinc finger-DNA complexes has been characterized structurally [9], and, in this case, there is no wild type structure available for direct comparison. Here we report systematic structural studies on a set of complexes resulting from selections performed by Rebar and Pabo [1]. The peptides selected were variants of Zif268, providing an excellent opportunity for careful structural analysis and comparison since the wild type Zif268-DNA complex has been solved and refined to 1.6 Å resolution [10, 11].

The Cys_2His_2 zinc finger proteins have a simple, modular structure. Each finger consists of about 30 amino acids and contains a short two-stranded antiparallel β sheet and an α helix. The sheet and the helix are held together by a small hydrophobic core and by a zinc ion that is coordinated by two conserved cysteines from the sheet region and two conserved histidines from the α helix. Crystallographic studies of the three-finger Zif268 peptide-DNA complex revealed that each of the fingers has a similar docking arrangement and that the fingers use residues from the amino terminal portion of the α helix to contact bases in the major groove. In the Zif268-DNA complex, most of the base contacts involve the guanine-rich strand of the Zif268 binding site (GCG/TGG/GCG), and each finger makes its primary contacts to a three base pair subsite.

The Zif268 structure revealed a characteristic pattern of contacts correlating certain residue positions along the α helix of each finger with certain base positions in that finger's subsite (Fig. 1a). Comparing the three fingers shows that base contacts are made by residues at positions -1, 2, 3, and 6 (numbering with respect to the start of each α helix). There is a general tendency for the residue at position -1 of a finger's α helix to contact the 3' base (on the primary strand) of that finger's subsite, for the residue at position 3 of the α helix to contact the central base of the subsite, and for the residue at position 6 of the α helix to contact the 5' base. The residue at position 2 of the α helix also projects directly into the major groove and sometimes contacts a base that is on the secondary strand of the DNA and just outside of the 3 bp subsite. (One can - for simplicity - describe the Zif268 complex in terms of 3 bp subsites that include the primary base contacts, or one can - to include the contacts made by residue 2 - describe the complex in terms of 4 bp subsites that overlap by one base pair at each finger/finger boundary.) Residues from these same four positions along the α helix (-1, 2, 3, and 6) also make critical base contacts in other zinc finger-DNA complex structures, and, in many cases, these contacts involve corresponding bases within a given finger's subsite [9, 12-14]. Such observations have led to much discussion about the prospects for deriving a "code" governing zinc finger-DNA interactions [e.g., 2, 15].

To test the versatility of the zinc finger motif in recognition and to explore potential patterns of side chain-base interactions, Rebar and Pabo [1] randomized positions -1, 2, 3, and 6 in finger one of Zif268 (leaving fingers two and three unchanged) and used phage display to select variants that bound to alternative DNA sites. In these target DNA sites (which

contained a full length binding site for the three-finger peptide), the region normally recognized by finger one was altered from the $\underline{GCG}^{G/}_{T}$ preferred by the wild type protein to either $\underline{GACC}$ or $\underline{GCAC}$. When selections were performed against the $\underline{GACC}$-containing site, the consensus amino acids obtained at positions -1, 2, 3, and 6 were Asp, Ser, Asn, and Arg (DSNR). Selections against the $\underline{GCAC}$-containing site were performed under two different sets of conditions, giving a consensus of Gln, Gly, Ser, and Arg (QGSR) at these positions when the wild type binding site was used as a competitor in the selections, and a consensus of Arg, Ala, Asp, and Arg (RADR) when nonspecific DNA was used as a competitor [1].

In this paper, we report a series of cocrystal structures (involving a total of seven different complexes) that allow detailed analysis of the contacts made by the DSNR, QGSR, and RADR variants at several different DNA sites. These structures provide important data about the adaptability and versatility of the zinc finger motif - revealing how alternative side chain-base and side-chain phosphate interactions can be accommodated in the zinc finger framework - and give us new perspectives on the prospects for a zinc finger-DNA recognition code.

Results

Sequences of the zinc finger peptides and of the binding sites used in this study are given in Figure 1b and 1c. In every case, cocrystals contained a full three-finger peptide bound to a duplex site that included a 10 bp region of double-stranded DNA. Since sequence changes are confined to finger one of the peptide and to its corresponding 3-4 bp subsite on the

DNA, our notation highlights these regions. Our designation for each 90 amino acid peptide (such as "DSNR") refers to the residues selected at positions -1, 2, 3, and 6 of the α helix in finger one. (These positions correspond to residues 18, 20, 21, and 24 of the three-finger peptide.) Our designation for each 10 bp duplex DNA site (such as "GACC") refers to the subsite for finger one, and we underline the 3 bp region that, by analogy with the wild type Zif268 complex, might be expected to be involved in the primary base contacts.

In the course of this project, we have solved and refined seven new cocrystal structures. Studies involved the variant peptides (DSNR, QGSR, and RADR) each crystallized with the target site that had been used in the selections. To gain a deeper understanding of the differential basis of specificity, we also studied these peptides with other sites and crystallized the wild type Zif268 peptide with one of the variant sites. In the following sections, we proceed to summarize each of these structures, focusing on the contacts made by residues -1, 2, 3, and 6 in the α helix of finger one. Other contacts seen in these complexes (such as the conserved contacts made by fingers two and three) are summarized in a later section.

The DSNR peptide with the targeted GACC site

The DSNR variant had been obtained via selections with a GACC binding site, and our 2.1 Å structure of this complex gives a very satisfying explanation for the specificity. Each of the selected residues contacts a different base pair (Fig. 2a). Both carboxylate oxygens of Asp18 (the residue at position -1 of the α helix) contact the exocyclic amine of cytosine 10 (2.9 Å and 3.3 Å). The hydroxyl group of Ser20 (position 2 of

107

the helix) is near two potential hydrogen bond acceptors: the O6 of guanine 11' (2.7 Å) and the O4 of thymine 12' (3.3 Å). Ser20 can donate a hydrogen bond to only one of these bases at a time; we presume that the serine usually hydrogen bonds to guanine 11' since it is closer to this base. Asn21, at position 3 of the helix, makes a pair of hydrogen bonds with adenine 9 (both 3.0 Å). Finally, Arg24, residue 6 of the helix, makes a pair of hydrogen bonds with guanine 8 (2.9 and 3.0 Å). These key residues also make several water-mediated contacts with the DNA. Ser20 and Asp18 serve as ligands for a water which contacts the O4 of thymine 12'. Arg24 makes one water-mediated contact to the O4 of thymine 9' and another to the phosphate of base 7.

The DSNR peptide with the wild type <u>GCG</u>T site

A complex of the DSNR peptide with the wild type <u>GCG</u>T site (to which it binds less tightly) was also solved (at 1.9 Å resolution) to help us understand the basis for specificity, and we find that the DSNR peptide makes fewer contacts with this site (Fig. 2b). Asp18 does not make any contacts to the DNA. Ser20 interacts with the phosphate of adenine 11' (3.5 Å), and a single water molecule bridges Ser20 to both the N6 and N7 of this base. The side chain oxygen of Asn21 makes a single hydrogen bond with the exocyclic amine of cytosine 9 (3.4 Å), while Arg24 makes a pair of hydrogen bonds with guanine 8 (2.9 and 3.0 Å). As in the previous complex, Arg24 also makes water-mediated contacts to the O6 of guanine 9' and to the phosphate of base 7.

The QGSR peptide with the targeted <u>GCA</u>C site

The QGSR variant had been selected for its ability to bind the <u>GCA</u>C target site in the presence of competing wild type site. For this particular

cocrystal, molecular replacement failed to give clear density for finger one, and multiple isomorphous replacement was used to solve the structure. Combining the MIR phases with phases from a partial model (containing fingers 2 and 3 with their subsites) gave an interpretable map at 1.6 Å resolution, but the resulting density suggested that finger one (up to about residue 22) is somewhat disordered in this crystal. However, the density for the side chains of Gln18, Ser21, and Arg24 was still readily interpretable, indicating that they have well-defined conformations and allowing us to model the key contacts.

The side chains of residues 18, 21, and 24 each make direct and water-mediated contacts with the bases, and our structure readily explains the specificity for the GCAC site (Fig. 3). Gln18, the residue immediately preceding the α helix, makes a pair of hydrogen bonds with adenine 10 (2.8 and 3.0 Å). Ser21, from position 3 of the α helix, accepts a hydrogen bond from the N4 of cytosine 9 (3.5 Å). Arg24 makes a pair of hydrogen bonds with guanine 8 (2.9 and 3.0 Å). In addition, the Cβ of Ser19, the first residue in this α helix, makes van der Waals contacts with the methyl group of thymine 12' (3.4 Å). (In the set of structures reported here, this is the only example of a base contact made by a residue that is not at position -1, 2, 3, or 6 of the α helix, although such contacts are seen in other zinc finger-DNA complexes, e.g., [12].) Gly20, the residue at position 2 of this α helix, does not make any base contacts or adopt any unusual $\Phi\Psi$ angles, and it is not yet clear why glycine occurs at this position.

Specificity of recognition in the QGSR complex may also be enhanced via several water-mediated contacts. The -NH$_2$ of the Gln18 side chain interacts with both the O6 of guanine 11' and the O4 of thymine 12'

through a single bridging water molecule. Ser21 makes a water-mediated contact to the phosphate of base 8, and Arg24 makes water-mediated contacts with the O6 of guanine 9' and the phosphate of base 7.

The RADR peptide with the targeted GCAC site

The RADR variant also had been selected for binding to the GCAC site, but in this case nonspecific DNA had been used as a competitor. Our 1.6 Å crystal structure shows that finger one of the RADR peptide makes relatively few base contacts with the targeted GCAC site. To our surprise, we find that Arg18, the residue at position -1 of the α helix, interacts with the phosphate of base 9 rather than contacting a base. Asp21, which is at position 3 of the helix, makes a bifurcated hydrogen bond to Arg18 but does not make any direct contacts with the DNA. The only direct base contacts from this finger are made by the residues at positions 2 and 6 of the α helix: Ala20 makes a van der Waals contact with the methyl group of thymine 12' (3.6 Å), and Arg24 makes a pair of hydrogen bonds with guanine 8 (2.9 and 3.1 Å).

There are also a few water-mediated contacts in this complex. Asp21 makes one water-mediated contact to the N4 of cytosine 9 and another water-mediated contact to the phosphate and O5' of base 8. Arg24 makes a water-mediated contact to the O6 of guanine 9'.

The RADR peptide with the wild type GCGT site

Binding studies had shown that the RADR peptide also binds very tightly to the wild type site, and we pursued this complex to help us understand the significance of the new phosphate contact made by the RADR peptide with its targeted GCAC site. In the 2.0 Å complex with the

wild type site, as in the GCAC complex, Ala20 makes a van der Waals contact to thymine 12' (3.5 Å) and Arg24 makes a pair of hydrogen bonds to guanine 8 (2.8 and 2.9 Å)(Fig. 4b). However, in this complex with the GCGT site, both Arg18 and Asp20 occupy two conformations. One conformation of each side chain is similar to that observed in the GCAC complex: the Arg contacts the phosphate of base 9 (2.7 and 3.2 Å), and the Asp makes a bifurcated hydrogen bond with the Arg. The alternate conformation of Arg18 makes a base contact, donating a pair of hydrogen bonds to guanine 10 (3.3 and 3.5 Å) and hydrogen bonding with the alternative conformation of Asp21. (Note that adenine occupies position 10 in the GCAC complex.) This alternative conformation of Arg18 also makes water-mediated interactions with the O4 of thymine 11 and with the N6 of adenine 11'. The other water-mediated contacts observed in this GCGT complex are ones that were seen in the complex with the targeted site: a water bridges the first conformation of Asp21 to the phosphate and O5' of base 8, and another water bridges Arg24 to the O6 of guanine 9'.

The RADR peptide with the (less favorable) GACC site

Binding studies had shown that the RADR peptide binds significantly less tightly to the GACC site, and we also studied this complex (at 1.9 Å resolution) to help analyze the differential basis for recognition. In this complex, Arg24 is the only residue from finger one that makes any base contacts (either direct or water-mediated) with the DNA (Fig. 4c). Arg24 (position 6) makes a pair of hydrogen bonds with guanine 8 (2.6 and 3.0 Å), but these have a somewhat different geometry than in the other complexes since this Arg is tilted such that it also makes a hydrogen bond to the O4 of thymine 9' (2.8 Å). Neither Arg18, Ala20, nor Asp21 make any contacts with the DNA, and there actually appear to be two unfavorable

interactions between the peptide and the GACC site: Arg24 is 3.3 Å away from the N6 of adenine 9, and Asp21 is 3.2 Å away from the phosphate of base 8.

The wild type RDER peptide with the GCAC site

To gain additional information about discrimination and specificity, we crystallized the wild type peptide (which prefers the GCGT site) with the less favorable GCAC site and solved this structure at 2.3 Å resolution. (Prompted by the RADR structures, we wondered what would happen to Arg18 when the guanine at position 10 was replaced with adenine.) In this complex, we find that Arg18, the residue at position -1, has two distinct conformations (Fig. 4d). In one conformation, Arg18 interacts with the phosphate of base 9 (2.9 and 3.0 Å) and makes a bifurcated hydrogen bond to Glu21. In the other conformation, Arg18 extends toward adenine 10 and makes a pair of hydrogen bonds with Asp20. This second conformation of Arg18 allows the NH1 of the side chain to form a hydrogen bond with the N7 of adenine 10 (3.3 Å), but it also places the side chain NH2 3.0 Å from the N6 of the adenine. There are no contacts with cytosine 8, and we do not observe any water-mediated contacts between finger one and the GCAC subsite. As observed in all of the other complexes, Arg 24 (position 6) makes a pair of hydrogen bonds with guanine 8 (2.8 and 3.0 Å).

Other contacts in this set of complexes

In describing these seven complexes, our discussion has focused on finger one since this is the region that was randomized and since our structures show that other regions of the complex are relatively well conserved. In all of the complexes described here, fingers two and three make the same base contacts that they do in the wild type Zif268 complex.

Arg46 and His49 (positions -1 and 3 of finger two) hydrogen bond to guanines in the central TGG subsite. Arg74 and Arg80 (positions -1 and 6 of finger three) hydrogen bond to guanines in the terminal GCG subsite. As observed in the high resolution structure of the wild type complex [11], the aspartic acids at position 2 in these helices also play a role in recognition. Asp48 (in finger two) contacts cytosine 8' on the secondary strand of the DNA. Asp76 (finger three) makes a water-mediated contact to cytosine 3 (in all seven of the complexes) and may have a weak favorable interaction with the N6 of adenine 5' (which is 3.2-3.6 Å away in the various structures).

In addition, several of the residues that interact with the phosphate backbone in the Zif268 complex make the same contacts in all seven variant complexes. His25 and His53 (conserved zinc ligands from fingers one and two) contact the phosphates of base 7 and base 4, respectively. Ser47 makes a water-mediated contact to the phosphate of base 9'. Thr52, the residue at position 6 in the helix of finger two, makes a water-mediated contact to the phosphate of base 4. Arg74 (the residue at position -1 of the helix in finger three) makes another water-mediated contact with the same phosphate. Although many contacts are fully conserved, some residues - especially in finger one - make phosphate contacts only in a subset of the complexes. For example, Arg3 interacts with the phosphate of base 8 in every complex except the QGSR/GCAC and RADR/GACC complexes, while Arg14 contacts the phosphate of base 7 in the wild type, DSNR/GACC, and RADR/GCAC complexes, but not in the other structures. On the whole, the structure and contacts of the finger two and finger three region are very well conserved in this set of complexes.

DNA conformation

The DNA in the wild type Zif268 complex, and in the other zinc finger-DNA complexes whose crystal structures have previously been reported, is a distinctive form of B-DNA [9-14, 16]. This conformation, called $B_{enlarged\ groove}$-DNA, is characterized by an unusually wide and deep major groove that results from a slight unwinding of the DNA and an increased displacement of the base pairs from the helical axis [16]. Analysis of the helical parameters for each of the seven structures reported here reveals that the DNA has a $B_{enlarged\ groove}$-DNA conformation in every complex, even those in which finger one makes relatively few contacts with the DNA. (The DNA conformation may change as the protein binds, as it appears to in the wild type Zif268 and Sp1 zinc finger complexes [11, 17], and this may be propagated as a cooperative structural change for the entire site.) In the seven complexes, the average helical twist angle ranges from 31.7° to 32.3° (corresponding to 11.1-11.4 base pairs per turn), and the average displacement of the base pairs from the helical axis ranges from 1.6 Å to 1.8 Å (as determined with NEWHELIX93 [18]). These values are very similar to those observed for the wild type Zif268 complex, where the average helical twist angle is 32.1° (corresponding to 11.2 base pairs per turn) and the average displacement is 1.6 Å.

Discussion

This set of structures provides a basis for addressing many fundamental questions about zinc finger-DNA recognition. One of the most

intriguing issues in the field involves the idea that there may be some type of simple "code" underlying zinc finger-DNA interactions (at least for the subfamily of Cys_2His_2 fingers most closely related to Zif268). Our data provide much new information about this question and about related issues involving structural plasticity and the physical/chemical basis for specificity. Given that we also have determined structures of peptides with suboptimal or alternative binding sites, we can examine the differential basis for site specific recognition.

Is there a simple zinc finger-DNA "recognition code"?

In trying to understand how the variant proteins recognize their targeted sites, we begin by considering base contacts that may explain specificity, and we discuss these in the context of ideas about a possible "recognition code" for zinc finger-DNA interactions. The crystal structure of the wild type Zif268-DNA complex had shown that residues at positions -1, 2, 3, and 6 of the α helix were especially important for site-specific recognition, providing the basis for randomizing these residues in the initial phage display experiments and for thinking about patterns of side chain-base interactions. Figure 5 summarizes the contacts made by these residues of finger one in the wild type Zif268 complex and in the seven new complexes described in this paper.

In some cases, the basis for specificity seems quite clear and is generally consistent with ideas about a recognition code. The contacts between the DSNR peptide and the targeted GACC site seem especially satisfying from this perspective. Each base pair in the subsite is contacted by one of the four residues that had been randomized. These contacts preserve the characteristic alignment of specific residue

positions in the α helix with specific base positions in the subsite (Fig. 1a), and the side chain-base interactions can readily be rationalized in terms of a recognition code. (Note, however, that serine may make a limited contribution to specificity, since serine can act as either a hydrogen bond donor or acceptor and all four bases have a hydrogen bond acceptor or donor at a similar position.) The idea that the set of contacts made by DSNR to the GACC site contributes to specificity was confirmed structurally by crystallizing this peptide with the wild type site. The peptide has approximately 100-fold lower affinity for this GCGT site, and we find that the DSNR peptide makes fewer contacts at this site (compare Fig. 5 a and b).

