

# Building zinc fingers by selection: Toward a therapeutic application

(gene therapy/transcriptional regulation/phage display/combinatorial libraries/human immunodeficiency virus)

HERREN WU, WEI-PING YANG, AND CARLOS F. BARBAS III\*

Department of Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037

Communicated by K. Barry Sharpless, The Scripps Research Institute, La Jolla, CA, September 16, 1994 (received for review August 19, 1994)

**ABSTRACT** A phage display approach was utilized to modify the specificity of each of the three fingers of the murine transcription factor Zif268. Selections were performed by using the consensus binding sequence of the natural protein and a conserved sequence in the genome of the type 1 human immunodeficiency virus. By using an extensive randomization strategy, the entire 3-bp specificity of a finger has been changed. Rapid analysis of selected zinc fingers was facilitated by the development of an immunoscreening assay for DNA binding and specificity. To investigate the mechanism of binding and specificity, the binding kinetics of Zif268 and 10 selected variants were determined in real time with an assay based on surface plasmon resonance. Differential mechanisms for sequence-specific recognition were observed. No evidence in support of a single general coding relationship between zinc finger and target DNA sequence was observed. The prospects for the development of this class of proteins in human therapy are considered.

The study of the interaction of proteins with nucleic acids remains a central theme in the area of molecular recognition. Within the known classes of DNA binding proteins, the zinc-finger motif is among the most fascinating. The zinc-finger domain was first observed in the transcription factor IIIA (TFIIIA) of *Xenopus laevis* (1). The TFIIIA zinc-finger sequence motif, Cys-(Xaa)<sub>2-4</sub>-Cys-(Xaa)<sub>3</sub>-Phe-(Xaa)<sub>5</sub>-Leu-(Xaa)<sub>2</sub>-His(Xaa)<sub>3-5</sub>-His, encodes a modular recognition domain stabilized by the tetrahedral coordination of a zinc ion with the conserved cysteine and histidine residues. This motif has since been identified in over 220 proteins, which together contain more than 1300 zinc fingers (2, 3).

The structure of this motif was first determined using NMR methods in 1989 (4) and shown to consist of an antiparallel  $\beta$ -ribbon packed against an  $\alpha$ -helix. Subsequent studies reveal the conservation of this fold in finger motifs of different sequences (5, 6). Naturally occurring proteins generally contain multiple repeats of the zinc-finger motif, which suggests a modular nature that is unique among the classes of DNA binding proteins. Polydactyl arrays of as many as 37 zinc-finger domains allow this modular recognition domain to recognize extended asymmetric sequences (2, 3). The x-ray crystal structures of the zinc fingers of the murine transcription factor Zif268 (7), the protein encoded by the human *GLI* oncogene (8), and the *Drosophila* transcription factor tramtrack (9) in complex with DNA have now been reported and give insight into polydactyl recognition. These structures reveal that the individual finger domains contact DNA in the major groove with specific contacts to the DNA from the amino-terminal part of the helix. This class of proteins is distinguished from other classes of DNA binding proteins not only in its modularity but also in its ability to recognize RNA as well as DNA (10). Together these features make zinc-finger proteins an interesting system to explore molecular recognition and a

possible alternative to antisense or catalytic RNA approaches in modulating gene expression.

In this report we investigate the malleability of this motif to recognize defined DNA targets. Combinatorial libraries of zinc fingers were constructed by randomizing six residues in the amino-terminal part of the recognition helix. Fingers with altered DNA recognition specificity were selected by phage display. Each of the three fingers of the murine transcription factor Zif268 has been examined. The prospects and potential applications of zinc fingers in therapy are considered. This report extends earlier insights into the recognition of DNA by this class of proteins as obtained by site-directed mutagenesis and phage display approaches (11–14).

## MATERIALS AND METHODS

**Reagents, Strains, and Vectors.** Heparin-Sepharose CL-6B medium was from Pharmacia. Oligonucleotides were from Operon Technologies (Alameda, CA) or were prepared on a Gene Assembler Plus (Pharmacia LKB) in the laboratory. pZif89 was a gift from N. P. Pavletich and C. O. Pabo (7). *Escherichia coli* BL21(DE3)pLysS and plasmid pET3a were from Novagen, *E. coli* XL1-Blue, phage VCSM13, the phagemid vector pComb3, and pAraHA are as described (15–17).

