The Design of Functional DNA-Binding Proteins Based on Zinc Finger Domains

Derek Jantz, Barbara T. Amann, Gregory J. Gatto, Jr., and Jeremy M. Berg*

Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205

Received June 11, 2003

Contents

1. Introduction 789
2. Zinc Finger Domains 789
   2.1. Zinc Finger Domain Peptides with Naturally Occurring Sequences 790
   2.2. Designed Zinc Finger Domain Peptides 791
3. DNA Recognition by Arrays of Zinc Finger Domains 792
4. Alteration of DNA-Binding Specificity of Zinc Finger Domain Arrays 793
   4.1. Structure-Based Design of Zinc Finger Domain Arrays 793
   4.2. Protein Selection by Phage Display Methods 794
   4.3. Use of Chemically Synthesized Zinc Finger Domains 795
   4.4. Arrays with More than Three Zinc Finger Domains 796
5. Applications of Designed Zinc Finger Proteins 796
   5.1. Artificial Transcription Factors 796
   5.2. Other Applications 797
6. Acknowledgments 798
7. References 798

1. Introduction

The sequencing of our complete genome represents a landmark in human endeavor.[1,2] The sequence provides the basis for many deep insights into human biology and into evolution. Among the analyses that have been completed to date is the determination of the protein domains that are encoded in large numbers within the genome. Perhaps the most common recognizable domain is the Cys_2His_2 zinc finger domain; more than 4000 such domains are present in over 700 proteins. These domains, first identified in the *Xenopus* protein transcription factor IIIA, are characterized by sequences that approximate the form (Tyr, Phe)-X-**Cys**-$X_{2,4}$-**Cys**-X_3-Phe-X_5-Leu-X_2-**His**-X_{3-5}-**His**. The conserved cysteine and histidine residues (shown in bold) coordinate a zinc ion in a tetrahedral fashion in the context of a structure consisting of a beta strand, a turn, a second strand, a turn, and a helix,[3,4] as shown in Figure 1.

* To whom correspondence should be addressed at the National Institute of General Medical Sciences, Bethesda, MD 20892. Phone: (301) 594-2172. Fax: (301) 402-0156. E-mail: bergj@mail.nih.gov.

The biological roles of specific zinc finger proteins have been extensively investigated although a tremendous amount of work remains to be done, given the vast number of zinc finger proteins encoded by the human and other eukaryotic genomes. In most cases, these proteins function by binding to nucleic acids (usually DNA) in a sequence-specific fashion. Some aspects of this research have been recently reviewed.[5−10] Furthermore, the Cys_2His_2 zinc finger domain is only one of more than 10 known types of zinc-binding domain that have been identified.[11−19] These domains participate in a range of molecular interactions, including protein−DNA, protein−RNA, protein−protein, and protein−lipid interactions. Structures have been determined for at least one member for most of these families.[11−19]

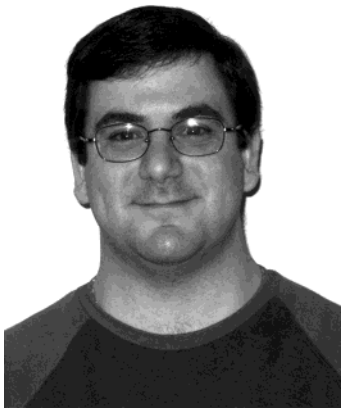
While progress has been made in utilizing some of these other classes of domains in biomimetic studies,[13] these will not be discussed further in this review. Instead, we will focus exclusively on Cys_2-His_2 zinc finger domains and their use for the design of sequence-specific DNA binding proteins. Key examples that reveal the approaches that have been taken will be discussed in considerable detail to reveal the impressive progress that has been made in the generation of proteins that reproduce many of the properties of naturally occurring transcription factors. We will also discuss properties of zinc finger domains that may be utilized for other applications. We begin with a brief review of the discovery of these domains and the structural studies that enabled protein design work.

2. Zinc Finger Domains

The existence of zinc finger domains was first proposed on the basis of studies of the transcription factor IIIA (hereafter, TFIIIA) from the African clawed toad *Xenopus laevis*. This protein had been purified and shown to contain bound zinc ions.[20] Furthermore, the bound zinc was shown to be required for the sequence-specific DNA-binding activity of this protein. The first indication that TFIIIA contained multiple small zinc-binding domains emerged from analysis of the deduced amino acid sequence of TFIIIA.[21,22] This sequence was shown to include nine tandem sequences that approximately match the zinc finger consensus sequence noted above. The presence of two conserved cysteine and two conserved histidine

10.1021/cr020603o CCC: $48.50
Published on Web 12/09/2003

Derek Jantz was born in Littleton, Colorado, and graduated from the University of Colorado at Boulder (B.A. 1996). He received his Ph.D. from the Johns Hopkins University School of Medicine in 2003 under the direction of Jeremy Berg. His research concerned the DNA-binding characteristics of designed Cys_2His_2 zinc finger proteins. He is presently working as a Postdoctoral Fellow with Homme Hellinga at Duke University Medical Center, focusing on computational protein design. He was selected as graduate student speaker at his graduation from Johns Hopkins, where he described graduate school as "the best period of indeterminate length of my life".



Gregory J. Gatto, Jr., was raised in East Brunswick, New Jersey. He received his A.B. in chemistry in 1994 from Princeton University, where he worked in the laboratory of Martin F. Semmelhack. In the spring of 2003, he received his M.D. and Ph.D. degrees from Johns Hopkins University School of Medicine. There, he worked in the laboratory of Jeremy Berg on the structural biology of designed zinc finger proteins and the molecular details of peroxisomal targeting signal recognition. He is currently an NIH postdoctoral fellow with Christopher T. Walsh at Harvard Medical School. He enjoys watching baseball and not completing crossword puzzles, and is an avid bowler.

Barbara T. Amann was born and raised in Swarthmore, Pennsylvania. She received her B.A. in chemistry from Mt. Holyoke College in 1983. She developed an interest in bioinorganic chemistry at the Pennsylvania State University, where she received her Ph.D. working with Bill Horrocks on lanthanide substitution of the calcium-binding protein calmodulin. In 1988, she joined Jeremy Berg's laboratory as a Postdoctoral Fellow in the Chemistry Department of Johns Hopkins University. She moved with the Berg laboratory to the Department of Biophysics and Biophysical Chemistry at Johns Hopkins School of Medicine and became a Research Associate. In this position, she has studied a range of zinc-binding domains, primarily by NMR, and has run the NMR facility of the department. She has enjoyed teaching other researchers NMR methods and assisting them with their projects. She loves spending time with her husband, two children, dog, and snake.

Jeremy M. Berg was born and raised in Stanford, California. He received his B.S. and M.S. degrees in chemistry from Stanford University. He completed his Ph.D. in chemistry from Harvard, working with Richard Holm. After a Postdoctoral Fellowship with Carl Pabo in the Department of Biophysics at Johns Hopkins School of Medicine, he started his independent career in the Department of Chemistry at Johns Hopkins in 1986. He moved back to the School of Medicine in 1990 as Director of the Department of Biophysics and Biophysical Chemistry. His research interests are centered on molecular recognition processes including the structures and metal-binding characteristics of metal-binding domains. He is the coauthor of two leading textbooks, *Principles of Bioinorganic Chemistry* with Stephen Lippard and *Biochemistry* with John Tymoczko and Lubert Stryer. In late 2003, he became the Director of the National Institute of General Medical Sciences in Bethesda. He enjoys time outside of work with his wife and three children.

residues per sequence, as well as the determination that TFIIIA-5S RNA complexes isolated from *Xenopus* oocytes contained 7−11 zinc ions per protein,[21] led to the proposal that each 30-amino acid sequence bound one zinc ion through the conserved cysteine and histidine residues. This hypothesis was supported by limited proteolysis studies that yielded fragments differing in length by about 3 kDa,[21] suggesting that these proposed metal-binding units form individually folded, structurally stable domains.