Comparisons of these variant complexes with known structures also are relevant when thinking about a code, and we note that finger five of the GLI peptide [12], which has DSSK at the key α helical positions, also binds a GACC subsite. Comparing this finger with finger one of the DSNR complex (Fig. 6a) shows that in both cases the 5' guanine is recognized by a basic residue at position 6 of the helix (Arg in finger one of DSNR and Lys in finger five of GLI). The middle base, adenine, is recognized by either an asparagine (DSNR) or a serine (GLI) from position 3 of the α helix. In both complexes the 3' cytosine is recognized by an aspartic acid, albeit in somewhat different ways. In finger five of GLI, the Asp immediately preceding the α helix makes a hydrogen bond to each of the cytosines in the subsite, whereas in finger one of DSNR the corresponding Asp makes a hydrogen bond only to the first cytosine. Finally, in each complex, the guanine on the opposite strand in the fourth position of the subsite is contacted by a serine from position 2 of the α helix.

The structure of the QGSR peptide with the targeted GCAC site also

provides a reasonable explanation for the specificity of this peptide (Fig. 5c), and the contacts observed can easily be rationalized in the context of a recognition code. Again, structural comparisons are very interesting in thinking about a code. It happens that the second finger of a designed peptide, which has QSDK at the key α helical positions, recognizes a GCAG subsite [9] that is very similar to the GCAC subsite recognized by finger one of QGSR. Both fingers use a basic residue at position 6 of the α helix to recognize the 5' guanine: Arg in finger one of QGSR, and Lys in finger two of the designed peptide (although this Lys also contacts an additional base (Fig. 6b)). Position 3 uses an Asp (designed peptide) or Ser (QGSR) to recognize the cytosine in the middle of the DNA subsite. In both complexes, the Gln at position -1 of the α helix makes a bidentate contact with the 3' adenine.

In summary, comparing our fingers with previous structures clearly reveals related patterns of side chain-base interactions (as expected for a code). However, these comparisons reveal that there is no simple one-to-one correspondence between the identity of the base at a given position of the subsite and the identity (or even location within the α helix) of the amino acid used to recognize that base. The idea of having an exhaustive code also is complicated by the observed sets of water-mediated contacts, by occasional secondary contacts (such as those made by the Asp in GLI and the Lys in the designed peptide) and by occasional contacts from other positions in the helix (such as those made by Ser19 of QGSR and by the Arg in GLI). Such results highlight a crucial difference between using a code to rationalize a set of contacts (which, as in the above cases, often proves to be satisfactory) and the much harder problem of trying to develop a code or algorithm that could predict

an optimal set of contacts.

The new phosphate contact seen in the complex between the RADR peptide and its targeted GCAC site illustrates another difficulty with developing an exhaustive recognition code and raises interesting questions about recognition and specificity. The structural results are quite surprising, since Arg18 (at position -1, supported by Asp21 at position 3) contacts a phosphate and there are very few base contacts. Three additional structures were solved to help us understand the basis for specificity in binding of the RADR peptide. 1) We solved the structure of the RADR peptide with the 140-fold less favorable GACC site, and this structure revealed that there are indeed significantly fewer contacts with this less favorable site (cf Fig. 5d and 5f). 2) Binding studies had shown that the RADR peptide binds very well to the wild type GCGT site (actually with a slightly greater affinity than for the targeted GCAC site). Our crystal structure of this complex shows that the Arg18/Asp21 pair has two conformations at the GCGT site. One conformation allows Arg18 to make a pair of hydrogen bonds with guanine 10 that are similar to the contacts seen in the wild type complex (cf Fig. 5e and 5h). The other conformation of the Arg18/Asp21 pair allows Arg18 to make a phosphate contact that is very similar to that seen in the complex of the RADR peptide with the targeted GCAC site (cf Fig. 5d and 5e). 3) To further explore the role of Arg18, we solved the structure of the wild type RDER peptide with the GCAC site (where binding is about 2-fold weaker than at the wild type site). Again multiple conformations occur, with an Arg18/Glu21 pair contacting the phosphate in one arrangement and an Arg18/Asp20 pair contacting adenine 10 in the alternative conformation (cf Fig. 5g with 5d and 5h).

Comparing these four structures with each other and with that of

118

the wild type complex reveals that the acidic residues at position 2 or 3 of the helix play a key role in orienting the Arg at position -1 and thus help determine what sort of contacts the Arg makes with the DNA. (The most favorable conformation for Arg18 is that which maximizes contacts with nearby acidic residues as well as with the DNA.)   Such interactions between side chains introduce an additional level of complexity that can not readily be incorporated into a recognition code. Similarly, phosphate contacts such as those observed in the RADR complexes would be difficult to predict with any existing code. In addition, in our structures of all the variant complexes, we see water-mediated contacts between the peptide and the DNA. These contacts may help enhance the specificity of binding, but it is not clear how they can be incorporated into a simple recognition code.

Structural adaptations in the zinc finger framework

A central issue in designing fingers with novel binding specificities and in thinking about a recognition code involves understanding how various side chain-base interactions can be accommodated in zinc finger-DNA complexes. How does the position and orientation of the polypeptide backbone (which will be determined by folding and docking of the whole domain) help to determine which side chain-base interactions are possible? How can a longer or shorter side chain be accommodated at a given position? Clearly, some structural plasticity is needed to accommodate new side chain-base interactions (since, for example, glutamine and arginine are not isosteric), but too much flexibility may allow interactions to occur at other, nonspecific sites and thereby actually reduce the specificity of recognition. Our structures provide an excellent basis for considering these issues since we can directly

compare the variant complexes with the wild type complex. We consider three examples to illustrate the range of structural variation observed.

The complex between the QGSR peptide and its targeted site provides our first example of how alternative contacts can be accommodated. In this complex, the orientation of finger one with respect to the DNA appears to be quite similar to that of finger one in the wild type complex (Fig 7a). However, there are changes in the conformation of finger one (rmsd=0.61 Å for the α carbons), particularly at the N-terminal end of the α helix, that appear to be necessary for the new contacts to occur.

The structure of the RADR peptide with its targeted site (which has a new phosphate contact from the Arg at position -1) illustrates another type of structural rearrangement. Here the conformation of finger one is very similar to that seen in the wild type complex (rmsd=0.31 Å for the α carbons). However, the orientation of finger one with respect to the DNA is rather different. It appears to have rotated away from the DNA, pivoting as a relatively rigid unit around a point near the C-terminal end of the α helix (Fig. 7b).

The complex between the DSNR peptide and its targeted site shows changes in both the conformation and docking of finger one (Fig. 7c). Here finger one rotates (as a unit) so that it is closer to the DNA. In addition, the conformation of the amino-terminal end of the α helix is slightly different in the DSNR variant than in the wild type peptide (rmsd=0.66 Å for the α carbons). Together, these changes permit the Asp at position -1 of the α helix (which is much shorter than the Arg of the wild type peptide) to reach the DNA. (Note: this Asp may also help stabilize the altered conformation of the α helix, since it makes a hydrogen bond to the

backbone amide of the third residue in the helix.) The DSNR/<u>GAC</u>C complex may be the extreme case (within this subfamily of closely related structures), but some alterations in the conformation and/or orientation of finger one are seen in all of the complexes. Recognizing alternative sites by varying side chains on the conserved zinc finger framework involves a fine balance between plasticity and rigidity of the zinc finger unit.

Modularity of zinc finger peptides is another key issue in recognition and design, and our structures show that in these Zif268 variant complexes most of the changes are limited to the region involving finger one and the corresponding 3-4 bp subsite: in each of the seven complexes, we find that fingers two and three and the corresponding regions of the DNA are very similar to those in the wild type Zif268 complex. (Of course, finger one variants that disrupted contacts made by fingers two and three would pay a severe energetic penalty and presumably would not have been selected in the phage display protocol.) The conserved structure of the finger two/finger three region is visually striking (three examples are shown in Figure 7) and is confirmed by noting rms distances for the superimposed complexes. (Superimposing α carbons for fingers two and three and phosphorous and C1' atoms for the corresponding base pairs (2-7) gives rms distances of 0.38 Å, 0.35 Å, and 0.49 Å when comparing the wild type complex with complexes of the QGSR, RADR, and DSNR peptides at their targeted sites.) Overall, we find that alterations in finger one had relatively little effect on the structure and docking of fingers two and three.

Conclusions

Our structures provide new data and new perspectives on several aspects of zinc finger-DNA recognition. The new side chain-base and side chain-phosphate interactions in these Zif268 variants are accommodated by relatively modest changes in the structure and docking of finger one. There are no substantive changes in the region involving fingers two and three. These two important observations provide a direct structural basis for understanding the versatility, modularity, and adaptability of the zinc finger motif.

The complexes of the DSNR and QGSR peptides with their targeted sites have overall patterns of contacts that can be rationalized in terms of a (degenerate) recognition code. However, our high resolution structures reveal many details that are not accounted for in any simple code. These include water-mediated contacts, details of the DNA conformation, and occasional contacts by residues at other positions along the α helix.

The complex between the RADR peptide and its targeted site reveals unexpected phosphate contacts made by the arginine residue immediately preceding the α helix. This surprising new contact is facilitated by side chain-side chain interactions and by subtle changes in the overall docking arrangement of finger one. This variant structure would not have been predicted by any existing recognition code, since features like interactions between neighboring side chains, alternative side chain conformations, and changes in the conformation and orientation of the finger are difficult to incorporate into any code.

In summary, a zinc finger-DNA recognition code - which typically allows several alternatives at a given position - can rationalize many side

chain-base interactions seen in this subfamily of zinc finger-DNA complexes. However, the problem of predicting an optimal zinc finger sequence for any desired DNA target site is much more difficult. Examining the structural complexity of the zinc finger-DNA interface - as illustrated in our set of structures - reveals the problems inherent in proceeding from a simple code to reliable three-dimensional models and energetic predictions. Proposed recognition codes [e.g., 2, 15] seem to summarize meaningful patterns and correlations of allowed side chain-base interactions, but they can not yet substitute for systematic optimization in finding the best finger to use at a given binding site or for detailed structural analysis to understand the full set of contacts.

Biological implications

Zinc fingers of the Cys_2-His_2 class constitute one of the most abundant and versatile DNA-binding motifs found in eukaryotes [19, 20]. Zinc fingers also have provided a framework for the structure-based design and selection of proteins with novel DNA-binding specificities [e.g., 1-8]. In one such selection, key residues in the first finger of the three-finger Zif268 peptide were randomized and variant peptides with altered DNA-binding specificities were selected by phage display methods [1]. We report here the structures of each of these three peptides bound to the target site used in its selection (the DSNR/GACC, QGSR/GCAC, and RADR/GCAC complexes). To understand the differential basis of site-specific recognition, we also solved four other combinations of peptides and binding sites (the DSNR/GCGT, RADR/GCGT, RADR/GACC, and RDER/GCAC complexes). Many of the contacts we observe in these complexes, particularly those between the DSNR and QGSR peptides and

their targeted sites, fit the pattern of interactions that has been observed in the wild type Zif268 complex [10, 11] and other zinc finger-DNA complexes [9, 12-14]: the residue immediately preceding the α helix tends to contact the 3' base in the finger's subsite, the third residue in the α helix the middle base, the sixth residue in the α helix the 5' base, and the second residue in the α helix a base on the opposite strand of the DNA in the preceding finger's subsite. However, not all of the observed contacts might have been predicted, and the new phosphate contacts in the RADR complexes were especially surprising. In general, we find that the new contacts in our variant complexes are facilitated by changes in both the conformation of finger one and its orientation with respect to the DNA. Such adaptations, which accommodate new side chain-base and side chain-phosphate interactions within a generally conserved structural framework, provide a basis for understanding the versatility of naturally occurring zinc finger proteins and may facilitate design, selection, and model-building of other zinc finger-DNA complexes.

Materials and methods

Purification of zinc finger peptides

The wild type Zif268 peptide was purified essentially as described [10], using a set of steps involving reversed-phase batch extraction, reversed-phase high performance liquid chromatography (HPLC), cation exchange chromatography, and a final reversed-phase HPLC column. The RADR, QGSR, and DSNR peptides were expressed as described [1], and the cells were lysed with a freeze/thaw protocol. Inclusion bodies containing the peptides were pelleted, then dissolved in 8 M urea and 50 mM HEPES at pH 7.5 and reduced with 150 mM dithiothrietol (DTT) for 30 minutes at 70°C. The peptides were then loaded on a Source15S cation exchange column (Pharmacia) in 8 M urea, 50 mM HEPES at pH 7.5, and 10 mM DTT and were eluted with a NaCl gradient. The peptides were next purified on a C4 reversed-phase column (Vydac), reconstituted with zinc, and further purified on a MonoS cation exchange column (Pharmacia). The final purification step involved a C4 reversed-phase column run as described [10]. Purified peptides were stored dried in aliquots in an anaerobic chamber (Coy Laboratory Products). The expected mass of each peptide was confirmed by electrospray ionization mass spectrometry (Harvard University Microchemistry Facility).

Crystallization

Each zinc finger-DNA complex was prepared by dissolving an aliquot of the dried peptide in water, adding 1.5 molar equivalents (per finger) of zinc chloride to the peptide, and adjusting the pH with buffer (to either 6.2 or 8.0, as indicated below). The folded peptide was then added to 1

molar equivalent of buffered duplex DNA binding site (oligos were synthesized and purified as described in [21]). The concentration of the zinc finger-DNA complex was about 1 mM; the complex was solubilized by the addition of NaCl (at concentrations indicated below). All crystals were grown by hanging drop vapor diffusion, using an anaerobic chamber to eliminate any risk of oxidation.

In the two noncognate complexes involving the wild type binding site (the DSNR/GCGT and RADR/GCGT complexes), the GCGT site is used since it was the binding site used in crystallization of the wild type Zif268 complex [10, 11]. Binding studies show that each of these two peptides has an equivalent affinity for the GCGT site used in this study as for the GCGC site used in [1] (M.E.-E., unpublished data).

Crystals of the QGSR peptide with the GCAC binding site were grown by mixing the complex, in 200 mM NaCl and 50 mM MES at pH 6.2, with an equal volume of the well buffer (27.5-35% PEG 3350, 0-200 mM NaCl, 100 mM Tris, pH 8.5). For this complex, three isomorphous derivatives were prepared by substituting 5-iodouracil for thymine in the oligonucleotide. Cocrystals were grown with a single substitution at position 5, a single substitution at position 12', and substitutions at both positions 5 and 12' (numbering as in Fig. 1b). Crystals of the RADR peptide with the GCAC site were obtained by doing two successive rounds of macroseeding into drops prepared with two volumes of this complex (in 600 mM NaCl and 75 mM MES, pH 6.2) and one volume of well buffer (22.5% PEG 3350, 500 mM NaCl, 25 mM MES, pH 6.2). Small crystals of the wild type Zif268-DNA complex [10, 11] were used as seeds in the first round; RADR/GCAC crystals produced at this stage were used for a second round of seeding. Crystals of the RADR peptide with the GCGT binding site were grown by mixing the complex, in 300 mM NaCl and 75 mM MES at pH 6.2, with an equal volume of

well buffer (35% PEG 3350, 200 mM NaCl, 25 mM MES, pH 6.2). Crystals of the RADR peptide with the GACC binding site were produced by mixing the complex, in 300 mM NaCl and 75 mM MES at pH 6.2, with an equal volume of well buffer (25% PEG 1450, 25 mM MES, pH 6.2). Crystals of the DSNR peptide with the GACC binding site were obtained by macroseeding with small crystals of the RADR/GCAC complex into drops containing two volumes of the DSNR/GACC complex (in 300 mM NaCl and 75 mM MES, pH 6.2) and one volume of well buffer (20% PEG 400, 200 mM $MgCl_2$, 100 mM HEPES, pH 7.5). Crystals of the DSNR peptide with the GCGT binding site were grown by mixing equal volumes of the complex, in 200 mM NaCl and 100 mM bis-tris propane (BTP) at pH 8.0, and the well buffer (25% PEG 1450, 200 mM NaCl, 25 mM BTP, pH 8.0). Crystals of the wild type Zif268 peptide with the GCAC binding site were obtained by mixing the complex, in 400 mM NaCl and 100 mM BTP at pH 8.0, with an equal volume of well buffer (35% PEG 1450, 300 mM NaCl, 25 mM BTP, pH 8.0). All of the complexes crystallized in space group $C222_1$; unit cell dimensions are given in Table 1.

Data collection

Diffraction data for each complex was collected from a single crystal under cryogenic conditions (at approximately 130K). In most cases, the mother liquor served as the cryoprotectant. However, the RADR/GCAC cocrystal was transiently transferred to 30% PEG 3350 and 25 mM MES at pH 6.2 before flash cooling, and the RADR/GACC crystal was soaked in 30% PEG 1450, 300 mM NaCl, and 25 mM MES at pH 6.2 for 30 minutes before flash cooling. Data sets were collected using a Rigaku RU-200 X-ray generator equipped with mirrors (Molecular Structure

Corporation) and an R-Axis IIC image plate system, and data were processed with DENZO and SCALEPACK [22]. Data collection statistics are summarized in Table 1.

Phasing, model building, and refinement

For each complex, an initial set of phases was obtained by molecular replacement. The 1.6 Å structure of the wild type complex [11] was used as a search model, but the side chains of residues 18-24 and all water molecules were deleted to minimize phase bias. Initial maps revealed readily interpretable density for finger one in six of the seven complexes. However, the initial map for the complex containing the QGSR peptide with the targeted GCAC site was not as clear, so this structure was solved by multiple isomorphous replacement. Each derivative data set was scaled to the native data by local scaling with MAXSCALE (M. A. Rould, personal communication). Heavy atom refinements and phase calculations were carried out in MLPHARE [23, 24]. Phasing statistics for the three derivatives, which gave an overall figure of merit of 0.55, are shown in Table 2. An interpretable electron density map was obtained by using SIGMAA [24, 25] to combine the MIR phases with phases from a partial model (containing fingers two and three and the corresponding DNA subsites from the wild type complex).

For each of the seven complexes, model building was done with the program O [26]. The automatic search routines PEAKMAX and WATPICK from the CCP4 package [24] facilitated modeling the water structure. Water molecules were included only if there was clear spherical density in the 2Fo-Fc and Fo-Fc maps and at least one hydrogen bond donor or acceptor within 3.5 Å. Each round of rebuilding was followed by simulated annealing or Powell minimization and by restrained individual B-factor

refinement in X-PLOR [27]. A bulk solvent correction [28] was applied, and the overall progress of refinement was monitored by following the R_{free}. As a final step, each model was checked for accuracy against simulated annealing omit maps, with five residues of the peptide at a time omitted from the calculation. Refinement statistics for all seven complexes are given in Table 3.

Accession numbers

Coordinates have been deposited with the Brookhaven Data Bank. The PDB ID codes are 1a1f (DSNR/GACC complex), 1a1g (DSNR/GCGT complex), 1a1h (QGSR/GCAC complex), 1a1i (RADR/GCAC complex), 1a1j (RADR/GCGT complex), 1a1k (RADR/GACC complex), and 1a1l (Zif268/GCAC complex).

Acknowledgements

References

1.   Rebar, E.J. & Pabo, C.O. (1994). Zinc finger phage: affinity selection of   fingers with new DNA-binding specificities. *Science* **263**, 671-673.

