Annu. Rev. Biophys. Biomol. Struct. 1999. 3:183–212
Copyright © 2000 by Annual Reviews. All rights reserved

DNA RECOGNITION BY Cys_2His_2 ZINC FINGER PROTEINS

Scot A. Wolfe, Lena Nekludova, and Carl O. Pabo
Howard Hughes Medical Institute, Department of Biology, 68-580, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; e-mail: pabo@mit.edu

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

Key Words recognition code, phage display, structure-based design, gene therapy, protein-DNA interactions

■ **Abstract** Cys_2His_2 zinc fingers are one of the most common DNA-binding motifs found in eukaryotic transcription factors. These proteins typically contain several fingers that make tandem contacts along the DNA. Each finger has a conserved $\beta\beta\alpha$ structure, and amino acids on the surface of the α-helix contact bases in the major groove. This simple, modular structure of zinc finger proteins, and the wide variety of DNA sequences they can recognize, make them an attractive framework for attempts to design novel DNA-binding proteins. Several studies have selected fingers with new specificities, and there clearly are recurring patterns in the observed side chain–base interactions. However, the structural details of recognition are intricate enough that there are no general rules (a "recognition code") that would allow the design of an optimal protein for any desired target site. Construction of multifinger proteins is also complicated by interactions between neighboring fingers and the effect of the intervening linker. This review analyzes DNA recognition by Cys_2His_2 zinc fingers and summarizes progress in generating proteins with novel specificities from fingers selected by phage display.

CONTENTS

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

INTRODUCTION

There are a number of different families of "zinc finger" proteins that contain multiple cysteine and/or histidine residues and use zinc coordination to stabilize their folds (9, 10, 24, 70). Cys_2His_2 zinc finger proteins were the founding members of this superfamily and were first noted as repeating domains in the TFIIIA sequence (14, 38, 53, 85). Proteins that contain Cys_2His_2 zinc fingers are quite common in eukaryotic organisms, with this domain used not only for protein-DNA interactions but also for protein-RNA and protein-protein interactions (81, 111). The DNA-binding activity of these fingers has been the major focus of research (20). A number of studies have tried to determine the principles of zinc finger–DNA recognition (30–33, 61, 87, 88, 121) and to create zinc fingers that recognize novel DNA sites (18, 19, 50, 56, 58, 59, 104, 110, 124, 125). The selection of fingers with new specificities was inspired by the hope that their assembly into multifinger DNA-binding domains might provide useful new tools for diagnostics, biochemical research, and gene therapy. There has been exciting progress in understanding these proteins, but even this simple motif is remarkably complex. There still are significant challenges in understanding natural zinc finger proteins and in developing design methods that are versatile and reliable enough to find widespread application in biochemical research and gene therapy.

STRUCTURE OF THE ZINC FINGER DOMAIN

The Cys_2His_2 zinc finger unit was first identified in TFIIIA, which contains nine tandem repeats of this approximately 30 amino acid motif (14, 85). As additional zinc finger sequences became available, it was clear that these fingers share the consensus sequence (F/Y)-X-C-X_{2-5}-C-X_3-(F/Y)-X_5-ψ-X_2-H-X_{3-5}-H, where X represents any amino acid and ψ is a hydrophobic residue. These sequences fold in the presence of zinc (45) to form a compact $\beta\beta\alpha$ domain (78, 94, 95; Figure 1). Each finger binds a single zinc ion that is sandwiched between the two-stranded antiparallel β-sheet and the α-helix; the zinc is tetrahedrally coordinated between two cysteines at one end of the β-sheet and two histidines in the C-terminal portion of the α-helix. Detailed structural studies of zinc fingers show that the "α-helix" often contains sections of 3_{10} helix, particularly in the region between the histidines when the fingers have a HX_3H sequence pattern (35, 41, 55, 62, 78, 91, 92, 95, 96, 126).

It is interesting to note that zinc fingers have a relatively small number of fully conserved residues. Most of the structural stability is provided by zinc coordination

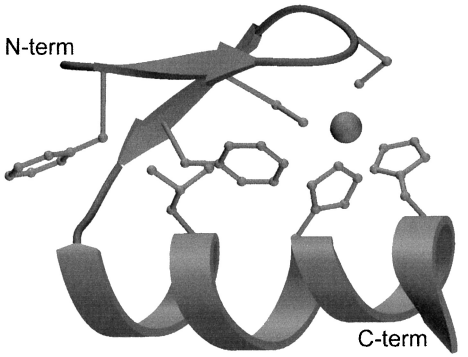

Figure 1 Diagram of the $\beta\beta\alpha$ motif from finger 2 of Zif268 (37, 95). The side chains of the conserved cysteines and histidines, which are involved in zinc coordination, and side chains of the three conserved hydrophobic residues are shown.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

and by the conserved hydrophobic core that flanks the zinc binding site (84, 113). The hydrophobic residues are well conserved, but there are some examples where their spacing within the domain can change (85, 93): For example, finger 6 of human ZFY has a sequence of the form (Y-X-C-X_2-C-X-F-X_7-L-X_2-H-X_4-H) where the second conserved aromatic residue is two residues closer to the cysteine than in the standard consensus sequence. Studies have demonstrated that an aromatic residue at either position on this β-strand can pack into the core of the $\beta\beta\alpha$ motif and stabilize the fold (123).

The stability of the $\beta\beta\alpha$ architecture is largely derived from the intrastrand "crosslinking" that zinc coordination provides. Fingers are unfolded in the absence of zinc (45), and substituting a residue other than cysteine or histidine at one of the ligand positions usually results in a loss of function (25, 120). Conservative substitutions (interchanging cysteine and histidine) of the zinc-coordinating residues is tolerated at some positions (48, 74, 86), with the final histidine being the most amenable to change (48, 83). Because of the stability provided by zinc coordination, this structure provides an excellent scaffold for presenting diverse peptide sequences in a helical conformation.

BIOLOGICAL ROLES OF ZINC FINGER PROTEINS

Proteins containing Cys_2His_2 zinc fingers are quite common in the genomes of eukaryotes. Approximately 0.7% of the genes in *Saccharomyces cerevisiae* and *Caenorhabditis elegans* contain Cys_2His_2 zinc fingers (12, 22), and it is estimated that a similar fraction will be found in humans (54). Zinc fingers occur in animals, plants, and fungi, but the typical number of finger repeats and the length of the linker between neighboring fingers varies greatly between the kingdoms (12, 22, 119). Cys_2His_2 zinc fingers are absent from the

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

genomes of *Escherichia coli* or *Methanococcus jannaschii* (22), but a potential zinc finger has been identified in *Synechococcus* PCC 7942 (11), suggesting that although this motif is uncommon in bacteria, it is not completely foreign to prokaryotes.

Cys_2His_2 zinc fingers that bind DNA have been studied in considerable detail. DNA recognition usually requires 2 to 4 tandemly arranged zinc fingers; when only one or two fingers are present, additional secondary structure elements are generally used to augment DNA recognition (13, 34, 40, 41, 90, 92). Zinc finger proteins can bind with sufficient specificity and affinity to function independently as the master regulator of a set of genes [TFIIIA (111) and NRSF/REST (109)], or like members of the SP1 family, they can work cooperatively with other DNA-binding proteins (75). Zinc fingers typically function in the context of a much larger protein, and certain other sequence motifs seem especially common in these proteins. The best characterized can be divided into four major classes: FAX (71), KRAB (8, 80), POZ (5), and FAR (69). These domains seem to have roles in transcriptional regulation or protein-protein interactions. It also has been shown that the zinc fingers themselves can be involved in protein-protein contacts, interacting directly with other transcription factors (49, 77, 82, 97).

A number of Cys_2His_2 zinc fingers have been identified that bind RNA, but aside from TFIIIA and p43, the biological significance of these interactions requires further study (2, 4, 15, 42, 43, 51, 72, 108, 111). Given the versatility and widespread distribution of this domain, it would not be surprising to find that many zinc fingers function as RNA-binding domains. The Wilms' tumor suppressor, WT1, is one good example. It was initially characterized only as a DNA-binding protein, but it now appears to function in gene regulation at the RNA level as well (15, 29, 39, 76).

DNA BINDING AND RECOGNITION

The crystal structure of Zif268 bound to DNA (95) has served as the prototype for understanding DNA recognition by this family of proteins (Figure 2). Zif268 contains three zinc fingers; the α-helical portion of each finger fits in the major groove of the DNA, and binding of successive fingers causes the protein to wrap around the DNA. Each finger has a similar docking arrangement and contacts an overlapping four base pair subsite (37, 95). However, the majority of base contacts occur in three base pair segments along one strand of the DNA (primary strand). Neighboring fingers are three base pairs apart: A helical motion that shifts the register of one finger by 3 base pairs superimposes neighboring fingers. The three fingers of Zif268 are oriented so that finger 1 is at the 3′ end of the primary strand and finger 3 is at the 5′ end. The DNA conformation is generally similar to that of B-form DNA, but the major groove is enough wider and deeper than normal that this seems to represent a distinctive DNA conformation (89). This