The nature of the metal binding sites in TFIIIA was directly probed by X-ray absorption spectroscopic methods. EXAFS analysis of the TFIIIA-5S RNA complex isolated from immature *Xenopus* oocytes was consistent with the proposed coordination with two sulfur atoms at 2.30 Å and two nitrogen atoms at 2.00 Å.[23] These distances matched reasonably closely those observed for zinc model complexes prepared by Koch and co-workers.[24,25]

2.1. Zinc Finger Domain Peptides with Naturally Occurring Sequences

The zinc finger hypothesis suggested a reductionist approach to the characterization of these domains.

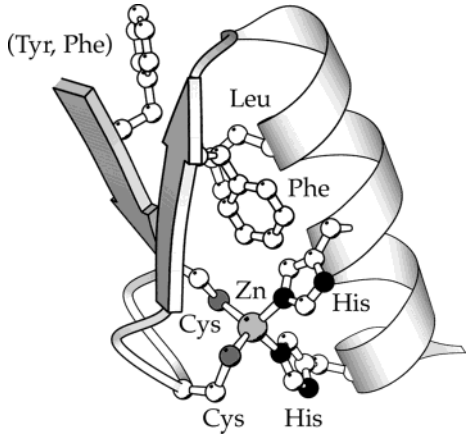


Figure 1. Structure of a zinc finger domain of the class first identified in transcription factor IIIA. The metal-binding and conserved hydrophobic residues characteristic of these domains are shown.

A peptide corresponding to a single 30-amino acid sequence (corresponding to the second putative zinc finger domain of TFIIIA) was produced.[26] In the absence of metal, this peptide behaved as a random coil. However, upon addition of zinc(II), the peptide folded into a stable structure, as indicated by changes in its circular dichroism spectrum and in its sensitivity to proteolysis. The peptide also bound cobalt(II). The spectroscopic properties of the cobalt(II)-substituted peptide suggested that the metal-binding site was, indeed, tetrahedral and were consistent with the proposed coordination site with two cysteinate and two histidine ligands.[26] The availability of a convenient spectroscopic handle facilitated more detailed metal-binding studies of this peptide. Titration of cobalt(II) into the peptide and subsequent displacement of the cobalt(II) with zinc(II) enabled determination of the affinities of this peptide for these two metal ions, with $K_d^{Co} = 3.8\ \mu M$ and $K_d^{Zn} = 2.8$ nM.[27] These studies laid the groundwork for future studies of designed peptides corresponding to single zinc finger domains.

As cDNAs for additional eukaryotic transcription factors were cloned and sequenced, evidence for the widespread occurrence of similar zinc finger domains appeared. For example, the cloning of the gene for the yeast transcription factor ADR1 revealed two zinc finger domains,[28] while the deduced sequence of the mammalian transcription factor Sp1 revealed the presence of three tandem zinc finger domains.[29] In addition to the independent discovery of proteins containing zinc finger domains, attempts were made to clone DNA fragments encoding these domains intentionally.[30−33] These efforts led to the identification of many other potential zinc finger proteins, including *Xenopus* Xfin, which includes a total of 37 zinc finger domains.[31] These studies foreshadowed the discovery of the vast zinc finger protein families encoded in the human and other eukaryotic genomes.

The proposal that single zinc finger domain peptides are largely autonomous suggested a strategy for determination of the three-dimensional structures of these domains. Single-domain peptides of approximately 30 amino acids appeared to be well suited for structure determination by NMR methods. Studies of a single zinc finger peptide from the yeast transcription factor ADR1 revealed the approximate structure of the domain,[34] including the presence of a helical structure that extended from before the conserved Leu residue through the two metal-binding His residues. Analysis of a peptide corresponding to the 31st domain from *Xenopus* Xfin revealed the zinc finger structure in more detail,[4] demonstrating the occurrence of the characteristic β-strand-turn−β-strand-turn−α-helix ($\beta\beta\alpha$) structure shown in Figure 1. This structure closely matched a proposed structure for these domains generated by combining recurring metal-chelating substructures from metalloproteins.[3]

2.2. Designed Zinc Finger Domain Peptides

By 1991, the number of known zinc finger domains identified at the amino acid sequence level had grown to 131 sequences from 18 different proteins. A 26-amino acid peptide was designed as the consensus sequence from this database; that is, the amino acid that occurred most frequently at each position in the database was incorporated into a peptide.[35] This peptide was termed Consensus Peptide 1 (CP-1) and has the amino acid sequence ProTyrLys**Cys**Pro-Glu**Cys**GlyLysSerPheSerGlnLysSerAspLeuValLys-**His**GlnArgThr**His**ThrGly. The rationale for this design was that amino acids involved in stabilizing the zinc finger structure should be present at high frequency in the database and, hence, an unusually stable zinc finger domain might be produced.

The success of the design of CP-1 was tested by two methods, metal-binding measurements and NMR spectroscopy. The fact that metal ion binding is required for zinc finger peptide folding implies that zinc finger peptides that have particularly stable folded structures will bind metal ions with high affinity. Cobalt(II) and zinc(II) bind to CP-1 with dissociation constants of 60 nM and 2 pM, respectively.[35] These affinities are approximately 2 orders of magnitude higher than those for the TFIIIA-derived peptide discussed above and, indeed, are higher than those for any natural sequence zinc finger peptide that has been characterized to date. It is not clear whether evolution has selected zinc-binding domains with relatively modest zinc affinities to facilitate zinc release or insufficient selective pressure has been present to drive zinc affinities to the higher levels demonstrated by CP-1.

The structure of CP-1 has been probed by NMR methods.[35,36] Like other zinc finger peptides, the addition of zinc(II) to the apopeptide resulted in dramatic shifts in the NMR spectrum. An alignment of the folded CP-1 structure[36] with those of ADR1a and Xfin shows that the backbone atoms have root-mean-squared differences of less than 1.5 Å, even though their sequences differ by more than 50%.

The similarity of these structures suggested that the characteristic zinc finger fold is determined by the position and identities of the metal-binding residues and the three conserved hydrophobic residues. To test this hypothesis, a "minimalist" zinc finger (hereafter referred to as MZF) was designed, synthesized, and characterized.[37] MZF has the se-