**Plasmid Construction.** Genes encoding wild-type zinc-finger proteins were placed under the control of the *araB* promoter by insertion of a DNA fragment amplified by PCR containing the wild-type Zif268 gene of pZif89 (7) into pAraHA. The resulting plasmid vector, pAra-Zif268, expressed the zinc-finger protein as a fusion with a hemagglutinin decapeptide tag at its carboxyl terminus, which could be detected with an anti-decapeptide monoclonal antibody (see Fig. 1A) (18). For phage display, the phagemid pComb3 was modified by removal of the antibody light-chain fragment, yielding plasmid pComb3.5. The Zif268 PCR fragment was inserted into pComb3.5 as above.

**Zinc-Finger Libraries.** Three zinc-finger libraries were constructed by PCR overlap extension using conditions previously described (19). Briefly, for finger 1 library, primer pairs Aseq (5'-GTCCATAAGATTAGCGGATCC-3') and Zf16rb [5'-CTGGCCTGTGTGGATGCGGATATG(MNN)<sub>5</sub>CGA-MNNAGAAAAGCGGCGATCGCAGGA-3', where N is A, T, G, or C and M is A or C], and Bseq (5'-GTGAGCGAG-GAAGCGGAAGAG-3') and Zf1f (5'-CATATCCGCATC-CACACAGGCCAG-3') were used to amplify fragments of the Zif268 gene using plasmid pAra-Zif268 as a template. The two PCR fragments were used as templates for overlap extension, and libraries were constructed in pComb3.5 as described (19). Finger 2 and 3 libraries were constructed in a similar manner except that the PCR primers Zf16rb and Zf1f used in finger 1 library construction were replaced by Zfnsi1b (5'-CATGCATATTCGACACTGGAA-3') and Zff2r6f [5'-CAGTGTGCAATATGCATGCGTAACTTC(NNK)<sub>6</sub>ACCA-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: BSA, bovine serum albumin; HIV-1, type 1 human immunodeficiency virus; TFIIIA, transcription factor IIIA.

\*To whom reprint requests should be addressed.

CCCACATCCGCACCCAC-3', where K is G or T] for finger 2 library and by Bzf3 (5'-GGCAAACCTTCCTCCCACAAAT-3') and Zf36k [5'-ATTTGTGGGAGGAAGTTTGCNNK-AGT(NNK)<sub>5</sub>CATACCAAATCCATTTA-3'] for finger 3 library.

**In Vitro Selection of Zinc Fingers.** A 34-nucleotide hairpin DNA containing either the consensus or an altered Zif268 binding site was used for zinc-finger selection (see Fig. 1B). The consensus binding site is denoted as Z268N (5'-CCTGCGTGGGCGCCCTTTTGGGCGCCCACGCAGG-3'). The altered site for finger 1 is TGT (5'-CCTGCGTGGT-GTCCCTTTTGGGACACCACGCAGG-3'), for finger 2 is TTG (5'-CCTGCGTTGGCGCCCTTTTGGGCGCCAACGCAGG-3'), and for finger 3 is CTG (5'-CCTCTGTGGGCGCCCTTTTGGGCGCCCACAGAGG-3'). The oligonucleotide was synthesized with a primary *n*-hexyl amino group at its 5' end. A DNA-bovine serum albumin (BSA) conjugate was prepared by mixing 30 μM DNA with 3 μM acetylated BSA in a solution containing 100 mM 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride and 40 mM *N*-hydroxysuccinimide at room temperature for 5 hr or overnight. Zif268 phage, 10<sup>12</sup> colony-forming units, in 50 μl of zinc buffer (10 mM Tris Cl, pH 7.5/90 mM KCl/1 mM MgCl<sub>2</sub>/90 μM ZnCl<sub>2</sub>/5 mM dithiothreitol) containing 1% BSA was applied to a microtiter well precoated with 4.9 μg of DNA-BSA conjugate in 25 μl of PBS buffer (10 mM potassium phosphate/160 mM NaCl, pH 7.4) per well. After a 2-hr incubation at 37°C, the phage was removed, and the plate was washed once with TBS buffer (50 mM Tris Cl/150 mM NaCl, pH 7.5) containing 0.5% Tween for the first round of selection. The plate was washed 5 times for round 2 and 10 times for three additional rounds. Subsequent steps were as described (15, 16).

**Immunoscreening.** Mutant zinc-finger genes selected after five rounds of panning were subcloned into the pAraHA vector. Specificity of unpurified protein in cell lysates could be assayed using an ELISA with immobilized DNA-BSA conjugate and alkaline phosphatase conjugated anti-decapeptide antibody.