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

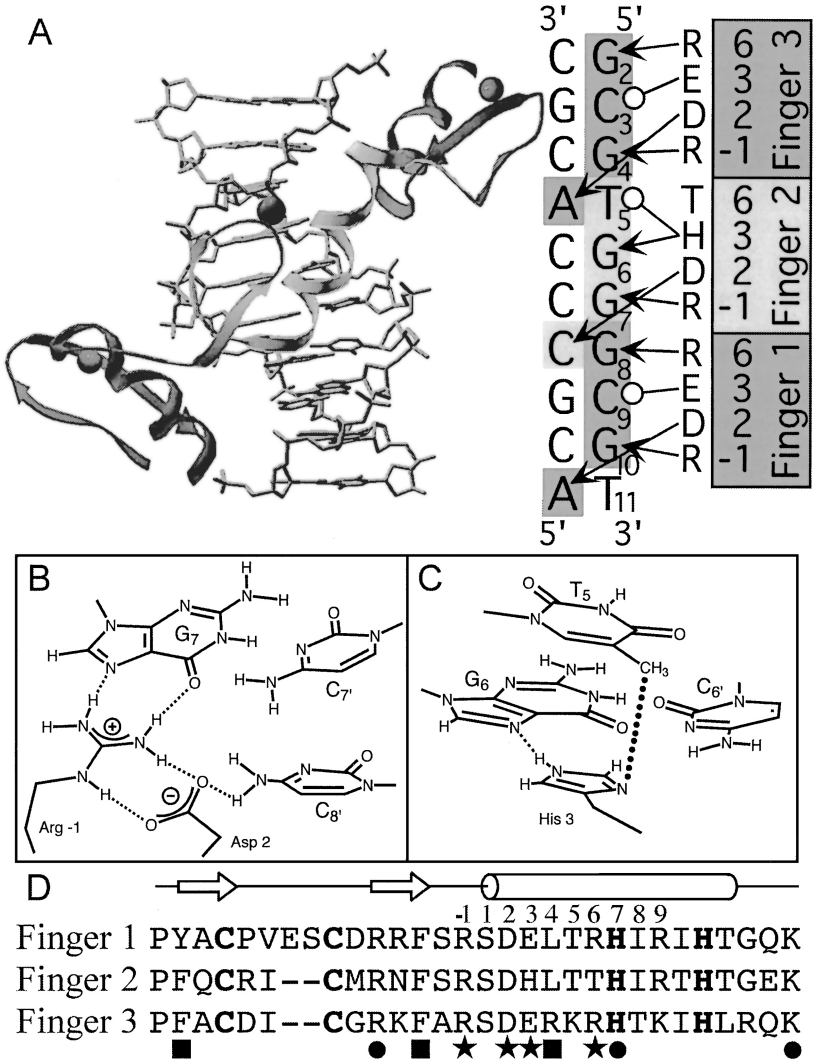

Figure 2 (*A*) Structure of the three fingers of Zif268 bound to DNA (37). Base contacts made from positions −1, 2, 3, and 6 of each α-helix are indicated schematically to the right of the structure. *Arrows* indicate contacts mediated by hydrogen bonds; *open circles* indicate hydrophobic interactions. For reference the base pairs are numbered (2 through 11) as in the original reference (95). (*B*) and (*C*) To give some idea of the structural details at the interface, this figure shows the base contacts made by finger 2 of Zif268. *Dashed lines* indicate hydrogen bonds and the **single** *dotted line* indicates a van der Waals contact of 3.4 Å. (*D*) The sequence of the three fingers of Zif268 is shown with the cysteines and histidines involved in zinc coordination indicated in *bold*. *Filled squares* below the sequence indicate the position of the conserved hydrophobic residues. *Filled circles* and *stars* indicate residue positions that are involved in phosphate and base contacts (respectively) in most of the fingers.

enlarged major groove is a common feature in the structures of most other zinc finger–DNA complexes and occurs in a number of other protein-DNA complexes (89).

Docking Arrangement and Base Recognition in Zif268

It may be useful to consider the structure of Zif268 in some detail because it provides a framework for the analysis and comparison of other zinc finger–DNA complexes, and because attempts to design novel zinc finger proteins have focused on fingers that have Zif-like docking arrangements. The base contacts in the Zif268 complex are made by amino acids in the N-terminal portion of the recognition helices. Each helix docks at a rather steep angle in the major groove, with the α-helical axis tipped at an angle of about 45° with respect to the double-helical axis (Figure 2). The amino acids at positions −1, 3, and 6 of the helix are well positioned to make contacts with bases in the primary DNA strand, whereas the amino acid at position 2 can make a contact to the complementary strand of the DNA. (In this numbering scheme, the −1 position is the residue immediately preceding the α-helix.) The base pair contacted by position 2 is just outside the "core" three base pair subsite, and this base also is typically recognized by position 6 of the neighboring N-terminal finger. (The Zif268 fingers recognize overlapping four base pair subsites, but neighboring fingers are only three base pairs apart.)

All three Zif268 fingers have identical residues at positions −1 and 2 (Arg and Asp), and these residues make coordinated DNA contacts (Figure 2*B*). The arginine at position −1 makes a pair of hydrogen bonds to the guanine at the 3′ position in the primary DNA strand of each binding site. This interaction is stabilized by the aspartate at position 2, which also makes two hydrogen bonds to the guanidinium group of arginine. The carboxylate group of the aspartate also forms a hydrogen bond with the exocyclic amine of adenine (or cytosine) on the complementary strand just outside of the primary three base pair subsite (although the geometry for these contacts does not always seem ideal; Figure 2*A*, *B*). The arginine and aspartate also form a number of water-mediated contacts with the bases and the phosphate backbone, but the contribution of these water contacts to DNA-binding specificity is unknown.

In the Zif268 structure, the remaining base contacts are mediated by residues at positions 3 and 6 of the α-helix. When glutamate is at position 3 (as in fingers 1 and 3) there appears to be a hydrophobic interaction between the Cγ and Cδ carbons of this residue and the C5-C6 edge of the neighboring cytosine (37). When histidine is at position 3 of the helix (as in finger 2; Figure 2*C*), it forms a hydrogen bond from Nε to the N7 (or O6) of guanine and simultaneously forms a van der Waals contact with the methyl group of the adjacent thymine. At position 6 of the helix (in fingers 1 and 3) there is an arginine that makes a pair of hydrogen bonds to guanine. Position 6 of finger 2 is a threonine, which does not make any direct contacts with the DNA.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

Docking Arrangements and Base Contacts in Other Zinc Finger–DNA Complexes

Analysis of zinc finger–DNA interactions often focuses on fingers that have a DNA-docking arrangement very similar to that of Zif268, but it is important to recognize that a variety of docking arrangements are observed in zinc finger–DNA complexes. Comparing the known structures reveals that the vast majority of the base-specific contacts in the zinc finger–DNA complexes are made from positions −1, 2, 3, and 6 of the α-helix (presumably because these residues are the most prominently exposed on the surface of the helix), but variations in the docking arrangement of the fingers allows these residues to make alternative patterns of base contacts in different complexes. Analysis of known structures (92a) allows a provisional division of fingers into two sets (Figure 3): (a) Canonical fingers have the same pattern of base contacts as Zif268. This group of fingers includes Tramtrack finger 2 (TTK; 41), two Zif268 finger 1 variants (DSNR and QGSR; 35), TFIIIA finger 3 (91, 126), all three fingers from Berg's designed protein (1MEY; 62), and finger 3 of YY1 (55). (b) Nonstandard fingers have several different patterns of base contacts. They often use residues at positions 3 and 6 of the helix to recognize bases on the primary strand of the DNA in a manner similar to that of Zif268, but they deviate from the canonical recognition pattern in contacts made by residues at positions −1 and 2. There is also considerable variation in the length and spacing of the subsites for these zinc finger–DNA complexes. This set of nonstandard fingers includes TTK 1; TFIIIA fingers 1, 2, and 5; YY1 fingers 1, 2, and 4; GLI fingers 2, 4, and 5 (96); GAGA (92); and one Zif268 finger 1 variant (RADR; 35).

Analyzing zinc finger–DNA complexes by treating each three base pair subsite and each finger as a rigid body provides one way to assess differences in the docking arrangements of the various fingers. Zif268 was used as a reference, and docking arrangements were compared (in a pairwise fashion) by aligning the region of DNA duplex recognized by each finger and then calculating the translation and rotation necessary to overlay the α-carbons of the helices from the two different fingers. The results of this comparison are shown in Figure 4. Fingers with canonical docking patterns tend to cluster near the origin of this graph, whereas the majority of "nonstandard" fingers require much larger translations and rotations to be aligned with the fingers of Zif268. It also is interesting that proteins that contain three tandem fingers each making two or more base contacts (Zif268, QGSR, DSNR, 1MEY, YY1 fingers 2–4, TFIIIA fingers 1–3) tend to dock in a manner similar to Zif268. The canonical docking arrangement seems to allow a very favorable set of base contacts from three consecutive fingers and also appears to make the zinc finger complexes quite modular. Much of the past modeling, design, and selection of fingers has focused on such canonical docking arrangements, but it is important to recognize that a variety of different docking arrangements have been observed with zinc fingers, and for the recognition of some DNA sequences it is possible that a noncanonical docking arrangement may provide superior specificity.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

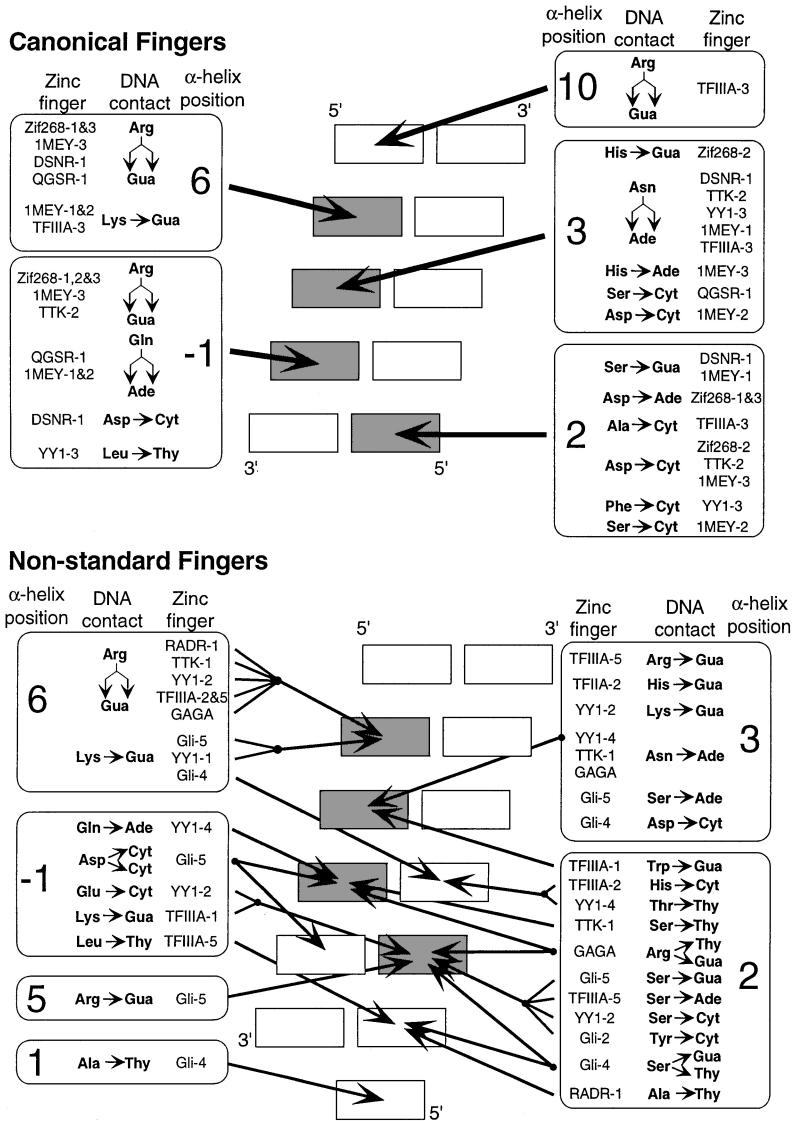

Figure 3 Summary of base contacts in various zinc finger–DNA complexes (35, 41, 55, 62, 91, 92, 95, 96, 126), showing how side chains at key positions along the α-helix contact bases in the respective subsites. Several complexes have a pattern of side chain–base interactions that are similar to those in Zif268 and thus are referred to as "canonical contacts." Other fingers (nonstandard) have somewhat different docking arrangements, showing a more diverse pattern of side chain–base interactions and sometimes contacting a larger subsite.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