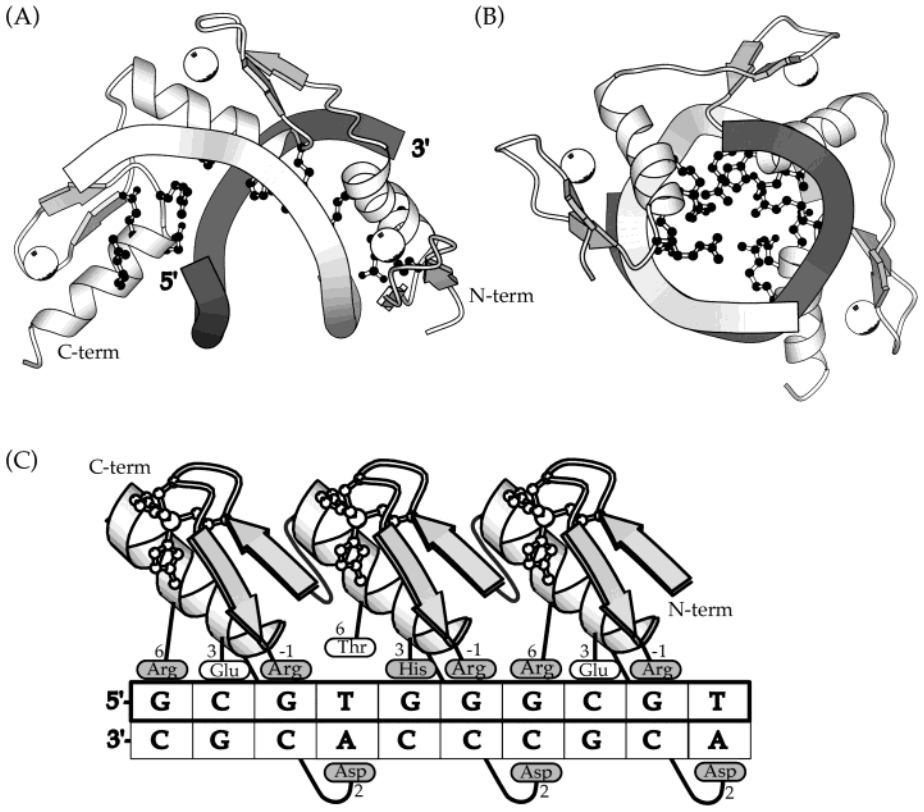


Figure 2. Ribbon diagram[108] of the three-zinc-finger domain array from Zif268 bound to DNA. (A) Schematic view showing the three zinc finger domains wrapping around the DNA, with the primary strand of the DNA shown in dark gray. (B) End view showing the side chains that make sequence-specific contacts extended into the DNA double helix. (C) Schematic view showing the amino acid residues involved in sequence-specific contacts. Amino acid residues involved in hydrogen-bonding interactions with the DNA are shown in gray boxes.

quence Lys**Tyr**Ala**Cys**AlaAla**Cys**AlaAlaAla**Phe**Ala-AlaLysAlaAla**Leu**AlaAla**His**AlaAlaAla**His**AlaLys, where the metal-binding and conserved hydrophobic residues are shown in bold. Three Lys residues were included to promote aqueous solubility. MZF bound cobalt(II) and zinc(II) with affinities comparable to those observed for natural-sequence zinc finger peptides. Analysis of the ^{1}H chemical shifts of the zinc-(II) complex of MZF demonstrated that this complex adopts a structure similar to those observed for other zinc finger peptides.

These studies demonstrate that zinc finger domains can tolerate substantial changes in amino acid sequence without significantly affecting three-dimensional structure and stability. This property has enabled these domains to be used as versatile building blocks for the construction of novel DNA-binding proteins, as we will discuss shortly.

3. DNA Recognition by Arrays of Zinc Finger Domains

DNA-binding proteins that utilize zinc finger domains to contact DNA almost always have tandem arrays of two or more fingers. The first structural characterization of such a multifinger protein came with the elucidation of the structure of the DNA-binding region of the murine Zif268 transcription factor[38] in complex with an oligonucleotide that corresponded to a favored DNA binding site by X-ray diffraction methods (Figure 2).[39,40] This structure provided a framework for understanding a wide set of zinc finger protein−DNA interactions. The DNA-binding region of Zif268 comprises three zinc fingers joined to one another by heptapeptide linkers (measured from the last His residue of one domain to the first Cys of the next domain). Individual fingers within the array adopt the same $\beta\beta\alpha$ architecture characteristic of single zinc finger domains. Each domain primarily interacts with three base pairs of DNA. These triplet subsites are directly adjacent to one another, so the complete core binding site is nine base pairs. Base contacts are made by residues in the helix of each zinc finger domain. The helices are inserted into the major groove with their amino terminal ends directed more toward the DNA. In the case of Zif268, most of the base contacts are made with one strand of the DNA (hereafter referred to as the primary strand). The protein binds such that it lies antiparallel relative to the primary strand; that is, the first (amino-terminal) zinc finger domain interacts with the 3′-most triplet subsite of the binding site, while the last (carboxyl-terminal) domain interacts with the 5′-most triplet subsite.

The Zif268 cocrystal structure revealed a simple pattern of base contacts within each subsite. The amino acid residues in positions −1, 3, and 6 of each zinc finger domain (numbered relative to the start of the helix) are positioned to make specific, one-to-one contacts with the 3′-, central, and 5′-bases of the DNA triplet subsite on the primary strand. The first and third zinc finger domains of Zif268 have Arg in

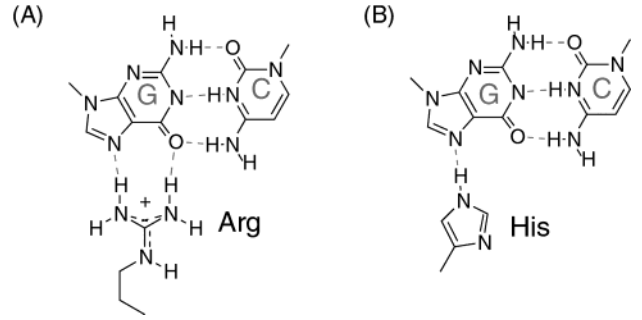


Figure 3. Contacts responsible for sequence-specific interactions from the Zif268 cocrystal structure. (A) Arginine contacting guanine. (B) Histidine contacting guanine.

position −1, Glu in position 3, and Arg in position 6. The guanidinium groups of these Arg residues participate in a pair of hydrogen bonds with guanine bases in the binding site (Figure 3). The Glu residues do not participate in hydrogen bonds with the DNA but sterically favor cytosine bases. The second zinc finger domain of Zif268 has Arg in position −1, His in position 3, and Thr in position 6. The Arg residue forms two hydrogen bonds with a guanine, while the His side chain binds guanine through a single hydrogen bond to N7 of the base. The Thr does not contact the DNA. In addition to these contacts, each zinc finger domain has an Asp residue in position 2. The carboxylate group of each Asp forms a hydrogen-bonded salt bridge with the Arg residue in position 1 and a hydrogen bond with an adenine or cytosine base on the complementary strand immediately preceding that zinc finger domain's subsite.[41] Given this interaction, adjacent zinc finger domains actually contact partially overlapping four-base-pair subsites.

In addition to these base contacts, Zif268 makes a number of base-independent contacts with the DNA phosphodiester backbone. The N−H group from the imidazole of the first metal-bound His residue from each domain participates in a hydrogen bond with an oxygen from a backbone phosphate. In addition, several Lys and Arg side chains within each domain may form direct or water-mediated hydrogen bonds with the backbone.

In the case of Zif268, the linkers joining one zinc finger domain to the next have the sequence **His**-ThrGly(Glu,Gln)LysPro(Tyr,Phe)X-**Cys**. Similar linker sequences are frequently found in many multidomain zinc finger proteins.[30] These linkers are flexible in the absence of DNA but become structured upon DNA binding.[42] While the linker region does not appear to make any direct contacts with the DNA, structural comparisons[43] and mutagenesis experiments[44,45] suggest that it plays a role in DNA binding by stabilizing the positioning of individual zinc finger domains relative to one another. Recently, it was proposed that this short stretch of amino acids may also act as a substrate for a cellular kinase which regulates DNA binding activity in a cell-cycle-dependent manner through phosphorylation of the conserved threonine residue.[46] As will be discussed below, the conservation of the length of this linker appears to limit the DNA-binding activities of arrays of tandem zinc finger domains with more than three fingers.