**Overexpression and Purification of Zinc-Finger Proteins.** Zinc-finger proteins were overproduced by using the pET expression system (20). The Zif268 gene was introduced after PCR into *Nde* I- and *Bam*HI-digested vector pET3a. Inclusion bodies containing zinc-finger protein were purified on a Heparin-Sepharose FPLC column (1.6 × 4.5 cm) equilibrated with zinc buffer. Proteins were eluted with a 0–0.7 M NaCl gradient. Protein was properly folded after elution from the column. The yield of purified protein was from 7 to 19 mg/liter of cell culture. Protein was over 90% homogeneous as determined by densitometry of SDS/PAGE gels.

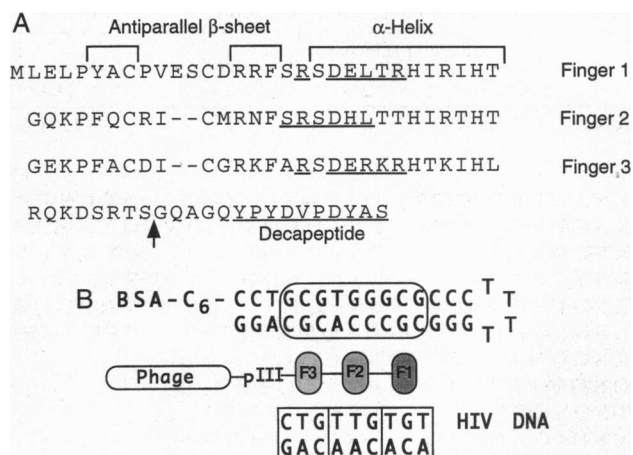
**Kinetic Analysis.** The kinetic constants for the interactions between Zif268 proteins and their DNA targets were determined by surface plasmon resonance-based analysis using the BIAcore instrument (Pharmacia Biosensor) (21). The surface of a sensor chip was labeled with affinity-purified streptavidin (Pierce). Oligonucleotides were then captured onto the chip by injection of 40 μl of biotinylated oligonucleotides (50 μg/ml) in 0.3 M sodium chloride. The association rate (*k*<sub>on</sub>) was determined by studying the rate of binding of the protein to the surface at five different protein concentrations ranging from 10 to 200 μg/ml in the zinc buffer. The dissociation rate (*k*<sub>off</sub>) was determined by increasing the flow rate to 20 μl/min after the association phase. The *k*<sub>off</sub> value is the average of three measurements. The *k*<sub>on</sub> and *k*<sub>off</sub> values were calculated using Biacore kinetics evaluation software.

**RESULTS**

**Library Design and Selection.** Phage display of the Zif268 protein was achieved by modification of the phagemid display system pComb3 (15). A test panning experiment showed that

Zif268-displaying phage bound its consensus target DNA sequence 9-fold over the TFIIIA sequence or BSA and demonstrated that sequence-specific binding of the finger complex is maintained during phage display. The level of background binding of phage to the TFIIIA sequence is similar to the nonspecific binding observed in the display of other proteins on the surface of phage (15, 16). A 4-fold reduction in phage binding was noted when Zn<sup>2+</sup> and dithiothreitol were not included in the binding buffer. Two reports published during the course of this work verify that Zif268 can be displayed on the phage surface (13, 14), although unusual growth conditions were utilized to ensure folding in the earlier report. Likewise, the first three fingers of TFIIIA were displayed on the surface of phage and also shown to retain specific binding activity.

Libraries of each of the three fingers of Zif268 were independently constructed using the previously described overlap PCR mutagenesis strategy (19). Randomization was limited to six positions due to constraints in the size of libraries that can be routinely constructed (22). Zinc-finger protein recognition of DNA involves an antiparallel arrangement of protein in the major groove of DNA; i.e., the amino-terminal region is involved in 3' contacts with the target sequence, whereas the carboxyl-terminal region is involved in 5' contacts (Fig. 1). Within a given finger-DNA subsite complex, contacts remain antiparallel where in finger 1 of Zif268, guanidinium groups of arginine at helix positions -1 and 6 hydrogen bond with the 3' and 5' guanine residues, respectively, of the GCG target sequence. Contact with the central base in a triplet subsite sequence by the side chain of the helix position 3 residue is observed in finger 2 of Zif268, fingers 4 and 5 of GLL, and fingers 1 and 2 of tramtrack. Within the three reported crystal structures of zinc finger-DNA complexes, direct base contact has been observed between the side chains of residues -1 to 6 with the exception of residue 4 (7–9). Based on these observations, residues corresponding to the helix positions -1, 2, 3, 4, 5, and 6 were randomized in the finger 1 and 3 libraries. The serine of position 1 was conserved in these experiments since it is well conserved at this position in zinc-finger sequences in general and completely conserved in Zif268 (2). In