2.   Desjarlais, J.R. & Berg, J.M. (1992). Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc. Natl. Acad. Sci. USA* **89**, 7345-7349.

3.   Desjarlais, J.R. & Berg, J.M. (1994). Length-encoded multiplex binding site determination: Application to zinc finger proteins. *Proc. Natl. Acad. Sci. USA* **91**, 11099-11103.

4.   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.

5.   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.

6.   Jamieson, A.C., Wong, H. & Kim, S.-H. (1996). A zinc finger directory for high-affinity DNA recognition. *Proc. Natl. Acad. Sci. USA* **93**, 12834-12839.

7.   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.

8.   Greisman, H.A. & Pabo, C.O. (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657-661.

9.   Kim, C.A. & Berg, J.M. (1996). A 2.2 Å resolution crystal structure of a designed zinc finger protein bound to DNA. *Nat. Struc. Biol.* **3**, 940-945.

10.   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.

11.   Elrod-Erickson, M., Rould, M.A., Nekludova, L. & Pabo, C.O. (1996). Zif268 protein-DNA complex refined at 1.6 Å: A model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171-1180.

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.   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.

14.   Houbaviy, H.B., Usheva, A., Shenk, T. & Burley, S.K. (1996). Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl. Acad. Sci. USA* **93**, 13577-13582.

15.   Choo, Y. & Klug, A. (1997). Physical basis of a protein-DNA recognition code. *Current Opinion in Structural Biology* **7**, 117-125.

16.   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.

17.   Shi, Y. & Berg, J.M. (1996). DNA unwinding induced by zinc finger protein binding. *Biochemistry* **35**, 3845-3848.

18.   Dickerson, R. (1993). *NEWHELIX93*. University of California, Los Angeles, CA.

19.   Jacobs, G.H. (1992). Determination of the base recognition positions of zinc fingers from sequence analysis. *EMBO J.* **11**, 4507-4517.

20.   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.

21.   Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. & Pabo, C.O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA

recognition with tethered DNA-binding modules. *Cell* **77**, 21-32.

22.   Otwinowski, Z. (1993). *DENZO: an oscillation data processing program for macromolecular crystallography.* Yale University, New Haven, CT.

23.   Otwinowski, Z., *Maximum likelihood refinement of heavy atom parameters,* in *Isomorphous replacement and anomalous scattering, proceedings of the CCP4 study weekend,* W. Wolf, P.R. Evans, andA.G.W. Leslie, Editors. 1991, SERC Daresbury Laboratory: Warrington. p. 80-86.

24.   CCP4 (1994). The CCP4 suite: Programs for protein crystallography. *Acta Cryst.* **D 50**, 760-763.

25.   Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* **A 42**, 140-149.

26.   Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods of building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A 47**, 110-119.

27.   Brunger, A.T. (1992). *XPLOR Manual version 3.1.* Yale University Press, New Haven, CT.

28.   Jiang, J.-S. & Brunger, A.T. (1994). Protein hydration observed by X-ray diffraction: Solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol.* **243**, 100-115.

29.   Evans, S.V. (1993). SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11**, 134-138.

30.   Kleywegt, G.J. & Brunger, A.T. (1996). Checking your imagination: Applications of the free R value. *Structure* **4**, 897-904.

Table 1. Data collection statistics

Peptide	DNA site	Resolution (Å)	Observations	Unique reflections	Completeness (%)	R_{sym}[1]	Unit cell lengths (Å)		
							a	b	c

Native data sets

Peptide	DNA site	Resolution (Å)	Observations	Unique reflections	Completeness (%)	R_{sym}[1]	a	b	c
DSNR	GACC	20-2.1	19,390	8,185	87.0 (76.0)[2]	6.4 (9.3)[3]	43.0	55.9	128.4
DSNR	GCGT	20-1.9	36,356	11,768	91.8 (53.5)	3.4 (17.0)	43.3	55.8	129.9
QGSR	GCAC	20-1.6	80,360	20,835	95.7 (87.8)	3.5 (14.7)	44.1	55.9	130.5
RADR	GCAC	20-1.6	45,579	21,212	96.4 (89.9)	4.9 (10.9)	43.2	56.3	133.8
RADR	GCGT	20-2.0	29,842	9,959	89.6 (82.9)	3.4 (5.4)	42.3	55.9	133.6
RADR	GACC	20-1.9	23,142	10,102	79.9 (63.8)	3.4 (25.8)	42.7	55.5	130.4
RDER	GCAC	20-2.3	28,164	6,831	91.2 (65.2)	5.5 (12.7)	43.3	55.9	132.5

Derivative data sets[4]

Peptide	DNA site	Resolution (Å)	Observations	Unique reflections	Completeness (%)	R_{sym}[1]	a	b	c
QGSR	GCAC								
	I-dU 5,12'	20-2.3	23,358	7,061	96.7 (90.4)	6.9 (23.9)	43.8	55.7	128.0
	I-dU 5	20-2.3	19,051	6,828	91.7 (62.1)	4.6 (6.8)	43.9	55.9	130.0
	I-dU 12'	20-2.3	22,898	6,325	86.0 (53.5)	4.8 (11.8)	43.5	55.9	129.6

[1] $R_{sym} = \sum_{i,(h,k,l)} \left| I_{i,(h,k,l)} - \langle I_{(h,k,l)} \rangle \right| / \sum_{i,(h,k,l)} \langle I_{(h,k,l)} \rangle$ where $\langle I_{(h,k,l)} \rangle$ is the statistically weighted average intensity of symmetry equivalent reflections.

[2,3] Statistics in parenthesis indicate the completeness and R_{sym} for the highest resolution shell.

[4] Positions of iodouracil indicated with numbering scheme of Figure 1c.

Table 2. Phasing statistics for QGSR multiple isomorphous replacement (20-2.3 Å)

Positions of iodines in GCAC site	R cullis centrics[1]	R cullis anomalous[2]	Phasing power[3]	
			acentrics	centrics
I-dU 5,12'	0.78	0.91	1.24	1.00
I-dU 5	0.76	0.87	1.22	0.95
I-dU 12'	0.81	0.96	0.77	0.73

[1]R cullis $= \Sigma_{(h,k,l)}\left\|F_{PH}\right|-\left|F_P+F_H\right\|/\Sigma_{(h,k,l)}\left|F_{PH}-F_P\right|$

[2]R cullis anom $= \Sigma_{(h,k,l)}\left\|F_{PH}^+-F_{PH}^-\right|_{obsd}-\left|F_{PH}^+-F_{PH}^-\right|_{calc}\right|/\Sigma_{(h,k,l)}\left|F_{PH}^+-F_{PH}^-\right|_{obsd}$

[3]Phasing power $= \left\langle\left|F_H\right|/\left\|F_{PH}\right|-\left|F_P+F_H\right\|\right\rangle$

Table 3. Refinement statistics

Peptide (Residues modeled)	DNA site	Resolution (Å)	Reflections used in refinement	Nonhydrogen atoms in complex	Waters	R_{cryst}[1] (%)	R_{free}[2] (%)	Rmsd bond lengths (Å)		Rmsd bond angles (deg)	
								Protein	DNA	Protein	DNA
DSNR (3-86)	GACC	20-2.1	8,173	1,146	88	23.0	27.0	0.006	0.003	1.3	0.67
DSNR (3-86)	GCGT	20-1.9	11,737	1,143	155	22.2	27.5	0.010	0.004	1.5	0.84
QGSR (3-87)	GCAC	20-1.6	20,810	1,162	149	23.9	27.7	0.010	0.004	1.5	0.87
RADR (3-87)	GCAC	20-1.6	21,186	1,160	255	19.4	21.6	0.008	0.004	1.3	1.1
RADR (3-86)	GCGT	20-2.0	9,959	1,160	153	21.0	26.2	0.008	0.003	1.3	0.84
RADR (3-87)	GACC	20-1.9	9,890	1,156	121	21.0	25.9	0.008	0.006	1.3	1.1
RDER (3-87)	GCAC	20-2.3	6,805	1,167	101	20.4	26.7	0.007	0.003	1.2	0.82

[1]$R_{cryst} = (\Sigma_{(h,k,l)} \| F_o | - | F_c \| / \Sigma_{(h,k,l)} | F_o |)$

[2]$R_{free} = (\Sigma_{(h,k,l) \in T} \| F_o | - | F_c \| / \Sigma_{(h,k,l) \in T} | F_o |)$ where T is the test set, the 10% of the observations omitted during refinement. For each of the seven complexes, the set of test reflections in each resolution shell was the same as that used in the refinement of the wild type Zif 268 complex [11] as suggested in [30].

Figure 1 **(a)** Schematic view of characteristic interactions between a Zif268-like zinc finger and its DNA subsite (not all contacts are made by every finger, and additional atypical contacts can occur). Residue positions in the finger are numbered with respect to the start of the α helix. Arrows indicate the most common pattern of side chain-base interactions, and the core 3 bp subsite is shaded in gray. **(b)** Sequences of the zinc finger peptides. Each of the Zif268 variants contains three zinc fingers and 90 amino acids, but these peptides differ at the four positions of finger one (indicated with open circles) that were randomized by Rebar and Pabo [1]. These residues correspond to positions -1, 2, 3, and 6 of the α helix, and residues selected at these positions are shown at the top of the figure. The three fingers are aligned to highlight conserved residues and secondary structure elements. The helix is indicated by a cylinder and the β strands by arrows. The conserved cysteine and histidine residues that are ligands for the zinc ions are highlighted in bold. (Adapted from [11].) **(c)** Sequences of the DNA binding sites used for cocrystallization, which contain 10 bp of duplex DNA and have a single overhanging nucleotide at each 5' end. The subsites under finger one are shown in boldface.

Figure 1

(a)

Primary strand
5'
6
3
2
-1
Secondary strand
5'

(b)

	-1	2 3	6	
Variant sequences	D	S N	R	Selected with the GACC site
	Q	G S	R	Selected with the GCAC site
	R	A D	R	Selected with the GCAC site
Wild type sequence	R	D E	R	

```
         1        5          10          15          20          25        30
         M E R P Y A C P V E S C D R R F S O S O O L T O H I R I H T G Q K
```

```
              35          40          45          50          55          60
         P F Q C R I - - C M R N F S R S D H L T T H I R T H T G E K
```

```
              65          70          75          80          85          90
         P F A C D I - - C G R K F A R S D E R K R H T K I H L R Q K D
```

(c)

GCGT site

```
 1   2 3 4   5 6 7   8 9 10  11
A  G C G   T G G   G C G   T
   C G C   A C C   C G C   A T
   2'  3' 4'  5' 6' 7'  8' 9' 10'  11'12'
```

GACC site

```
 1   2 3 4   5 6 7   8 9 10  11
A  G C G   T G G   G A C   C
   C G C   A C C   C T G   G T
   2'  3' 4'  5' 6' 7'  8' 9' 10'  11'12'
```

GCAC site

```
 1   2 3 4   5 6 7   8 9 10  11
A  G C G   T G G   G C A   C
   C G C   A C C   C G T   G T
   2'  3' 4'  5' 6' 7'  8' 9' 10'  11'12'
```
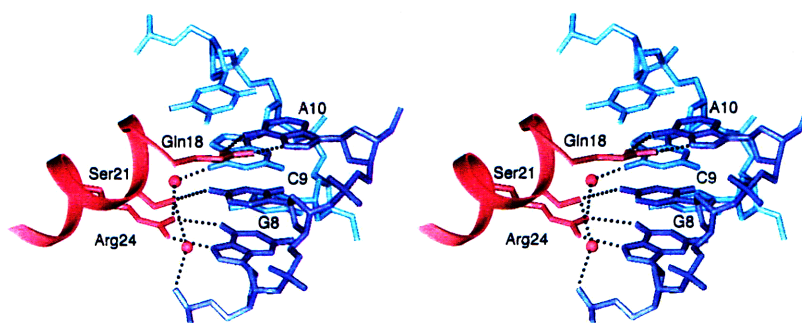
Figure 2 Stereo view of the contacts made by finger one in the complex between **(a)** the DSNR peptide and the targeted GACC binding site and **(b)** the DSNR peptide and the wild type GCGT binding site. The side chains of residues 18, 20, 21, and 24 (positions -1, 2, 3, and 6 of the α helix) and the peptide backbone are shown in green. Water molecules are represented as green spheres; only those waters that mediate interactions between the peptide and base pairs 8-11 are shown. The DNA is color-coded by strand, with the primary strand (as denoted in Fig. 1a) in purple and the secondary strand in blue. Fingers two and three are not shown since the structure of this region is very similar in all complexes. This figure was made with Setor [29].

Figure 2

a.
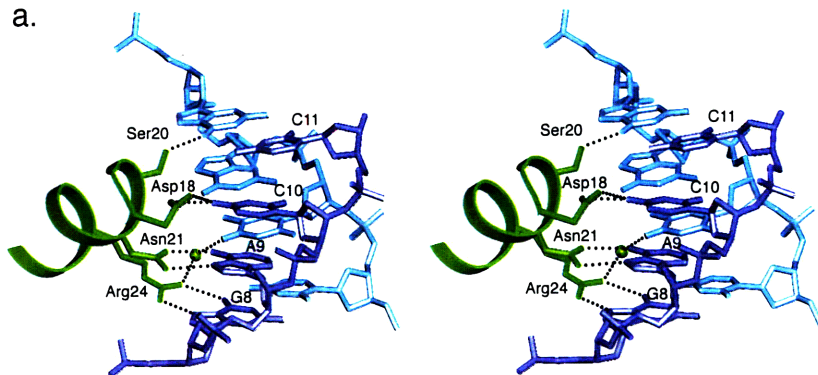
b.
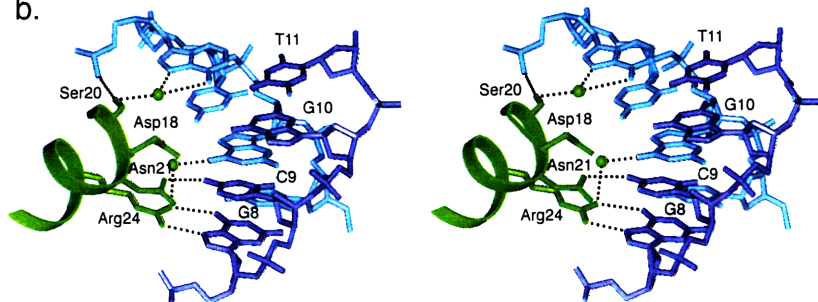
Figure 3 Stereo view of the contacts made by finger one in the complex between the QGSR peptide and the targeted GCAC binding site. The side chains of residues 18, 21, and 24 (positions -1, 3, and 6 of the α helix) and the peptide backbone are shown in salmon. Water molecules are represented as salmon spheres; only those waters that mediate interactions between the peptide and base pairs 8-10 are shown. The DNA is color-coded by strand, with the primary strand in purple and the secondary strand in blue. As before, fingers two and three are not shown. This figure was made with Setor [29].
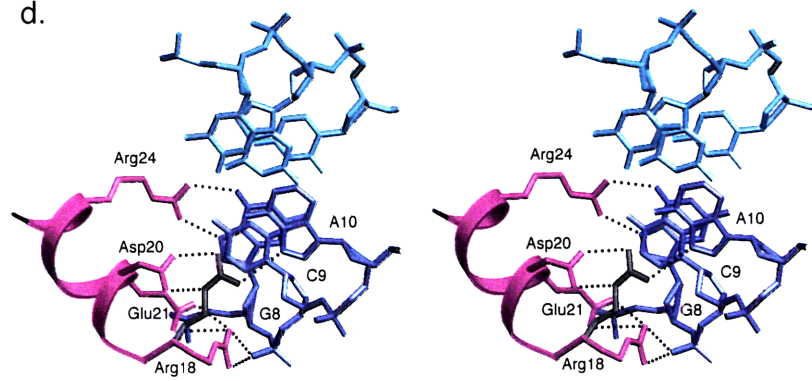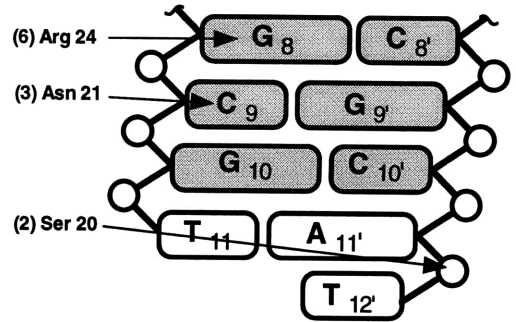
Figure 3

Figure 4 Stereo view of the contacts made by finger one in the complex between **(a)** the RADR peptide and the targeted GCAC binding site, **(b)** the RADR peptide and the wild type GCGT binding site, **(c)** the RADR peptide and the GACC binding site, and **(d)** the wild type Zif268 peptide and the GCAC binding site. The side chains of residues 18, 20, 21, and 24 (positions -1, 2, 3, and 6 of the α helix) and the peptide backbone are shown in gold for the RADR peptide, with alternate conformations indicated in gray. The wild type peptide is shown in pink, with alternate conformations indicated in gray. Water molecules are represented as spheres; only those waters that mediate interactions between the peptide and base pairs 8-10 are shown. The DNA is color-coded by strand, with the primary strand in purple and the secondary strand in blue. As before, fingers two and three are not shown. This figure was made with Setor [29].