4. Alteration of DNA-Binding Specificity of Zinc Finger Domain Arrays

Even before the determination of the Zif268 cocrystal structure, mutagenesis studies had demonstrated[47] that it is possible to change the DNA-binding specificity of zinc finger proteins by substituting different amino acids into positions within what is now known to be the recognition helix.[48−52] These observations had tremendous implications, suggesting that it might be possible to construct a series of zinc finger domains with different residues in the DNA-contacting positions and use these engineered arrays to recognize a wide range of DNA sequences. Implementation of this strategy requires several stages. First, the relationships between the amino acid residues in the recognition helix of a given zinc finger domain and the sequence of its preferred DNA-binding subsite must be established. Second, the importance of array context in modulating these DNA-binding subsite preferences must be deduced. In the simplest case, the subsite preferences of a given zinc finger domain would be independent of the DNA-binding activities of adjacent zinc fingers within an array; that is, each zinc finger domain would behave in a completely modular fashion.

4.1. Structure-Based Design of Zinc Finger Domain Arrays

Initial experiments directed toward the generation of novel zinc finger domain-based proteins were cast in terms of the possible existence of a relatively simple design "code".[49,50] In these experiments, the DNA-contacting residues suitable for recognition of a particular DNA triplet were selected one at a time on the basis of the collection of known DNA contacts in naturally occurring proteins. While this approach did yield modest success, it quickly became apparent that the DNA-contacting residues within a particular recognition helix do not function independently of one another and, as such, cannot effectively be selected one at a time. For example, as was demonstrated in the Zif268 cocrystal structure, Arg in position −1 appears to specify guanine at the appropriate position along the primary strand of the binding site. Other zinc finger domains were characterized that had Gln in position −1 but preferred adenine or thymine in this position. With Arg in position −1, His in position 3 leads to a preference for guanine or adenine in the central position of the binding subsite. On the basis of a simple code, a zinc finger domain with Gln in position −1 and His in position 3 would be expected to prefer adenine or thymine in the 3′-most position and guanine or adenine in the central position. However, experimental examination of domains containing Gln_{-1}−His_3 revealed significant decreases in both DNA-binding affinity and base discrimination relative to similar designed proteins.[49,50] This behavior can be rationalized in terms of side-chain length. In the context of a long side chain in position −1, such as Arg, both this residue and His in position 3 are able to contact the DNA productively. However, with a shorter side chain such as Gln in position −1, the presence of a relatively large side chain such as His

in position 3 pushes the zinc finger domain away from the DNA, so both residues cannot simultaneously make productive contacts. These arguments are well supported by X-ray crystallographic analyses of designed proteins, as discussed below.[53] Thus, the design of biomimetic zinc finger proteins is more effectively cast in terms of recognition by an entire helix rather than by a set of completely independent recognition residues.

The modularity of individual zinc finger domains within an array was investigated using a zinc finger domain swapping approach. Previous mutagenesis experiments had produced a series of variants of the three-finger DNA-binding unit from the human transcription factor Sp1.[49−51] These variants all had alterations in the DNA-recognition helices of the middle zinc finger domain that changed the DNA-binding preferences conferred by that domain. The mutated recognition helices from these Sp1 variants were then incorporated into zinc finger sequences that otherwise matched the CP-1 sequence discussed above.[52] Zinc finger proteins containing three such domains were then constructed. The first protein to be examined had the following residues at key positions within the recognition helices:

zinc finger domain 1: $Gln_{-1}Ser_{2}Asp_{3}Arg_{6}$

zinc finger domain 2: $Arg_{-1}Asp_{2}Glu_{3}Arg_{6}$

zinc finger domain 3: $Arg_{-1}Asp_{2}His_{3}Arg_{6}$

This protein was anticipated to bind 5′-G(G,A)G-GCG-GC(A,T)-3′. Gratifyingly, DNA-binding site selection studies using a randomized pool of DNA sequences demonstrated that this protein preferred the DNA site 5′-GGG-GCG-GCT-3′ with levels of discrimination comparable to those observed within the Sp1 context. The dissociation constant for the optimal binding site was approximately 2 nM. For comparison, a second protein was prepared that had the first and second zinc finger domains interchanged. Assuming modular behavior, this protein was anticipated to prefer 5′-GGG-GCT-GCG-3′ over 5′-GGG-GCG-GCT-3′. This preference was, in fact, observed, although the second protein showed a smaller degree of discrimination than did the first. These studies provided considerable evidence that the modular design of multidomain zinc finger proteins was likely to be a productive strategy for the generation of DNA-binding units recognizing a range of nucleic acid sequences.

The success of modular zinc finger protein design was probed directly by examining the three-dimensional structure of a designed protein bound to a preferred DNA site.[53] All three zinc finger domains in the protein were based on the CP-1 scaffold and differed only in the identities of the residues in positions −1, 2, 3, and 6 of the DNA recognition helices. The docking arrangement of the designed protein is remarkably similar to that of Zif268, and all of the DNA-contacting residues recognize their cognate bases through anticipated patterns of hydrogen bonds with only minor variations (Figure 4). While the Zif268 cocrystal complex was dominated

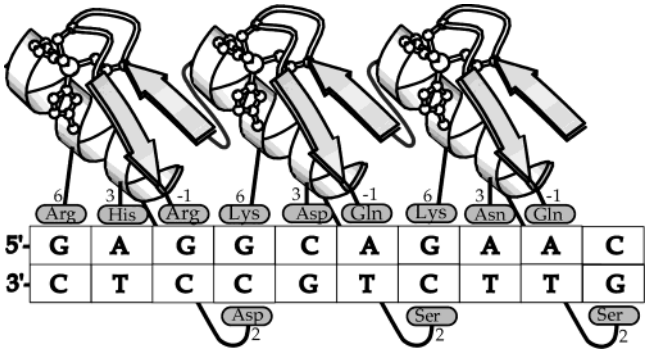


Figure 4. Schematic view of a designed three-zinc-finger domain bound to DNA. Amino acid residues involved in hydrogen-bonding interactions with the DNA are shown in gray boxes.

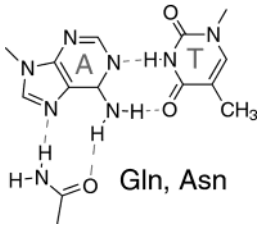


Figure 5. Interactions involving the carboxamide side chains of Gln and Asn with adenine bases.

by Arg−guanine interactions, the designed protein utilized a more varied set of residues for base recognition. In particular, the carboxamide side chains of Gln and Asn are frequently involved in adenine recognition (Figure 5). Serine residues in position 2 participate in hydrogen-bonding interactions with bases on the secondary strand, analogous to those involving Asp residues in position 2 observed in the Zif268 crystal structure. Moreover, this structure demonstrated the adaptability of zinc finger domains to differences in the lengths of their DNA-contacting side chains. For example, the first zinc finger domain had Gln in position −1, Asn in position 3, and Lys in position 6, while the third zinc finger domain had Arg in position −1, His in position 3, and Arg in position 6. The shorter side-chain lengths in the first domain require that the entire domain lie relatively close to the DNA to make productive contacts. The relatively flexible linkers that connect zinc finger domains allow this to occur; the first domain is positioned between 1 and 2 Å closer to the DNA than is the third domain.