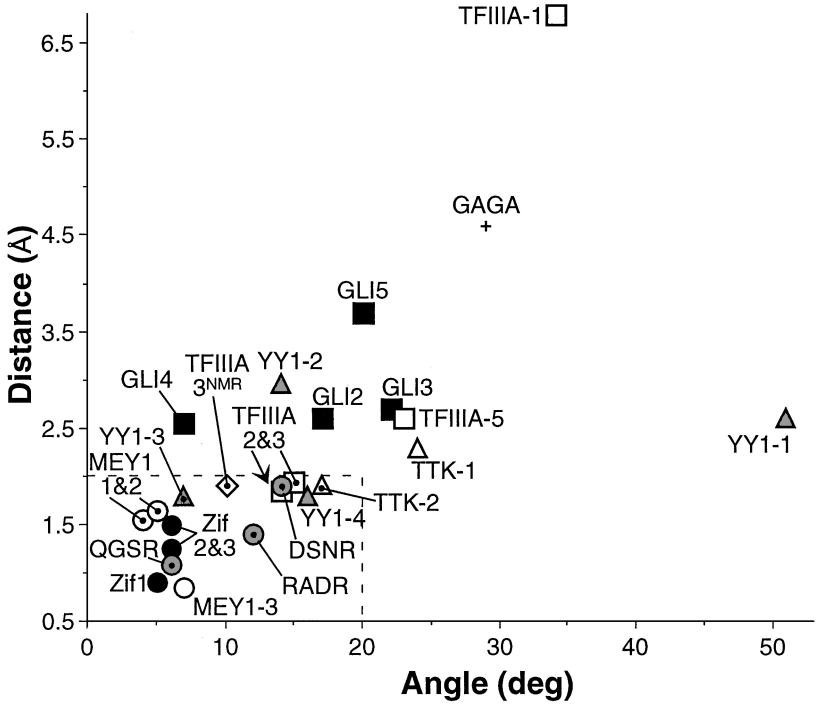

Figure 4 Comparison of the docking arrangement of individual zinc fingers in the major groove of DNA using Zif268 as a reference. Our comparison focuses on the position and orientation of the α-helix within the major groove, and complexes are compared in a pairwise manner by 1) aligning their DNA binding sites, and then 2) calculating the difference in docking arrangement between these two fingers. This difference is described by considering: (*a*) the distance between the center of mass for each α-helix and (*b*) the angle of rotation required to superimpose the first 8 C_α atoms of the helices. Each finger was compared successively with fingers 1, 2, and 3 of Zif268; the plot displays the average for the comparison of each finger with all three fingers of Zif268. In this figure, the "canonical" zinc fingers—which are most similar to Zif268—tend to cluster near the origin (within the 2.0 Å, 20° box). [Note: For finger 3 of TFIIIA both the NMR (*open diamond*; 126) and X-ray (*open squares*; 91) structures are indicated since they are rather different. Since these methods give good agreement about the orientation of fingers 1 and 2, only the parameters for the X-ray structure are plotted for these fingers.]

Phosphate Contacts

There are a number of phosphate contacts in the structures of zinc finger–DNA complexes, and these presumably are important in the energetics of zinc finger–DNA recognition, but their role in determining sequence specificity is not yet known. The majority of the phosphate contacts are made to the primary strand of the DNA, thereby securing the finger to the strand that also receives the majority

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

of the base contacts. However, there are only three phosphate contacts that are conserved in the majority of the structures, and there is only a modest correlation between the presence (or absence) of a particular phosphate interaction and the docking arrangement (canonical or nonstandard) of the finger (117). The most conserved phosphate contact (observed ~80% of the time) is made by the histidine (through Nδ) at position 7 of the helix to the phosphate just to the 5′ side of the finger binding site. Because zinc is coordinated through Nε of the same residue, this interaction brings the core of the finger in close proximity to the DNA strand that is to be recognized. The fingers in which this contact is absent (TFIIIA finger 1, GLI fingers 3 and 5, TTK finger 1, and YY1 finger 4) tend to have large deviations from the canonical helix docking geometry (Figure 4), and all display nonstandard base recognition patterns (GLI finger 3 does not even interact with the DNA bases). However, we note that there are other nonstandard fingers that do make this phosphate contact, and thus its presence or absence does not provide any simple way of classifying fingers. The next most conserved phosphate contact (present in ~60% of the structures) involves the lysine or arginine at position 1 of the second β-strand (two residues beyond the second cysteine). A positively charged residue at this position of the β-strand also is strongly conserved among known zinc finger sequences (57). This phosphate contact also involves the primary strand of the DNA, and this basic residue contacts either the same phosphate contacted by histidine 7 or the phosphate just to the 3′ side of this position. In those structures where a direct contact is not observed, this lysine or arginine still is near the phosphate backbone, and it may contribute to binding. The third conserved phosphate contact is made from the linker region between fingers. As discussed below, when consecutive fingers contact the DNA the intervening linker tends to have the consensus sequence TGEKP, and the lysine (or arginine) at the fourth position of this linker usually makes a direct or water-mediated contact with a phosphate, typically one on the primary strand of the DNA (37).

Phosphate contacts are made by other positions in the zinc finger structure, but these contacts are poorly conserved among the various complexes. When tyrosine occupies either of the conserved aromatic positions, its hydroxyl group can make hydrogen bonds the phosphate backbone. Amino acids at positions −2, 1, and 5 of the α-helix also can make phosphate contacts, although contacts from positions 1 and 5 require long side chains when the fingers dock in a canonical arrangement. Amino acids at positions −1 and 2, which typically interact with the DNA bases, also can make direct or water-mediated contacts with the phosphate backbone. Even this list of phosphate contacts is not exhaustive: Because a large proportion of the $\beta\beta\alpha$ structure is so close to the DNA, and because a number of different docking arrangements are observed, phosphate contacts can be made from a number of positions on the zinc finger.

Linkers

The linker region that connects neighboring Cys_2His_2 zinc fingers is an important structural element that helps control the spacing of the fingers along the DNA

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

site. The most common linker arrangement has five residues between the final histidine of one finger and the first conserved aromatic of the next finger. In the Transcription Factor Database (47), we find that roughly half of the fingers with this linker length match a consensus sequence of the form TGEKP. Among these linkers the consensus actually is so strong that the degenerate DNA sequence that encodes this linker has been used to identify new zinc finger proteins by hybridization or RT-PCR (1, 54). Mutagenesis studies of the TGEKP linkers in ADR1 (25, 120) and TFIIIA (17, 23, 107) have demonstrated that they are important for high-affinity DNA binding. Some point mutations result in 10–100-fold reductions in DNA-binding affinity when measured in vitro, and mutations in the TGEKP linker also can result in loss of function in vivo (28, 46).

NMR studies indicate that the TGEKP linker between fingers is flexible in the free protein, but becomes more rigid upon binding DNA (13, 44, 126). The TGEKP linker is actually well ordered and similarly organized in the structures of most zinc finger–DNA complexes. With one exception, the TGEKP linkers in the structures (Zif268 and its variants, YY1, 1MEY, GLI, and TFIIIA; 13 in all) overlay with a RMS deviation between 0.15 Å and 0.50 Å and have a very well conserved length falling between 13.9 Å and 14.5 Å. The lone exception is the linker between fingers 4 and 5 of GLI (SNEKP), which contains two changes from the consensus sequence and has a RMS deviation of >1 Å and a length of 11.8 Å. (These differences may be related to the HX_4H pattern that occurs in finger 4 immediately preceding the linker.) In the Zif268 structure, the TG(E/Q)KP linkers have crystallographic B-values (for side chain and backbone) that are similar to the fingers themselves; in the Tramtrack structure there is a nonconsensus linker, KRNVKV, which has very high B-value relative to the fingers, and presumably this reflects a higher degree of disorder and thermal motion.

Examining the conformation of the TGEKP linker in the various protein-DNA complexes allows us to assign a structural role to each of the residues. The linker caps the C terminus of the preceding finger's helix using an α_L motif (3). Threonine provides the C-cap, while glycine assumes a positive ϕ angle that also is needed to complete this cap. As discussed in the next paragraph, glutamate can play a distinctive role in stabilizing finger-finger contacts. The following positively charged residue (lysine or arginine) makes a direct or water-mediated contact to the phosphate backbone. Proline probably rigidifies the connection between the linker and the first β-strand of the subsequent finger, and the proline stacks on the first highly conserved aromatic residue of the next finger. This aromatic subsequently stacks on the main-chain atoms at the N terminus of the α-helix (at position -1), thereby helping to define their conformation.

The docking of adjacent fingers is further stabilized by a contact involving the side chain from position 9 of the preceding finger's helix (typically involving the sequence HXRXH) and the backbone carbonyl or side chain at position -2 of the subsequent finger. This contact appears to be correlated with the use of a canonical linker: When a TGEKP linker occurs between fingers in the Transcription Factor Database (47), there almost always (470 out of 475 examples) are three residues between the two histidines of the preceding finger, and in 80% of these

cases there is an arginine or lysine at position 9. When an arginine occurs at the corresponding position in the known zinc finger–DNA complexes (in the context of a TGEKP linker), it invariably makes an interfinger contact to the backbone carbonyl at position −2. In some structures, the conformation of this arginine is stabilized by interactions with the glutamate from the linker. The highly conserved nature of the TGEKP linker and the interfinger contact from position 9 implies that interfinger organization is important in DNA recognition.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

Biochemical Information about Zinc Finger–DNA Interactions

Cys_2His_2 zinc finger proteins often bind their DNA target sites with high affinity and specificity. In general, observed DNA affinities increase as the number of fingers increases from one to two to three. Proteins containing three fingers, such as Zif268 and SP1, bind their preferred sequences with dissociation constants typically between 10^{-8} M and 10^{-11} M (depending on the buffer conditions and assay methods) (7, 36, 50, 104, 113, 118, 124). These proteins also display good specificity for their binding sites (as determined by DNA site selections or by competition with nonspecific DNA) (113, 118, 124), and the arginine ⇒ guanine contacts often provide highly specific interactions. Substituting alanine for arginine at either position −1 or 6 in finger 1 of Zif268 revealed that each arginine contributes about 3 kcal/mol of binding energy (36). Changing the aspartate at position 2 to alanine or the glutamate at position 3 to alanine results in much smaller changes to the affinity of the protein (the aspartate-to-alanine change is actually energetically favorable), but these acidic residues still do play a role in determining specificity.

