In Vitro Selection of Zinc Fingers with Altered DNA-Binding Specificity†

Andrew C. Jamieson,‡,§ Sung-Hou Kim,‡ and James A. Wells*,§

Department of Chemistry and Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, and Department of Protein Engineering, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080

Received November 19, 1993; Revised Manuscript Received March 3, 1994®

ABSTRACT: We have used random mutagenesis and phage display to alter the DNA-binding specificity of Zif268, a transcription factor that contains three zinc finger domains. Four residues in the helix of finger 1 of Zif268 that potentially mediate DNA binding were identified from an X-ray structure of the Zif268–DNA complex. A library was constructed in which these residues were randomly mutated and the Zif268 variants were fused to a truncated version of the gene III coat protein on the surface of M13 filamentous phage particles. The phage displayed the mutant proteins in a monovalent fashion and were sorted by repeated binding and elution from affinity matrices containing different DNA sequences. When the matrix contained the natural nine base pair operator sequence 5′-GCG-TGG-GCG-3′, native-like zinc fingers were isolated. New finger 1 variants were found by sorting with two different operators in which the singly modified triplets, GTG and TCG, replaced the native finger 1 triplet, GCG. Overall, the selected finger 1 variants contained a preponderance of polar residues at the four sites. Interestingly, the net charge of the four residues in any selected finger never deviated more that one unit from neutrality despite the fact that about half the variants contained three or four charged residues over the four sites. Measurements of the dissociation constants for two of these purified finger 1 variants by gel-shift assay showed their specificities to vary over a 10-fold range, with the greatest affinity being for the DNA binding site for which they were sorted. We were unable to enrich for clones that bind to five other binding sites (ACG, CCG, CGC, ATA, and TAT), suggesting modification of just these four residues in finger 1 may not allow it to adapt to all DNA binding sites. The studies show it is possible to isolate zinc fingers by phage display that distinguish operator sequences that differ by a single base change. Moreover, such selection methods should aid in clarifying rules for zinc finger–DNA recognition.

An understanding of the forces that drive protein–DNA interactions is essential to the design of molecules that bind DNA. The ability to design these molecules has many potential applications including the directed control of gene expressions or the development of biochemical reagents, such as tools for gene mapping.

The zinc finger DNA-binding domain was first observed in the transcription factor TFIIIA from the oocytes of the African clawed toads *Xenopus laevis* (Miller et al., 1985). A single finger is approximately 30 amino acid residues in length and can be identified by the zinc-chelating sequence motif Cys-$(X)_{2-4}$-Cys-$(X)_3$-Phe-$(X)_5$-Leu-$(X)_2$-His-$(X)_{3-5}$-His. To date, over 1600 zinc fingers have been found in over 200 eukaryotic intracellular proteins, where they may occur singly or in sets of up to 30 [reviewed in Berg (1990)]. Several NMR studies on isolated single fingers have shown that the peptide consists of an α helix containing the two invariant histidine residues, coordinated through zinc to the two cysteines of a single β turn (Lee et al., 1989; Klevit et al., 1990; Omichinski et al., 1990).

The X-ray crystal structure of a complex between DNA and Zif268, a three-finger peptide from a murine transcription factor (Pavletich & Pabo, 1991), shows that the α helix of each finger lies in the major groove of DNA. Residues in the amino-terminal part of the helix make direct base contacts. The specific interactions between the helix and DNA are relatively simple: each finger in Zif268 recognizes a three base pair nucleotide sequence on one strand of the DNA duplex only (Figure 1). There is a one-to-one correspondence between side chains that contact individual DNA bases, and among the fingers only three positions on the helix are involved in making these contacts. Zif268 is therefore especially modular in DNA binding, suggesting the possibility for developing base recognition rules (Pavletich & Pabo, 1991; Desjarlais & Berg, 1992, 1993). More recent data for the human glioblastoma zinc finger–DNA complex show in this case that side-chain interaction with particular bases is more complex and varies with changes in the precise docking of the finger and its binding site (Pavletich & Pabo, 1993).

In this paper we investigate the DNA recognition properties of Zif268 by random mutagenesis and selection by monovalent phage display (Bass et al., 1990). The Zif268 zinc finger peptide was displayed in a functional form on filamentous phage particles, and it was possible to generate fingers with altered DNA-binding specificity from a library in which the amino-terminal finger (finger 1) was randomized at its base contact positions.

EXPERIMENTAL PROCEDURES

Subcloning of Zif268 Zinc Finger Peptide. The zinc finger peptide coding sequence was amplified by PCR from a plasmid containing *zif268* cDNA (Christy et al., 1988) obtained from the ATCC[1] (No. 63027). Appropriate restriction sites were introduced during the amplification to facilitate subcloning the PCR product into a phagemid vector, phGHamg3,

† This work was supported by grants from the NIH (AI30725) and the Office of Health and Environmental Research, DOE.

* To whom correspondence should be addressed.
‡ University of California.
§ Genentech, Inc.

® Abstract published in *Advance ACS Abstracts,* April 15, 1994.

0006-2960/94/0433-5689$04.50/0

described for hGH monovalent phage display (Bass et al., 1990; Lowman et al., 1991). The final construct (designated pZF005) contained residues 333–420 from the *zif268* cDNA clone. The gene was preceded by the promoter for alkaline phosphatase and the StII signal sequence (Chang et al., 1987) and followed by an amber codon at residue 249 of the gene for the M13 gIII coat protein (Lowman et al., 1991). The amber codon permits expression of the zinc finger peptide either as a secreted peptide or as a coat protein fusion simply by switching from a nonsuppressor to a suppressor (*supE*) strain, respectively. Single phenylalanine residues at the amino and carboxy termini were introduced during cloning, such that the predicted peptide sequence is 90 amino acid residues in length (Figure 2).