4.2. Protein Selection by Phage Display Methods

These results demonstrated that direct zinc finger protein design could be successful, at least in some contexts. However, it quickly became apparent that there were limitations on the range of DNA sequences to which designed proteins could be targeted. Furthermore, this approach to protein construction did not ensure that the proteins produced were optimized in terms of their affinity and specificity for the intended target sites. The recognition of a broader range of DNA sites required the development of experimental selection strategies capable of identifying proteins that recognized particular DNA sequences without preconceived notions about the

specific contacts utilized. Phage display,[54−56] the most successful approach of this type, utilizes pools of proteins expressed on the surface of phage particles that have been randomized at certain positions within their amino acid sequences.[57−62] These phage particles are then segregated on the basis of their ability to bind immobilized DNA fragments that include the desired target site. Over several rounds of selection and amplification, the pool becomes enriched for phage-displaying zinc finger proteins that bind with high affinity to the target DNA sequence. This technique has been used to construct zinc finger domains recognizing a relatively broad range of different DNA sites, many of which do so through contacts that deviate considerably from previously observed modes of recognition observed in the Zif268 and other structures.[57−62]

As an example of the power of phage display methods and the versatility of zinc finger protein−DNA recognition surfaces, consider the structure of a zinc finger protein that was selected to recognize the DNA site 5′-GCTATAAAAG-3′, a sequence characteristic of those recognized by the TATA-box-binding protein.[63] The three zinc finger domain protein was constructed using a protocol involving the selection of successive zinc finger domains within the context of the Zif268 framework. Each selected domain was randomized at positions −1, 1, 2, 3, 5, and 6, corresponding to $20^6 = 6.4 \times 10^7$ possible sequences per zinc finger domain. Several proteins were identified that bound the target site with affinities within a factor of 100 of the affinity of Zif268 for its preferred site under the same binding conditions. Binding site selection studies revealed clear preferences for most of the bases within the targeted binding site. The overall consensus sequence was found to be 5′-GCTATAAANN-3′. Thus, phage display selections had yielded several proteins that showed both high affinity and high specificity for the targeted site.

The crystal structures of two of the TATA-box-directed zinc finger proteins in complex with oligonucleotides containing the target binding site have been reported.[64] These structures reveal that the overall arrangement of the zinc finger domains is quite similar to that observed for Zif268, with root-mean-squared deviations of less than 1.5 Å on backbone atoms. An examination of the residues involved in sequence-specific contacts with the DNA, however, shows striking differences (Figure 6). First, approximately the same number of contacts are made to the primary and secondary strands of the DNA, in contrast to the strong preference for the primary strand in Zif268 and in the designed protein discussed above. Some of these interactions involve residues in positions 1 and 2 in addition to the residues in positions −1, 3, and 6. Second, more side chain−side chain interactions occur in the TATA-box-binding site-directed proteins. These interactions buttress the residues involved in contacts with the DNA and thereby contribute indirectly to binding specificity. Despite these differences, many contacts are observed that are similar to those made by Zif268 and the designed protein discussed above, including

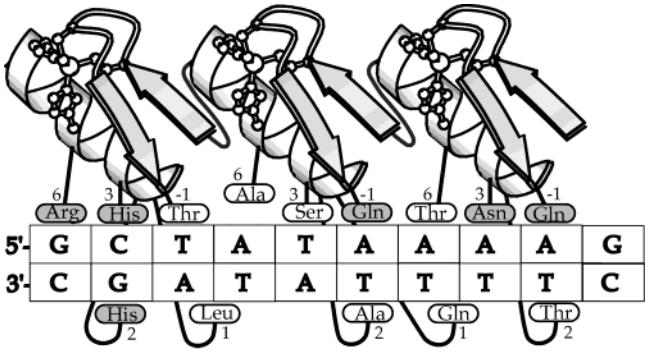


Figure 6. Schematic view of a three-zinc-finger domain protein selected by phage display methods bound to the target site. Note the occurrence of a number of hydrogen-bonding (with amino acids shown in gray boxes) and other interactions with both strands of the DNA.

Arg−guanine and Gln−adenine. These structural results demonstrate that the binding surfaces possible within the zinc finger framework are considerably more versatile than imagined from the Zif268 structure. While this complicates the design of DNA-binding proteins based on simple chemical bonding arguments, it clearly expands the number of DNA sequences for which specific biomimetic zinc finger domain-based proteins can be constructed.

In addition to those already discussed, a number of Cys_2His_2 zinc finger arrays have been structurally characterized in complex with DNA. Many of these proteins contain zinc finger domains that dock with the DNA in a manner that is remarkably similar to that of Zif268, recognizing individual DNA bases primarily through the same pattern of residues in positions −1, 2, 3, and 6.[42,65−68] There are, however, numerous examples of Cys_2His_2 zinc finger proteins that adopt different docking arrangements upon complexation with their DNA sites. The structures of TFIIIA,[68,69] GLI,[70] Tramtrack,[65] GAGA,[71] and YY1,[66] for example, show a wide range of docking orientations with the DNA and frequently recognize individual bases through contacts that deviate considerably from those observed for Zif268.

4.3. Use of Chemically Synthesized Zinc Finger Domains

Even with phage display and other selection methods, many DNA sequences have proven to be difficult targets for zinc finger proteins comprised entirely of the 20 amino acids that occur naturally in proteins. The development of powerful methods for chemical protein synthesis and semisynthesis[72−74] has enabled the preparation of zinc finger proteins that contain amino acids that do not occur in proteins normally. The first example of the use of this approach involved the production of a semisynthetic three-zinc-finger protein in which the third zinc finger was made synthetically and incorporated into the protein by expressed protein ligation methods.[75] The amino acid citrulline was placed in position 6 of the recognition helix of the synthetic domain. This amino acid has the same overall length as Arg, but has functional groups that are more similar to those of Gln, potentially facilitating a specific interaction with adenine.

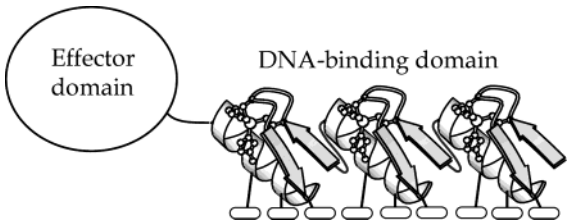


Figure 7. Schematic view of a zinc finger protein with an appended effector domain. Note that effector domains can be attached to either the carboxyl-terminal (shown) or amino-terminal end of the zinc finger region and the zinc finger-based DNA-binding domain can have any number of zinc finger domains.

Binding studies demonstrated that the semisynthetic protein does, indeed, prefer adenine in the expected position, while the specificities at the other positions within the triplet subsite are the same as those observed when Arg is in position 6. For comparison, Gln in position 6 of an otherwise identical protein showed little ability to discriminate between bases, presumably because the Gln side chain is too short to reach the DNA in this context. By overcoming limitations in the repertoire of side-chain lengths and functionalities available with the 20 proteinogenic amino acids, these synthetic approaches should enable the production of zinc finger proteins recognizing a broad range of DNA sites.

4.4. Arrays with More than Three Zinc Finger Domains

Thus far, we have discussed proteins comprising three zinc finger domains. One of the appealing properties of zinc finger proteins is the ability to vary the number of domains within an array. The inclusion of additional domains would be expected to increase the binding site size, with associated increases in DNA-binding affinity and specificity. This proved to be more difficult than first envisioned. The generation of six domain proteins by fusion of two three-domain proteins with the canonical **His**-ThrGlyGluLysProTyrX-**Cys** linker led to proteins that bound larger sites, but the DNA-binding affinity and specificity were not increased to nearly the extent anticipated.[76] This appears to be due to the fact that the repeat of a zinc finger protein is not truly 3.0 base pairs, so that strain is introduced into the DNA as more zinc finger domains bind simultaneously. Nonetheless, six zinc finger domain proteins with standard linkers have shown impressive activities in some applications, as will be discussed below. Furthermore, alternative linkers have been examined, and some proteins with very high DNA-binding affinities have been generated.[6,77,78]

5. Applications of Designed Zinc Finger Proteins

The ability to generate zinc finger proteins that recognize preselected DNA sites has numerous potential applications. By fusing these custom DNA-binding domains to an appropriate ancillary domain that has other functional characteristics, it is possible to produce proteins with specific activities targeted to particular DNA sites (Figure 7).