**FIG. 1.** Amino acid sequence of the Zif268 protein and the hairpin DNA used for phage selection. (A) The Zif268 protein is aligned to show the conserved features of each zinc finger. The  $\alpha$ -helices and antiparallel  $\beta$ -sheets are indicated. The six amino acid residues underlined in each finger sequence were randomized in library constructions. The carboxyl-terminal end of Zif268 protein was fused with a fragment containing a decapeptide tag. The position of fusion is indicated by an arrow. (B) The hairpin DNA containing the 9-bp consensus binding site (5'-GCGTGGGCG-3'; enclosed) of wild-type Zif268 was used for affinity selection of phage-displayed zinc-finger proteins. In addition, the 3-bp subsites (boxed) of consensus HIV-1 DNA sequence were substituted for wild-type Zif268 3-bp subsites for affinity selection.

the finger 2 library, helix positions -2, -1, 1, 2, 3, and 4 were randomized to explore a different mutagenesis strategy where the -2 position is examined since both Zif268 and GLI structures reveal this position to be involved in phosphate contacts and since it will have a context effect on the rest of the domain. Residues 5 and 6 were fixed since the target sequence TTG retained the 5' thymidine of the wild-type TGG site. Introduction of ligated DNA by electroporation resulted in the construction of libraries consisting of  $2 \times 10^9$ ,  $6 \times 10^8$ , and  $7 \times 10^8$  independent transformants for finger libraries 1, 2, and 3, respectively. Each library results in the display of the mutagenized finger in the context of the two remaining fingers of wild-type sequence.

To examine the potential of modifying zinc fingers to bind defined targets and to ultimately examine their potential in gene therapy, a conserved sequence within the type 1 human immunodeficiency virus (HIV-1) genome was chosen as a target sequence. The 5' leader sequence of the HIV-1 HXB2 clone at positions 106-121 relative to the transcriptional initiation start site represents one of several conserved regions within HIV-1 genomes (23, 24). For these experiments we have targeted the 9-bp region, 113-121, shown in Fig. 1B.

**Sequence Analysis of Selected Fingers.** Expression of the selected proteins in pAraHA resulted in the fusion of the mutant Zif268 proteins with a peptide tag sequence recognized by a monoclonal antibody (Fig. 1A). Binding was determined in an ELISA format using crude cell lysates. A qualitative assessment of specificity can also be achieved with this methodology, which is sensitive to at least 4-fold differences in affinity (Tables 1 and 2). Several positive clones from each selection were sequenced and are given in Table 1. Finger 1 selection with the consensus binding site GCG revealed a strong selection for lysine at position -1 and arginine at position 6. Clone C7 was preferentially enriched in the selection based on its occurrence in 3 of the 12 clones sequenced. Selection using the HIV-1 target sequence in this region, TGT, revealed a diversity of sequences with a selection for residues with hydrogen-bonding site chains in position -1 and a modest

selection for glutamine at position 3. Finger 2 selection using the consensus TGG subsite showed a selection for an aromatic residue at -1, whereas selection using the HIV-1 target TTG demonstrated a selection for a basic residue at this position. The preference for serine at position 3 may be relevant in the recognition of thymidine. Contact of thymine with serine has been observed in the GLI and tramtrack structures (8, 9). Selections were performed utilizing a *supE* strain of *E. coli*, which resulted in the reading of the amber codon TAG as a glutamine during translation. Of the 53 sequences presented in Table 1, 14 clones possessed a single amber codon. No clones possessed more than one amber codon. Selection for suppression of the amber stop codon in *supE* strains has been noted in other DNA-binding protein libraries (25). For further characterization, high-level expression of zinc-finger proteins was achieved using the T7 promoter (20). Clones C10, F8, and G3 each possessed an amber codon that was converted to CAG to encode for glutamine prior to expression in this system.