Expression and Purification of Zif268 Zinc Finger Peptide. We found that the most efficient production of the mature Zif268 peptide occurred when the growth medium was supplemented with 0.1 mM $ZnCl_2$. The zinc finger peptide was secreted into the periplasm of *Escherichia coli* 16C9 (a nonsuppressor strain) under conditions of phosphate starvation to activate the alkaline phosphatase (AP) promoter (Chang et al., 1987). Briefly, overnight cultures were inoculated in LB medium with 50 μg/mL carbenicillin. The cells were diluted 1:50 in 25 mL of AP medium supplemented with 0.1 mM $ZnCl_2$ and grown for 24 h in shaker flasks at 37 °C. Cells were harvested by centrifugation and the pellets stored for at least 1 h at −80 °C.

To purify the Zif268 peptide, the pellets were resuspended in 0.5 mL of low salt buffer containing 10 mM Tris-HCl (pH 8.0) for 1 h on ice which released most soluble periplasmic proteins. The peptide remained associated with the cell pellet despite the fact it was expressed with a signal peptide which was properly processed. The high p*I* of the peptide (>11) may have caused a tight association with the negatively charged membrane during secretion. This was advantageous because the peptide was easily extracted from the cell pellet with high salt buffer containing 10 mM Tris-HCl and 0.5 M NaCl (pH 8.0) for 30 min on ice. The high salt extract contained secreted peptide at up to 75% purity, as estimated by SDS–PAGE (Figure 3). The peptide was blotted onto nitrocellulose, and the N-terminal residues were sequenced to verify that the signal peptide was properly processed. The peptide was purified further for use as a concentration standard by reverse-phase chromatography on a Vydac C18 column using an acetonitrile gradient (0–60%) with 0.1% TFA. The purified product was submitted for total amino acid analysis to determine its concentration. Subsequent concentration determinations were by laser densitometry scanning of high salt extracts separated by 14% SDS–PAGE and stained with Coomassie blue, as described (Cunningham & Wells, 1989). Peptide concentrations were in the 5–10 μM range, corresponding to yields of 1–2 mg/L.

A minor clipped form of the peptide was occasionally detected by SDS–PAGE. Mass spectrometry of a reverse-phase purified sample containing the mature peptide and the smaller form showed that the clip occurred immediately prior to an arginine residue located four residues from the peptide carboxy terminus and outside the third finger domain (Figure 2). The shortened molecule had essentially the same binding affinity as the cleaved form (data not shown).

[1] Abbreviations: ATCC, American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD; BSA, bovine serum albumin; hGH, human growth hormone; PCR, polymerase chain reaction; PEG, poly(ethylene glycol); SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; STII, the *Escherihia coli* heat-stable enterotoxin II; TFA, trifluoroacetic acid; Tris, tris(hydroxymethyl)aminomethane. Oligonucleotide sequences are designated in a 5′ to 3′ orientation.

Table 1: Binding of Zif268 and Zif268–Phagemid Particles to the Zif268 Operator or Nonspecific DNA[a]

sample	cpm bound, Zif268 operator	nonspecific DNA
Zif268–peptide	4999	702
Zif268–phagemid	1014	249
	953	104
non-Zif268–phagemid	136	152
	77	133

[a] The Zif268 operator duplex corresponds to GATCC-GCG-TGG-GCG-CTGCAGA and its complement. The nonspecific DNA duplex corresponds to GGTCTGTCTAAAGGTTGCTTCGGCCTGAAAC-TGGATCGTATCGGA and its complement. DNA was end-labeled and incubated with protein or phagemid samples as described in Experimental Procedures. Protein–DNA complexes were bound to nitrocellulose filters, and radioactivity was measured. The non-Zif268–phagemid contained the *hGH* gene instead of the *zif268* gene (Lowman et al., 1991).

Zinc Finger Phage Display. Phagemid particles displaying the Zif268 zinc fingers were propagated in *E. coli* XL-1 Blue (Stratagene) that contains an amber suppressor mutation to permit expression of the zinc finger–M13 gIII fusion protein. Phagemid particles were propagated by growing the cells in 25 mL of 2YT medium containing 50 μg/mL carbenicillin and 0.1 μM $ZnCl_2$ overnight at 37 °C, with approximately 10^{10} KO7 helper phage (Vierra & Messing, 1987; Lowman et al., 1991). Phagemid stocks were purified by precipitation with 3.3% poly(ethylene glycol) 8000 plus 0.4 M NaCl and diluted to titers between 10^{13} and 10^{14} colony forming units (cfu)/mL.

Specific DNA binding to the Zif268 three-finger binding site GCG-TGG-GCG was quantified by a filter binding assay (Yarus & Berg, 1970). A 21-mer oligonucleotide containing the binding site (GATCC-GCG-TGG-GCG-CTGCAGA, designated OZF 001) was end-labeled with [γ-^{32}P]ATP and annealed to its unlabeled complementary strand (TCTGCAG-CGC-CCA CGC-GGATC, designated OZF003). The Zif268 peptide (1 nM) or the phage displaying Zif268 (2×10^{12} particles /mL) was incubated in 50 μL containing 50 nM end-labeled DNA in 10 mM Tris (pH 8.0) and 10 mM NaCl at 25 °C for 10 min. The amount of DNA bound was quantified by binding the complexes to nitrocellulose filters, washing with same buffer minus DNA, and scintillation counting. The purified Zif268 peptide and the Zif268 display phage both bound five to eight times more of the end-labeled cognate operator (OZF001/OZF003) than a noncognate operator (Table 1). Moreover, phage not containing the Zif268 gene fusion gave the same background level of binding for each operator. From the number of phagemid particles and the counts specifically retained, we calculate that about 5% of the phagemid particles display a functional finger domain. This is comparable to the level of display seen for a similar monovalent hGH construct (Bass et al., 1990; Lowman & Wells, 1991).