5.1. Artificial Transcription Factors

The most natural application of this approach is the generation of biomimetic transcription factors. Here, the effector domains utilized are those associated with transcription activators or transcriptional repressors. Early experiments of this type used reporter systems present on plasmids to show that a zinc finger domain-based transcription factor could dramatically affect the expression levels of specific genes.[59,79] Innovative work by Klug and co-workers provided suggestive evidence that a three-zinc-finger protein could inhibit the expression of a stably integrated gene in a mouse cell line.[59]

More recent studies have demonstrated that endogenous genes can be activated and repressed in cell lines. A striking example involves the regulation of the erbB-2 and erbB-3 genes[80−85] in human cell lines.[86] A six-zinc-finger domain DNA-binding unit was designed to target the sequence 5′-GGG-GCC-GGA-GCC-GCA-GTG-3′ present upstream of the translational start site of the erbB-2 gene.[76,87] Note that the 5′-most base in each triplet subsite is guanine, facilitating zinc finger protein design. The framework of the six-zinc-finger domain protein was based on Consensus Peptide 1, with canonical **His**-ThrGlyGluLysProTyrLys-**Cys** linkers between all fingers. This protein was found to bind to the target site with an affinity[87] near 1 nM and discriminated against a site that differed in three positions[86] by a factor of 15. This engineered DNA-binding domain was then fused to a nuclear localization signal and a transcription activation domain derived from the herpes simplex virus protein VP16.[88] The resulting artificial transcription factor was intended to increase the expression level of the erb-B2 protein on the cell surface when introduced into the human carcinoma cell line A431. The retroviral vector used to introduce the transcription factor also supported the expression of green fluorescent protein (GFP) so that cells that had been successfully infected could be identified.[86]

The expression levels of erb-B2 on the cell surface were monitored with the use of a specific antibody in conjunction with fluorescence-activated cell sorting.[86] Two populations of cells were identified. Approximately 25% of the cells expressed GFP and showed increased levels of erb-B2, whereas the remainder had no detectable GFP and normal levels of erb-B2. Thus, those cells that had been productively infected so that they expressed the designed transcription factor did, indeed, show increased erb-B2 levels. Levels of the related cell surface proteins erb-B1 and erb-B3 were unaffected. These results provided strong evidence that the DNA-binding specificity associated with the zinc finger region had allowed specific up-regulation of the expression of a targeted endogenous gene. Interestingly, engineered transcription factors that contained only three of the original six zinc finger domains were not able to promote erb-B2 up-regulation, even though they exhibited relatively high affinity for the intended binding site in vitro.[86] The reasons for this striking difference in activity in vivo remain to be fully established.

The same six-zinc-finger unit used to up-regulate erb-B2 expression was then fused with a transcriptional repression peptide called a KRAB domain that is frequently found in natural zinc finger-containing proteins.[89] As before, the gene encoding this designed repressor protein was introduced into A431 cells using a retroviral vector that also expressed GFP. Cell sorting again revealed two classes of cells. Approximately 60% of the cells expressed GFP and showed essentially no erb-B2 expression. The remainder of the cells did not express GFP and expressed erb-B2 at normal levels. For both classes of cells, the levels of erb-B1 and erb-B3 were normal. These results demonstrated that specific gene repression was also possible using designed zinc finger proteins.[86]

As a final demonstration of the specificity provided by the zinc finger domains, a second six-zinc-finger unit was designed[86] to target the DNA sequence 5′-GG**A**-GCC-GGA-GCC-G**G**A-GT**C**-3′ found in the promoter of the related erb-B3 gene. This sequence differs in three positions from that from the erb-B2 promoter site as indicated by the bases shown in bold. Correspondingly, the designed protein targeted to this sequence was different in three of the six recognition helices from that directed to the erb-B2 site. The designed protein was constructed and found to bind the anticipated site with nanomolar affinity and 30-fold discrimination against the erb-B2 site. Transcriptional activators and repressors were constructed and analyzed as discussed above for the erb-B2-directed protein. Specific changes in cell surface erb-B3 protein levels were observed with no significant effects on erb-B1 or erb-B2.

Since the publication of this seminal work, several reports have further demonstrated the potential of biomimetic zinc finger transcription factors for specific regulation of endogenous genes in cultured mammalian cells[78,90,91] or plants.[92] However, not all such attempts proved successful. One factor that clearly plays a significant role in this type of directed gene regulation is the accessibility of the targeted DNA sites within the context of chromatin. Recent work[91,93,94] has demonstrated that engineered transcription factors are far more likely to be effective when they are targeted to sites in DNaseI-sensitive regions of the promoter of interest.

A particularly exciting recent report demonstrated the effectiveness of engineered zinc finger transcription factors for the directed regulation of gene expression in an animal model.[95] In this set of experiments, a designed transcription factor was used to induce angiogenesis in mice through the up-regulation of vascular endothelial growth factor (VEGF-A) expression. A series of zinc finger proteins were first designed to bind sequences in the murine *vegfa* gene promoter that appeared to be in accessible chromatin structures.[91] One of these constructs, a three-zinc-finger protein directed toward the sequence 5′-GCTGGGGGC-3′, was linked to a VP16-based activation domain and cloned into an adenoviral delivery system. When the adenoviral vector encoding the transcription factor was injected subcutaneously into the ears of mice, increased blood vessel development was visually apparent.[95] Infection with the zinc finger protein-encoding virus more than doubled the overall vasculature in the ears of these mice relative to those injected with a control virus. These results provide real hope that designed transcription factors will find applications in clinical medicine.

5.2. Other Applications

We have focused on the use of engineered zinc finger domains for the generation of biomimetic transcription factors. Several other applications have also been demonstrated through the use of different ancillary domains. For example, novel site-specific DNA methylases have been generated through the fusion of designed zinc finger DNA-binding domains with *S*-adenosyl methionine-dependent methylase domains.[96,97] Similarly, Kim and Chandrasegaran fused a designed DNA-binding domain composed of three Cys_2His_2 zinc fingers to the DNA hydrolysis domain from the type IIS restriction enzyme *Fok*I.[98] This produced a restriction enzyme with a novel cleavage site dictated by the engineered zinc finger domains.[98−101] The nuclease domains from *Fok*I dimerize, leading to double-stranded DNA cleavage at appropriate sites flanked by zinc finger protein recognition sequences. These chimeric restriction enzymes were designed with the goal of producing restriction enzymes with sequence specificities that could be manipulated for in vitro application. However, recent discoveries have suggested that these chimeric proteins may have more powerful applications. In particular, these zinc finger-based chimeric restriction enzymes were recently found to promote site-specific gene recombination when expressed in eukaryotic cells.[102,103] With these chimeric enzymes, targeting of specific genes for recombination could be increased by factors up to several 1000-fold over background rates. These designed proteins thus have tremendous potential to allow gene replacement in eukaryotic (including human) cells at levels of efficiency that may be useful for real therapeutic applications. These results highlight how the ability to design specific DNA-binding modules based on zinc finger domains can be harnessed with appropriate effector domains to produce molecules that can facilitate both research and novel approaches to clinical problems. As our understanding of the properties of these and other effector domains increases, additional applications of zinc finger-based proteins can be anticipated.

Finally, zinc finger domain peptides have proven to be useful tools for the examination of properties unrelated to DNA binding. Variation of the metal-binding residues themselves has provided insights into the basis for metal ion-binding affinity and specificity. The removal of one of the metal-binding amino acids from a zinc finger peptide sequence has led to domains that can bind exogenous ligands.[104,105] Furthermore, the fact that these domains fold stably only when they bind metal ions has been used to probe several aspects of protein folding.[106,107] Work with these versatile metal-based domains is likely to provide additional insights into a variety of fundamental biological processes.