It is not yet known how the stability of a zinc finger may affect the affinity and specificity of DNA binding, but a study by Shi & Berg (113) suggests that this may have an effect. They altered the sequence of the fingers of SP1 (except for those residues involved in DNA recognition) to match a consensus sequence that had been developed from a database of fingers (73) and that coordinates zinc with a higher affinity than does finger 3 of SP1 (73, 101). Since zinc coordination is coupled with folding, this suggests that the consensus backbone is more stable than finger 3 of SP1. Shi & Berg found that this new protein displayed improved affinity (sixfold) and specificity relative to SP1. Conversely, reduced affinity and specificity were observed if the sequence at most positions in a single finger was changed to alanine (retaining only the conserved and DNA-binding amino acids). This reduces the affinity of the finger for zinc and can result in some heterogeneity in its coordination (84).

CD studies have demonstrated that binding of Zif268 induces a conformational change in the DNA (37) that is consistent with induction of the enlarged-major-groove conformation observed in the structures of the zinc finger–DNA complexes (37, 89). [Analysis of the crystal structures shows that the enlarged major groove

results from a combination of negative base pair displacement and unwinding of the DNA (89).] Biochemical studies of supercoiling levels also have shown that zinc finger binding unwinds the DNA by approximately 18° per finger (114). Unwinding may limit the number of fingers that contact neighboring subsites: With TGEKP linkers, binding energy tends to plateau after three fingers (103, 112), but further studies are needed to understand the basis of this effect.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

SELECTING ZINC FINGERS THAT RECOGNIZE NOVEL DNA SITES

One of the most striking features observed in the Zif268 complex (95) involved the conserved pattern of base contacts in the tandemly linked fingers (using residues at positions −1, 2, 3, and 6 of the helix). Given that this family of proteins was known to recognize a variety of different sequences and that one might be able to "mix and match" fingers for new sites (88), these proteins provided an attractive framework for design efforts. Initial attempts to rationally alter the specificity of zinc finger proteins were based on sequence and structural comparisons of zinc fingers (30–33, 61, 87, 121). This approach met with some success. Selection by phage display (6, 116) provided another potential method for finding sequences (from a library of randomized fingers) that might recognize a desired target site. Because this method begins from an "unbiased" library (by fully randomizing key recognition positions), it can provide new information about DNA recognition by zinc fingers while revealing which amino acid sequences are best for a given site. Using the Zif268 framework and randomizing potential base-contacting residues, many of the initial studies succeeded in recovering fingers with novel specificities (18, 19, 58, 59, 104, 125). These results provided new information about the best finger sequences for recognizing a given DNA subsite, but these successes only involved purine-rich sites. Recent selection studies have focused on the recognition of a broader range of sequences and on generating functional proteins that recognize entirely novel sites.

Greisman & Pabo developed a sequential selection protocol that changes all three fingers of a protein, selecting one finger at a time while "walking" across the binding site in three stages (50; Figure 5): *Stage A*) In the first step, finger 1 is selected over the 3′ portion of the binding site while held in place by two Zif268 anchor fingers that recognize a DNA sequence fused to the target site. *Stage B*) In the second step, one anchor finger is discarded and an additional random finger 2 library is attached to the selected finger 1 clones. Finger 2 is then selected to bind to the central portion of the target site. *Stage C*) In the third step, the remaining anchor finger is discarded from the finger 1–finger 2 clones. A random finger 3 library is attached to these fingers and finger 3 is selected to recognize the 5′ portion of the target site. This process attempts to ensure the compatibility of neighboring fingers in DNA recognition by carrying a small pool of clones from one stage to the next (such that one finger can be reoptimized as the next is added), and

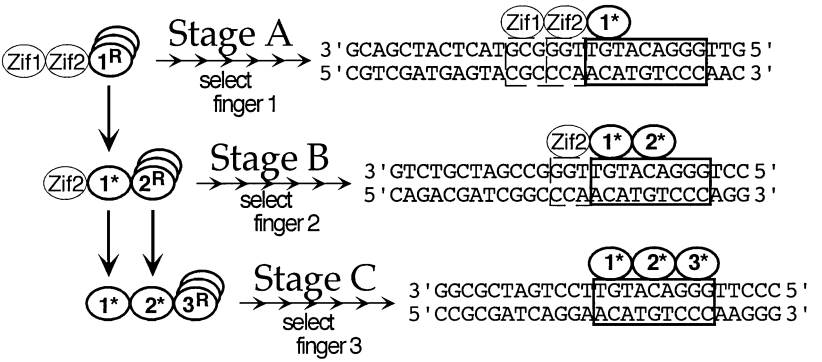

Figure 5 Overview of the sequential selection protocol (stages A, B, and C) that successively optimizes fingers 1, 2, and 3 to create a new zinc finger protein (50). The *left side* of the diagram indicates the constructs that are displayed in the phage libraries, and the *right side* shows fingers remaining after multiple rounds of selection and amplification (indicated with *small horizontal arrows*). *Zif1* and *Zif2* denote wild-type Zif268 fingers; the superscript *R* denotes a randomized finger library; and an *asterisk* denotes the set of selected sequences. The nine base pair recognition sequence that the fingers are selected against is *boxed*, as is the subsite for each Zif268 anchor finger. The set of fingers selected in one stage is incorporated into the phage libraries used in the next stage of selection, allowing a final optimization of previously selected fingers in their new context.

the process helps ensure that each finger is selected in the most relevant structural context. Using this method, proteins were selected for three biologically important control sites with very different G·C content. Subsequent studies have confirmed that these proteins have the desired DNA-binding specificity (124), demonstrating that the Zif268 framework can be adapted for recognition of many different DNA sequences (59).

Segal et al used a different strategy to address concerns about context dependence, choosing to focus on the development of a set of fingers that recognize each of the 16 possible 5′-GNNG-3′ sequences (110). When sites of this form are combined to give any extended site of the form GNNGNNGNN, the overlap of neighboring subsites is always at a G·C base pair, and the residue at position 2 will readily be able to make the expected contacts with the flanking G·C base pair in the neighboring subsite. (Thus the four base pair subsites will always mesh with the three base pair repeat.) After the initial selection of these fingers by phage display, other variants were tested in an attempt to further improve the specificity of the fingers, and most of the resulting fingers display good discrimination against other sequences. Because of the large number of sites that were successful targeted, the sequences of these fingers provide further information about the preferred residues for DNA recognition at positions −1, 1, 2, and 3 (at least in the context of the arginine that is retained at position 6). There is also some evidence for cooperation between these residues in DNA recognition. Beerli et al have shown that fingers from this library can be assembled to target many G·C-rich sites (7).

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

Isalan et al took a rather different approach to deal with the subsite/subsite interface: Rather than fixing the identity of the base at this position, they randomized residues on both sides of the finger/finger interface that could contact this region (56). Most previous studies had been influenced by the use of Zif268 anchor fingers: Because these fingers contain aspartate at position 2 they tend to create a preference for G or T under position 6 of the preceding finger due to the partial overlap of recognition sites. Isalan et al prepared a library in which position 6 of finger 2 and positions −1, 1, 2, and 3 of finger 3 were randomized (again using a Zif268-based construct), and then used this library of selected proteins for all 16 possible dinucleotide sequences under position 6 of finger 2 and position –1 of finger 3. They recovered fingers that were specific for 15 of the 16 possible combinations, and specificity for the remaining junction sequence has been obtained by Greisman & Pabo (50, 124). Thus it appears that the specificity of zinc fingers is not inherently limited by structural requirements at the subsite/subsite interface (although more data still are needed on possible variations in affinity of these proteins). Because of the large number of sites used in this study, the resulting data also helped to clarify which amino acids at positions −1 and 6 define a given sequence specificity. Unfortunately the role of position 2 in sequence specificity is still poorly understood, and there is not yet any simple correspondence between amino acid type and observed sequence preference.

PREDICTING ZINC FINGER SPECIFICITY

As the body of data from zinc finger selections continues to grow, it becomes increasingly important to compile it in a manner that would facilitate the design of new zinc finger proteins. If relevant patterns could be recognized and such sequence motifs readily reused, the need for time-consuming phage display selection methods when creating new zinc finger proteins might be reduced. In principle, comparisons should be simplified by the fact that most selections have used a Zif268-like framework and have focused on variations of many of the same key residues.

The first attempts to predict zinc finger specificity focused on the idea of a recognition code that would correlate specific residues in the recognition helix with specific bases in the subsite (18, 30, 31). Past successes in altering finger specificity, and the analysis of natural zinc finger proteins, revealed some significant patterns in the observed side chain–base interactions. These patterns (which assume a canonical binding geometry for each finger) have been compiled into a "recognition code" that attempts to break down the contacts between the finger and DNA into a chart of 1:1 interactions between specific positions on the helix and specific base pairs in the finger-recognition site (Figure 6). Clearly, approximations are involved when analyzing the data in this way: Structural studies, mutagenesis, statistical analysis of sequences, and design studies all show that the amino acids at positions −1, 2, 3, and 6 do not play fully independent roles in DNA recognition (20, 30, 31). Nonetheless, the recognition code contains useful

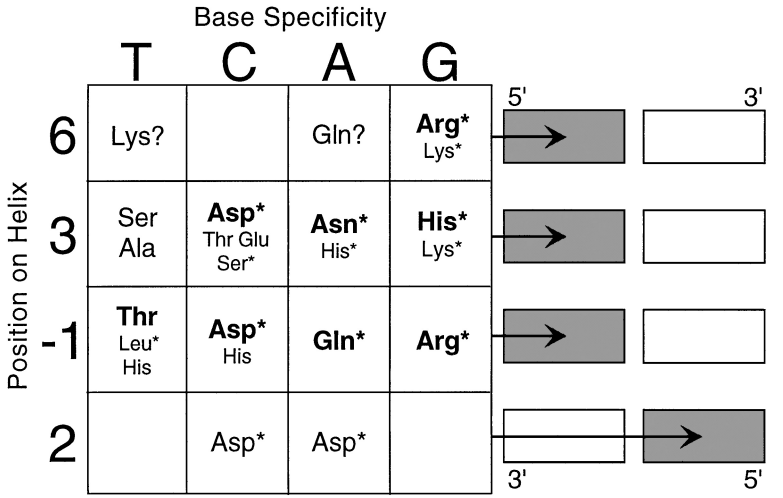

Figure 6 Pattern of side chain–base interactions that provide an approximate "recognition code" for zinc fingers that have a canonical binding mode. This chart describes contacts between residues at key positions in the α-helix (-1, 2, 3, and 6) and bases at the corresponding positions in the canonical subsite (cf *upper panel* of Figure 3). *Boldface type* highlights amino acids that occur most frequently in phage display selections when a particular base specificity is desired, and an *asterisk* indicates contacts that have been observed in structural studies. Question marks (?) indicate that the specificity of the respective amino acid/base contact is uncertain. Positions for which base specificity is largely undefined are left blank. This way of representing the contacts ignores critical side chain–base interactions (such as the Asp(2) $\Rightarrow$ Arg(-1) interaction in Zif268) that are not taken into account in this chart.

information and there are cases in which it has strong predictive power (an arginine at position -1 usually does specify guanine). In the absence of other information, the recognition code provides a good place to start when attempting to predict or design the specificity of a new finger. Proteins containing three fingers have even been designed using the recognition code (26, 27), and they do bind their target sequence, although their specificity has not always been optimal.