Construction of the Phagemid Library. The Zif268 phagemid library was constructed by site-directed mutagenesis according to Kunkel et al. (1991). As a mutagenesis template, we used a frame-shifted derivative of the zinc finger phagemid constructed by introducing a termination codon at residue 15 to avoid background contamination with the wild-type sequence. A 47-mer oligonucleotide (TCC TGC GAT CGT CGA TTT TCT NNS TCG NNS NNS CTT ACC NNS CAT ATC CGC ATC, designated OZF011, where N is any of the four possible bases and S is either G or C) was used to randomize residues 18, 20, 21, and 24 in the Zif268 peptide (Figure 2). These residues correspond to the helix positions −1, 2, 3, and 6 in finger 1 that interact with GCG underlined in the sequence GCG-TGG-GCG (Figure 1). The library

was electroporated into *E. coli* XL-1 Blue cells (Stratagene). The number of independent transformants was estimated at 5×10^7, approximately 50-fold greater than the maximum number of DNA sequences in the library (32^4).

Affinity Matrix Construction and Phagemid Selection. An affinity matrix was constructed by binding biotinylated DNA to the wells of microtiter plates coated with streptavidin. The 21-mer oligonucleotides containing either the Zif268 or altered finger 1 binding sites were used for affinity selection. The oligonucleotide OZF001 and its complementary 5′-biotinylated strand OZF003 were used to select for the Zif268 finger 1 binding site, GCG-TGG-GCG. Selection for altered finger 1 binding sites was with similar oligonucleotides, in which the GCG binding site was replaced with ATA, TAT, CGC, GTG, ACG, CCG, and TCG.

Microtiter plate wells (Nunc Maxisorp) were filled with 100-μL aliquots of a 10 μg/mL solution of streptavidin (Pierce) in 50 mM carbonate buffer (pH 9.0) and incubated overnight at 4 °C. The wells were rinsed and refilled with 50 mM carbonate buffer, pH 8.0. DNA targets were prepared by annealing 2 μM solutions of oligonucleotides containing the binding sequence to 1 μM solutions of complementary 5′-biotinylated oligonucleotides. Aliquots (2 μL) of annealed DNA were added to the wells and the solutions incubated overnight at 4 °C. The wells were rinsed and incubated with blocking solution (5 mg/mL acetylated BSA, 50 mM carbonate, pH 8.0) and finally rinsed with a low salt buffer (10 mM Tris-HCl, pH 8.0, 10 mM NaCl). Acetylated BSA was prepared from globulin-free BSA (Sigma) in water (10 mg/mL) with acetic anhydride (20 μL/mL). After 1 h at room temperature, the solution was dialyzed overnight at 4 °C in 100 volumes of 0.2 M NaCl and 10 mM Tris-HCl (pH 8.0) and stored at −20 °C. We were unable to sort efficiently without using the acetylated BSA blocking agent.

Aliquots (10 μL) of Zif268 phagemid stocks (approximately 5×10^{11} cfu) were added to the rinsed wells containing 90-μL aliquots of 10 mM NaCl and 10 mM Tris-HCl (pH 8.0) and incubated at 4 °C for at least 3 h to establish binding equilibrium. The phagemid solutions were removed and the wells rinsed 15–20 times with the same low salt buffer. Bound phagemids were eluted by incubating each well with 100 μL of 0.5 M NaCl and 10 mM Tris-HCl (pH 8.0) for 1 h at room temperature. Aliquots (10 μL) were used to titer the eluted phagemid particles, and the remaining 90 μL was incubated with 1 mL of log phase *E. coli* XL-1 Blue cells (approximately 10^8 cells) for 1 h at 37 °C. The specific enrichment for Zif268 phagemids per round was between 1 and 2 orders of magnitude over background levels. We found enrichments were higher when biotinylated oligonucleotides were bound to streptavidin from Pierce rather than Sigma. Background levels were calculated from negative control wells that contained duplex DNA in which the binding site sequence was replaced with TAT-GTT-TAT. Clones from the enriched phagemid pool were sequenced (Sanger et al., 1977) after various rounds of selection by preparing single-stranded template from colonies isolated from microtiter plates (Vierra & Messing, 1987). Sequencing reactions were performed using a Sequenase kit (USB) according to the manufacturer's instructions.

K_d Determinations. The relative binding affinities of peptides for different base substitutions in the finger 1 binding site were determined from band-shift assays (Fried & Crothers, 1981). The fractions of complexed and free DNA present were measured under varying concentrations of peptides. A set of 11 oligonucleotides was constructed on the basis of oligonucleotide OZF001 (GATCC-GCG-TGG-GCG-CTG-CAGA) by replacing the finger 1 binding site (the underlined GCG) with GNG, GCN, or GTN (where N represents a single base substitution for each oligonucleotide). Each oligonucleotide was end-labeled with γ-^{32}P (6000 Ci/mmol; Du Pont) and annealed to its respective 5′-biotinylated complementary strand. The unlabeled strands were present in approximately 10-fold excess.

Peptide–DNA complexes were formed in 20 μL of binding buffer [1 × TB (80 mM Tris–borate, pH 7.9), 0.1 M KCl, 1 mM $MgCl_2$, 1 mM DTT, 50 nM $ZnCl_2$, 4.5% Ficoll, 0.05% BSA, and 0.2 nM end-labeled duplex DNA] for 15 min at room temperature. Aliquots (10 μL) were then loaded onto 16% polyacrylamide gels containing 0.5 × TB buffer and electrophoresed at 100 V for 40 min. The gels were dried under vacuum onto Whatman paper and exposed to X-ray films (Figure 4). The relative proportions of complexed and free DNA were determined by laser densitometry scanning of autoradiographs (LKB 2202 ultroscan) over a range in which the complex was partially dissociated. Scans were repeated incrementally over the entire width of the band and then averaged. K_d determinations were made from the linear plot, $C/F = [P_f]/K_d$, where C/F is the ratio of complexed to free DNA and $[P_f]$ is the free peptide concentration. $[P_f]$ was substituted by $[P_T]$, the total peptide concentration, when in large excess over the DNA concentration. The average root mean square error for these determinations was estimated at 30%, where 85% of the data points were within 1 standard deviation from their calculated mean.

