THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 277, No. 6, Issue of February 8, pp. 3850–3856, 2002

Printed in U.S.A.

Validated Zinc Finger Protein Designs for All 16 GNN DNA Triplet Targets*

Received for publication, November 6, 2001
Published, JBC Papers in Press, November 28, 2001, DOI 10.1074/jbc.M110669200

Qiang Liu, ZhenQin Xia, and Casey C. Case‡

From Sangamo BioSciences Inc., Point Richmond Technical Center, Richmond, California 94804

The Cys_2-His_2-type zinc finger DNA-binding proteins can be engineered to bind specifically to many different DNA sequences. A single zinc finger typically binds to a 3–4-base pair DNA subsite. One strategy for design is to identify highly specific fingers that recognize each of the 64 possible DNA triplets. We started with a subgroup of the 64 triplets, the GNN-binding fingers. The GNN-binding fingers have been examined in several studies, but previous studies did not produce specific fingers for all of the 16 GNN triplets. These previous studies did not provide any information on the possible positional or context effects on the performance of these fingers. To identify the most specific design and take the possible positional effects into consideration, we did a large-scale site selection experiment on our GNN designs. From this study, we identified very specific fingers for 14 of the 16 GNN triplets, demonstrating for the first time a clear positional dependence for many of the designs. Further systematic specificity study reveals that the *in vivo* functionality of these zinc finger proteins in a reporter assay depends on their binding affinities to their target sequences, thus giving a better understanding of how these zinc finger proteins might function inside cells.

Transcriptional regulation is achieved mainly by transcription factors that bind to the cis-response elements in target gene promoters. Transcription factors are often composed of two modular domains, a DNA-binding domain and a functional domain. The DNA-binding domain confers the specificity of a transcription factor, because it determines which genes are affected. If the DNA-binding domains could be manipulated to bind any desired sequence, then it would be possible to construct novel transcription factors in which engineered DNA-binding domains are combined with various functional domains. Of the DNA-binding domains that have been manipulated by design or selection, the THIIIA class Cys_2-His_2 zinc-finger proteins (ZFPs)[1] have shown the greatest potential to be engineered into tailor-made transcription factors (1, 2). The Cys_2-His_2 zinc finger is ~30 amino acids in length and contains both an anti-parallel β-sheet and a recognition helix that interacts directly with bases in the major groove of DNA. The best studied zinc finger domain in terms of its DNA base recognition properties is the murine transcription factor Zif268 (3). Zif268 contains three zinc fingers, with each finger interacts mainly with a three to four nucleotide subsite. The human transcription factor, Sp1 is also a three-finger zinc finger protein and a member of the Cys_2-His_2 zinc finger family (4), and it binds to a 9-bp target just as Zif268 does (5). Zinc finger proteins can be altered rationally or selected to bind many different DNA sequences. The importance of amino acids-1, 2, 3, and 6 of the zinc finger recognition helix was demonstrated by experiments using a phage display to select zinc finger proteins with new sequence specificities (6–10). From these studies and from the crystal structures of the Zif268-DNA complex (13, 14) several proposed codes have emerged for the binding of zinc finger proteins to DNA targets (6, 7).

The DNA recognition codes predict many side-chain-base contacts. Still, there are many such interactions that cannot be explained by the codes (11, 12). The codes do not permit the choice of similar ZFP designs for a particular DNA triplet nor do they address finger positional and context issues. The crystal structure of the Zif268-DNA complex suggests possible finger context dependence (13, 14). In addition to contacting 3 basepairs along one strand, the +2 position of the DNA recognition helix can make a cross-strand contact to a base on the opposite strand creating a 4-bp subsite (15–17).