**Characterization of Affinity and Specificity.** To gain insight into the mechanism of altered specificity or affinity we have determined the kinetics of binding using real-time changes in surface plasmon resonance (21). The kinetic constants and calculated equilibrium dissociation constants of 11 proteins are shown in Table 2. The calculated equilibrium dissociation constants for Zif268 binding to its consensus sequence in the form of the designed hairpin or a linear duplex lacking the tetrathymidine loop are virtually identical, suggesting that the conformation of the duplex sequence recognized by the protein is not perturbed in conformation within the hairpin. The value of 6.5 nM for Zif268 binding to its consensus sequence is in the range of 0.5 to 6 nM reported using electrophoretic mobility shift assays for this protein binding to its consensus sequence within oligonucleotides of different length and sequence (7, 13, 14). As a measure of specificity, the affinity of each protein was determined for binding to the native consensus sequence and a mutant sequence in which one finger subsite had been changed. Clone C7 is improved 13-fold in affinity for binding the wild-type sequence GCG. The major

Table 1. Amino acid sequences of selected zinc-finger proteins

Finger 1 selection		Finger 2 selection		Finger 3 selection	
GCG	TGT	TGG	TTG	GCG	CTG
<b>R S D E L T R</b> (WT)		<b>S R S D H L</b> (WT)		<b>R S D E R K R</b> (WT)	
-1 1 2 3 4 5 6	-1 1 2 3 4 5 6	-2 -1 1 2 3 4	-2 -1 1 2 3 4	-1 1 2 3 4 5 6	-1 1 2 3 4 5 6
K S A D L K R (C7)*†	Q S T A S K A (F8)*‡	T Y L N T P	G V T M Q P (G3)*	R S D L A N S	N S V G D K P*
K S C V R G R (C9)*‡	P S T H L Q T (F15)*‡	G Y R A A P	P Q P L S D	S S G Q W W R (A14)*†	S S W I C G I
K S C D R G R	P S E R T Q P	L Y R Y H L	R E Q V S R (G4)	S S L L V I A	I S A W M E L
K S Y C R T R	T S S E A D H	P T L V N A	T H M W M I	V S S V R G L	I S M M T F F
K S Q L P W T (C10)*	S S E Q R Y P*	V R P H Q R	Q R M R T L (G5)		R S E C R M L
K S N S Q H P*	H S Q Q N K P	P F C P Y R	Q R V G L F		I S A L L D T
K S C Q M D S*	R S G Q G M A		L R T G N Y (G6)		N S V Q G L R
Q S Q V T R T*	R S A R Q T G		E R E F S L		
T S Q S Q S P*	E S N S F T D		E K E S R G		
V S H I Q A N*	N S V M G H D		E G V R K N		
	N S R G Q R K		T G V N S I		
	S S R P S Q W		T Q A R P P		
	T S S E A D H		T H M W M I		

The six randomized residues of fingers 1 and 3 at positions -1, 2, 3, 4, 5, and 6 in the  $\alpha$ -helical region and of finger 2 at -2, -1, 1, 2, 3, and 4 are shown. The serine at position 1 in the finger 2 and finger 3 selections was not mutated and is shown in boldface type. The three nucleotides denote the binding site used for affinity selection of each finger. Proteins studied in detail are indicated with a clone designation. The zinc-finger proteins in this table bind to their BSA-conjugated DNA target with a signal 2- to 15-fold higher than to BSA as determined by immunoscreening. Some were further characterized by immunoscreening to test their ability to bind nontarget DNA as stated below. WT, wild type.

\*These zinc-finger proteins bind to their DNA target with a signal 2- to 8-fold higher than to a nontarget DNA, TF3N (5'-TACTTGGATGGGAGACCGCCTTTTGGCGTCTCCCATCCAAGTA-3').

†These zinc-finger proteins bind to their DNA target with a signal 2- to 8-fold higher than to a DNA with a 3-bp change compared to the target DNA.

‡These zinc-finger proteins bind to their DNA target with a signal 1.2- to 1.8-fold higher than to a DNA with a 3-bp change compared to the target DNA.

Table 2. Kinetic and calculated equilibrium dissociation constants of zinc-finger proteins