RESULTS

Design of the Zif268 Phage Display Library. The Zif268 peptide consists of three fingers that interact with a single strand of DNA in the binding site. This strand runs 3′ to 5′ relative to the amino–carboxy orientation of the peptide, such that finger 1 interacts with the GCG triplet underlined in the duplex binding site GCG-TGG-GCG. The guanidinium groups of Arg side chains at positions 1 and 6 contact the 3′ and 5′ guanines, respectively (Figure 1). Finger 2 is the only finger in Zif268 in which contacts are made with the middle base of its triplet, where His at position 3 contacts a guanine.

The phagemid library was constructed by randomizing finger 1 at four positions, corresponding to positions −1, 2, 3, and 6 on the recognition helix (indicated by asterisks in Figure 2). Positions −1 and 6 were targeted because of their direct interactions with the DNA, and position 2 was randomized since Asp2 appears to form a salt bridge with the Arg−1 side chain. Position 3 in finger 1 was randomized to see if it was possible to select for altered binding sites containing substitutions at the middle base.

Sorting with the Wild-Type Zif268 DNA Binding Site. We selected variants from the Zif268 phagemid library that bound to the wild-type operator sequence (GCG-TGG-GCG) to test the substitution tolerance at these four positions and to look for other variants capable of binding the operator. After three rounds of binding selection, about 10 times more phagemid particles bound to the Zif268 binding site compared to a negative control sequence (TAT-GTT-TAT). The library was sorted for up to six rounds (with about 50-fold enrichment) and in a separate experiment was sorted for nine rounds.

Clones were sequenced from rounds 3–6 and 9 to follow the evolution of finger 1 and to evaluate consensus features of the selectants (Table 2). All clones contained Arg at position 6, except for one containing His. Position 3 showed a strong preference for Asp or Glu and a few contained small neutral substitutions. Asp predominated in the latter rounds. In a separate library in which residues 3–5 were randomized, we found that Asp3 was selected exclusively (data not shown).

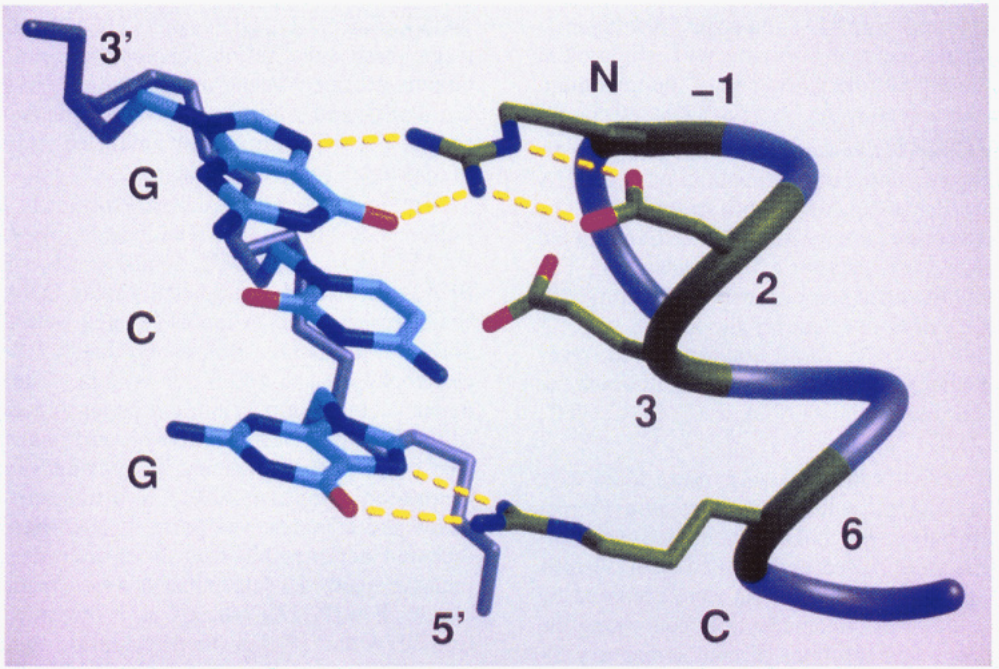

FIGURE 1: Complex between Zif268 finger 1 and its trinucleotide binding site 5′GCG3′. Nitrogen atoms are shown in blue and oxygen atoms are shown in red. Hydrogen bonds are shown in yellow. The numbers –1, 2, 3, and 6 refer to positions on the zinc finger recognition helix that are involved in base contacts in the Zif268 cocrystal (Pavletich & Pabo, 1991). In finger 1, these correspond to Arg, Asp, Glu, and Arg, respectively. Glu at position 3 does not contact the middle base, but His at the corresponding position on finger 2 contacts guanine as the middle base. N and C designate the amino and carboxy orientation of the molecule, respectively.

```
F E R P Y A C P V E S C D R R F S R S D E L T R H I R I H T G Q
                   10                 *  * *  *          30

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

K P F A C - - D I C G R K F A R S D E R K R H T K I H L R Q K F
                 70                    80                ^
```

FIGURE 2: Amino acid sequence of the Zif268 peptide. Residues 2–89 correspond to residues 333–420 of the *zif268* coding sequence (Christy et al., 1988). The sequence is aligned for comparison of each of the three fingers. The conserved histidine and cysteine side chains that coordinate zinc and the recognition helices of each finger are underlined. Residues 18, 20, 21, and 24 correspond to helix positions –1, 2, 3, and 6 on finger 1 (Figure 1) and are marked with asterisks. The circumflex shows the clip site of the minor form of the peptide occasionally detected by SDS–PAGE.