Of all 64 possible 3-bp triplets, the group of 5′-GNN-3′ triplet-binding fingers is the most studied of the zinc finger domains. This is partly because the GNN-binding fingers are particular strong DNA-binding fingers. The amino acid Arg at +6 can make two hydrogen bonds with the 5′ base guanine (13). Recently 16 GNN-binding fingers have been selected and characterized at the finger 2 (F2) position from phage display libraries (12, 18). Not all of the 16 GNN binders are shown to be specific at binding to their target sites. The others showed various binding preferences to non-target sequences. For example, it is clear from the data that the selected +3 Ser and +3 Thr do not specify the middle base cytosine and thymine in the GCG and GTG triplets. The paper concluded that highly specific zinc fingers could not be obtained from phage display alone but only in combination with rational design. Also because of the possible context-dependent effects, these GNN binders selected from finger 2 libraries may not work well at other finger positions. For some GNN triplets, different fingers may be needed for each finger position.

We used both rational designs based on the DNA recognition codes and selections to construct ZFPs for gene regulation. To make sure that our designs were specific for their targets and taking the contextual influences into consideration, we systematically performed site selection experiments on all of our GNN designs. By selecting from a random pool of oligonucleotides, the site selection method has been used by many laboratories to study ZFP-DNA interactions (11, 19) and seems to be an unbiased means of determining ZFP specificity. Here we present a complete GNN finger directory with carefully tested designs for

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

‡ To whom correspondence should be addressed. Tel.: 510-970-6000, ext. 202; Fax: 510-236-8951; E-mail: ccase@sangamo.com.

[1] The abbreviations used are: ZFP, zinc finger protein; N, A/C/G/T; EP2C, erythropoietin gene-binding protein 2C; F1, F2, and F3, finger 1, finger 2, and finger 3 positions, respectively; oligo, oligonucleotide.

ZFP1	F3: RSDNLAR			F2: RSDELRT			F1: DRSNLTR			**Target**		**K_d (nM)**
Target	**G**	**A**	**G**	**T**	**C**	**G**	**G**	**A**	**C**	Cognate	GAGTCGGAC	50
5'-3'	**G14**	**A12**	**G11**	**G10**	**T6**	**G11**	**G12**	**A10**	**C11**	Mt-1	GAG**GT**GGAC	12.5
		G2	C2	T3	C4	T1	T1	G2	G2	Mt-2	GAGTCGGA**G**	> 400
			T1	C1	G4	A1	C1	T1	A1			
						C1		C1				
ZFP2	F3: RSDHLSR			F2: QSSHLAR			F1: QSSDLRR			**Target**		**K_d (nM)**
Target	**G**	**G**	**G**	**G**	**G**	**T**	**G**	**C**	**T**	Cognate	GGGGGTGCT	1
5'-3'	**G11**	**G7**	**G11**	**G12**	**G10**	**A5**	**G10**	**C12**	**T6**	Mt-3	GGGGG**A**GCT	0.5
	T1	T3	T1		A2	T3	T1		C2	Mt-4	GGGGG**C**GCT	1
		A2				C3	C1		G2			
						G1			A2			

FIG. 1. **Selection and gel shift results of ZFP1 and ZFP2.** The target and selected sequences of ZFP1 and ZFP2 are listed from 5′ to 3′direction, and the selected consensus sequences are aligned beneath the 9-bp intended target sequences. The frequencies at each base (the four bases are A, C, G, and T) are listed in descending order. The residues from the −1 to +6 position of each finger are presented above its target sequence. The dissociation constants (K_d) of ZFP1 and ZFP2 to their cognate and mutant target sequences are given in nanomolar (nM) along each target sequence. The mutant sequences that differ from the cognate target sequences are shown by *boldface, underlined letters*.

each of the 16 GNN triplets. The establishment of a GNN directory will allow us to better understand the interaction between the ZFP and DNA, as well as the positional effects of the finger. The highly specific ZFPs constructed using the GNN directory will enhance the practical application of the zinc finger technology.