Figure 4
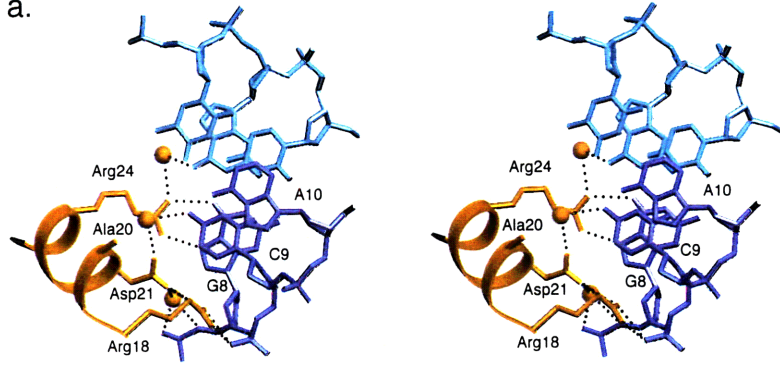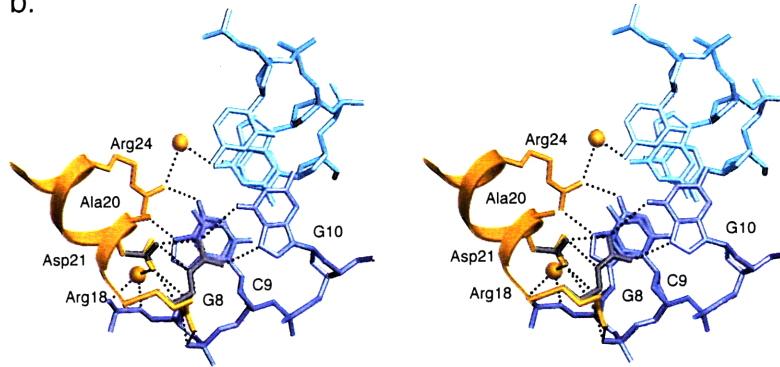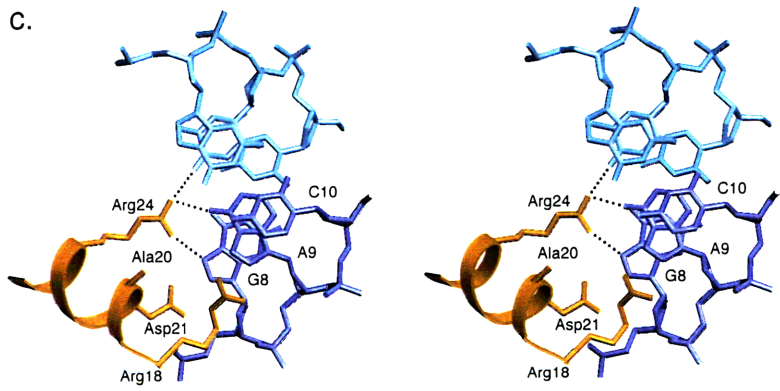
Figure 5 Summary of direct base and phosphate contacts made by the side chains at positions -1, 2, 3, and 6 of the helix in finger one. Only the base pairs contacted by these residues are represented; base pairs 1-7, which are contacted by fingers two and three, are not shown in this figure. Arrows indicate hydrogen bonds; the dotted arrow represents a hydrogen bond with marginal geometry. Lines ending in filled circles represent van der Waals interactions.
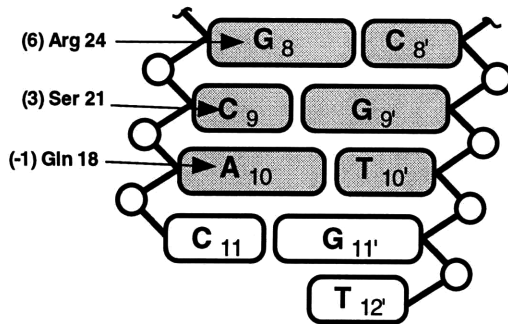
Figure 5

a. DSNR with the targeted <u>GA</u>CC site
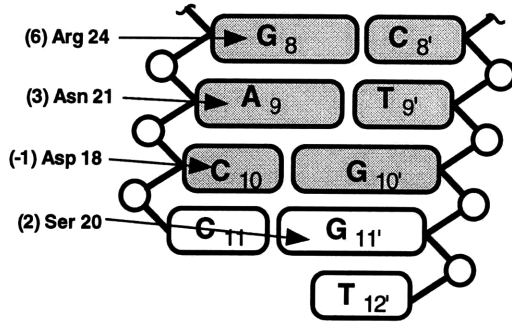
b. DSNR with the wild type <u>GCG</u>T site

c. QGSR with the targeted <u>GCA</u>C site

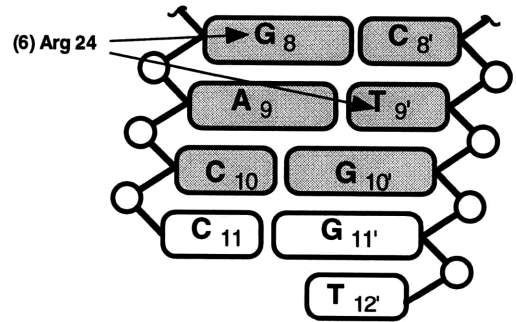d. RADR with the targeted <u>GCA</u>C site
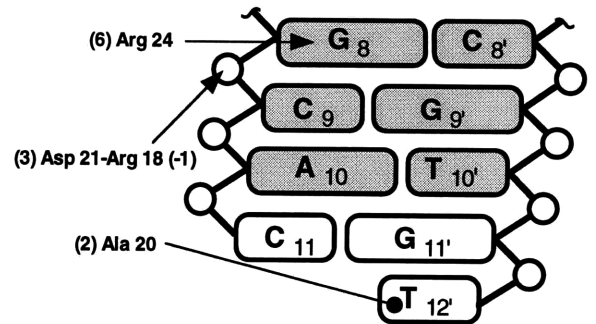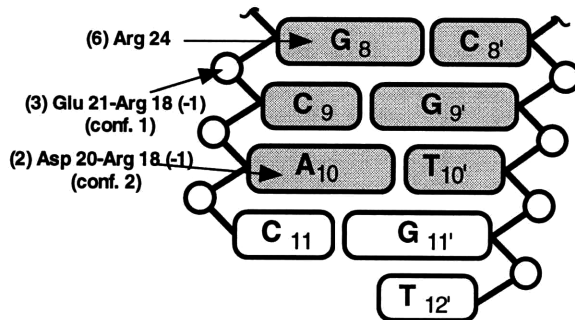

e. RADR with the wild type <u>GCG</u>T site
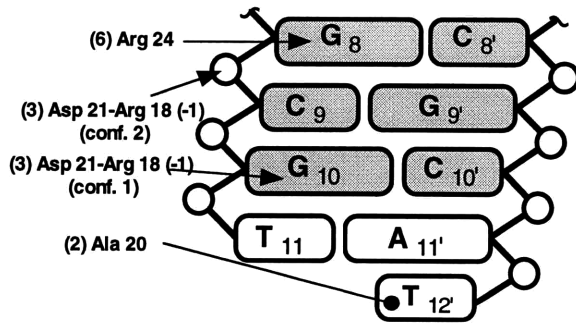
f. RADR with the <u>GA</u>CC site

g. Zif268 (RDER) with the <u>GCA</u>C site

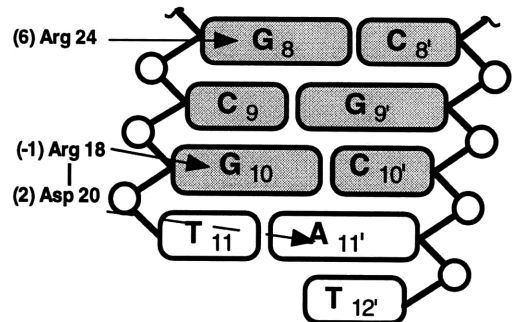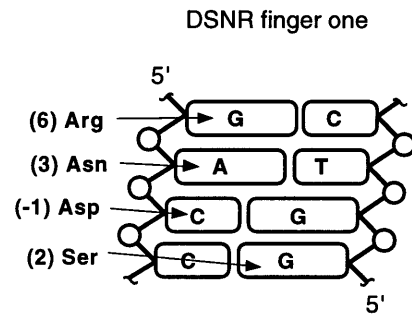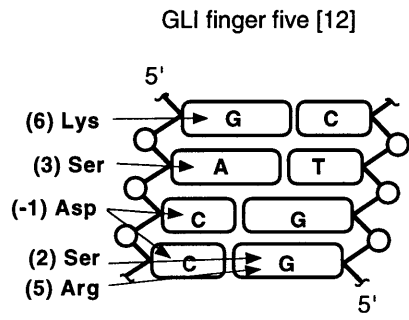h. Zif268 (RDER) with the wild type <u>GCG</u>T site [10,11]

Figure 6 (a) Comparison of the contacts made at matching GACC subsites by finger five of GLI [12] and by finger one of the DSNR peptide. **(b)** Comparison of the contacts made at a GCAG subsite by finger two of the designed peptide [9] and at a closely related GCAC subsite by finger one of the QGSR peptide.
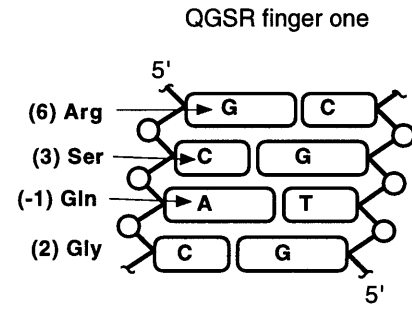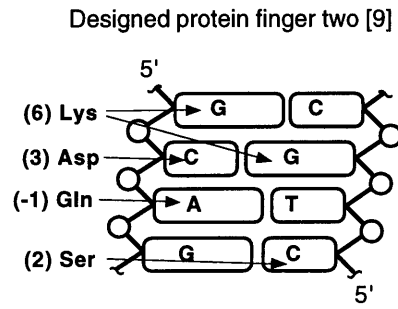
Figure 6

a.

GLI finger five [12]

(6) Lys — G C
(3) Ser — A T
(-1) Asp — C G
(2) Ser — C G
(5) Arg

DSNR finger one

(6) Arg — G C
(3) Asn — A T
(-1) Asp — C G
(2) Ser — C G

b.

Designed protein finger two [9]

(6) Lys — G C
(3) Asp — C G
(-1) Gln — A T
(2) Ser — G C

QGSR finger one

(6) Arg — G C
(3) Ser — C G
(-1) Gln — A T
(2) Gly — C G

Figure 7 Overview of three variant complexes superimposed on the wild type Zif268 structure (by aligning the α carbons of fingers two and three and the phosphorus and C1' atoms of their corresponding GCG/TGG subsites). Panels show **(a)** the QGSR peptide with the targeted GCAC site; **(b)** the RADR peptide with the targeted GCAC site; and **(c)** the DSNR peptide with the targeted GACC site. This figure was made with Setor [29].
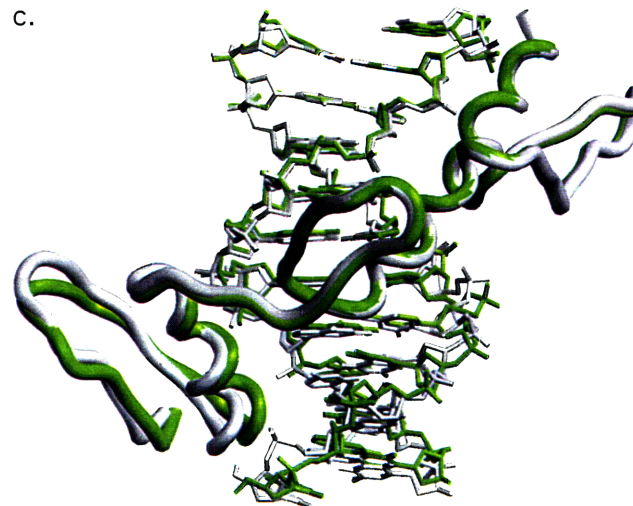
Figure 7

a.
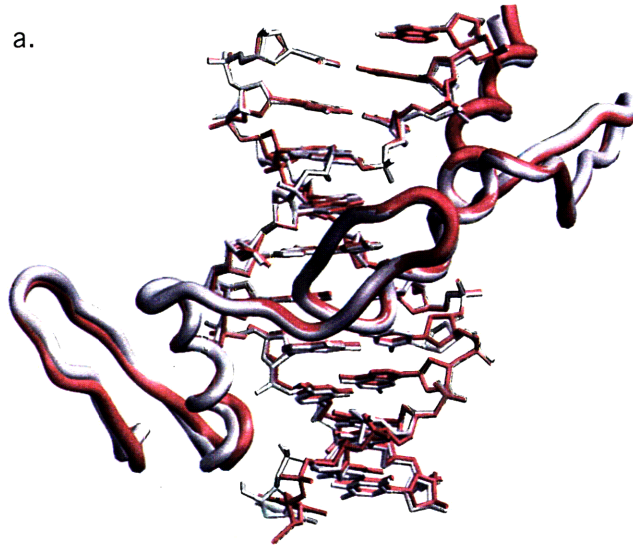
b.
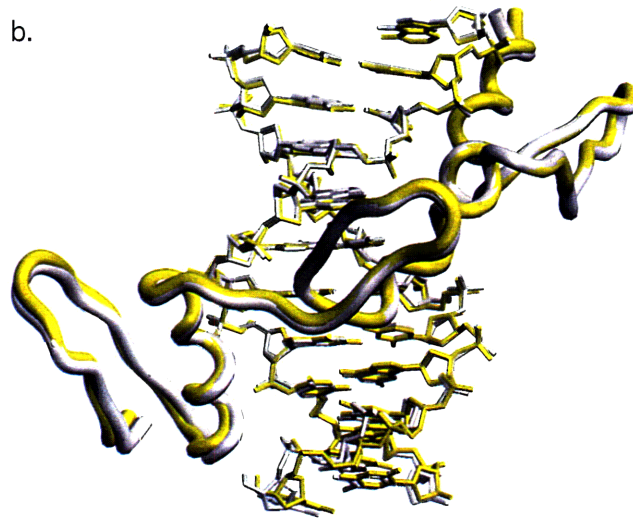
c.

Chapter Four

Crystal Packing Induces A-DNA Conformation in Zif268 Binding Site

Abstract

Previous crystallographic analysis of the Zif268-DNA complex revealed that the Zif DNA adopts a B-type conformation with an unusually deep major groove. In an attempt to determine whether any features of this distinctive conformation are intrinsic properties of the free DNA, we crystallized the Zif DNA site and also used circular dichroism to characterize its solution conformation. Circular dichroism shows that the Zif DNA adopts a B-type structure in solution, but the crystal structure reveals canonical A-DNA. Our data show that the process of crystallization is responsible for the transition of the Zif DNA to A-form, providing a dramatic demonstration of the effect that crystal packing forces can have on DNA structure. Analysis of the crystal structure reveals extensive packing contacts, involving 1000 Å^2 of buried surface area per DNA duplex. Evidently the energy derived from these extensive crystal packing contacts, which predominantly involve hydrophobic groups, drives the B-to-A transition.

Introduction

Structural studies of protein-DNA complexes suggest that sequence-dependent features of DNA structure - such as local changes in the helical parameters - may play a significant role in protein-DNA recognition (reviewed in ref. 1). This possibility has received considerable attention in the case of zinc finger-DNA interactions (2-4), and comparing the Zif268, GLI, and tramtrack zinc finger-DNA complexes suggested that these proteins may recognize a variant form of B-DNA that has an unusually deep major groove (5). To determine whether any aspects of the distinctive DNA conformation observed in the Zif268-DNA complex are inherent characteristics of the free DNA, we have crystallized and solved the structure of the Zif268 binding site. Surprisingly, the DNA has a dramatically different structure in this crystal than it does in the protein-DNA complex: the crystallized DNA is nearly canonical A-form. However, circular dichroism (CD) data shows that the Zif DNA has a B-type structure in solution, indicating that crystallization of the free DNA induces a B-DNA to A-DNA transition. Analysis of the DNA crystals suggests that the loss of solvent accessible surface area, via crystal packing contacts, may provide the energy which drives this transformation.

Results

We have crystallized and determined the structure of the Zif268 binding site at 2.6 Å resolution. Our DNA duplex has the same sequence as the binding site used for cocrystallization of the zinc finger-DNA complex

(AGCGTGGGCGT/TACGCCCACGC; ref. 6-7). However, the DNA structures are radically different: the 10 bp duplex portion in the DNA crystals is nearly canonical A-form DNA (Fig. 1; Table 1). The overhanging base, which had been added to facilitate crystallization of the zinc finger-DNA complex, is unpaired and mobile in the DNA crystals. Zif DNA crystallizes in the same space group with nearly identical cell dimensions as two previously-solved A-form decamer DNAs in the Nucleic Acid Database (8). These DNA oligos have the sequences ACCGGCCGGT (9) and GCGGGCCCGC (13). The structures of all three DNA duplexes are essentially isomorphous: any two of the duplexes superimpose with an rms under 0.6 Å (using P atoms, C1' atoms, or P and C1' atoms). All three of these crystal structures have extensive - and essentially identical - crystal packing contacts, with the exposed bases on the ends of the duplex packing against the minor groove of symmetry-related molecules.

How does the A-form structure that we observe in the Zif DNA crystals compare with the structure of the DNA in solution and with the structure in the protein-DNA complex? Analysis of the DNA conformation in the Zif268 cocrystals (6-7) revealed that this DNA is a variant form of B-DNA with an unusually deep major groove, which has been called $B_{enlarged\ groove}$-DNA (5). The crystal structure for the free DNA is dramatically different, since in this case the DNA adopts a classical A-form structure. Differences in the crystal structures of the free and complexed DNAs are readily apparent in Figure 1, and least-squares superposition of these structures (Table 2) reveals an rms distance of 3.4 Å. Comparison of the helical parameters on a base-by-base level does not reveal any similarities or common structural patterns, and modeling confirms that the zinc fingers could not dock against the major groove of the A-form

DNA (the major groove is too narrow to accommodate the recognition helix).

Circular dichroism studies were used to obtain information about the solution structure of the Zif DNA. Figure 2 shows CD spectra for the Zif DNA; spectra for calf thymus DNA and for the Dickerson-Drew dodecamer (14), which serve as B-form reference spectra; and the spectrum of "Zif RNA" (an RNA duplex which has the same sequence as the Zif DNA site and which provides an A-form reference spectrum). Comparison of these spectra indicates that the Zif DNA does not adopt an A-form structure in solution. At this point we cannot determine whether the solution structure of the Zif binding site is like canonical B-DNA or whether it has some features of the $B_{enlarged\ groove}$ conformation seen in the crystal structure of the complex. [CD studies have shown that the DNA undergoes a conformational change upon protein binding, indicating that the conformation of the free DNA is not exactly like that seen in the complex (7).]

Since the DNA does not adopt an A-form structure in solution and the spectrum of the Zif DNA does not change when measured in the crystallization buffer (Fig. 2), we conclude that the process of crystallization itself must be responsible for the transition to A-form. Indeed, the packing of a duplex terminus in the minor groove requires that the phosphate backbone be displaced to an A-like conformation. This gives a sufficiently wide minor groove for the end of the duplex to form van der Waals contacts with the sugars and bases and to minimize the solvent accessible surface of the hydrophobic face of the terminal base pair. Analysis of the Zif DNA crystals using the algorithm of Lee and Richards (15) shows that 1000. Å^2 of solvent accessible surface area per duplex

(26% of the total) is buried upon crystallization (Fig. 3). This approaches the amount of surface area buried when the Zif268 complex forms (1346 $\mathring{A}^2$ for the DNA) or the amount buried in many protein-protein interactions (typically 1200-2000 $\mathring{A}^2$, ref. 16). With the Zif DNA, 80% of the buried surface results from sugar and base desolvation, and burying these uncharged surfaces may provide a significant energetic contribution to crystallization. The large surface area involved in these intermolecular, crystal packing contacts may drive the transition to A-DNA.

Discussion

The interpretation of crystallographic studies is usually based on the assumption that the crystal structure will accurately reflect the solution structure, and a large body of data indicates that this is typically true for protein structures. Thus, comparing different crystal forms of the same protein or comparing results from a crystallographic study with those from an NMR study usually reveals only minor differences in the structures. The situation with DNA is less clear, and there still is active debate in the literature about the extent to which crystal packing forces affect DNA conformation. However, several reports indicate that crystal packing can influence DNA structure (e.g., 13, 17-21), in some cases resulting in a fundamentally different structure in the crystal than in solution (this study and refs. 22-25).