At position –1, virtually all the selectants contained hydrogen-bonding groups including Arg, Lys, Asn, Glu, Asp, Thr, and Ser. Position 2 was much more variable but changed systematically with position –1. For example, we never found two like charges at these positions and often found complementary charges. Usually the positive charge was at position –1 and the negative charge at position 2, as for the wild-type finger 1. However, for two clones in round 9 the charges were reversed. In addition, we often found many large and small side-chain combinations at these two positions.

The amino acid preferences at the four positions sorted for binding the GCG triplet can also be compared statistically by calculating the number of standard deviation units for each substitution above its expected frequency for random selection in the NNS library (Table 3). This analysis suggests that for GCG selection Arg6 is the least tolerant residue to substitution in finger 1. Asp is as strongly preferred at position 3 as Arg is at position 6, and either a positive or negative charge usually appears at position –1. A general correlation that emerges from these data is that like the wild type there is a strong preference for charged residues at all four positions, yet the overall charge is usually 0 (Tables 2 and 3). Although the wild-type finger sequence was not recovered, many were found that resemble it. Virtually all selectants contained Arg6 and many Glu3. Asp at position 2 coselected with Arg at position 1 on three occasions in the context of Arg–1 Asp2 Asp3 Arg6. Another close relative to the wild-type finger 1, which appeared after sorts 5 and 6, contained Arg–1 Ser2 Glu3 Arg6.

Sorting with Altered DNA Binding Sites. To determine how adaptable finger 1 in Zif268 is to other binding sites, we sorted the library with operator sites in which the GCG triplet was altered by a single base pair to GTG, TCG, or ACG. For sorting with GTG, specific enrichment was evident after four rounds and clones were sequenced after nine rounds (Table 2). Again there was exclusive preference for Arg at position 6. However, for GTG either Glu or Thr was found at position –1, instead of Arg–1 commonly found when sorting with the GCG. When Glu was present at position –1, a positive charge was present at position 2, whereas when Thr was present at position –1, Gly covaried at position 2. Thus, like the GCG sort covariation occurred between positions –1 and 2 but for different pairs. Unlike the GCG sort, position 3 was always small and neutral, possibly due to the larger thymine in GTG.

The library was sorted for binding to TCG in the binding site GCG-TGG-TCG. After three rounds, 10 times more phage bound to this site than the noncognate site, TAT-GTT-TAT. Sequences of these clones after rounds 3 and 4 showed a virtual consensus for a single clone, Lys–1 Asn2 Asp3 Lys6 (Table 2). While Asp3 and Lys6 were completely invariant, a few homologous replacements were found for Lys–1 and Asn2. The net charge over these four positions was +1 despite their containing three charged residues.

We also sorted the library with a binding site containing ACG. However, after four rounds we saw no preference for

Table 2: Zif268 Variants Identified after Sorting a Zif268–Phagemid Library[a]

site	sort	helix position −1	2	3	6	no. of charged residues, 4	net charge, pH 7, 0
GCG (wild type)	3	K	S	E	R	3	+1
		K	E	D	R	4	0
		S	Y	D	R	2	0
		T	T	G	R	1	+1
		K	G	D	R	3	+1
		R	D	D	R	4	+1
		A	W	E	R	2	0
	4	E	T	G	R	2	0
		G	K	D	R	3	+1
		K	G	E	R	3	+1
		K	D	D	R	4	0
	5	N	R	D	H	2	0
		R	E	G	R	3	+1
		T	W	E	R	2	0
		R	S	E	R	3	+1
		N	S	D	R	2	0
		R	D	D	R	4	0
	6	D	Y	T	R	2	0
		N	A	D	R	2	0
		R	E	D	R	4	0
		N	R	D	R	3	+1
		S	S	D	R	2	0
		D	Y	G	R	2	0
		D	Y	T	R	2	0
		R	S	E	R	3	+1
		K	G	E	R	3	+1
		R	D	D	R	4	0
	9	D	K	D	R	4	0
		T	A	D	R	2	0
		T	A	D	R	2	0
		D	R	D	R	4	0
		R	D	D	R	4	0
GTG	9	E	R	A	R	3	+1
		T	G	G	R	1	+1
		T	G	A	R	1	+1
		E	K	S	R	3	+1
		T	G	G	R	1	+1
		E	K	S	R	3	+1
TCG	3	K	N	D	K	3	+1
		K	N	D	K	3	+1
		Q	E	D	R	3	−1
		K	N	D	K	3	+1
		R	S	D	K	3	+1
		K	N	D	K	3	+1
		K	N	D	K	3	+1
	4	K	N	D	K	3	+1
		R	N	D	K	3	+1
		K	N	D	K	3	+1
		K	N	D	K	3	+1
		K	N	D	K	3	+1
		K	N	D	K	3	+1
		K	N	D	K	3	+1
		K	N	D	K	3	+1
		K	N	D	K	3	+1
ACG	4	D	R	D	R	4	0
		G	T	E	K	2	0
		E	R	R	T	3	+1
		R	G	Q	E	2	0
		Y	E	E	R	3	−1
		E	G	S	R	2	0

[a] The positions −1, 2, 3, and 6 that were randomized in the library correspond to residues R18, D20, E21, and R24 (RDER) of the wild-type Zif268 peptide. The trinucleotide sequences refer to the finger 1 binding sites used for affinity selection, as described in Experimental Procedures. Clones isolated after sort 9 for the GCG binding site were from an additional experiment (see text for details). Some clones isolated which had the same protein sequence also had the same DNA sequence. This occurred for clones in which there was only one DNA sequence that could possibly code for the protein sequence in the NNS library (e.g., all KNDK clones isolated in the TCG sort had the same DNA sequence).

phage binding to GCG-TGG-ACG over the noncognate site. Nonetheless, we sequenced six clones. There was a trend for a positive charge at position 6, and the charge over the four sites stayed near neutrality. However, unlike the other triplet sites, positions −1, 2, and 3 were highly variable and showed no obvious consensus.