Zinc-finger protein	Binding site	$k_{\text{on}} \times 10^{-4}$ , $\text{M}^{-1}\text{s}^{-1}$	$k_{\text{off}} \times 10^4$ , $\text{s}^{-1}$	$K_{\text{d}}$ , nM	$K_{\text{d}}/K_{\text{d}}$ (target)
<i>Finger 1</i>					
WT	<b>GCG</b>	$3.1 \pm 0.04$	$2.0 \pm 0.1$	6.5	1
	TGT	$1.1 \pm 0.2$	$9.0 \pm 1.0$	81.8	12.6
C7	<b>GCG</b>	$2.4 \pm 0.4$	$1.5 \pm 0.7$	6.3	1
	TGT	$8.0 \pm 0.7$	$0.4 \pm 0.1$	0.5	108.8
C9	<b>GCG</b>	$2.0 \pm 0.2$	$1.3 \pm 0.3$	6.5	1
	TGT	$0.9 \pm 0.1$	$23.0 \pm 3.0$	255.6	39.3
C10	<b>GCG</b>	$1.8 \pm 0.1$	$4.5 \pm 2.0$	25.0	1
	TGT	$0.3 \pm 0.002$	$1.4 \pm 0.1$	46.7	1.9
F8	<b>TGT</b>	$3.7 \pm 1.0$	$11.0 \pm 1.5$	29.7	1
	<b>GCG</b>	$4.8 \pm 0.1$	$52.0 \pm 0.9$	108.3	3.6
F15	<b>TGT</b>	$1.9 \pm 0.1$	$7.9 \pm 1.0$	41.6	1
	<b>GCG</b>	$0.9 \pm 0.03$	$17.0 \pm 1.7$	188.9	4.5
<i>Finger 2</i>					
WT	<b>TGG</b>	$3.1 \pm 0.04$	$2.0 \pm 0.1$	6.5	1
	TTG	$1.2 \pm 0.2$	$9.9 \pm 0.02$	83.7	12.9
G3	<b>TTG</b>	$1.7 \pm 0.2$	$2.7 \pm 0.2$	15.9	1
	TGG	$2.7 \pm 0.3$	$6.0 \pm 0.2$	22.2	1.4
G4	<b>TTG</b>	$3.3 \pm 0.2$	$2.1 \pm 0.1$	6.4	1
	TGG	$2.5 \pm 0.6$	$5.7 \pm 0.2$	22.8	3.6
G5	<b>TTG</b>	$0.8 \pm 0.1$	$2.2 \pm 0.02$	27.5	1
	TGG	$1.9 \pm 0.2$	$9.1 \pm 0.1$	47.9	1.7
G6	<b>TTG</b>	$10.0 \pm 1.0$	$4.6 \pm 0.3$	4.6	1
	TGG	$0.7 \pm 0.1$	$1.4 \pm 0.1$	20.0	4.3
<i>Finger 3</i>					
WT	<b>GCG</b>	$3.1 \pm 0.04$	$2.0 \pm 0.1$	6.5	1
	CTG	$0.9 \pm 0.1$	$8.8 \pm 0.03$	101.0	15.5
A14	<b>GCG</b>	$1.3 \pm 0.1$	$1.7 \pm 0.0$	13.1	1
	CTG	$0.2 \pm 0.0$	$10.0 \pm 0.4$	500.0	38.2

Each zinc finger protein studied is indicated by a clone designation (see Table 1). The target DNA site used for selection of each finger is indicated in boldface type. The consensus binding site for the wild-type (WT) protein is also shown in boldface type. The non-hairpin duplex DNA (underlined) was prepared by annealing two single-stranded DNAs.  $K_{\text{d}}$  was calculated from  $K_{\text{d}} = k_{\text{off}}/k_{\text{on}}$ .

contribution to this improvement in affinity is a 5-fold slowing of the dissociation rate of the complex. Specificity of the C7 protein is also improved 9-fold with respect to the HIV-1 target sequence. This result suggests that additional or improved contacts are made in the complex. Studies of protein C9 demonstrate a different mechanism of improved specificity. In this case, the overall affinity of C9 for the GCG site is equivalent to Zif268, but the specificity is improved 3-fold over Zif268 for binding to the TGT target site by an increase in the off rate of this complex. Characterization of proteins F8 and F15 demonstrates that the 3-bp recognition subsite of finger 1 can be completely changed to TGT and that new fingers can be selected to bind this site.

Characterization of proteins modified in the finger 2 domain and selected to bind the TTG subsite reveals that the specificity of this finger is amenable to modification. Proteins G4 and G6 bind an oligonucleotide bearing the new subsite with affinities equivalent to Zif268 binding its consensus target. Specificity of these proteins for the target on which they were selected to bind is demonstrated by an  $\approx$ 4-fold better affinity for this oligonucleotide as compared to the native binding site, which differs by a single base pair. This level of discrimination is similar to that reported for a finger 1 mutant (14). The finger 3 modified protein A14 was selected to bind the native finger 3 subsite and binds this site with an affinity that is only 2-fold lower than Zif268. Note that protein A14 differs radically in sequence from the native protein in the recognition subsite. Sequence specificity in 10 of the 11 proteins characterized was provided by differences in the stability of the complex. Only a single protein, G6, achieved specificity by a dramatic change in on rate.