Crystallographic studies of DNA structure can clearly be complicated by the existence of multiple minima on the potential energy surface that describes DNA structure and by the confounding effects of crystal packing forces. Figure 4 shows a highly simplified, abstract

representation of a potential energy surface and illustrates how crystal packing forces can lead to dramatic changes in three dimensional structure. Global structural changes (such as the B-DNA to A-DNA transition indicated in the figure) can occur whenever crystal packing forces are larger than the energy differences between the wells that characterize these distinct conformations. [Note: Crystal packing forces may also induce local structural changes, and these could be represented with a more detailed diagram of the energy landscape. In this case, rather than using a smooth harmonic well to represent B-DNA, one would have jagged walls to indicate local minima that could become accessible via crystal packing forces.]

In DNA crystallography, packing forces can be significant because of a relatively large surface-to-volume ratio (especially for short DNA duplexes) and because of the end effects inherent in studying isolated DNA fragments. A short DNA oligo is (artificially) free from any conformational constraints that might be imposed by flanking DNA in the biologically relevant context. In addition, a short DNA oligo has exposed hydrophobic ends that are free to participate in adventitious crystal packing contacts. We presume that these features explain our results, and they emphasize the caution that must be used in analyzing crystal structures of DNA duplexes.

Depending on the crystal packing contacts and on the shape of the energy landscape (Fig. 4) for a particular DNA duplex, crystal packing forces will either have a minimal effect on the DNA structure, introduce local structural changes, or introduce global structural changes. Thinking carefully about this energy landscape reminds us that there is no inherent contradiction between the observation that crystallization can sometimes induce dramatic global changes (as in this study and as reported in

references 22-25) and the argument (26) that it may, in other instances, give meaningful information about local structural preferences. The real problem in interpreting a DNA crystal structure is the question of how we possibly can know - in the absence of an a priori knowledge of the structure/energy relationships - which situation pertains and whether crystallization has changed the structure. In general, we expect that perturbations induced by crystal packing may be smaller when the DNA duplexes are stacked end-to-end, since in this case the stacking forces that stabilize the crystal mimic the environment in duplex DNA. However, additional data may be needed to determine whether there are significant distortions in any particular duplex. It may be necessary to determine the structures of several closely related sequences, to compare the structures in several different crystal packing environments, or to obtain some information about the solution structure of the DNA.

Materials and Methods

Crystallization and Structure Determination

DNA oligos AGCGTGGGCGT and TACGCCCACGC were synthesized, purified and annealed as in the study of the Zif268 cocrystals (6-7). Crystals were grown by hanging drop vapor diffusion over a reservoir of 23% (v/v) 2-methyl-2,4-pentanediol, 10 mM cobalt hexamine, and 25 mM bis-tris-propane (BTP) pH 7.0. Diffraction data were collected to 2.6 Å on an R-Axis image plate system and reduced with DENZO/SCALEPACK (Z. Otwinowski). Merging 3642 observations from 20-2.6 Å gave 1015 unique reflections, with an R_{merge} on intensity of 3.2% (all data less 6 rejected observations). The space group appeared to be $P6_122$ or its enantiomorph, with cell dimensions a=b=39.2Å, c=78.0Å. A search of the Nucleic Acid Database (8) revealed that the same space group and cell dimensions had been reported for a previously solved A-form decamer DNA structure (access number ADJ022, ref. 9). Using that model and simply changing the base sequence to match that of the duplex region of Zif DNA gave an R-factor, before any refinement, of 37.9% for all data from 20-2.6 Å. However, since the asymmetric unit of this crystal contains only one strand, with the other strand related by crystallographic symmetry, this implies that the Zif DNA packs in the crystal in two directions: in some asymmetric units the sequence reads 5'-GCGTGGGCGT and in the other asymmetric units the sequence reads 5'-ACGCCCACGC. The resulting electron density is thus a superposition of the DNA rotationally averaged in these two arrangements. This "directional scrambling" of the DNA has been observed in other DNA crystal structures (10), and this was confirmed in our case via difference Fouriers with an iodouracil

derivative. Since there are insufficient constraints to refine both directions simultaneously (as indicated by monitoring the free R), the Zif DNA model used for refinement is a "consensus" self-complementary strand, ACGCGCGCGT. The overhanging base is disordered and was not included in our model. The crystallographic R value for this consensus model (202 atoms, with no solvent molecules) refined with XPLOR (11) is 25.0% (all data from 6 to 2.6 Å), with a free R of 24.5% and deviations from ideal bond lengths, angles and improper dihedral angles of 0.010 Å, 1.0° and 1.0° (using XPLOR stereochemical dictionary PARNDBX.DNA; ref. 12). Tightly-restrained individual B-factors were refined, with a final RMS difference in B-factor between bonded atoms of 1.0 $Å^2$.

Circular Dichroism

CD spectra were taken at 25°C on an Aviv 60DS spectropolarimeter with 1.5 nm bandwidth and 1 sec averaging time, and were sampled every 1 nm from 320 to 220 nm. DNA samples were at 0.5 mg/ml in a 0.1 cm path length cuvette; RNA was at 0.05 mg/ml in a 1.0 cm cuvette. Each spectrum is the average of two baseline-corrected scans, smoothed in 5 nm windows. Due to strong absorption by cobalt hexamine in the 220-236 range, the spectrum of the DNA in the crystallization buffer is only shown from 238-320 nm.

Acknowledgements

References

1. Travers, A.A. (1992) *Curr. Opin. Struct. Biol.*, **2**, 71-77.

2. McCall, M., Brown, T., Hunter, W.N. and Kennard, O. (1986) *Nature*, **322**, 661-664.

3. Fairall, L., Martin, S. and Rhodes, D. (1989) *EMBO J.*, **8**, 1809-1817.

4. Huber, P.W., Morii, T., Mei, H.-Y. and Barton, J.K. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10801-10805.

5. Nekludova, L. and Pabo, C.O. (1994) *Proc. Natl. Acad. Sci USA,* **91**, 6948-6952.

6. Pavletich, N.P. and Pabo, C.O. (1991) *Science,* **252**, 809-817.

7. Elrod-Erickson, M., Rould, M.A., Nekludova, L. and Pabo, C.O. (1996) *Structure*, **4**, 1171-1180.

8. Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S., Srinivasan, A.R. and Schneider, B. (1992) *Biophys. J.*, **63**, 751-759.

9. Frederick, C.A., Quigley, G.J., Teng, M.K., Coll, M., Van der Marel, G.A., Van Boom, J.H., Rich, A. and Wang, A.H. (1989) *Eur. J. Biochem.*, **181**, 295-307.

10. DiGabriele, A.D. and Steitz, T.A. (1993) *J. Mol Biol,* **231**, 1024-1039.

11. Brunger, A.T. (1992) *Xplor Manual Version 3.1.* Yale University Press, New Haven, CT.

12. Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. and Berman, H.M. (1996) *Acta Cryst. D,* **52**, 57-64.

13. Ramakrishnan, B. and Sundaralingam, M. (1993) *Biochemistry,* **32**, 11458-11468.

14. Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.E. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 2179-2183.

15. Lee, B. and Richards, F.M. (1971) *J. Mol. Biol.,* **55**, 379-400.

16. Janin, J. (1995) *Proteins,* **21**, 30-39.

17. Jain, S. and Sundaralingam, M. (1989) *J. Biol. Chem.,* **264**, 12780-12784.

18. Jain, S., Zon, G. and Sundaralingam, M. (1991) *Biochemistry,* **30**, 3567-3576.

19. Lipanov, A., Kopka, M.L., Kaczor-Grzeskowiak, M., Quintana, J. and Dickerson, R.E. (1993) *Biochemistry,* **32**, 1373-1389.

20. Ramakrishnan, B. and Sundaralingam, M. (1993) *J. Mol. Biol.,* **231**, 431-444.

21. Shakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolow, F., V.P. and Rabinovich, D. (1989) *Nature,* **342**, 456-460.

22. Wang, Y., Thomas, G.A. and Peticolas, W.L. (1987) *Biochemistry,* **26**, 5178-5186.

23. Wang, Y., Thomas, G.A. and Peticolas, W.L. (1987) *J. Biomolecular Structure and Dynamics,* **5**, 249-274.

24. Clark, G.R., Brown, D.G., Sanderson, M.R., Chwalinski, T., Neidle, S.N., Veal, J.M., Jones, R.L., Wilson, W.D., Zon, G., Garman, E. and Stuart, D.I. (1990) *Nucleic Acids Res.,* **18**, 5521-5528.

25. Reid, D.G., Salisbury, S.A., Bellard, S., Shakked, Z. and Williams, D.H. (1983) *Biochemistry,* **22**, 2019-2025.

26. Dickerson, R.E., Goodsell, D.S. and Neidle, S. (1994) *Proc. Natl. Acad. Sci. USA,* **91**, 3579-3583.

Table 1. Average helical parameters[a].

|	Rise(Å)	Twist(°)	Xdisp(Å)	Inclin(°)
Ideal B-form DNA	3.38	36.0	+0.23	-5.7
Complexed Zif DNA	3.25	32.1	-1.68	+8.6
Unbound Zif DNA	2.58	33.8	-4.16	+17.7
Ideal A-form DNA	2.56	32.7	-4.49	+20.0

a Parameters for the Zif DNA crystallized alone and in the Zif268 zinc finger-DNA complex are shown, along with reference values for ideal A- and B-form DNA as represented in INSIGHT (Biosym Technologies). Values for Zif DNA were calculated with NEWHEL93 (R. Dickerson), using a helix axis defined by the C1' atoms of the duplex decamers. This table shows the mean rise per residue, the helical twist per residue, the radial displacement of the base pairs from the helix axis, and the inclination of the base pairs relative to this axis. Examining other parameters for the unbound Zif DNA also confirms that it is essentially ideal A-form DNA (the riboses are in the C3' endo configuration, the minor groove width is 9.7 Å, and the major groove width is 2.7 Å).

Table 2. Residual distances in Å from least-squares superpositions of the C1' and P atoms of the unbound Zif DNA, of the DNA in the Zif268 zinc finger-DNA complex (6), and of ideal A-DNA and B-DNA.

	Ideal A-DNA	Complexed Zif DNA	Ideal B-DNA
Unbound Zif DNA	1.03	3.42	5.32
Ideal A-DNA	0	3.66	5.59

Figure 1. Views down the helix axes and stereo sideviews of **a)** canonical A-DNA; **b)** the identical DNA duplex as it appears when crystallized alone; **c)** the DNA as it appears in the complex (6) with Zif268 zinc finger protein; and **d)** canonical B-DNA. All DNA duplexes have the same length and sequence and are drawn to the same scale. Canonical A-DNA and B-DNA models were prepared using INSIGHT (Biosym Technologies).
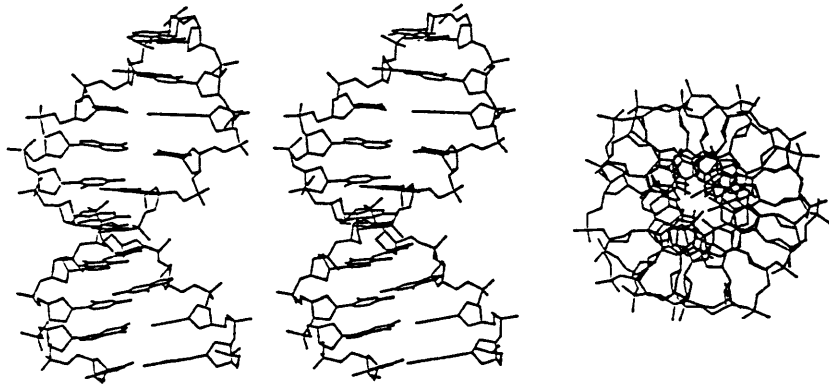
A
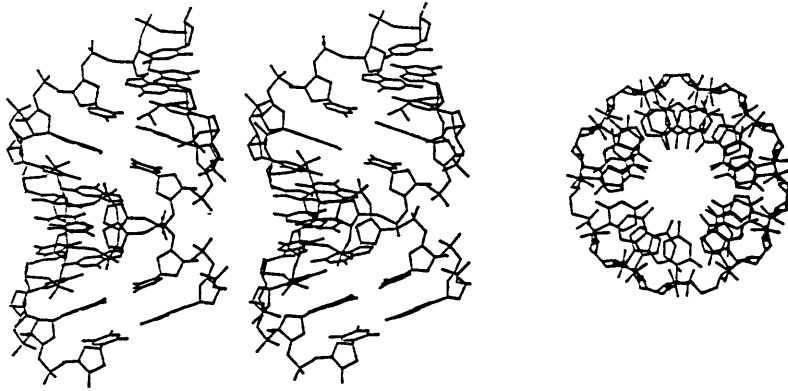
B
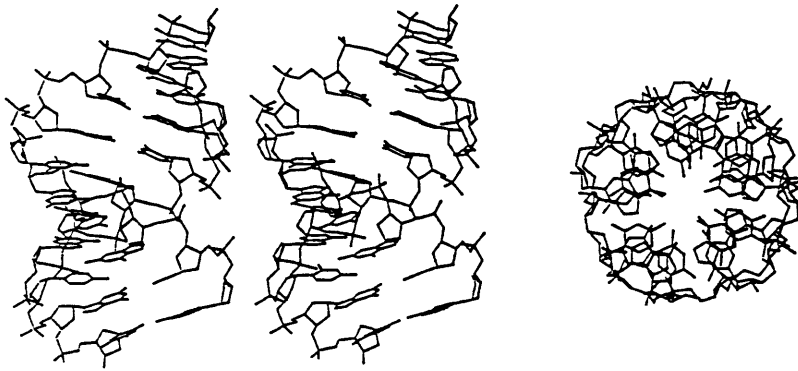
C

D
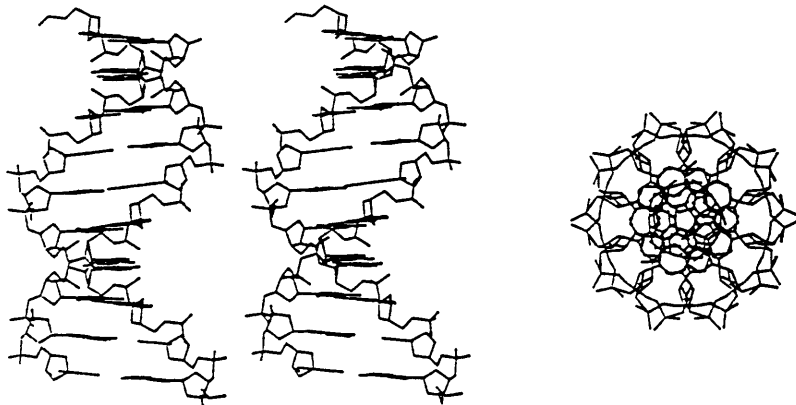
Figure 2. Circular dichroism spectra of the Zif DNA, calf thymus DNA, the Dickerson-Drew dodecamer (14), and Zif RNA, each in 25 mM BTP pH 7, and of Zif DNA in the crystallization buffer, plotted as the molar circular dichroic extinction coefficient per nucleotide ($\Delta\varepsilon$, in $M^{-1}cm^{-1}$).

Legend:
- Zif DNA
- Dickerson-Drew dodecamer
- calf thymus DNA
- Zif RNA
- Zif DNA in crystallization buffer

Axes: Δε (vertical, from -6 to 6), wavelength (nm) (horizontal, from 220 to 320)

Figure 3. Crystal packing environment of the Zif DNA duplex. The coloring scheme is intended to highlight the extensive contacts that each DNA duplex makes w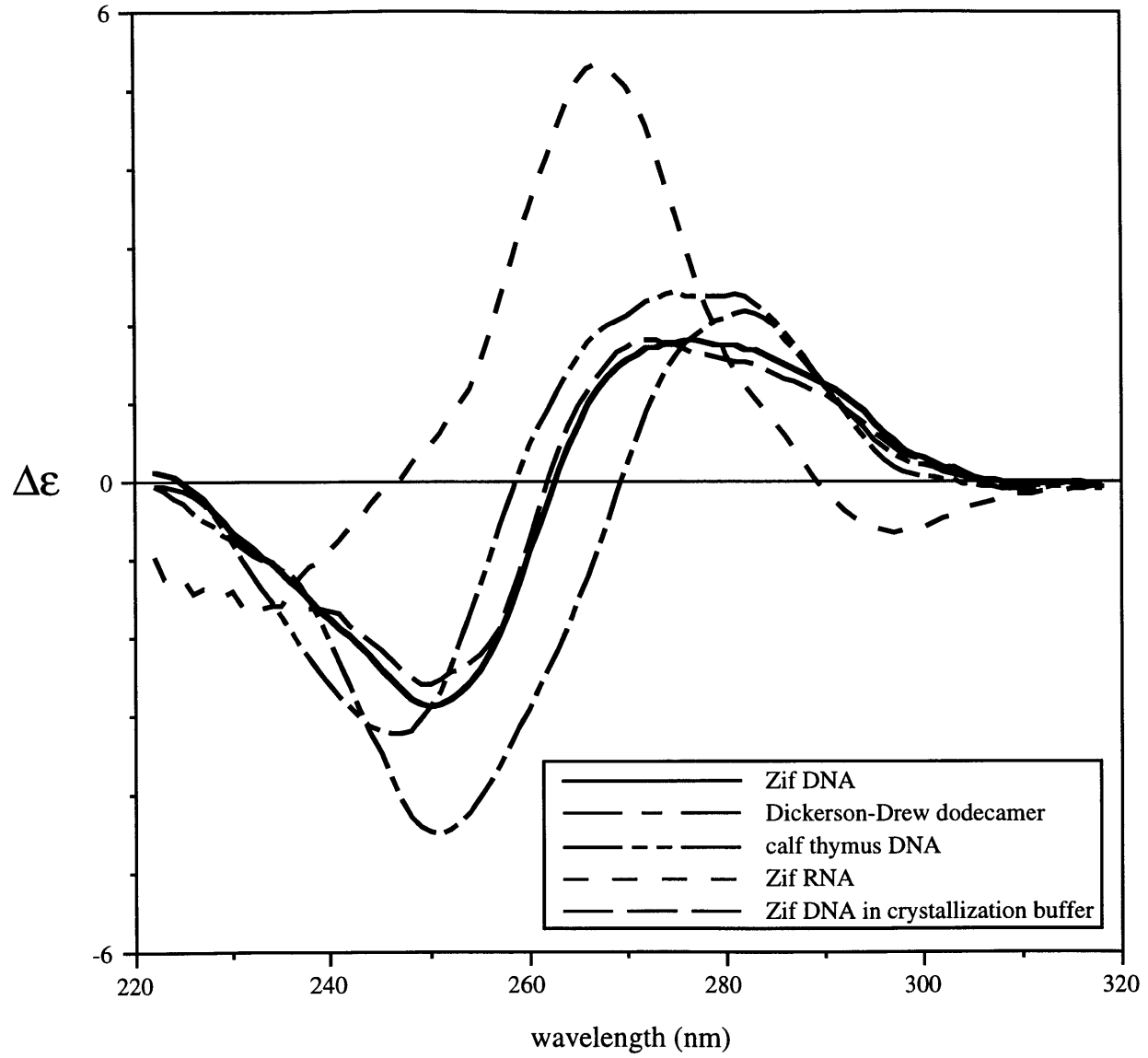ith the four surrounding symmetry-related oligos (shown in blue). Contacts involve the terminal base pairs of each DNA duplex and a region in the minor groove near the center of each site, and these contacts bury 1000 Å^2 of surface area per DNA duplex.
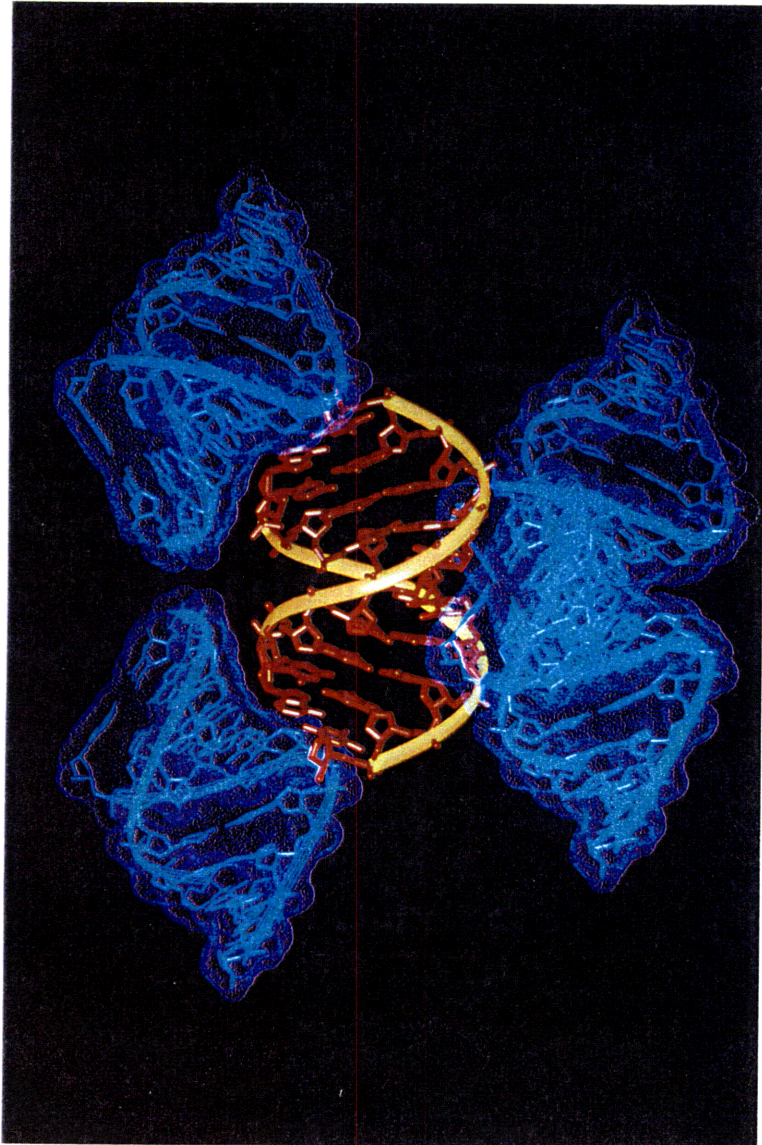
Figure 4. Diagram illustrating how the energy landscape affects the possibility of dramatic structural changes during crystallization of a DNA duplex. The solid line suggests the relative stabilities of the B-DNA and A-DNA forms in solution. (Our CD spectra of the Zif DNA show that the B-DNA form is more stable than the A-DNA form in solution.) The arrow indicates how crystal packing forces (such as those illustrated in Fig. 3) might - by adding a new set of intermolecular interactions - change the relevant energy surface and thus induce A-DNA formation.