The applications of Cys_2His_2 zinc finger technology are only as limited as our creativity. The use of zinc finger transcription factors may soon allow us to regulate uncharacterized genes to assist in determining their functions. Several recent demonstrations of zinc finger-based gene control in animal and plant cells suggest diverse applications for this technology in medicine and agriculture. As key issues relating to protein design and gene delivery are resolved in the coming years, the ability to target specific functionalities to discreet locations in the genome should have significant impacts on medicine and biological research.

6. Acknowledgments

We thank the members of the Berg laboratory, past and present, for contributions to the research described herein and the National Institute of General Medical Sciences for support.

7. References

(1) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Gabor Miklos, G. L.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C. *Science* **2001**, *291*, 1304.

(2) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J. P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendl, M. C.; Delehaunty, K. D.; Miner, T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton, S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J. F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M. *Nature* **2001**, *409*, 860.

(3) Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 99.

(4) Lee, M. S.; Gippert, G. P.; Soman, K. V.; Case, D. A.; Wright, P. E. *Science* **1989**, *245*, 635.

(5) Klug, A. *J. Mol. Biol.* **1999**, *293*, 215.

(6) Choo, Y.; Isalan, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 411.

(7) Wolfe, S. A.; Nekludova, L.; Pabo, C. O. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 183.

(8) Pabo, C. O.; Peisach, E.; Grant, R. A. *Annu. Rev. Biochem.* **2001**, *70*, 313.

(9) Segal, D. J.; Barbas, C. F., III. *Curr. Opin. Biotechnol.* **2001**, *12*, 632.

(10) Laity, J. H.; Lee, B. M.; Wright, P. E. *Curr. Opin. Struct. Biol.* **2001**, *11*, 39.

(11) Omichinski, J. G.; Clore, G. M.; Schaad, O.; Felsenfeld, G.; Trainor, C.; Appella, E.; Stahl, S. J.; Gronenborn, A. M. *Science* **1993**, *261*, 438.

(12) Perez-Alvarado, G. C.; Miles, C.; Michelsen, J. W.; Louis, H. A.; Winge, D. R.; Beckerle, M. C.; Summers, M. F. *Nat. Struct. Biol.* **1994**, *1*, 388.

(13) Berg, J. M.; Shi, Y. *Science* **1996**, *271*, 1081.

(14) Borden, K. L.; Freemont, P. S. *Curr. Opin. Struct. Biol.* **1996**, *6*, 395.

(15) Srinivasan, N.; Bax, B.; Blundell, T. L.; Parker, P. J. *Proteins* **1996**, *26*, 217.

(16) De Guzman, R. N.; Wu, Z. R.; Stalling, C. C.; Pappalardo, L.; Borer, P. N.; Summers, M. F. *Science* **1998**, *279*, 384.

(17) Kumar, R.; Thompson, E. B. *Steroids* **1999**, *64*, 310.

(18) Matthews, J. M.; Kowalski, K.; Liew, C. K.; Sharpe, B. K.; Fox, A. H.; Crossley, M.; MacKay, J. P. *Eur. J. Biochem.* **2000**, *267*, 1030.

(19) Amann, B. T.; Worthington, M. T.; Berg, J. M. *Biochemistry* **2003**, *42*, 217.

(20) Hanas, J. S.; Hazuda, D. J.; Bogenhagen, D. F.; Wu, F. Y.; Wu, C. W. *J. Biol. Chem.* **1983**, *258*, 14120.

(21) Miller, J.; McLachlan, A. D.; Klug, A. *EMBO J.* **1985**, *4*, 1609.

(22) Brown, R. S.; Sander, C.; Argos, P. *FEBS Lett.* **1985**, *186*, 271.

(23) Diakun, G. P.; Fairall, L.; Klug, A. *Nature* **1986**, *324*, 698.

(24) Corwin, D. T., Jr.; Fikar, R.; Koch, S. A. *Inorg. Chem.* **1987**, *26*, 3079.

(25) Corwin, D. T., Jr.; Gruff, E. S.; Koch, S. A. *J. Chem. Soc., Chem. Commun.* **1987**, *13*, 966.

(26) Frankel, A. D.; Berg, J. M.; Pabo, C. O. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 4841.

(27) Berg, J. M.; Merkle, D. L. *J. Am. Chem. Soc.* **1989**, *111*, 3759.

(28) Hartshorne, T. A.; Blumberg, H.; Young, E. T. *Nature* **1986**, *320*, 283.

(29) Kadonaga, J. T.; Carner, K. R.; Masiarz, F. R.; Tjian, R. *Cell* **1987**, *51*, 1079.

(30) Schuh, R.; Aicher, W.; Gaul, U.; Côté, S.; Preiss, A.; Maier, D.; Seifert, E.; Nauber, U.; Schröder, C.; Kemler, R.; Jäckle, H. *Cell* **1986**, *47*, 1025.

(31) Ruiz i Altaba, A.; Perry-O'Keefe, H.; Melton, D. A. *EMBO J.* **1987**, *6*, 3065.

(32) Chowdhury, K.; Deutsch, U.; Gruss, P. *Cell* **1987**, *48*, 771.

(33) Pellegrino, G. R.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 671.

(34) Parrága, G.; Horvath, S. J.; Eisen, A.; Taylor, W. E.; Hood, L.; Young, E. T.; Klevit, R. E. *Science* **1988**, *241*, 1489.

(35) Krizek, B. A.; Amann, B. T.; Kilfoil, V. J.; Merkle, D. L.; Berg, J. M. *J. Am. Chem. Soc.* **1991**, *113*, 4518.

(36) Bianchi, E.; Folgori, A.; Wallace, A.; Nicotra, M.; Acali, S.; Phalipon, A.; Barbato, G.; Bazzo, R.; Cortese, R.; Felici, F.; Pessi, A. *J. Mol. Biol.* **1995**, *247*, 154.

(37) Michael, S. F.; Kilfoil, V. J.; Schmidt, M. H.; Amann, B. T.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4796.

(38) Christy, B. A.; Lau, L. F.; Nathans, D. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7857.

(39) Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809.

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

(41) Isalan, M.; Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 5617.

(42) Wuttke, D. S.; Foster, M. P.; Case, D. A.; Gottesfeld, J. M.; Wright, P. E. *J. Mol. Biol.* **1997**, *273*, 183.

(43) Laity, J. H.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2000**, *295*, 719.

(44) Choo, Y.; Klug, A. *Nucleic Acids Res.* **1993**, *21*, 3341.

(45) Clemens, K. R.; Zhang, P.; Liao, X.; McBryant, S. J.; Wright, P. E.; Gottesfeld, J. M. *J. Mol. Biol.* **1994**, *244*, 23.

(46) Dovat, S.; Ronni, T.; Russell, D.; Ferrini, R.; Cobb, B. S.; Smale, S. T. *Genes Dev.* **2002**, *16*, 2985.

(47) Nardelli, J.; Gibson, T. J.; Vesque, C.; Charnay, P. *Nature* **1991**, *349*, 175.

(48) Nardelli, J.; Gibson, T.; Charnay, P. *Nucleic Acids Res.* **1992**, *20*, 4137.

(49) Desjarlais, J. R.; Berg, J. M. *Proteins* **1992**, *12*, 101.

(50) Desjarlais, J. R.; Berg, J. M. *Proteins* **1992**, *13*, 272.

(51) Desjarlais, J. R.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7345.

(52) Desjarlais, J. R.; Berg, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2256.