## DISCUSSION

Naturally occurring zinc fingers have been characterized, which specifically bind to DNA target sites of diverse sequence composition (7–9). We have demonstrated in this report that each of the three zinc fingers of Zif268 is amenable to modification by using an *in vitro* selection strategy based on phage display. Libraries of each of the three fingers of Zif268 were targeted for random mutagenesis of the amino-terminal part of the recognition helix. To take into account context effects, wherein each amino acid may have an effect on the structural disposition of adjacent residues, six residues in this region were targeted for mutagenesis. Previous reports have demonstrated some success in modifying finger specificity by site-directed mutagenesis (11, 12) or phage display by targeting the four core-recognition residues at helix positions –1, 2, 3, and 6 (13, 14). With our more extensive mutagenesis strategy, the entire 3-bp recognition site of a zinc finger has been modified. Ideally, all eight residues between the conserved phenylalanine and the first histidine in the consensus sequence would be randomized. Unfortunately a complete survey of this region would exceed the capacity of the phage display approach (22).

The crystal structure of the Zif268–DNA complex revealed a simple coding relationship between protein and DNA target (7). To explore the issue of a coding relationship, selections were performed by using the native consensus sites as well as new target sites. Our results have implications for the development of a coding relationship between protein sequence and DNA subsite sequence. Selection of the finger 1 library for binding to the wild-type binding site GCG produced a strong

selection for arginine at position 6. This result is in agreement with the wild-type protein and the structure of the protein-DNA complex, which reveals direct hydrogen bonding contacts with the 3' guanine of the target site. Position -1, after selection, produced a strong selection for lysine even though the wild-type protein has an arginine at this position, which contacts a 5' guanine via two hydrogen bonds. In the structure of the Zif268 complex, the arginine at position -1 is stabilized by an intramolecular hydrogen bond to aspartic acid at position 2. In our experiment, covariation of residues at positions -1 and 2 was observed in the finger 1 GCG selection. Predominately, residues of complementary charge were selected at these positions, as suggested by the structure of the complex. However, the protein of the highest affinity and specificity, C7, did not follow this trend. Characterization of protein A14 derived from the finger 3 GCG selection demonstrated that a finger sequence radically different from the wild type can bind the wild-type target with only a 2-fold reduction in affinity. Perhaps the most surprising feature of the selected proteins is the observation that a number of clones containing proline have survived the selection. In  $\alpha$ -helical regions, proline residues are generally preferred prior to the start of the helix and act as helix initiators (26). Their appearance within helical structures is less common, and within the  $\alpha$ -helical region of zinc-finger proteins in particular, their appearance at positions beyond residue 2 has not been observed (2). The secondary structure of these proteins in the contact region may be different from those previously studied. Collectively these data suggest that a single general coding relationship between protein and target DNA sequence does not exist. Multiple coding relationships may exist, however, which define subsets of finger-DNA sequence.

Each of the 11 proteins studied in detail demonstrated a higher affinity for the target on which it was selected. Studies of the binding kinetics of this class of proteins to DNA were facilitated using real-time changes in surface plasmon resonance (21). These studies revealed that changes in the specificity of the fingers are governed in almost all cases by changes in the stability of the protein-DNA complex as reflected in the  $k_{off}$  values. Furthermore, different mechanisms for the improvement or modification of Zif268 proteins for the target were observed. Based on off-rate measurements, the  $t_{1/2}$  of the Zif268-DNA complex is improved from 58 min for the natural protein to 289 min for the C7 protein. Measurements of half-lives of DNA-protein complexes using electrophoretic mobility shift assays are difficult and have led to some disagreement in the literature (27, 28). A previous report has emphasized electrostatic constraints in the selection of fingers where the net charge of the region under selection is constrained between +1 and -1 (14). We observe a variation in net charge over the selected region of +3 to -2, with a bias toward a net positive charge.

Modularity of the zinc-finger domain has been demonstrated previously by the construction of additional three-finger complexes (12). Should this observation prove to be general, the tethering of individually selected domains would allow for the specific recognition of any DNA sequence. This would lead to a new and potentially useful class of therapeutic and research molecules, which could be utilized as both positive and negative genetic switches. A unique site in the human genome could be targeted by a six-finger complex of the Zif268 type. Control at the DNA level is advantageous since only a single site needs to be occupied as compared to targeting the many RNA messages that might be produced. Recently, the three zinc fingers of Egr-1 have been demonstrated to inhibit transformed cell growth, a result that gives insight into possible future applications (29). We suggest that

zinc fingers may be developed as specific antiviral or antitumor agents wherein the finger would be delivered genetically or through a delivery vehicle as the protein. As inhibitors of transcription, it should be possible to target not only promoter elements but also conserved coding regions of proteins where high-affinity binding should block the processivity of the polymerase. Blocking RNA polymerase elongation by targeting conserved coding regions, while more difficult, may allow for specific gene regulation since the problem of unique recognition within more common promoter elements might be avoided. Such a strategy should be of utility in HIV-1 infection to produce resistant cell populations or to inhibit viral replication. The ability of this motif to bind RNA could also be utilized to facilitate this goal.