Table 3: Amino Acid Preferences at Positions −1, 2, 3, and 6 of Finger 1 after Sorting the Zif268–Phagemid Library for the Finger 1 Binding Site, GCG[a]

position	residue	P_e	P_f	σ_{32}	$(P_f - P_e)/\sigma_{32}$
−1	Arg	0.094	0.25	0.052	3.0
	Asn	0.031	0.13	0.030	3.3
	Asp	0.031	0.16	0.030	4.3
	Lys	0.031	0.19	0.030	5.3
2	Asp	0.031	0.16	0.030	4.3
	Tyr	0.031	0.13	0.030	3.3
3	Asp	0.031	0.6	0.030	18.9
	Glu	0.062	0.22	0.043	3.7
6	Arg	0.094	0.97	0.052	16.8

[a] The frequencies of amino acids are shown on the basis of their fractional representation (P_f) among all clones isolated after three to nine rounds of sorting. Expected frequencies (P_e) were calculated from the number of corresponding NNS codons, assuming a completely random starting library. The standard deviation (σ) for each amino acid was calculated as $\sigma = [P_e(1 - P_e)/n]^{1/2}$, with $n = 32$. Only residues for which the fraction found exceeded the expected fraction by at least 2σ are shown.

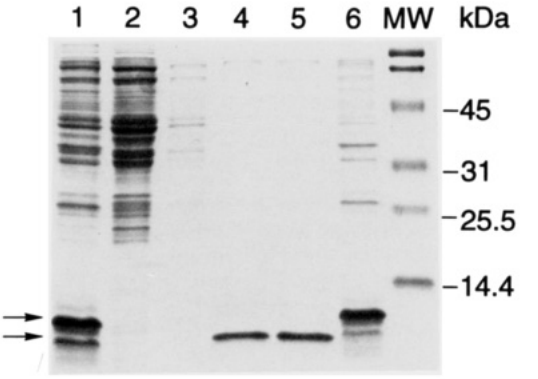

FIGURE 3: Expression and purification of the Zif268 DNA-binding domain. Cell extracts were separated by 16% SDS–PAGE and stained with Coomassie blue. Lane 1: whole cell extracts of *E. coli* 16C9 grown under conditions of phosphate starvation (Chang et al., 1987). Zinc finger bands are marked with arrows. The lower band corresponds to the processed form of the peptide after removal of the STII signal sequence during secretion into the periplasm and was only detected when cells were grown in the presence of zinc (0.1 mM $ZnCl_2$). Lane 2: the soluble protein fraction released from the periplasm after osmotic shock. Lane 3: the residual protein fraction released after the ruptured cells were extracted with 0.2 M NaCl and 10 mM Tris-HCl (pH 8.0) on ice for 30 min. Lanes 4 and 5: protein fractions containing the Zif268 finger peptide after reextraction with either 0.5 M NaCl or 0.75 M NaCl and 10 mM Tris-HCl (pH 8.0) on ice for 30 min, respectively. Lane 6: the remainder of the cell contents after extraction with the high salt buffer. The overexpressed band corresponding to that marked with the top arrow in lane 1 most likely represents the unprocessed form of the peptide.

The library was sorted against the binding site containing CCG and three other triplets that differed from the wild-type site at all three base pairs, CGC, ATA, and TAT. Even after nine rounds of binding selection, none of these phagemid pools showed enrichments for binding to the target DNA over the nonspecific DNA, and clones from them were not sequenced.

Affinity and Specificity for Selected Zinc Fingers. To analyze the affinity and specificity of a few of the isolated finger 1 variants, we measured their dissociation constants with various DNA binding sites by a gel-shift assay (Figure 4). Under approximately physiological salt conditions the affinity of the wild-type Zif268 for the native operator was about 0.5 nM (Table 4). When the middle base or 3′-terminal base was changed to any of the other three nucleotides, affinity

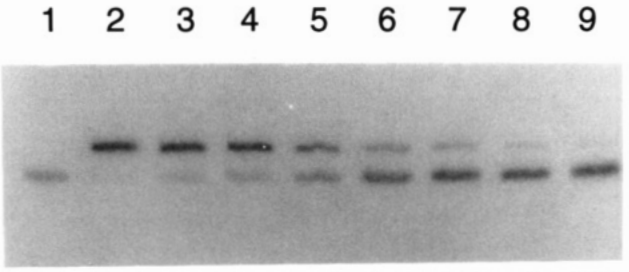

FIGURE 4: Typical band-shift assay showing the titration of the Zif268 peptide with DNA containing an altered finger 1 binding site, GCC. Complexes were formed in 20 μL of 80 mM Tris–borate (pH 7.9), 0.1 M KCl, 1 mM $MgCl_2$, 1 mM DTT, 50 nM $ZnCl_2$, 4.5% Ficoll, 0.05% BSA, and 0.2 nM end-labeled duplex DNA with 2-fold serial dilutions of peptide. After 15 min at room temperature, 10-μL aliquots were separated on 16% acrylamide gels containing 40 mM Tris–borate (pH 7.9) at 100 V for 40 min (lanes 2–9). The peptide concentration in lane 2 was 32 nM. The ratio of free to complexed DNA was determined over a range in which the complex was partially dissociated (lanes 4–7) as described in the Experimental Procedures. Lane 1 is a control lane containing end-labeled DNA only.