energy surface for isolated
Zif DNA duplex

energy

energy from burying 1000 Å^2
in crystal packing contacts

B-DNA

A-DNA

conformational space
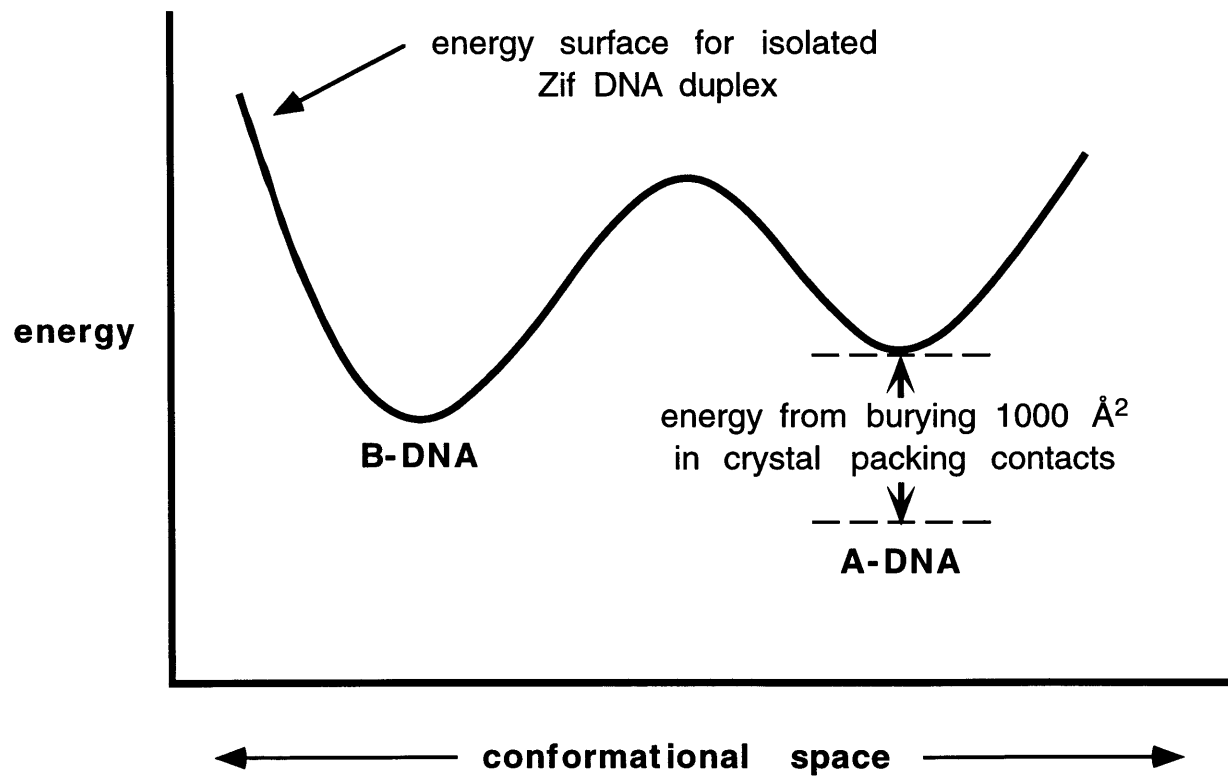
Chapter Five

Binding Studies with Mutants of Zif268: Contribution of Individual Side Chains to Binding Affinity in the Zif268 Zinc Finger-DNA Complex

Introduction

Although the Zif268 zinc finger-DNA complex has been very well characterized structurally [1, 2] and has served as the basis for numerous modeling, design, and selection studies [3-10], little biochemical data is available for this complex. Binding site selection data indicate that TGCGT(G/A)GGCG(G/T) is the most favored binding site for Zif268 [11], and competition experiments using mutated oligonucleotides confirm this [11, 12]. Methylation interference data [12] are consistent with the contacts seen in the structure of the complex [1, 2]. The dissociation constants that have been measured for Zif268 range from 0.01-6.5 nM, depending on the conditions [1, 6-9, 11, 13]. Kinetic constants have also been reported, with k_{on} ranging from 3.1×10^4 to $>7.0 \times 10^8$ $M^{-1}s^{-1}$ [8, 9] and k_{off} from 1.2×10^{-4} to 1.4×10^{-2} s^{-1} [7-9], corresponding to a half-life of 39 seconds to 96 minutes.

How much individual contacts observed in the Zif268-DNA complex structure contribute to binding affinity and specificity, however, has been unclear. Although a few mutants with changes in the base-contacting residues have been constructed, the binding studies that have been performed with these mutants used peptides expressed on the surface of phage [14, 15] (such measurements may not be as reliable as measurements using purified peptides). To address key issues regarding recognition by Zif268, we have constructed five mutants with alterations in the base-contacting residues of finger one. Here we report the equilibrium dissociation constants measured for these five mutant peptides.

Results and discussion

We have constructed mutants of Zif268 in which each of the base-contacting residues in finger one - Arg 18, Asp 20, Glu 21, and Arg 24 - has been mutated individually to alanine. Since the side chains of Arg 18 and Asp 20 interact with each other, we have also constructed an RA18/DA20 double mutant. Each of these mutant peptides has been overexpressed and purified, and its affinity for a Zif268 binding site (Figure 1b) has been determined by a gel mobility shift assay. The apparent dissociation constants (K_d's) measured for these mutants and for wild type Zif268 in parallel experiments are listed in Table 1.

Wild type Zif268 binds the oligonucleotide binding site used in these studies with a K_d of 0.17 nM (Figure 2), which is within the range of K_d's observed in previous studies. Mutating Arg 18 (at position -1 of the helix) to alanine results in a 100-fold loss of affinity (K_d =17 nM; Figure 3). Mutating both Arg 18 and Asp 20 (at position 2 of the helix) simultaneously produces a peptide that binds with a K_d of 4.4 nM (Figure 4), 26-fold less tightly than wild type Zif268 but about four times as tightly as the RA18 single mutant. The unexpected observation that the double mutant binds more tightly than the RA18 single mutant suggests that the loss of binding affinity observed for the RA18 mutant does not result solely from the loss of the hydrogen bonds Arg 18 normally makes with a guanine. Instead, the mutation must also result in some unfavorable interaction (for example, unfavorable electrostatic interactions between the acidic residues at positions 2 and 3 of the helix and the DNA).

The DA20 peptide binds with slightly greater affinity than does the wild type peptide (K_d=0.10 nM; Figure 5). This observation is surprising;

the Asp 20 to alanine mutation was expected to result in a decrease in binding affinity, since this mutation eliminates the hydrogen bonds between Asp 20 and Arg 18 that presumably help orient the arginine side chain. The cause of the increased affinity displayed by the DA20 peptide is unclear, but one intriguing possibility is that, in the absence of the Asp 20/Arg 18 interaction, the arginine side chain might interact with Glu 21 and contact the phosphate backbone instead of the base (a similar arrangement has been observed in the structure of wild type Zif268 with a suboptimal binding site, for example [Chapter 3]).

The EA21 mutant binds with an affinity approximately equal to that of the wild type protein (K_d=0.19 nM; Figure 6), consistent with the observation that the glutamic acid makes van der Waals contacts but no hydrogen bonds with the DNA in the wild type complex. The largest effect on binding affinity is caused by mutation of Arg 24 to alanine: the RA24 peptide binds approximately 400-fold less tightly than does wild type Zif268 (K_d=7.0 nM; Figure 7).

The data presented here allow us to estimate the relative energetic contribution to binding affinity made by the residues at positions -1, 2, 3, and 6 of the α helix in finger one of Zif268, but many questions remain. For example, do the residues at corresponding positions in the other fingers make similar contributions to affinity? Although Glu 21 does not appear to make a significant contribution to binding affinity, does it play a role in determining binding specificity by excluding bases other than cytosine from the middle position of finger one's subsite? How much do contacts between the fingers and the phosphate backbone contribute to binding affinity? What structural changes and adjustments occur in these

mutant complexes? Clearly, further experiments will be needed to resolve such remaining questions about DNA recognition by the Zif268 zinc fingers.

Materials and methods

Cloning

Mutations were introduced into pZif89 (the expression construct which encodes the three fingers of Zif268 and is described in [1]) via the four primer PCR method [16]. The primers used in the mutagenesis are shown in Figure 8. For each mutant, the L-out primer and the R-mutagenic primer were used to amplify a fragment from pZif89; in a parallel reaction, the R-out primer and the L-mutagenic primer (partially overlapping and complementary to the R-mutagenic primer) were used to amplify a second, overlapping fragment. The products of the first two reactions were gel purified (Qiagen) and combined for another round of PCR amplification using the L- and R-outside primers. The resulting 400 bp product, which included the entire zinc finger coding region, was gel purified (Qiagen), digested with NdeI and BamHI, and ligated into NdeI-BamHI-digested pET3a (RA18, DA20, RA18/DA20, and RA24) or pET21a (EA21). The sequences of all five mutant genes were verified by dideoxy sequencing (performed by the Biopolymers Laboratory, Howard Hughes Medical Institute, Massachusetts Institute of Technology).

Protein expression and purification

The expression constructs were transformed into *E. coli* BL21(DE3) cells containing either the pLysE plasmid (wild type Zif268, RA18, DA20, RA18/DA20, and RA24) or the pLysS plasmid (EA21), and expression was induced as recommended (Novagen). The peptides were purified by reversed-phase batch extraction on Sep-Pak C18 cartridges (Waters) and reversed-phase high-performance liquid chromatography on a C4 column

(Vydac) essentially as described [1]. Purified peptides were folded in an anaerobic chamber (Coy Laboratory Products) by dissolving them in water and then adding $ZnSO_4$ to 2.75 mM and bis-tris-propane, pH 6.8, to 50 mM [6]. The folded peptides were stored in aliquots at -80°C. The peptides were about 95% (EA21 and RA24) or at least 98% (wild type, RA18, DA20, and RA18/DA20) pure, as estimated from examination of silver stained SDS-polyacrylamide gels.

Determination of apparent dissociation constants

The double stranded oligonucleotide binding site used in these studies is shown in Figure 2. The individual oligonucleotides were synthesized, gel purified, annealed, and end-labeled using [γ^{32}P]ATP and T4 polynucleotide kinase [17]. Binding assays were performed at room temperature in degassed binding buffer (50 mM NaCl, 5 mM $MgCl_2$, 10 μM $ZnSO_4$, 5% (v/v) glycerol, 0.1 mg/ml acetylated bovine serum albumin, 0.1% (w/v) Igepal-CA630, and 15 mM Hepes at pH 7.8) [6]. Binding reactions were equilibrated for three hours before being electrophoresed on 10% polyacrylamide gels in 0.5x TB. Dried gels were exposed to a PhosphorImager screen (Molecular Dynamics) overnight.

To derive the apparent dissociation constant for each peptide, the labeled binding site was mixed with increasing amounts of peptide. The binding reactions contained 2.5 pM (wild type Zif268, DA20, and EA21) or 25 pM (RA18, RA18/DA20, and RA24) labeled oligonucleotide and 14.7 μg/ml poly (dI-dC)-poly (dI-dC) in degassed binding buffer. K_d's were determined by fitting the data to the equation $\Theta = \dfrac{1}{1 + \frac{K_d}{[P]}}$, where Θ

represents the fraction of the DNA bound by the peptide. [P], the free protein concentration, was approximated by the total protein concentration (since the DNA concentration used was considerably below the K_d in each case). The active protein concentration was determined by titrating each peptide against either 150 nM (wild type Zif268, DA20, and EA21), 500 nM (RA18/DA20), or 1 μM (RA18 and RA24) of the oligonucleotide binding site. Each experiment (K_d determination and measurement of protein concentration) was performed at least twice, using a freshly thawed aliquot of peptide each time.

Acknowledgements

References

1.   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.

2.   Elrod-Erickson, M., Rould, M.A., Nekludova, L. & Pabo, C.O. (1996). Zif268 protein-DNA complex refined at 1.6 Å: A model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171-1180.

3.   Nardelli, J., Gibson, T. & Charnay, P. (1992). Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis. *Nuc. Acids Res.* **20**, 4137-4144.

4.   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.

5.   Pomerantz, J.L., Sharp, P.A. & Pabo, C.O. (1995). Structure-based design of transcription factors. *Science* **267**, 93-96.

6.   Rebar, E.J. & Pabo, C.O. (1994). Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**, 671-673.

7.   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.

8.   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.

9.   Kim, J.-S. & Pabo, C.O. (1998). Getting a handhold on DNA: Design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl. Acad. Sci. USA* **95**, 2812-2817.

10.   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.

11.   Swirnoff, A.H. & Milbrandt, J. (1995). DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Mol. Cell. Biol.* **15**, 2275-2287.

12.   Christy, B. & Nathans, D. (1989). DNA binding site of the growth factor-inducible protein Zif268. *Proc. Natl. Acad. Sci. USA* **86**, 8737-8741.

13.   Greisman, H.A. & Pabo, C.O. (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657-661.

14.   Choo, Y. (1998). End effects in DNA recognition by zinc finger arrays. *Nuc. Acids Res.* **26**, 554-557.

15.   Isalan, M., Choo, Y. & Klug, A. (1997). Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc. Natl. Acad. Sci. USA* **94**, 5617-5621.

16.   Cormack, B. (1997). Directed mutagenesis using the polymerase chain reaction, in *Current protocols in molecular biology,* F.M. Ausubel, *et al.,* Editors. John Wiley & Sons, Inc.: p. 8.5.1-8.5.10.

17.   Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
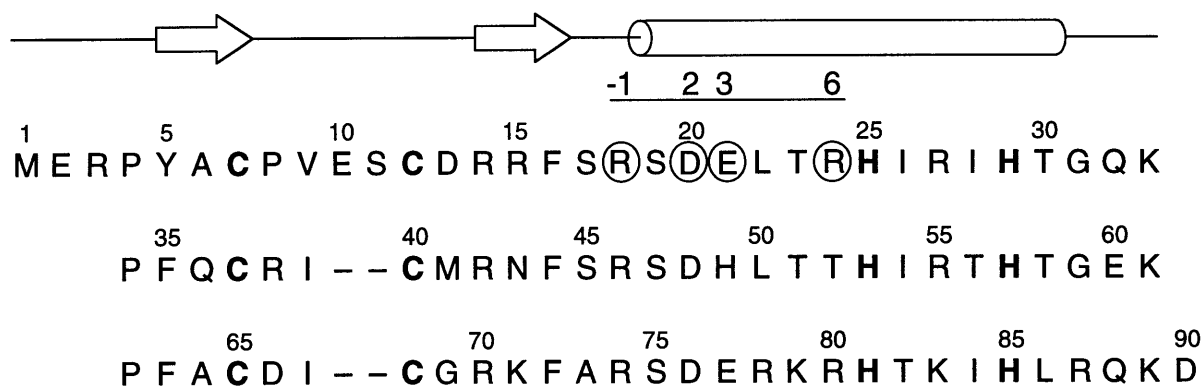
Table 1. Apparent dissociation constants measured for wild type Zif268 and for the five mutants.

protein	K_d (M)	$K_d/K_{d\ wild\ type}$	$\Delta G_{mut}-\Delta G_{wild\ type}$ (kcal/mol)
Zif268	$1.7\ (\pm 0.10)\times 10^{-10}$	1	-
RA18	$1.7\ (\pm 0.08)\times 10^{-8}$	100	-2.7
DA20	$1.0\ (\pm 0.01)\times 10^{-10}$	0.6	3.0×10^{-1}
RA18/DA20	$4.4\ (\pm 0.42)\times 10^{-9}$	26	-1.9
EA21	$1.9\ (\pm 0.97)\times 10^{-10}$	1.1	-5.6×10^{-2}
RA24	$7.0\ (\pm 1.21)\times 10^{-8}$	412	-3.6

Figure 1. Sequences of the zinc fingers peptides and of the oligonucleotide binding site. **(a)** Sequence of the wild type Zif268 zinc finger peptide. The residues at positions -1, 2, 3, and 6 of the α helix, which have been mutated singly or in pairs to alanine, are circled. The three fingers are aligned to highlight conserved residues and secondary structure elements. The helix is indicated by a cylinder and the β strands by arrows. The conserved cysteine and histidine residues that are ligands for the zinc ions are highlighted in bold. (Adapted from [2].) **(b)** Sequence of the oligonucleotide binding site used in the gel shift assay. The Zif268 binding site is highlighted in bold.