EXPERIMENTAL PROCEDURES

Zinc Finger Protein Synthesis and Gel Mobility Shift Assays—ZFPs used in this study were designed based on the DNA binding codes and the zinc finger data base of Sangamo BioSciences Inc. The designed ZFPs were then assembled using the human transcription factor Sp1 (amino acids 532–624) (20) as the backbone. To create the synthetic genes encoding ZFPs, we developed a PCR-based assembly procedure that utilizes six overlapping oligonucleotides (21). The PCR products were cloned directly into the *Tac* promoter vector pMal-c2 (New England Biolabs) using the *Kpn*I and *Bam*HI restriction sites. The fusion maltose binding protein ZFPs were purified following the manufacturer's procedures (New England Biolabs). Gel mobility shift assays were performed according to the conditions described (21).

Site Selection Experiment—A complete randomized double strand N_{14} library was synthesized by annealing SBLIB1B, one of the two primer oligos, SBLIB1F (5′-ATCCGAACTCGTTCAATA-3′) or SBLIB1B (5′-GCATGGATCCATTGCAAT-3′) to the N_{14} degenerated oligo (5′-ATCCGAACTCGTTCAATA NNNNNNNNNNNNNN ATTGCAATGGATCCATGC-3′). Then the annealed oligos were filled using a standard Klenow (New England Biolabs) filling reaction. End labeling with polynucleotide kinase used the following reaction: the filled-in library at 5 nM, 2.5 μl of 10× PNK buffer (New England Biolabs; 2 μl of γ-^{32}P, 0.5 μl of T4 polynucleotide kinase), and 19 μl of H_2O. After incubating at 37 °C for 30 min, the reaction volume was increased to 50 μl and purified by passing through a Pharmacia ProbeQuant G-50 column. 5 nmol of the labeled degenerate oligo library were incubated with ZFP protein (10× ZFP K_d) following the established gel mobility shift assay protocol (21). After exposure of the dried polyacrylamide gel to an x-ray film, the position of the shifted ZFP-DNA complex was identified. The same incubation and gel running conditions were used to do the site selection experiment, except a cold library at 5 nmol was used with the amount of protein about 4 times that of the ZFP K_d in the first-round site selection. Using the previously developed x-ray film or prestained rainbow protein marker, the shifted gel band that contains the bound oligos was cut out and pulverized by centrifuging in an Amicon column at 14,000 rpm for 10 min. DNA was eluted from the gel bits again using 10 μl of Tris-EDTA (1/10 strength) following the instructions of the Amicon Gel Extraction Kit (Millipore). We took 4 μl of the DNA elute and, using SBLIB1F and SBLIB1B primer sets, PCR-amplified the selected oligo pool (25 cycles). The PCR product (5 nM) was used to continue the second round of selections. With each round of selection, we decreased the protein amount by 2-fold. After 3–5 rounds of selection, the PCR-amplified products were cloned into the TOPO TA cloning vector (Invitrogen), and about 24 clones were picked and sequenced. All of the sequencing results were compiled and aligned to reflect the consensus sequence.

Luciferase Reporter Gene Assays—After characterization of the DNA binding specificities of EP2C, this ZFP was used to construct a stable Tet-inducible EP2C ZFP cell line (T-Rex-293™, Invitrogen) as described (21). To generate reporter constructs, three tandem copies of the various EP2C target sites were annealed and inserted in front of the SV40 promoter of the pGL3 promoter vector (Promega) between the *Mlu*I and *Bgl*II sites. All of these reporter constructs were confirmed by DNA sequencing. Luciferase reporter assays were performed by co-transfection of luciferase reporter DNA (200 ng) and pCMV-βgal (100 ng, used as an internal control) into the stable EP2C expression cells seeded in 6-well plates. The expression of EP2C was induced with doxycycline (0.05 μg/ml) 24 h after the transfection of reporter constructs. Cell lysates were harvested 40 h post-transfection, and the luciferase activities were measured by the Dual-Light luciferase and β-galactosidase reporter assay system (Tropix).