Table 4: Relative Affinities of Zif268 Finger 1 Variants[a]

finger 1	5′ binding site 3′	K_d (nM)	rel affinity ($K_d/K_{d\ target}$)
RDER	GAG	2.8 ± 0.4	5.6
	GCG	0.5 ± 0.14	1
	GGG	5.6 ± 0.6	11.2
	GTG	3.4 ± 1.7	6.8
	GCA	2.4 ± 0.5	4.8
	GCC	3.0 ± 0.35	6.0
	GCT	3.7 ± 0.4	7.4
RDDR	GAG	3.0 ± 1.1	7.5
	GCG	0.4 ± 0.1	1
	GGG	3.7 ± 0.8	9.2
	GTG	4.0 ± 0.9	10.0
ERAR	GAG	1.5 ± 0.5	2.5
	GCG	1.5 ± 0.2	2.5
	GGG	1.8 ± 0.4	3.0
	GTG	0.6 ± 0.06	1
	GTA	1.7 ± 0.3	2.8
	GTC	1.8 ± 0.5	3.0
	GTT	2.1 ± 0.3	3.5

[a] The single-letter code corresponds to the following: RDER, Arg–1 Asp2 Glu3 Arg6 (Zif268 finger 1); RDDR, Arg–1 Asp2 Asp3 Arg6; ERAR, Glu–1 Arg2 Ala3 Ala6. The relative affinities for each finger were determined by varying the finger 1 binding site to either GNG, GCN, or GTN. The binding site for RDER and the target site for which each variant was selected are underlined. K_d measurements were from band-shift assays (see Figure 4 and Experimental Procedures for details). The variation for each K_d measurement was calculated as the root mean square error of the data points.

was reduced from 5- to 10-fold depending on the replacement. We noted in sorting the library for GCG binders that Asp usually replaced Glu at position 3, especially in the later rounds. Because position 3 may interact with the middle base, we analyzed the K_d's for the Glu3Asp variant for binding sites that contained each of the four bases at the middle position. The affinity of this variant for the GCG binding site was virtually the same as the wild-type Zif268. Changing the middle base caused about a 10-fold reduction in affinity as was seen for the wild type. Finally, we analyzed one other mutant, Glu–1 Arg2 Ala3 Arg6 which was isolated by sorting with GTG. Indeed this variant had the highest affinity for its selected binding site (GTG). This affinity was comparable to wild type against its preferred site (GCG). This mutant showed a broader specificity where modification of either the middle or terminal base caused only about a 3-fold reduction in affinity. The ratio of the K_d values between this mutant and either the wild type, or the wild-type-like variant shows that there has been a clear change in specificity. For example, the Glu–1 Arg2 Ala3 Arg6 binds ~6 times tighter to GTG and ~3 times weaker to GCG than the wild type.

DISCUSSION

Our experiments suggest that phage display is a powerful tool for sorting libraries of zinc finger proteins for binding to different operator sequences. Specific enrichments were seen after several rounds of sorting with some operator sequences such as GCG, TCG, and to a lesser extent GTG. When clones were analyzed after several rounds of selection for a given operator, they often had similar sequences and differed from clones isolated using different operator sequences. Moreover, some operator sequences gave no enrichments for binding phage, and when sequences from one of these (ACG) were analyzed, little consensus was observed. These data show that operators differ in their ability to select finger variants and suggest that selection is driven by specific DNA binding. Indeed, when the affinities of two of these mutants were analyzed, they had the highest affinity for the site for which they were selected to bind.

Although the specificity advantages determined for the few clones analyzed are mild (only 3–6-fold), the clones were still powerfully selected. Similar results were obtained for sorting variants of hGH for improved affinity for its receptor (Lowman et al., 1991; Lowman & Wells, 1993). Selected hGH variants from libraries containing quartets of randomized residues yielded only 2–6-fold improvements in affinity despite their showing considerable consensus in sequence.

There are limitations of the phage display method that need to be emphasized [for review, see Wells and Lowman (1992)]. Although many of the finger 1 variants isolated in sorting with the wild-type operator sequence were native-like, the wild-type finger 1 (Arg–1 Asp2 Glu3 Arg6) was not recovered in our selection. We think the wild-type sequence should have been present because wild-type residues were isolated at each of the four positions among the selectants. It is possible that there are a number of variants whose affinity for the operator is comparable to or greater than the wild type and that not enough clones were sequenced to isolate it. In addition, the selected variants could bind with comparable affinity but generally express better or be more stable than the wild type and so be overrepresented in the phage pool. The conditions used for sorting the phage were not the same as those used in the gel-shift assay to measure affinity, and this may explain the difference in selection versus affinity. Moreover, very small differences in affinity can lead to a large selective advantage, and the affinity measurements may not be precise enough to determine the true difference between these clones. Any of these could explain why the Asp3 variant competed out the Glu3 wild type in the GCG sort despite their having comparable affinities for various operators in the gel-shift assay.

Zinc Finger Coding Relationships. The ability to alter zinc finger binding specificity has previously been demonstrated by site-directed mutagenesis of a related three-finger peptide, Sp1 (Desjarlais & Berg, 1993). Finger 2 in this molecule has the sequence Arg–1 Asp2 Glu3 Arg6 and recognizes GCG. To bind GCT required the Arg–1Asn substitution plus two additional ones, Asp2Ser and Glu3Ala. It was concluded that the side chains do not function independently in base recognition. This property has also been suggested to explain diminished specificities introduced by site-directed mutagenesis into another closely related three-finger molecule, Krox 20 (Nardelli et al., 1992). The obvious advantage to studying these interactions by phage display is that the entire repertoire of side chains involved in recognition

can be randomized simultaneously to screen for binding specificity.