Figure 1

A


```
             -1   2 3      6
1        5         10          15        20        25        30
M E R P Y A C P V E S C D R R F S Ⓡ S Ⓓ Ⓔ L T Ⓡ H I R I H T G Q K

        35          40          45          50          55          60
        P F Q C R I - - C M R N F S R S D H L T T H I R T H T G E K

                65          70          75          80          85        90
        P F A C D I - - C G R K F A R S D E R K R H T K I H L R Q K D
```

B

```
AGCAGCTGA GCGTGGGCGT AGTGAGCT
TCGTCGACT CGCACCCGCA TCACTCGA
```

Figure 2. Equilibrium binding curve for the wild type Zif268 zinc finger peptide binding to the site shown in Figure 1. The dashed line represents a theoretical curve with K_d=0.17 nM.
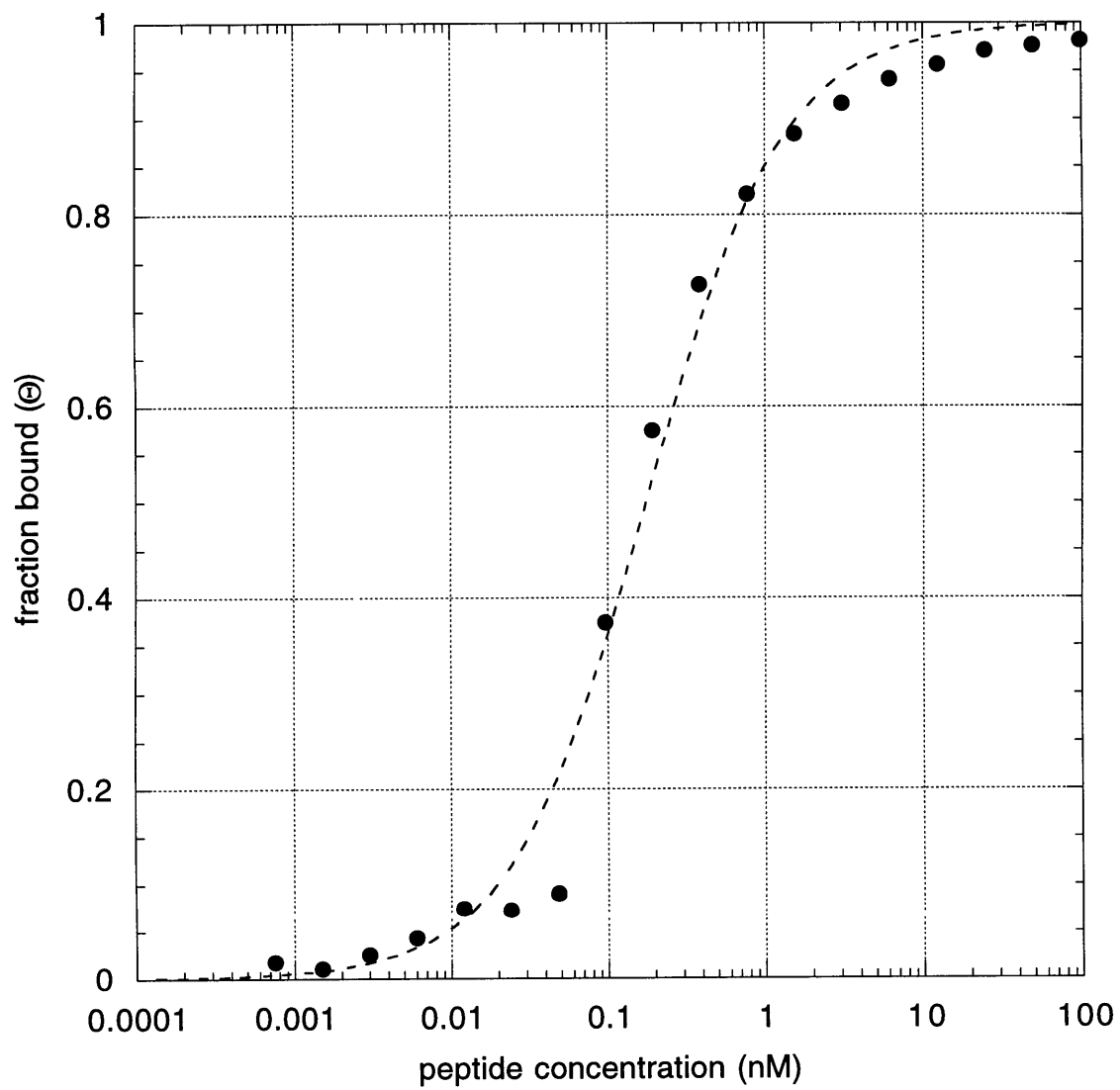
Figure 2

Figure 3. Equilibrium binding curve for the RA18 mutant peptide binding to the site shown in Figure 1. The dashed line represents a theoretical curve with K_d=17 nM.
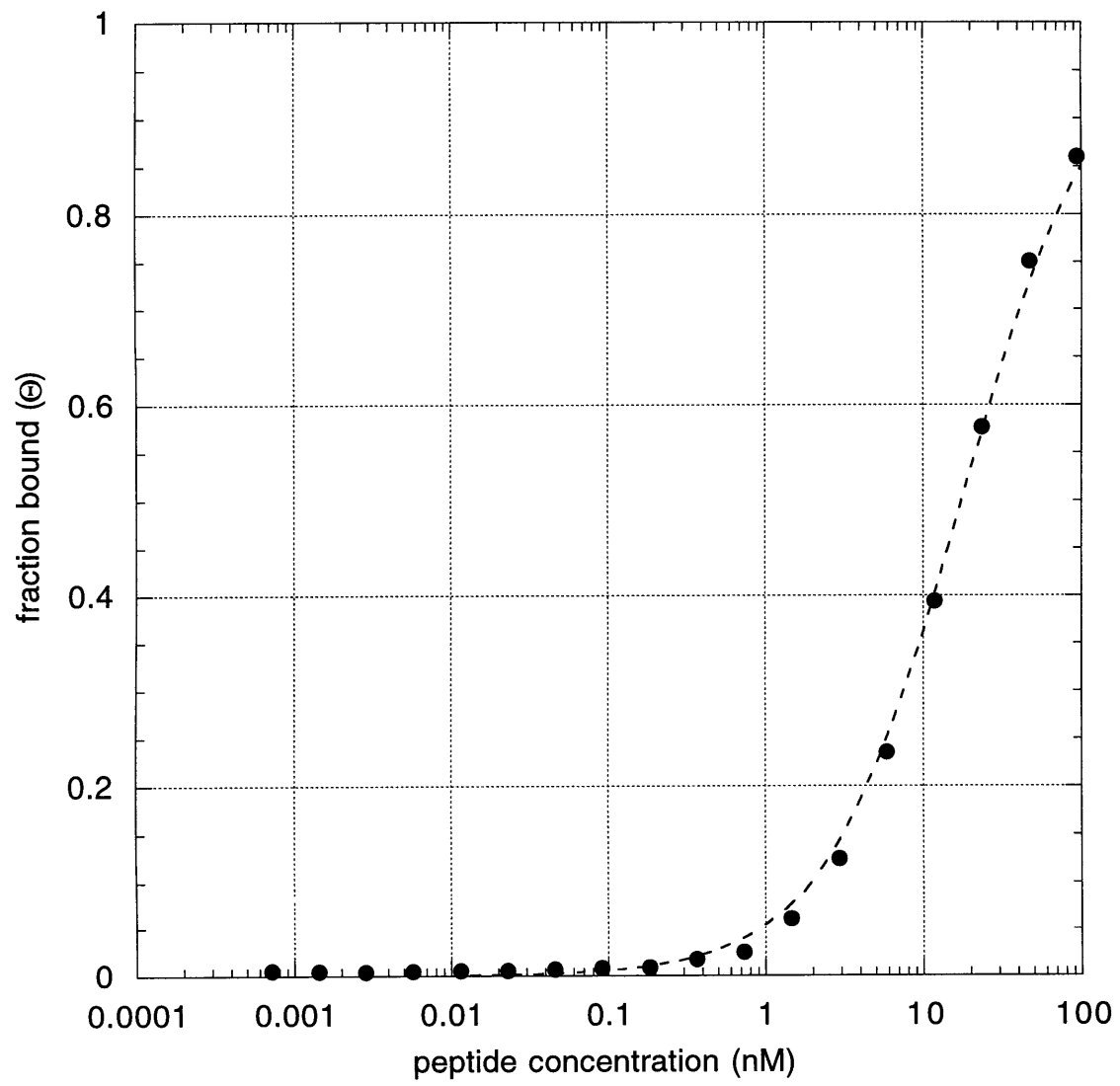
Figure 3

Figure 4. Equilibrium binding curve for the RA18/DA20 mutant peptide; the dashed line represents a theoretical curve. This peptide binds to the site shown in Figure 1 with a K_d=4.4 nM.
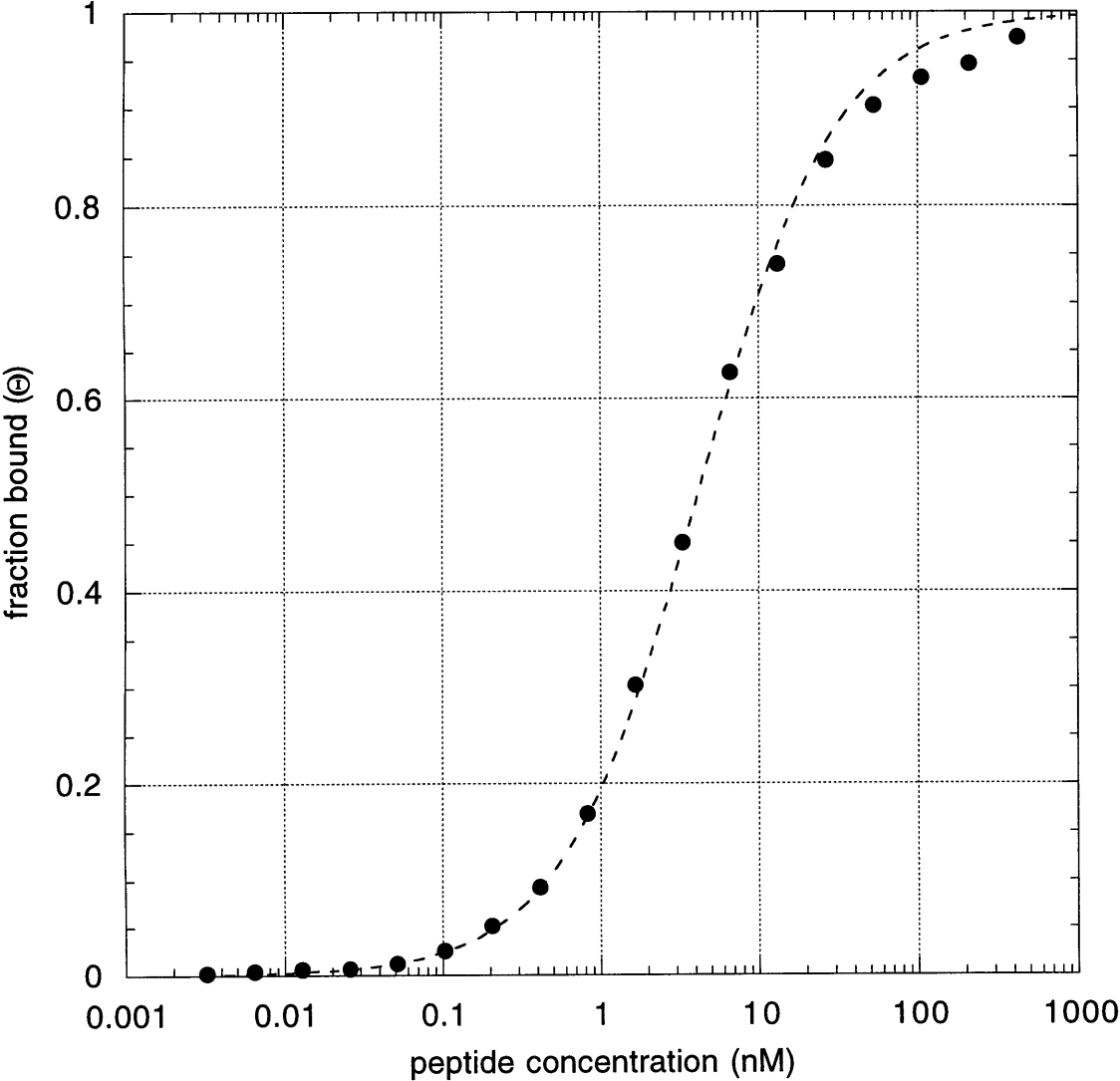
Figure 4

Figure 5. Equilibrium binding curve for the DA20 mutant peptide binding to the site shown in Figure 1. The dashed line represents a theoretical curve with K_d=0.10 nM.

Figure 5

Figure 6. Equilibrium binding curve for the EA21 mutant peptide; the dashed line represents a theoretical curve. This peptide binds to the site shown in Figure 1 with a K_d=0.19 nM.

Figure 6
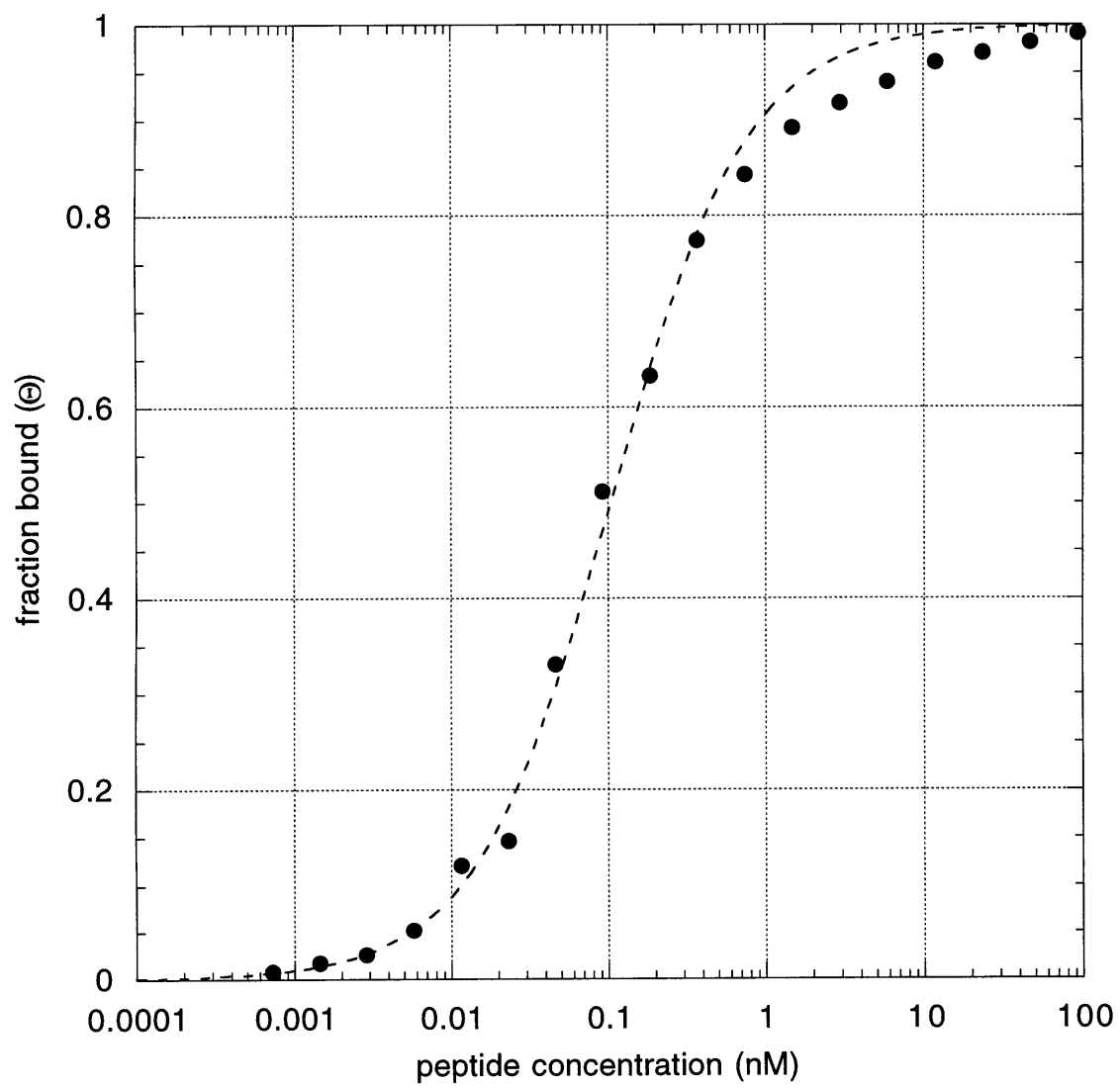
Figure 7. Equilibrium binding curve for the RA24 mutant peptide; the dashed line represents a theoretical curve. This peptide binds to the site shown in Figure 1 with a K_d=7.0 nM.