Interpreting or evaluating the recognition code becomes more difficult in cases where several different amino acids might be used to recognize a particular base. Residues that appear in the same block of the code (Figure 6) are not always interchangeable, and their utility may depend on the context. For example, Segal et al found many situations in which a histidine at position 3 provides the best specificity for guanine, but in one instance lysine clearly gave better discrimination (110). The factors that determine such preferences are not yet known, and this complicates the use of a recognition code in design. The code may not fully account for factors such as: (a) side chain–side chain interactions, (b) sequence-dependent conformational flexibility of DNA, (c) the role of water in recognition,

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

(d) how particular contacts may subtly affect the docking arrangement, and (e) the effects of neighboring fingers and subsites on recognition.

A more conservative approach for the design of novel zinc finger proteins might employ a database that correlates an entire finger sequence with a given four base pair subsite (20). In principle, such a database would list fingers suitable for each of the 256 four base pair subsites, and such fingers might be combined to give new proteins for a desired nine-to-ten base pair site. This database might avoid some of the approximations inherent in a simple "recognition" code because side chain–side chain interactions, water-mediated contacts, and subtle changes in the docking arrangement could more readily be accounted for. There is some evidence that fingers displaying specificity for a given subsite in one context will also function similarly in other contexts. Thus many of the fingers that were obtained using the sequential selection protocol are similar in sequence to natural fingers that recognize very similar or identical four base pair subsites (50, 124). One excellent example of this conservation of finger specificity comes from finger 3 of a zinc finger protein (NRE), which was selected against the sequence 5′-AAGG-3′ (50). The consensus sequence of the finger generated by phage display is identical at the base-recognition positions to finger 2 of Tramtrack, which happens to recognize the same DNA sequence. The modularity of fingers is also a fundamental assumption underlying the approach of Segal et al: Their project involved developing a "database" of fingers that will recognize each of the 16 possible 5′-GNNG-3′ sequences.

Although some type of database may eventually prove more useful than a "code," there still are practical limits in implementing this strategy. Even if we simplify the problem by focusing on the core triplets (ignoring the contribution of the residue at position 2 for the moment) one should note that the most of the reported fingers with defined specificities recognize GNN triplets. (Most of the successful selections have involved sites of this form.) For the majority of the remaining 48 triplets, no finger with the desired specificity has been reported. It is not yet known to what extent this bias represents an intrinsic preference of fingers for GNN sequences or just the limits in the range of sites that have been tested. Assigning the specificity of an individual finger from a protein (so that it can be properly entered into the database) can also be complicated: Distinguishing the contribution of specific residues at the finger interface (separating the role of positions 6 and 2 from neighboring fingers) can be problematic since the determinants of specificity for position 2 are as yet poorly defined. Furthermore, the sequence specificity and affinity of fingers in the database may vary to a large degree, and it is important to recognize that not all fingers or triplets may make equal contributions to specificity.

Even the idea of a consistent finger/subsite correlation may have limits; there are some examples in which the same finger sequences seem to have context-dependent site preferences. For example, the selection (by Greisman & Pabo) of a zinc finger protein that would recognize the p53 site generated a finger containing the sequence QGTR (positions −1, 2, 3, and 6) that recognizes the triplet ACA (50, 124).

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

However, based on the recognition code, an arginine at position 6 would be expected to specify guanine rather than adenine. Indeed, in another context the QGTR finger does indeed recognize the triplet GCA (110). Both fingers were selected as the middle finger in a three finger protein, but the recognition helices on the neighboring fingers are much different. Presumably a difference in context, most likely the amino acid at position 2 of finger 3 or a somewhat different docking arrangement (which may be allowed by the Greisman & Pabo selections), is responsible for the divergent specificity of these fingers. Another interesting example of this phenomenon involves the recognition sequence RSDELVR (positions −1 through 6) engineered by Segal et al to bind to the triplet GTG while discriminating against GCG (110). This same recognition sequence occurs in WT1 in a different context (finger 4 instead of finger 2) and modestly prefers GCG over GTG (52). A very similar sequence (RSDELTR) also occurs in Zif268 (finger 1) and GLKF (finger 2), and these proteins also display a preference for GCG over GTG (115, 118, 124). More data will be needed to understand these effects, but these examples suggest that the context provided by neighboring fingers and subsites may affect the specificity of a finger. This may put inherent limits on the use of a database as a design tool. Thus a database may ultimately prove more useful than a 1:1 code, but it still may be necessary to try several different combinations or use some selection steps to optimize affinity.

The influence of context dependence on recognition (involving a finger's position within a protein and the sequence of any neighboring fingers) has been examined for two fingers that were created using the sequential selection protocol (124). During this protocol, finger 1 is initially selected to recognize its sequence as the C-terminal finger in a three finger construct, but it must function in the final construct as an N-terminal finger (Figure 5). For two of the proteins originally generated by this protocol, finger 1 was randomized and reselected (now as an N-terminal finger) to explore the possibility of context-dependent effects in the selection protocol. In both cases, the preferred residues at positions −1, 1, and 2 changed, implying that the best residues for DNA recognition can depend on the position of a finger in the protein or the effect of neighboring fingers. Although only a modest improvement in DNA affinity (∼8-fold) was observed even in the more dramatic case, this type of improvement could be important for biomedical applications.

CREATING NOVEL TRANSCRIPTION FACTORS FROM ZINC FINGER PROTEINS

Cys_2His_2 zinc finger proteins can function as the DNA-binding domains of constructs designed to serve as biochemical or biomedical tools, and such proteins may eventually prove useful for gene therapy (98, 106, 127). Because of the wide variety of sequences that can be targeted with zinc fingers, they appear to be the most promising DNA-binding domains for this purpose. Attaching additional domains—for activation, repression, or enzymatic activity—should allow such

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

proteins to carry out the desired function in a site-specific manner. However, several key issues must be addressed before this type of zinc finger chimera can be used for human gene therapy. These issues involve: (a) delivery, (b) adequate specificity in vivo for a desired target sequence, (c) ability to function effectively in vivo, and (d) evasion of immune system surveillance. It would also be useful if one could control the activity of the zinc finger construct via a bioavailable compound.

Preliminary studies involving issues of specificity and in vivo functioning seem quite promising. Thus it has been shown that Cys_2His_2 zinc finger proteins containing three designed or selected fingers can display sufficient affinity and specificity to function in vivo (7, 16, 21, 26, 64, 65, 79). Discrimination at the single-base pair level, in a transient transfection assay, can be achieved if the expression level of the protein is carefully controlled (16). However, it is not yet known whether three fingers will provide sufficient specificity for applications in gene therapy. (Statistically, one expects that there will be ~10,000 identical nine base pair sites present in 3 billion base pairs of the human genome, and "side effects" caused by action at many different sites may be a significant problem.) In principle, the degree of discrimination within the genome can be improved by increasing the number of fingers (via covalent linkage or dimerization), using several three finger proteins simultaneously for synergistic activation, or by using fingers that dock in a nonstandard manner (such as finger 4 of GLI) to specify more base pairs per finger.

Increasing the number of covalently linked fingers would seem to be the obvious solution for improving specificity and targeting a particular site, but there appears to be a limit to the number of fingers that can be connected with TGEKP linkers and still bind DNA with a canonical docking arrangement. Proteins that contain three fingers (Zif268, its variants, and 1MEY) have a very regular arrangement on the DNA, but this does not seem to be the case for proteins with more than three fingers (YY1, GLI, TFIIIA), and even in these larger proteins there are, at most, three consecutive fingers that dock with the DNA in a canonical manner. Several groups have found that proteins with four or five fingers linked by the canonical TGEKP sequence display only very slight improvements in affinity relative to a three finger subset of the same construct (103, 112). Proteins containing six fingers that all are connected by TGEKP linkers display modest improvements in affinity (~70-fold) over the three finger constructs, but this increased binding energy falls far short of the anticipated gains based on ideas about the chelate effect and simple effective concentration calculations (7, 68, 79). Even extending to a nine finger protein with the canonical linker appears to provide little improvement in affinity over just three fingers (60).

It appears that the affinity of such six finger proteins can be improved by including one longer linker at the center of the construct (65). Thus, by using either a LRQKDGERP or LRQKDGGGSERP linker between the two three finger proteins, a 6000–90,000-fold improvement in binding affinity (over either individual three finger protein) was achieved. These constructs are about 70-fold more

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

specific than Zif268 when their discrimination against nonspecific DNA is compared (specificity ratio ~1 in 10^7). These constructs should provide superior specificity to a three finger protein when targeting a site in the human genome, but they still may bind to many other sites (one would like specificity at the level of 1 in 3×10^9). Binding at even a small number of alternative sites may still be a problem for gene therapy.