Several general features about finger 1 arise from the limited number of selections and analysis we have performed. First, virtually all isolates contained Arg6 when libraries were selected with the two binding sites containing a 5′ G (either GTG or GCG). This is consistent with the need for a charged hydrogen bond donor to interact with the guanine as seen in the X-ray structure (Figure 1). Moreover, when the 5′ G is changed to a 5′ T (sorting with GCG versus TCG in the binding site), there is complete consensus for Lys6. Thus, position 6 is very sensitive to the nature of the 5′ base in the triplet. In the X-ray structure of the complex the 3′ G is seen to hydrogen bond with Arg−1, which in turn forms a salt bridge with Asp2. Indeed these residues covaried in all our enrichments; this was also noted earlier from sequence analysis of natural zinc finger proteins [see Berg, (1990) for review]. We generally found complementary charge pairs, or charged and neutral pairs, and we never found like charges at these positions, all of which is consistent with there being important intramolecular interactions between positions −1 and 2. In addition, position 3 is closest to the middle base, and when we sorted with GTG over GCG, we isolated small side chains, suggesting the methyl group of thymine excludes side chains larger than Ser. Thus, changing the middle base in the triplet causes a change in the selective pressure at position 3. Interestingly, we found selection here despite the fact that in the X-ray structure Glu3 in finger 1 does not make direct contact with the middle C. Moreover, positions −1 and 2 are also very different between the GCG and GTG selectants even though the X-ray structure shows only contacts between the Arg−1 and the 3′ G. Thus, it would seem that DNA sequence changes that are nearby, but not in direct contact, can influence the amino acid preference at these positions. Additional structural and binding experiments should help to clarify this.

One of the most striking correlations we found was that the overall charge was close to neutrality despite a strong preference for charged residues at each of the four randomized positions. Over the four positions there were commonly three to four charged residues, yet the overall charge was always between +1 and −1 (usually +1 or 0). This electrostatic constraint may be important either for the structural integrity of finger 1 or for it to bind effectively to its operator sequence. Almost all other substitutions were neutral and polar amino acids. These were generally limited to small side chains, particularly glycine, which may be preferred since it is unlikely to be disruptive or may accommodate buried water. Understanding the general features of these selections may lead to more efficient design of zinc finger phage display libraries, since the diversity of the libraries can be constrained to the sets of amino acids preferred by finger 1 by limiting the codon choices in the random DNA.

We were unable to obtain specific enrichments for five other operator sequences containing altered finger 1 triplets including ACG, CCG, CGC, ATA, or TAT. This suggests that finger 1 mutated at positions −1, 2, 3, and 6 cannot adapt with wild-type affinity to all DNA binding sites. It does not indicate that it is impossible to engineer a finger to bind these sites or that if finger 1 were otherwise modified, it could not be made to bind these sequences. Overall, we are encouraged by the possibilities of using phage display to tailor proteins (especially zinc fingers) to bind different DNA binding sites.

As this paper was going to press, we received a preprint of the work performed by Rebar and Pabo (1994). They also sorted a Zif268 phage display library randomized at positions −1, 2, 3, and 6 in finger 1 but for binding to different operator triplets than we used, namely, GAC, GCA, and CCT. They isolated variants having altered DNA binding affinity and specificity. The changes in specificity reported were comparable to those reported here for distinguishing operators with single base changes. These studies should further stimulate the use of this technology to alter zinc finger binding specificity.

ACKNOWLEDGMENT

The authors thank Tim Clackson for critical reading of the manuscript, Henry Lowman, Brian Cunningham, and David Jackson for useful advice, the oligonucleotide group in Bioorganic Chemistry for supply of oligonucleotides, and Kathy O'Connell and Chris Grimley for assistance with mass spectrometry and peptide sequencing.

REFERENCES

Bass, S., Greene, R., & Wells, J. A. (1990) *Proteins 8*, 309–314.
Berg, J. M. (1990) *Annu. Rev. Biophys. Biophys. Chem. 19*, 405–421.
Chang, C. N., Rey, M., Bochner, B., Heyneker, H., & Gray, G. (1987) *Gene 55*, 189–196.
Christy, B. A., Lau, L. F., & Nathans, D. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7857–7861.
Cunningham, B. C., & Wells, J. A. (1989) *Science 244*, 1081–1085.
Desjarlais, J. R., & Berg, J. M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7345–7349.
Desjarlais, J. R., & Berg, J. M. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90, 2256–2260.
Fried, M., & Crothers, D. M. (1981) *Nucleic Acids Res. 9*, 6505–6525.
Kissinger, C. R., Liu, B., Martin-Bianco, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell 63*, 579–590.
Klevit, R. E., Herriot, J. R., & Horvath, S. J. (1990) *Proteins 7*, 215–226.
Kunkel, T. A., Bebenek, K., & McClary, J. (1991) *Methods Enzymol. 204*, 125–139.
Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A., & Wright, P. E. (1989) *Science 245*, 635–637.
Lowman, H. B., & Wells, J. A. (1993) *J. Mol. Biol. 234*, 564–578.
Lowman, H. B., Bass, S. H., Simpson, N., & Wells, J. A. (1991) *Biochemistry 30*, 10832–10838.
Miller, J., McLachlan, A. D., & Klug, A. (1985) *EMBO J. 4*, 1609.
Nardelli, J., Gibson, T., & Charnay, P. (1992) *Nucleic Acids Res. 20*, 4137–4144.
Omichinski, J. G., Clore, G. M., Apella, E., Sakaguchi, K., & Gronenborn, A. M. (1990) *Biochemistry 29*, 9324–9334.
Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J., & Gronenborn, A. M. (1993) *Science 261*, 438–445.
Pavletich, N., & Pabo, C. O. (1991) *Science 252*, 809–817.
Pavletich, N., & Pabo, C. O. (1993) *Science 261*, 1701–1707.
Rebar, E. J., & Pabo, C. O. (1994) *Science 263*, 671–673.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Sauer, R. T., Jordan, S. R., & Pabo, C. O. (1990) *Adv. Protein Chem. 40*, 1–61.
Vierra, J., & Messing, J. (1987) *Methods Enzymol. 153*, 3–11.
Wells, J. A., & Lowman, H. B. (1992) *Curr. Opin. Struct. Biol. 2*, 597–604.
Yarus, M., & Berg, P. (1970) *Anal. Biochem. 35*, 450–465.