Figure 7

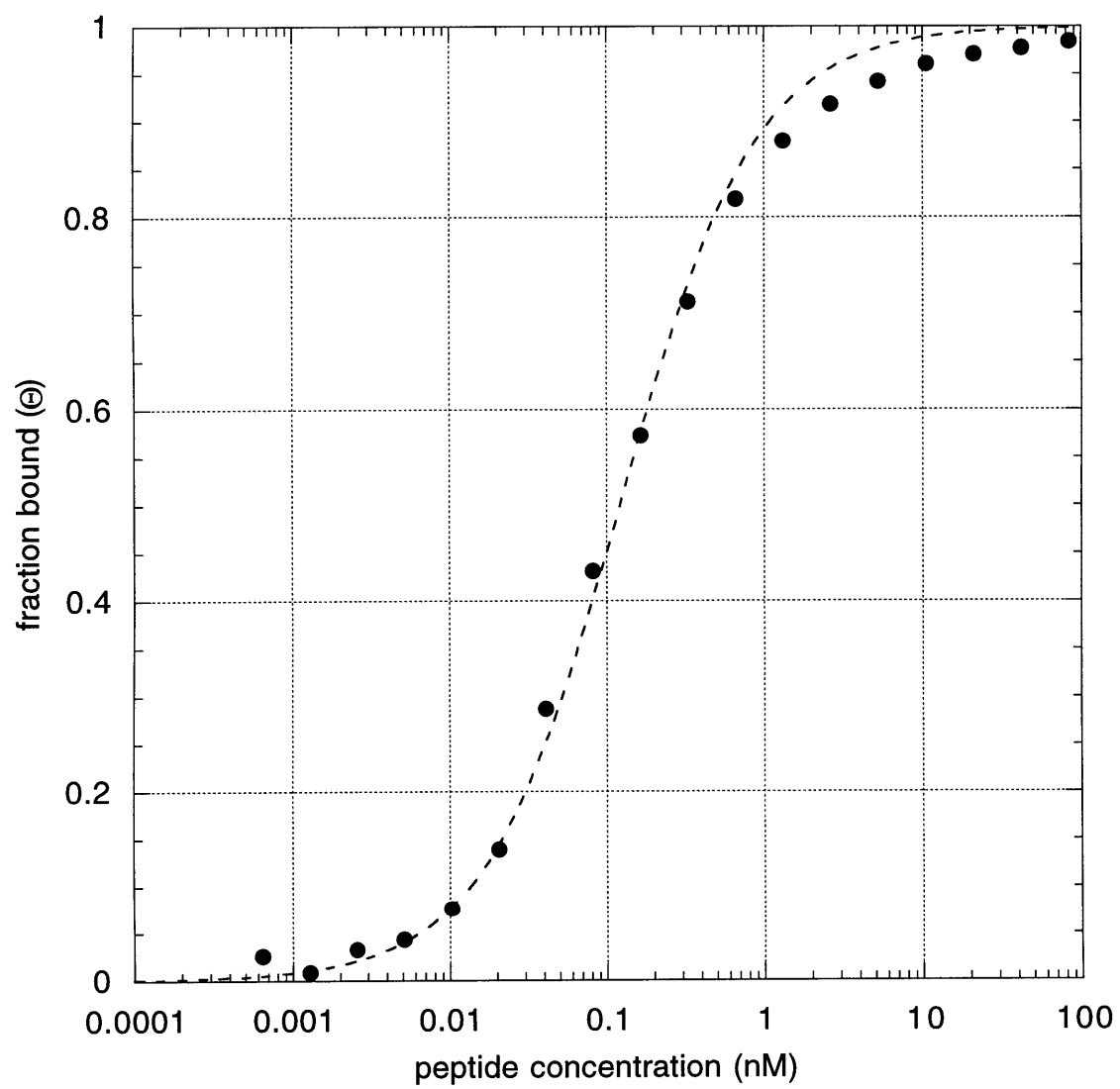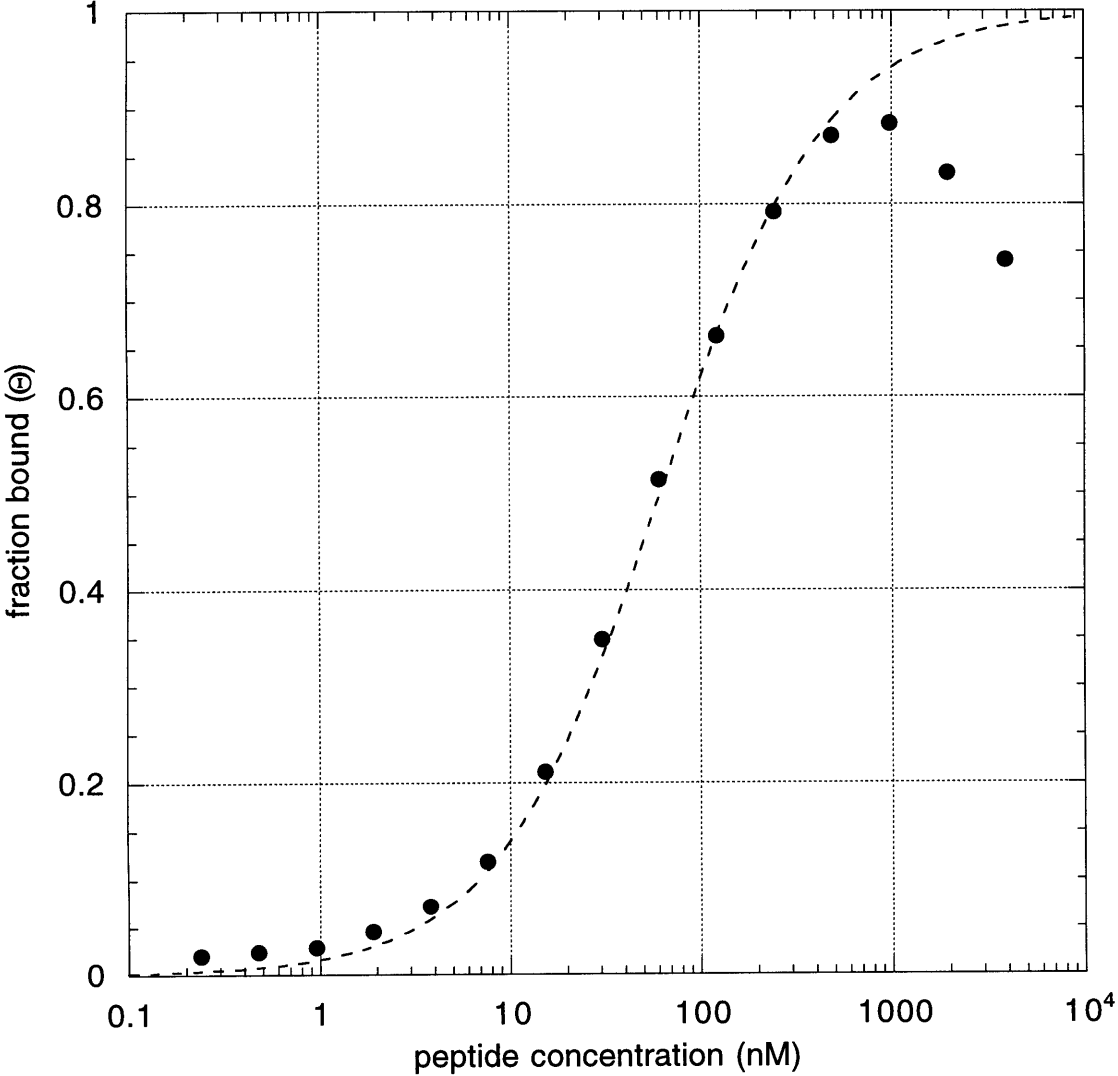**Figure 8.** Oligonucleotide primers used to introduce the mutations. The L- and R-out primers were used in the construction of all the mutants; the L- and R-RA18 primers were used only for the RA18 mutant, and so on. All sequences are written 5' to 3'.

Figure 8

L-out	GATGCGTCCGGCGTAGAGGATCGAG
R-out	GGTGGCAGCAGCCAACTCAGCTTCC
L-RA18	TTTCTGCCTCGGATGAGCTTACCCG
R-RA18	TCCGAGGCAGAAAAGCGGCGATCGC
L-DA20	GCTCGGCTGAGCTTACCCGCCATAT
R-DA20	AGCTCAGCCGAGCGAGAAAGCGGC
L-RA18/DA20	TTCTGCCTCGGCTGAGCTTACCCGC
R-RA18/DA20	CTCAGCCGAGGCAGAAAAGCGGCGA
L-EA21	CGGATGCGCTTACCCGCCATATCCG
R-EA21	GTAAGCGCATCCGAGCGAGAAAGC
L-RA24	CTTACCGCCCATATCCGCATCCACA
R-RA24	GATATGGGCGGTAAGCTCATCCGAG

Chapter Six

Overview and Future Directions

Overview

The structural and biochemical studies described in this thesis have provided much new information about zinc finger-DNA interactions. The high resolution structures of the wild type Zif268 zinc finger-DNA complex described in Chapter 2 and of the seven variant complexes described in Chapter 3 have revealed several features of zinc finger-DNA recognition. 1) Interactions between neighboring side chains may play an important role in recognition. In particular, the interactions between the arginine at position -1 of the α helix and the acidic residues at positions 2 and 3 appear to be important in determining whether the arginine contacts a base or the phosphate backbone. 2) There are many water-mediated hydrogen bonds between the zinc fingers and the DNA that could play a role in affinity and specificity. 3) Many of the direct base contacts observed in these structures fit the pattern of interactions that has been observed in other zinc finger-DNA complex structures [1-7] (i.e., the residue immediately preceding the α helix tends to contact the 3' base in the finger's subsite, the third residue in the α helix the middle base, the sixth residue in the α helix the 5' base, and the second residue in the α helix a base on the opposite strand of the DNA in the preceding finger's subsite). 4) However, a simple recognition code that could accurately predict the binding site recognized by a given zinc finger peptide or that could predict the optimal combination of side chains to insert into a designed protein to recognize a novel binding site does not appear to exist. The complexity of the interactions between the zinc fingers and the DNA seen in our structures explains why sequential optimization has been a more successful means of generating novel zinc finger proteins with

desired specificities than has code-directed design and supports the need for continued selection studies. 5) New side chain-base or side chain-phosphate interactions can be accommodated by localized changes in the conformation and/or orientation with respect to the DNA of one finger, without affecting the docking of the remaining fingers. Such adaptability may explain the versatility of the zinc finger motif.

The circular dichroism studies reported in Chapters 2 and 4 and comparison of the conformation of the Zif268 binding site in the complex crystal structure (Chapter 2) with that in crystals containing only the DNA (Chapter 4) indicate that, like many DNA sequences, the Zif268 binding site can adopt multiple conformations under different conditions. The process of recognition appears to exploit such flexibility, since the DNA changes conformation when the Zif268 zinc fingers bind.

Future directions

The studies reported in this thesis have answered many questions about DNA recognition by zinc fingers, but other questions remain to be addressed. This last section outlines some possible future directions for further research concerning zinc finger-DNA interactions in particular and protein-DNA recognition in general.

Mutagenesis and binding studies

The experiments described in Chapter 5 began to assess the contribution that individual side chains make to affinity and specificity in the Zif268 complex, but many further such experiments suggest

themselves. For example, the glutamic acid residue at position 3 of the α helix in finger one does not make a significant contribution to binding affinity, but does it nonetheless contribute to specificity (the proximity of its side chain to the middle base of the finger's subsite could help exclude bases other than cytosine from this position)? Or are other factors (such as base-stacking interactions or water-mediated hydrogen bonds from other residues in the finger to this cytosine) more important determinants of specificity at this position in the finger's subsite? This issue could be addressed by determining whether the EA21 mutant peptide displays reduced specificity for cytosine at the middle position of finger one's subsite relative to the wild type peptide.

The binding studies in Chapter 5 involve mutations only in the potential base-contacting residues of finger one: do contacts from the corresponding positions in fingers two and three make equivalent contributions? Binding studies involving peptides expressed on the surface of phage suggest that disrupting the peptide-DNA interactions at each end of the complex (i.e., the contacts made by the residue at position -1 of the α helix in finger one and the residue at position 6 in finger three) may have less of an effect than disrupting the analogous contacts in the middle of the complex [8]; will such end effects be apparent when the binding studies are performed with purified peptides? In finger one of Zif268, the aspartic acid at position 2 of the helix appears to play an important role in orienting the arginine at position -1 such that it contacts a base rather than the phosphate backbone. Do the corresponding aspartic acid residues in fingers two and three play an equally important role, or does having an amino-terminal neighboring finger bound to the DNA help restrict the orientation of these fingers and therefore of the

corresponding arginines?

How much do the contacts with the phosphate backbone observed in the Zif268 complex contribute to affinity? Similarly, do the observed water-mediated contacts make a significant contribution to affinity? The residues that make water-mediated hydrogen bonds with the bases also make direct contacts, so the contribution of these water-mediated interactions to affinity and specificity is difficult to assess. However, several residues make water-mediated contacts solely to the phosphate backbone, providing an opportunity to assess the energetic contribution of such contacts. Ser 17 (from finger one) makes a water-mediated phosphate contact, while Ser 45 (from the corresponding position of finger two) contacts the phosphate backbone directly; determining the effect of mutating each of these residues would thus allow a direct comparison of the energetic contribution made by water-mediated vs. direct phosphate contacts.

There are few contacts between adjacent fingers in the Zif268 complex, but a conserved arginine forms a hydrogen bond with a backbone carbonyl in the next finger (Arg 27 from finger one hydrogen bonds to the carbonylof Ser 45 from finger two, and Arg 55 from finger two interacts with the carbonyl of Ala 73 from finger three [1]). Would mutating either of these arginines, and thus eliminating the corresponding inter-finger contact, result in a loss of affinity?

Structural studies of the Zif268 binding site

A structure of the free Zif268 binding site for comparison with the conformation of the binding site in the Zif268 complex crystal structure would provide a better understanding of the structural changes that occur in the DNA upon complex formation. Crystals of the free DNA might be

produced by varying the length of the duplex oligonucleotide or the crystallization conditions more extensively than has previously been attempted (Chapter 4 and unpublished observations). Alternatively, the Zif268 binding site might be crystallized in the context of a different protein-DNA complex as part of the flanking sequence for some other DNA-binding protein's recognition site. Ideally, the structure of the DNA in at least two different environments would be solved, such that comparison of the structures would indicate whether any distinctive features observed in the DNA were likely to be inherent characteristics of the sequence or effects of crystal packing forces.

Structural studies of other zinc finger complexes

There are many zinc finger-DNA complexes whose structures would add to our knowledge of zinc finger-DNA recognition. For example, the structure of a complex involving the DA20 mutant peptide (Chapter 5) would reveal what happens to the arginine at position -1 of the helix in finger one when its interaction with the aspartic acid at position 2 is eliminated, providing an indication of how important this interaction is in positioning the arginine to make its usual contact with a guanine base. If this interaction proves to be vital in finger one, the structure of a complex with an analogous mutation in finger two or three (DA48 or DA76) would reveal whether the corresponding aspartic acids play equally important roles in these other fingers.

The variant complexes whose structures were described in Chapter 3 resulted from selections in which four amino acids in finger one of Zif268 were randomized [9]. Later selections, in which six positions of each finger in turn were randomized, have produced three variants of Zif268 in which all three fingers have altered specificities [10]. The structure of

the complex between each of these three variant peptides and its targeted site should provide answers to such questions as: Will the side chain-base contacts observed in these complexes be similar to those predicted from known structures [10]? Will the orientation of these fingers with respect to the DNA be similar to that seen in the wild type complex or the single finger variant complexes, or will the randomization of more positions in each finger permit a wider range of orientations?

Similarly, the structures of other naturally occurring zinc finger-DNA complexes could provide new insights into recognition. For example, four fingers from the Drosophila protein hunchback recognize a site that is more A-T rich than that found in any of the complexes that have been solved to date [11-13]. A structure of the ternary complex between the Sp1 zinc fingers, the YY1 fingers, and DNA would reveal the basis for the interaction that has been demonstrated between these two zinc finger regions [14]. Similarly, a structure of the ternary complex between the Sp1 fingers, the GATA-1 DNA-binding domain, and DNA [15], or between the YY1 fingers, ATFa2, and DNA [16], would show how fingers can be used to mediate interactions with members of other families of DNA-binding motifs. It would also be interesting to compare the way in which zinc fingers are used to bind RNA (in a complex between the TFIIIA fingers and 5S ribosomal RNA [17], for example) with the ways in which they are used in zinc finger-DNA complexes.

To achieve specificity, proteins must bind to their target sites more tightly than to noncognate DNA sequences; being able to compare specific and nonspecific protein-DNA complexes would therefore aid in understanding recognition. However, structures of nonspecific complexes have been obtained in only a limited number of instances (in the glucocorticoid receptor DNA binding domain-DNA complex, for example,

where one half of the dimer binds specifically to a consensus DNA target sequence and the other half interacts nonspecifically with a noncognate sequence [18]). Zinc finger proteins may provide a good system for visualizing nonspecific interactions, in complexes where one or two fingers bind to their target sites and the remaining fingers bind nonspecifically (one such semispecific complex was described in Chapter 3). The sequence of the nonspecifically bound fingers and of the corresponding region of the DNA could be varied to see how a range of fingers interact with different noncognate sequences.

References

1.   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.

2.   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.

3.   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.

4.   Houbaviy, H.B., Usheva, A., Shenk, T. & Burley, S.K. (1996). Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl. Acad. Sci. USA* **93**, 13577-13582.

5.   Kim, C.A. & Berg, J.M. (1996). A 2.2 Å resolution crystal structure of a designed zinc finger protein bound to DNA. *Nat. Struc. Biol.* **3**, 940-945.

6.   Wuttke, D.S., Foster, M.P., Case, D.A., Gottesfeld, J.M. & Wright, P.E. (1997). Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: Determinants of affinity and sequence specificity. *J. Mol. Biol.* **273**, 183-206.

7.   Nolte, R.T., Conlin, R.M., Harrison, S.C. & Brown, R.S. (1998). Differing roles for zinc fingers in DNA recognition: Structure of a six-finger transcription factor IIIA complex. *Proc. Natl. Acad. Sci. USA* **95**, 2938-2943.

8.   Choo, Y. (1998). End effects in DNA recognition by zinc finger arrays. *Nuc. Acids Res.* **26**, 554-557.

9.   Rebar, E.J. & Pabo, C.O. (1994). Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**, 671-673.

10.   Greisman, H.A. & Pabo, C.O. (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657-661.

11.   Stanojevic, D., Hoey, T. & Levine, M. (1989). Sequence-specific DNA-binding activities of the gap proteins encoded by *hunchback* and *Kruppel* in *Drosophila. Nature* **341**, 331-335.

12.   Treisman, J. & Desplan, C. (1989). The products of the *Drosophila* gap genes *hunchback* and *Kruppel* bind to the *hunchback* promoters. *Nature* **341**, 335-337.

13.   Zhang, C.-C., Muller, J., Hoch, M., Jackle, H. & Bienz, M. (1991). Target sequences for *hunchback* in a control region conferring *Ultrabithorax* expression boundaries. *Development* **113**, 1171-1179.

14.   Lee, J.-S., Galvin, K.M. & Shi, Y. (1993). Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proc. Natl. Acad. Sci. USA* **90**, 6145-6149.

15.   Merika, M. & Orkin, S.H. (1995). Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol. Cell. Biol.* **15**, 2437-2447.

16.   Zhou, Q., Gedrich, R.W. & Engel, D.A. (1995). Transcriptional repression of the c-*fos* gene by YY1 is mediated by a direct interaction withATF/CREB. *J. Virol.* **69**, 4323-4330.

17.   Clemens, K.R., *et al.* (1993). Molecular basis for specific recognition of both RNA and DNA by a zinc finger protein. *Science* **260**, 530-533.

18.   Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R. & Sigler, P.B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497-505.