Dimerization offers an interesting alternative to covalent linkage as a means for assembling more than three fingers at a target site. The cooperative association of two three-finger proteins on a DNA-binding site has several potential advantages over their covalent linkage: this may provide (a) a faster rate of equilibration with sites on genomic DNA (due to a lower inherent nonspecific affinity of each monomer), (b) a sharper transition as a function of protein concentration between the fully bound and unbound states, and (c) perhaps a greater degree of specificity (102). An artificial dimerization construct, which allows two different sets of zinc fingers to bind cooperatively to an asymmetric DNA site, was made by fusing the C-terminal coiled-coil domain of Gal4 to different sets of zinc fingers (100). N-terminal peptide sequences also have been selected that allow pairs of fingers to cooperatively bind DNA (122), and these might be used to allow recognition of extended sites. In short, it appears that dimerization may provide an interesting alternative to covalent linkage, but further data will be needed to see which methods are most useful in specific contexts.

Structure-based design (which begins by assembling known structural modules on a computer graphics system) has been used to create ZFHD1, a chimeric DNA-binding domain that shows promise for use in human gene therapy constructs. ZFHD1 was created by fusing fingers 1 and 2 of Zif268 to the Oct-1 homeodomain (99). The fusion protein specifically recognizes a composite binding site and displays good specificity in vitro and in vivo. ZFHD1 has been incorporated as the DNA-binding module in a prototype gene therapy system that allows drug-regulated expression of an added gene (containing upstream ZFHD1 binding sites). This drug-regulated expression system functions in human cells that have been introduced into mice (106) and has recently been used in an adeno-associated virus gene delivery system in monkeys (127). These exciting advances already suggest that zinc finger DNA-binding proteins may be useful in human gene therapy.

Chimeric proteins constructed using structure-based design also have been used as biochemical tools. Thus the ZFHD1 fusion protein has been used to study the role of the Oct-1 homeodomain in C1 complex assembly by herpes simplex virus (98). (ZFHD1 allows stable binding of the homeodomain in the absence of the POU-specific domain of Oct-1, and thus allows the roles of those domains to be clearly distinguished.) In other studies, zinc fingers have been linked to the cleavage domain of Fok I to create a restriction enzyme with sequence specificity that is defined by the zinc fingers (66, 67), and the fusion of zinc fingers to TBP has been used to create a chimera that could provide a novel method to regulate the expression of desired endogenous genes (63). Given that most eukaryotic transcription factors seem to be assembled from independently functioning domains, the use

of zinc fingers in a structure-based design approach should provide a powerful method for the creation of new DNA-binding proteins with novel functions.

UNANSWERED QUESTIONS

Exciting progress has been made in the study of zinc finger proteins, but important questions remain. As noted in the introduction, relatively little is known about zinc finger–RNA interactions or about protein-protein interactions involving zinc fingers. More information is needed about how often zinc fingers adopt such roles. What is the structural basis for such contacts? What are the biological roles for these interactions? Do such zinc finger–zinc finger and zinc finger– RNA interactions tend to occur in larger polyfinger proteins that also have DNA-binding fingers? Are certain fingers specialized for such roles? Are there fingers with multiple alternative roles in different contexts? Are there other zinc fingers that have lost binding function over evolutionary history and now just play a more passive structure role?

Many of the studies of zinc finger–DNA interactions have focused on analyzing base contacts that are made by three finger Zif-like proteins. For these designs, perhaps the most important remaining question involves the potential limits of sequence specificity that can be achieved by varying residues in the recognition helix. To date, the majority of fingers created by selection or design recognize GNN triplets. While it is clear that zinc fingers can recognize a wide variety of different sequences, the limits are not yet known, and certain sites may be inherently more suitable than other sites. In particular, pyrimidine-rich sequences seem especially problematical, and it may not always be suitable to simply use the other binding orientation (thereby selecting for contacts along the purine-rich strand). Since zinc fingers seem to prefer a distinctive DNA conformation, sequence-dependent aspects of DNA structure could also affect the affinity and specificity of the zinc fingers that can be selected for a particular sequence. Although it is hard to interpret failures, there are good examples of thoughtful selection efforts that have failed to obtain desired finger specificities (110, 124). This may reflect the inherent difficulties with certain sequences or just indicate that other residues need to be randomized. More information is needed about the potential roles of the backbone residues, phosphate contacts, and the linker in defining the specificity of a finger. Thus experiments (113) with the consensus backbone demonstrate that other residues not directly involved in DNA recognition can have an impact on specificity and affinity. Changing other parts of the finger structure or using fingers (like GLI) with nonstandard recognition sites may improve the specificity for some of these "difficult sequences." It also will be interesting to see how readily one can create mixed proteins that have fingers with canonical and noncanonical docking arrangements. The issue of DNA recognition by polyfinger proteins also remains a challenging problem. What is the source of the energetic penalty that seems to limit the affinity when there are TGEKP linkers connecting more

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

than three consecutive fingers? Perhaps zinc fingers connected by TGEKP linkers adopt a helical arrangement (when bound to DNA) that does not quite match the helical pitch of the DNA, and strain accumulates as more fingers are added. Using a longer linker between consecutive finger segments appears to alleviate some of this penalty, but these polyfinger proteins still do not bind as tightly as simple physical/chemical calculations would suggest.

Even for base contacts involving canonical docking arrangements, there is a considerable difference in the degree to which we understand the basis of DNA recognition from different positions in the helix, and there are questions about the utility of any simple "recognition code." Some of the greatest uncertainties involve the role of the residue at position 2, and the corresponding questions about interactions between neighboring fingers and subsites. There are many selection studies that have not obtained a consensus amino acid at position 2, and there may be cases in which the residue at this position plays only a peripheral role in DNA recognition. It is interesting that aspartate at position 2 is important in defining the finger specificity of Zif268 (36), but we note that serine, which does not appear to display a preference for any particular nucleotide, occurs most frequently at this position in natural proteins (57). There may be other cases in which the contribution of position 2 to finger specificity will be less direct, but still could be quite important. Thus a number of selections have found a preference for glycine at this position in certain DNA sequence contexts (50, 104, 110, 124). In the structure of QGSR (35), glycine at position 2 has α-helical ϕ/ψ values (-58 and -38; 105), implying that flexibility at this position is not important. Instead it appears that the absence of a β-carbon at position 2 is beneficial for binding at this site, allowing the main chain of the α-helix to approach the DNA more closely.

There also are important questions about the range of useful contacts that can be made from position 6 with a canonical docking arrangement, and it is still unclear how this position can be used to specify a base other than guanine (Figure 6). (A relatively small number of successful selections have a base other than G at the 5′ position.) In part, understanding specificity at this position may also be complicated by the overlapping contacts that are made by residue 2 of the next (C-terminal) finger. However, there may also be geometric limitations imposed by the position and the orientation of the α-helix in the canonical docking arrangement. Thus the average distance (among the structures of canonical fingers) from the Cα of position 6 to the nearest heavy atom of the nucleotide that is contacted in the canonical docking arrangement is 8.8 ± 0.8 Å, a distance too large to be spanned by the majority of amino acids. This may limit the utility of position 6 in specifying bases other than guanine (which can readily be contacted with lysine or arginine). More structural data also are needed, since to date no contact between a residue at position 6 and a base other than guanine has been observed in a three-dimensional structure.

In general, DNA recognition by residues at positions -1 and 3 has been studied much more thoroughly than at the other positions, but even for these more carefully explored positions there are paradoxes that remained unexplained. For

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

instance two selection studies have shown that serine at position 3 seems to prefer thymine as the central base (18, 19, 110), but in the QGSR Zif variant, the serine at position 3 specifies cytosine (EI Ramm, SA Wolfe, & CO Pabo, unpublished results). There also are important questions in cases where the "code" seems to suggest several side chains that might be used (this happens very frequently for position 3). Do these side chains represent relatively iso-energetic alternatives, or will the choice of residue depend on the structural context? Similar questions arise with respect to many other context-dependent effects. How often can fingers, which were designed or selected in one context, be "mixed and matched" to generate new specificities? Does the position of a finger within a protein affect the preferred residues for DNA recognition, or is context dependence purely a function of the neighboring finger and DNA sequences? Will a "database" approach, which tries to list fingers that are suitable for given three base pair or four base pair subsites, prove fundamentally more reliable than a "recognition code" (Figure 6) that tries to correlate specific residues in the recognition helix with particular bases in the site? Much remains to be learned, and the problems become even more challenging and interesting if we try to expand our analysis to include all the possibilities that become accessible via the addition of variant backbone structures, altered phosphate contacts, and nonstandard docking arrangements.

SUMMARY

Structural and biochemical studies have given us a wealth of information about zinc finger–DNA interactions. These fingers have a conserved sequence pattern that stabilizes the $\beta\beta\alpha$ fold, but they can present very different sequences on the surface of the α-helix and also can dock in the major groove with a variety of different orientations. Most design efforts have focused on fingers with a canonical (Zif-like) docking arrangement, and fingers with new specificities have been selected by phage display or designed using information about known interactions. These methods have allowed the construction of entire proteins with novel DNA-binding specificities, and phage display studies also have provided a considerable body of data about side chain–base interactions. Zinc fingers appear to provide a very powerful framework for the selection and design of new proteins, but phage display may still be required for finding the optimal contacts. The complexity of the protein-DNA interface suggests that no simple "recognition code" will ever provide a reliable, general method for designing proteins with optimal affinity for new sites. Developing a database of fingers that recognize particular subsites may be more useful, but much additional information is needed, and there still are important questions about the significance of context-dependent effects.

As answers to these and other unresolved questions are obtained, they will facilitate the ease with which zinc fingers are fused into site-specific transcription factors designed for a specific purpose, whether that is as part of a gene therapy

construct or biochemical tool. As selection methods are improved, and as further biochemical and structural information is obtained, it seems quite possible that zinc finger proteins with almost any desired DNA specificity could be obtained, and this would have very exciting implications for biological research and gene therapy.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

ACKNOWLEDGMENTS

We would like to thank B Wang, J Miller, and JK Joung for their comments on this manuscript. SAW was supported by the Howard Hughes Medical Institute and as a Special Fellow of the Leukemia Society of America; COP and LN were supported by the Howard Hughes Medical Institute.