We thank Richard A. Lerner for support and encouragement of this project, Joel Gottesfeld and Peter Wright for insightful discussions, Kim Green for protein purification, and Denise Hoekstra and J. S. Koh for early contributions. C.F.B. is a Scholar of the American Foundation for AIDS Research. This work was supported by an Investigator Award from the Cancer Research Institute.

1. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4**, 1609.
2. Jacobs, G. H. (1992) *EMBO J.* **11**, 4507-4517.
3. Rhodes, D. & Klug, A. (1993) *Sci. Am.* **268**, 56-65.
4. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989) *Science* **245**, 635-636.
5. Omichinski, J. G., Clore, G. M., Appella, E., Sakaguchi, K. & Gronenborn, A. M. (1990) *Biochemistry* **29**, 9324-9334.
6. Klevit, R. E., Herriott, J. R. & Horvath, S. J. (1990) *Proteins Struct. Funct. Genet.* **7**, 215-226.
7. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252**, 809-817.
8. Pavletich, N. P. & Pabo, C. O. (1993) *Science* **261**, 1701-1707.
9. Fairall, L., Schwabe, W. R., Chapman, L., Finch, J. T. & Rhodes, D. (1993) *Nature (London)* **366**, 483-487.
10. Clemens, K. R., Wolf, V., McBryant, S. J., Zhang, P., Liao, X., Wright, P. E. & Gottesfeld, J. M. (1993) *Science* **260**, 530-533.
11. Nardelli, J., Gibson, T. & Charnay, P. (1992) *Nucleic Acids Res.* **20**, 4137-4144.
12. Desjarlais, J. R. & Berg, J. M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2256-2260.
13. Rebar, E. & Pabo, C. O. (1994) *Science* **263**, 671-673.
14. Jamieson, A. C., Kim, S.-H. & Wells, J. A. (1994) *Biochemistry* **33**, 5689-5695.
15. Barbas, C. F., III, Kang, A. S., Lerner, R. A. & Benkovic, S. J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7978-7982.
16. Barbas, C. F., III, & Lerner, R. A. (1991) *Methods Companion Methods Enzymol.* **2**, 119-124.
17. Barbas, C. F., III, & Wagner, J. (1994) *ImmunoMethods*, in press.
18. Field, J., Nikawa, J.-I., Broek, D., MacDonald, B., Rodgers, L., Wilson, I. A., Lerner, R. A. & Wigler, M. (1988) *Mol. Cell. Biol.* **8**, 2159-2165.
19. Barbas, C. F., III, Bain, J. D., Hoekstra, D. M. & Lerner, R. A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4457-4461.
20. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60-89.
21. Malmqvist, M. (1993) *Curr. Opin. Immunol.* **5**, 282-286.
22. Barbas, C. F., III (1993) *Curr. Opin. Biotechnol.* **4**, 526-530.
23. Yu, M., Ojwang, J., Yamada, O., Hampel, A., Rapaport, J., Looney, D. & Wong-Staal, F. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6340-6344.
24. Myers, G., Korber, B., Berzofsky, J. A., Smith, R. F. & Pavlakis, G. N. (1992) *Human Retroviruses and AIDS 1992* (Theoretical Biology and Biophysics, Los Alamos, NM).
25. Huang, L., Sera, T. & Schultz, P. G. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 3969-3973.
26. Richardson, J. S. & Richardson, D. C. (1988) *Science* **240**, 1648-1652.
27. Liao, X., Clemens, K. R., Tennant, L., Wright, P. E. & Gottesfeld, J. M. (1992) *J. Mol. Biol.* **223**, 857-871.
28. Choo, Y. & Klug, A. (1993) *Nucleic Acids Res.* **21**, 3341-3346.
29. Huang, R.-P., Darland, T., Okamura, D., Mercola, D. & Adamson, E. D. (1994) *Oncogene* **9**, 1367-1377.