(53) Kim, C. A.; Berg, J. M. *Nat. Struct. Biol.* **1996**, *3*, 940.

(54) Scott, J. K.; Smith, G. P. *Science* **1990**, *249*, 386.

(55) Smith, G. P.; Scott, J. K. *Methods Enzymol.* **1993**, *217*, 228.

(56) Hoess, R. H. *Curr. Opin. Struct. Biol.* **1993**, *3*, 572.

(57) Rebar, E. J.; Pabo, C. O. *Science* **1994**, *263*, 671.

(58) Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11163.

(59) Choo, Y.; Sanchez-Garcia, I.; Klug, A. *Nature* **1994**, *372*, 642.

(60) Jamieson, A. C.; Kim, S. H.; Wells, J. A. *Biochemistry* **1994**, *33*, 5689.

(61) Wu, H.; Yang, W. P.; Barbas, C. F., III. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 344.

(62) Jamieson, A. C.; Wang, H.; Kim, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12834.
(63) Greisman, H. A.; Pabo, C. O. *Science* **1997**, *275*, 657.
(64) Wolfe, S. A.; Grant, R. A.; Elrod-Erickson, M.; Pabo, C. O. *Structure* **2001**, *9*, 717.
(65) Fairall, L.; Schwabe, J. W.; Chapman, L.; Finch, J. T.; Rhodes, D. *Nature* **1993**, *366*, 483.
(66) Houbaviy, H. B.; Usheva, A.; Shenk, T.; Burley, S. K. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13577.
(67) Elrod-Erickson, M.; Benson, T. E.; Pabo, C. O. *Structure* **1998**, *6*, 451.
(68) Nolte, R. T.; Conlin, R. M.; Harrison, S. C.; Brown, R. S. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2938.
(69) Foster, M. P.; Wuttke, D. S.; Radhakrishnan, I.; Case, D. A.; Gottesfeld, J. M.; Wright, P. E. *Nat. Struct. Biol.* **1997**, *4*, 605.
(70) Pavletich, N. P.; Pabo, C. O. *Science* **1993**, *261*, 1701.
(71) Omichinski, J. G.; Pedone, P. V.; Felsenfeld, G.; Gronenborn, A. M.; Clore, G. M. *Nat. Struct. Biol.* **1997**, *4*, 122.
(72) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705.
(73) Dawson, P. E.; Kent, S. B. *Annu. Rev. Biochem.* **2000**, *69*, 923.
(74) Hofmann, R. M.; Muir, T. W. *Curr. Opin. Biotechnol.* **2002**, *13*, 297.
(75) Jantz, D.; Berg, J. M. *J. Am. Chem. Soc.* **2003**, *125*, 4960.
(76) Liu, Q.; Segal, D. J.; Ghiara, J. B.; Barbas, C. F., III. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 5525.
(77) Kim, J. S.; Pabo, C. O. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 2812.
(78) Bartsevich, V. V.; Juliano, R. L. *Mol. Pharmacol.* **2000**, *58*, 1.
(79) Choo, Y.; Castellanos, A.; Garcia-Hernandez, B.; Sanchez-Garcia, I.; Klug, A. *J. Mol. Biol.* **1997**, *273*, 525.
(80) Hynes, N. E.; Stern, D. F. *Biochim. Biophys. Acta* **1994**, *1198*, 165.
(81) Altiok, N.; Bessereau, J. L.; Changeux, J. P. *EMBO J.* **1995**, *14*, 4258.
(82) Lee, K. F.; Simon, H.; Chen, H.; Bates, B.; Hung, M. C.; Hauser, C. *Nature* **1995**, *378*, 394.
(83) Kraus, M. H.; Fedi, P.; Starks, V.; Muraro, R.; Aaronson, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2900.
(84) Alimandi, M.; Romano, A.; Curia, M. C.; Muraro, R.; Fedi, P.; Aaronson, S. A.; Di Fiore, P. P.; Kraus, M. H. *Oncogene* **1995**, *10*, 1813.
(85) Siegel, P. M.; Ryan, E. D.; Cardiff, R. D.; Muller, W. J. *EMBO J.* **1999**, *18*, 2149.
(86) Beerli, R. R.; Dreier, B.; Barbas, C. F., III. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1495.
(87) Beerli, R. R.; Segal, D. J.; Dreier, B.; Barbas, C. F., III. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14628.
(88) Sadowski, I.; Ma, J.; Triezenberg, S.; Ptashne, M. *Nature* **1988**, *335*, 563.
(89) Margolin, J. F.; Friedman, J. R.; Meyer, W. K.; Vissing, H.; Thiesen, H. J.; Rauscher, F. J., III. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4509.
(90) Zhang, L.; Spratt, S. K.; Liu, Q.; Johnstone, B.; Qi, H.; Raschke, E. E.; Jamieson, A. C.; Rebar, E. J.; Wolffe, A. P.; Case, C. C. *J. Biol. Chem.* **2000**, *275*, 33850.
(91) Liu, P. Q.; Rebar, E. J.; Zhang, L.; Liu, Q.; Jamieson, A. C.; Liang, Y.; Qi, H.; Li, P. X.; Chen, B.; Mendel, M. C.; Zhong, X.; Lee, Y. L.; Eisenberg, S. P.; Spratt, S. K.; Case, C. C.; Wolffe, A. P. *J. Biol. Chem.* **2001**, *276*, 11323.
(92) Ordiz, M. I.; Barbas, C. F., III; Beachy, R. N. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13290.
(93) Wolffe, A. P.; Hayes, J. J. *Nucleic Acids Res.* **1999**, *27*, 711.
(94) Urnov, F. D. *Biochem. Cell Biol.* **2002**, *80*, 321.
(95) Rebar, E. J.; Huang, Y.; Hickey, R.; Nath, A. K.; Meoli, D.; Nath, S.; Chen, B.; Xu, L.; Liang, Y.; Jamieson, A. C.; Zhang, L.; Spratt, S. K.; Case, C. C.; Wolffe, A.; Giordano, F. J. *Nat. Med.* **2002**, *8*, 1427.
(96) Xu, G. L.; Bestor, T. H. *Nat. Genet.* **1997**, *17*, 376.
(97) McNamara, A. R.; Hurd, P. J.; Smith, A. E.; Ford, K. G. *Nucleic Acids Res.* **2002**, *30*, 3818.
(98) Kim, Y. G.; Cha, J.; Chandrasegaran, S. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 1156.
(99) Kim, Y. G.; Shi, Y.; Berg, J. M.; Chandrasegaran, S. *Gene* **1997**, *203*, 43.
(100) Smith, J.; Berg, J. M.; Chandrasegaran, S. *Nucleic Acids Res.* **1999**, *27*, 674.
(101) Smith, J.; Bibikova, M.; Whitby, F. G.; Reddy, A. R.; Chandrasegaran, S.; Carroll, D. *Nucleic Acids Res.* **2000**, *28*, 3361.
(102) Porteus, M. H.; Baltimore, D. *Science* **2003**, *300*, 763.
(103) Bibikova, M.; Beumer, K.; Trautman, J. K.; Carroll, D. *Science* **2003**, *300*, 764.
(104) Merkle, D. L.; Schmidt, M. H.; Berg, J. M. *J. Am. Chem. Soc.* **1991**, *113*, 5450.
(105) Lin, Q.; Barbas, C. F., III; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 612.
(106) Kim, C. A.; Berg, J. M. *Nature* **1993**, *362*, 267.
(107) Blasie, C. A.; Berg, J. M. *Biochemistry* **1997**, *36*, 6218.
(108) Kraulis, P. J. *J. Appl. Crystallogr.* **1991**, *24*, 946.

CR020603O