Visit the Annual Reviews home page at www.AnnualReviews.org

LITERATURE CITED

1. Agata Y, Matsuda E, Shimizu A. 1998. Rapid and efficient cloning of cDNAs encoding Kruppel-like zinc finger proteins by degenerate PCR. *Gene* 213:55–64
2. Andreazzoli M, De Lucchini S, Costa M, Barsacchi G. 1993. RNA binding properties and evolutionary conservation of the *Xenopus* multifinger protein Xfin. *Nucleic Acids Res.* 21:4218–25
3. Aurora R, Rose GD. 1998. Helix capping. *Protein Sci.* 7:21–38
4. Bardeesy N, Pelletier J. 1998. Overlapping RNA and DNA binding domains of the wt1 tumor suppressor gene product. *Nucleic Acids Res.* 26:1784–92
5. Bardwell VJ, Treisman R. 1994. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev.* 8:1664–77
6. Bass S, Greene R, Wells JA. 1990. Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins* 8:309–14
7. Beerli RR, Segal DJ, Dreier B, Barbas CF 3rd. 1998. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl. Acad. Sci. USA* 95:14628–33
8. Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, Martial JA. 1991. The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc. Natl. Acad. Sci. USA* 88:3608–12
9. Berg JM, Godwin HA. 1997. Lessons from zinc-binding peptides. *Annu. Rev. Biophys. Biomol. Struct.* 26:357–71
10. Berg JM, Shi Y. 1996. The galvanization of biology: a growing appreciation for the roles of zinc. *Science* 271:1081–85
11. Bird AJ, Turner-Cavet JS, Lakey JH, Robinson NJ. 1998. A carboxyl-terminal Cys2/His2-type zinc-finger motif in DNA primase influences DNA content in *Synechococcus* PCC 7942. *J. Biol. Chem.* 273: 21246–52
12. Bohm S, Frishman D, Mewes HW. 1997. Variations of the C2H2 zinc finger motif in the yeast genome and classification of yeast zinc finger proteins. *Nucleic Acids Res.* 25:2464–69
13. Bowers PM, Schaufler LE, Klevit RE. 1999. A folding transition and novel zinc finger accessory domain in the transcription factor ADR1. *Nat. Struct. Biol.* 6:478–85

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

14. Brown RS, Sander C, Argos P. 1985. The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS Lett.* 186:271–74
15. Caricasole A, Duarte A, Larsson SH, Hastie ND, Little M, et al. 1996. RNA binding by the Wilms tumor suppressor zinc finger proteins. *Proc. Natl. Acad. Sci. USA* 93:7562–66
16. Choo Y, Castellanos A, Garcia-Hernandez B, Sanchez-Garcia I, Klug A. 1997. Promoter-specific activation of gene expression directed by bacteriophage-selected zinc fingers. *J. Mol. Biol.* 273: 525–32
17. Choo Y, Klug A. 1993. A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA. *Nucleic Acids Res.* 21:3341–46
18. 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* 91:11168–72
19. 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* 91:11163–67
20. Choo Y, Klug A. 1997. Physical basis of a protein-DNA recognition code. *Curr. Opin. Struct. Biol.* 7:117–25
21. Choo Y, Sanchez-Garcia I, Klug A. 1994. In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* 372:642–45
22. Clarke ND, Berg JM. 1998. Zinc fingers in *Caenorhabditis elegans*: finding families and probing pathways. *Science* 282:2018–22
23. Clemens KR, Zhang P, Liao X, McBryant SJ, Wright PE, Gottesfeld JM. 1994. Relative contributions of the zinc fingers of transcription factor IIIA to the energetics of DNA binding. *J. Mol. Biol.* 244:23–35
24. Coleman JE. 1992. Zinc proteins: enzymes, storage proteins, transcription factors, and replication proteins. *Annu. Rev. Biochem.* 61:897–946
25. Cook WJ, Mosley SP, Audino DC, Mullaney DL, Rovelli A, et al. 1994. Mutations in the zinc-finger region of the yeast regulatory protein ADR1 affect both DNA binding and transcriptional activation. *J. Biol. Chem.* 269:9374–79
26. Corbi N, Libri V, Fanciulli M, Passananti C. 1998. Binding properties of the artificial zinc fingers coding gene Sint1. *Biochem. Biophys. Res. Commun.* 253:686–92
27. 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
28. Crozatier M, Kongsuwan K, Ferrer P, Merriam JR, Lengyel JA, Vincent A. 1992. Single amino acid exchanges in separate domains of the *Drosophila* serendipity delta zinc finger protein cause embryonic and sex biased lethality. *Genetics* 131:905–16
29. Davies R, Moore A, Schedl A, Bratt E, Miyahawa K, et al. 1999. Multiple roles for the Wilms' tumor suppressor, WT1. *Cancer Res.* 59:S1747–50
30. Desjarlais JR, Berg JM. 1992. Redesigning the DNA-binding specificity of a zinc finger protein: a data base-guided approach. *Proteins* 12:101–4
31. Desjarlais JR, Berg JM. 1992. Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc. Natl. Acad. Sci. USA* 89:7345–49
32. Desjarlais JR, Berg JM. 1993. Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA* 90:2256–60
33. Desjarlais JR, Berg JM. 1994. Length-encoded multiplex binding site determination: application to zinc finger proteins. *Proc. Natl. Acad. Sci. USA* 91:11099–103
34. Dutnall RN, Neuhaus D, Rhodes D. 1996.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

The solution structure of the first zinc finger domain of SWI5: a novel structural extension to a common fold. *Structure* 4:599–611

35. Elrod-Erickson M, Benson TE, Pabo CO. 1998. High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* 6:451–64
36. Elrod-Erickson M, Pabo CO. 1999. Binding studies with mutants of Zif268. *J. Biol. Chem.* 274:19281–85
37. Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO. 1996. Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure* 4:1171–80
38. Engelke DR, Ng S-Y, Shastry BS, Roeder RG. 1980. Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell* 19:717–28
39. Englert C, Vidal M, Maheswaran S, Ge Y, Ezzell RM, et al. 1995. Truncated WT1 mutants alter the subnuclear localization of the wild-type protein. *Proc. Natl. Acad. Sci. USA* 92:11960–64
40. Fairall L, Harrison SD, Travers AA, Rhodes D. 1992. Sequence-specific DNA binding by a two zinc-finger peptide from the Drosophila melanogaster Tramtrack protein. *J. Mol. Biol.* 226:349–66
41. Fairall L, Schwabe JWR, Chapman L, Finch JT, Rhodes D. 1993. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc finger/DNA recognition. *Nature* 366:483–87
42. Finerty PJ JR, Bass BL. 1997. A *Xenopus* zinc finger protein that specifically binds dsRNA and RNA-DNA hybrids. *J. Mol. Biol.* 271:195–208
43. Finerty PJ Jr, Bass BL. 1999. Subsets of the zinc finger motifs in dsRBP-ZFa can bind double-stranded RNA. *Biochemistry* 38:4001–7
44. Foster MP, Wuttke DS, Radhakrishnan I, Case DA, Gottesfeld JM, Wright PE. 1997. Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. *Nat. Struct. Biol.* 4:605–8
45. Frankel AD, Berg JM, Pabo CO. 1987. Metal-dependent folding of a single zinc finger from transcription factor IIIA. *Proc. Natl. Acad. Sci. USA* 84:4841–45
46. Gaul U, Redemann N, Jackle H. 1989. Single amino acid exchanges in the finger domain impair the function of the *Drosophila* gene Kruppel (Kr). *Proc. Natl. Acad. Sci. USA* 86:4599–603
47. Ghosh D. 1993. Status of the transcription factors database (TFD). *Nucleic Acids Res.* 21:3117–18
48. Green A, Sarkar B. 1998. Alteration of zif268 zinc-finger motifs gives rise to non-native zinc-co-ordination sites but preserves wild-type DNA recognition. *Biochem. J.* 333:85–90
49. Gregory RC, Taxman DJ, Seshasayee D, Kensinger MH, Bieker JJ, Wojchowski DM. 1996. Functional interaction of GATA1 with erythroid Kruppel-like factor and Sp1 at defined erythroid promoters. *Blood* 87:1793–801
50. Greisman HA, Pabo CO. 1997. Sequential optimization strategy yields high-affinity zinc finger proteins for diverse DNA target sites. *Science* 275:657–61
51. Grondin B, Bazinet M, Aubry M. 1996. The KRAB zinc finger gene ZNF74 encodes an RNA-binding protein tightly associated with the nuclear matrix. *J. Biol. Chem.* 271:15458–67
52. Hamilton TB, Borel F, Romaniuk PJ. 1998. Comparison of the DNA binding characteristics of the related zinc finger proteins WT1 and EGR1. *Biochemistry* 37:2051–58
53. Hanas JS, Hazuda DJ, Bogenhagen DF, Wu FY-H, Wu C-W. 1983. *Xenopus* transcription factor A requires zinc for binding to the 5 S RNA gene. *J. Biol. Chem.* 258:14120–25
54. Hoovers JMN, Mannens M, John R, Bliek J, van Heyningen V, et al. 1992. High-resolution localization of 69 potential human zinc finger protein genes: a number

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

are clustered. *Genomics* 12:254–63
55. Houbaviy HB, Usheva A, Shenk T, Burley SK. 1996. Cocrystal structure of YY1 bound to the adeno-associated virus P5 initiator. *Proc. Natl. Acad. Sci. USA* 93:13577–82
56. Isalan M, Klug A, Choo Y. 1998. Comprehensive DNA recognition through concerted interactions from adjacent zinc fingers. *Biochemistry* 37:12026–33
57. Jacobs GH. 1992. Determination of the base recognition positions of zinc fingers from sequence analysis. *EMBO J.* 11:4507–17
58. Jamieson AC, Kim SH, Wells JA. 1994. In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry* 33:5689–95
59. Jamieson AC, Wang H, Kim SH. 1996. A zinc finger directory for high-affinity DNA recognition. *Proc. Natl. Acad. Sci. USA* 93:12834–39
60. Kamiuchi T, Abe E, Imanishi M, Kaji T, Nagaoka M, Sugiura Y. 1998. Artificial nine zinc-finger peptide with 30 base pair binding sites. *Biochemistry* 37:13827–34
61. Kim CA, Berg JM. 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
62. Kim CA, Berg JM. 1996. A 2.2 Å resolution crystal structure of a designed zinc finger protein bound to DNA. *Nat. Struct. Biol.* 3:940–45
63. Kim JS, Kim J, Cepek KL, Sharp PA, Pabo CO. 1997. Design of TATA box-binding protein/zinc finger fusions for targeted regulation of gene expression. *Proc. Natl. Acad. Sci. USA* 94:3616–20
64. Kim JS, Pabo CO. 1997. Transcriptional repression by zinc finger peptides. Exploring the potential for applications in gene therapy. *J. Biol. Chem.* 272:29795–800
65. Kim JS, Pabo CO. 1998. Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc. Natl. Acad. Sci. USA* 95:2812–17
66. Kim YG, Cha J, Chandrasegaran S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. USA* 93:1156–60
67. Kim YG, Shi Y, Berg JM, Chandrasegaran S. 1997. Site-specific cleavage of DNA-RNA hybrids by zinc finger/FokI cleavage domain fusions. *Gene* 203:43–49
68. Klemm JD, Pabo CO. 1996. Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. *Genes Dev.* 10:27–36
69. Klocke B, Koster M, Hille S, Bouwmeester T, Bohm S, et al. 1994. The FAR domain defines a new *Xenopus laevis* zinc finger protein subfamily with specific RNA homopolymer binding activity. *Biochim. Biophys. Acta* 1217:81–89
70. Klug A, Schwabe JW. 1995. Protein motifs 5. Zinc fingers. *FASEB J.* 9:597–604
71. Knochel W, Poting A, Koster M, el Baradi T, Nietfeld W, et al. 1989. Evolutionary conserved modules associated with zinc fingers in *Xenopus laevis*. *Proc. Natl. Acad. Sci. USA* 86:6097–100
72. Koster M, Kuhn U, Bouwmeester T, Nietfeld W, el-Baradi T, et al. 1991. Structure, expression and in vitro functional characterization of a novel RNA binding zinc finger protein from *Xenopus. EMBO J.* 10:3087– 93
73. Krizek BA, Amann BT, Kilfoil VJ, Merkle DL, Berg JM. 1991. A consensus zinc finger peptide: design, high-affinity metal binding, a pH-dependent structure, and a His to Cys sequence variant. *J. Am. Chem. Soc.* 113:4518–23
74. Krizek BA, Merkle DL, Berg JM. 1993. Ligand variation and metal ion binding specificity in zinc finger peptides. *Inorg. Chem.* 32:937–40
75. Lania L, Majello B, De Luca P. 1997. Transcriptional regulation by the Sp family

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

proteins. *Int. J. Biochem. Cell Biol.* 29: 1313–23

76. Larsson SH, Charlieu JP, Miyagawa K, Engelkamp D, Rassoulzadegan M, et al. 1995. Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. *Cell* 81:391–401
77. Lee JS, Galvin KM, Shi Y. 1993. Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proc. Natl. Acad. Sci. USA* 90:6145–49
78. Lee MS, Gippert GP, Soman KV, Case DA, Wright PE. 1989. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* 245:635–37
79. Liu Q, Segal DJ, Ghiara JB, Barbas CF 3rd. 1997. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc. Natl. Acad. Sci. USA* 94:5525–30
80. Losson R. 1997. KRAB zinc finger proteins and nuclear receptors: a possible cross-talk. *Biol. Chem.* 378:579–81
81. Mackay JP, Crossley M. 1998. Zinc fingers are sticking together. *Trends Biochem. Sci.* 23:1–4
82. Merika M, Orkin SH. 1995. Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol. Cell. Biol.* 15:2437–47
83. Merkle DL, Schmidt MH, Berg JM. 1991. Design and characterization of a ligand-binding metallopeptide. *J. Am. Chem. Soc.* 113:5450–51
84. Michael SF, Kilfoil VJ, Schmidt MH, Amann BT, Berg JM. 1992. Metal binding and folding properties of a minimalist Cys_2His_2 zinc finger peptide. *Proc. Natl. Acad. Sci. USA* 89:4796–800
85. Miller J, McLachlan AD, Klug A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* 4:1609–14
86. Miura T, Satoh T, Takeuchi H. 1998. Role of metal-ligand coordination in the folding pathway of zinc finger peptides. *Biochim. Biophys. Acta* 1384:171–79
87. Nardelli J, Gibson T, Charnay P. 1992. Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis. *Nucleic Acids Res.* 20:4137–44
88. Nardelli J, Gibson TJ, Vesque C, Charnay P. 1991. Base sequence discrimination by zinc-finger DNA-binding domains. *Nature* 349:175–78
89. Nekludova L, Pabo CO. 1994. Distinctive DNA conformation with enlarged major groove is found in Zn-finger-DNA and other protein-DNA complexes. *Proc. Natl. Acad. Sci. USA* 91:6948–52
90. Neuhaus D, Nakaseko Y, Schwabe JW, Klug A. 1992. Solution structures of two zinc-finger domains from SWI5 obtained using two-dimensional ^1H nuclear magnetic resonance spectroscopy. A zinc-finger structure with a third strand of beta-sheet. *J. Mol. Biol.* 228:637–51
91. Nolte RT, Conlin RM, Harrison SC, Brown RS. 1998. Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc. Natl. Acad. Sci. USA* 95:2938–43
92. Omichinski JG, Pedone PV, Felsenfeld G, Gronenborn AM, Clore GM. 1997. The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. *Nat. Struct. Biol.* 4:122–32

92a. Pabo CO, Nekludova L. 2000. Geometric analysis and comparisonj of protein-DNA interfaces: Why is there no simple code for recognition? Submitted

93. Page DC, Mosher R, Simpson EM, Fisher EM, Mardon G, et al. 1987. The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* 51: 1091–104
94. Parraga G, Horvath SJ, Eisen A, Taylor WE, Hood L, et al. 1988. Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science* 241:1489–92

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

95. Pavletich NP, Pabo CO. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* 252:809–17
96. Pavletich NP, Pabo CO. 1993. Crystal structure of a five-finger GLI-DNA complex: new perspectives on Zn fingers. *Science* 261:1701–7
97. Perkins ND, Agranoff AB, Pascal E, Nabel GJ. 1994. An interaction between the DNA-binding domains of RelA(p65) and Sp1 mediates human immunodeficiency virus gene activation. *Mol. Cell. Biol.* 14:6570–83
98. Pomerantz JL, Pabo CO, Sharp PA. 1995. Analysis of homeodomain function by structure-based design of a transcription factor. *Proc. Natl. Acad. Sci. USA* 92:9752–56
99. Pomerantz JL, Sharp PA, Pabo CO. 1995. Structure-based design of transcription factors. *Science* 267:93–96
100. Pomerantz JL, Wolfe SA, Pabo CO. 1998. Structure-based design of a dimeric zinc finger protein. *Biochemistry* 37:965–70
101. Posewitz MC, Wilcox DE. 1995. Properties of the Sp1 zinc finger 3 peptide: coordination chemistry, redox reactions, and metal binding competition with metallothionein. *Chem. Res. Toxicol.* 8:1020–28
102. Ptashne M. 1992. A *Genetic Switch.* Cambridge, MA: Blackwell Sci. Cell
103. Rebar EJ. 1997. *Selection studies of zinc finger–DNA recognition.* PhD thesis. MIT, Cambridge, MA
104. Rebar EJ, Pabo CO. 1994. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* 263:671–73
105. Richardson JS, Richardson DC. 1989. In *Principles and Patterns of Protein Conformation,* ed. GD Fasman, pp. 1–98. New York: Plenum
106. Rivera VM, Clackson T, Natesan S, Pollock R, Amara JF, et al. 1996. A humanized system for pharmacologic control of gene expression. *Nat. Med.* 2:1028–32
107. Ryan RF, Darby MK. 1998. The role of zinc finger linkers in p43 and TFIIIA binding to 5S rRNA and DNA. *Nucleic Acids Res.* 26:703–9
108. Sands MS, Bogenhagen DF. 1991. Two zinc finger proteins from *Xenopus laevis* bind the same region of 5S RNA but with different nuclease protection patterns. *Nucleic Acids Res.* 19:797–803
109. Schoenherr CJ, Anderson DJ. 1995. Silencing is golden: negative regulation in the control of neuronal gene transcription. *Curr. Opin. Neurobiol.* 5:566–71
110. Segal DJ, Dreier B, Beerli RR, Barbas CF 3rd. 1999. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences. *Proc. Natl. Acad. Sci. USA* 96:2758–63
111. Shastry BS. 1996. Transcription factor IIIA (TFIIIA) in the second decade. *J. Cell. Sci.* 109:535–39
112. Shi Y. 1995. *Molecular mechanisms of zinc finger protein-nucleic acid interactions.* PhD thesis. John Hopkins Univ., Baltimore, MD
113. Shi Y, Berg JM. 1995. A direct comparison of the properties of natural and designed zinc-finger proteins. *Chem. Biol.* 2:83–89
114. Shi Y, Berg JM. 1996. DNA unwinding induced by zinc finger protein binding. *Biochemistry* 35:3845–48
115. Shields JM, Yang VW. 1998. Identification of the DNA sequence that interacts with the gut-enriched Krüppel-like factor. *Nucleic Acids Res.* 26:796–802
116. Smith GP. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228:1315–17
117. Suzuki M, Gerstein M, Yagi N. 1994. Stereochemical basis of DNA recognition by Zn fingers. *Nucleic Acids Res.* 22:3397–405

118. Swirnoff AH, Milbrandt J. 1995. DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Mol. Cell. Biol.* 15:2275–87
119. Takatsuji H. 1998. Zinc-finger transcription factors in plants. *Cell. Mol. Life Sci.* 54:582–96
120. Thukral SK, Morrison ML, Young ET. 1991. Alanine scanning site- directed mutagenesis of the zinc fingers of transcription factor ADR1: residues that contact DNA and that transactivate. *Proc. Natl. Acad. Sci. USA* 88:9188–92
121. Thukral SK, Morrison ML, Young ET. 1992. Mutations in the zinc fingers of ADR1 that change the specificity of DNA binding and transactivation. *Mol. Cell. Biol.* 12:2784–92
122. Wang B, Pabo C. 1999. Dimerization of zinc fingers mediated by peptides evolved in vitro from random sequences. *Proc. Natl. Acad. Sci. USA* 96:9568–73
123. Weiss MA, Keutmann HT. 1990. Alternating zinc finger motifs in the male-associated protein ZFY: defining architectural rules by mutagenesis and design of an "aromatic swap" second-site revertant. *Biochemistry* 29:9808–13
124. Wolfe SA, Greisman HA, Ramm EI, Pabo CO. 1999. Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J. Mol. Biol.* 285:1917–34
125. Wu H, Yang WP, Barbas CF 3rd. 1995. Building zinc fingers by selection: toward a therapeutic application. *Proc. Natl. Acad. Sci. USA* 92:344–48
126. Wuttke DS, Foster MP, Case DA, Gottesfeld JM, Wright PE. 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
127. Ye X, Rivera VM, Zoltick P, Cerasoli F Jr, Schnell MA, et al. 1999. Regulated delivery of therapeutic proteins after in vivo somatic cell gene transfer. *Science* 283:88–91

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

Annual Review of Biophysics and Biomolecular Structure
Volume 29, 2000

CONTENTS

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.

Annu. Rev. Biophys. Biomol. Struct. 2000.29:183-212. Downloaded from arjournals.annualreviews.org
by State University of New York - Buffalo on 02/01/06. For personal use only.