RESULTS

Site Selection Results Reflect Affinity Differences—To test how dependably the site selection results reflect the actual binding affinities of the ZFP to various related targets, we did site selection experiments on two three-finger ZFPs and later quantitative gel mobility shift assays to measure the binding affinities of some of these ZFP-target interactions (Fig. 1). The site selection results showed that most of our designs preferentially selected the intended target sequences, like fingers 1 and 3 of ZFP1 and ZFP2. The second finger of both ZFPs seemed to prefer an alternative sequence instead of the intended sequence (GTG instead of TCG for ZFP1 and GGA *versus* GGT for ZFP2).

To confirm the site selection results, we used gel shift assays to determine the binding affinities of ZFPs to some related individual targets suggested by the site selection results. We designed two mutant oligos for each of the ZFPs: Mt-1 and Mt-2 for ZFP1; Mt-3 and Mt-4 for ZFP2 (Fig. 1). Overall the binding affinities of the ZFPs to the various target sites supported the site selection data. For ZFP1, the selected consensus sequence, GAG GTG GAC, differs from the intended target sequence, GAG TCG GAC, by two bases. The consensus sequence binds 4 times more tightly to ZFP1 than the originally intended target sequence (K_d 12.5 *versus* 50 in Fig. 1). Besides this TCG finger, other ZFP1 finger designs were very specific for their intended targets. The F1 DRSNLTR was very specific for the GAC triplet. A cytosine to guanine (C → G) change caused the binding affinity to decrease at least 8-fold (Fig. 1). For ZFP2, we tested the degeneracy of the 3′ base of the GGT triplet. The K_d values produced by gel shift experiments supported the selection results, indicating that ZFP2 F2 bound the 3′ base adenine best and then thymine and cytosine with K_d (nM) 0.5, 1, and 1, respectively, as reflected by the occurrences of A_5, T_3, and C_3 at

the sixth base position in the site selection experiment (Fig. 1). Taken together, the strong correlation of site selection results with binding affinities of ZFPs to each of the variant target sequences prompted us to use the site selection method to systematically test all of our GNN finger designs.

A Complete GNN Binding Directory Validated by Site Selection Experiments—We expanded our selection efforts to validate all of our designs for GNN triplets and for each finger position. The goal was to establish a directory filled with highly specific zinc fingers for future rapid construction of sequence specific ZFPs. To take the potential positional effects into consideration, we felt it necessary to test each finger design at different positions. We successfully selected consensus sequences from more than 110 three-finger ZFPs with 4–6 rounds of site selection experiments. The selection results of all 16 GNN binders tested at all 3 positions are listed in Fig. 2. For each GNN triplet, we generally had 2–4 designs and tested all of these variant designs by site selection experiments. Many different designs worked equally well for each triplet, but in Fig. 2 we chose only one design for each triplet.

The GNG Finger Designs—Most of our designs for the GNG-type fingers were very specific, and the same or very similar designs worked for all finger positions (Fig. 2). The common helix motif RSD*X*L*X*R (position −1 to +6) worked for all four GNG triplets, with Asn^{+3} for the middle base adenine, His^{+3} for guanine, Ala^{+3} for thymine, and Asp^{+3} for cytosine. For the +5 position, we tested Ala, Thr, Ser, and Gln; they all showed similar specificity profiles, and these residues seemed to be interchangeable. Also, contrary to previous report (19), we found that the two natural GCG fingers RSDELTR and RSDERKR, from zif268, were not GCG-specific. They selected almost equal numbers of GCG *versus* GTG sequences. Later, gel shift experiments confirmed that the natural zif268 F3, RSDERKR, indeed binds GCG and GTG equally well (Fig. 5). To improve the specificity of the natural GCG finger, we made a change at the +3 position, using Asp^{+3} to replace Glu^{+3}. This change drastically increased the finger's specificity for the middle base cytosine and made RSDDLTR highly specific for the GCG triplet. Finger RSDDLTR (or RSDDLQR) was more specific than the natural GCG fingers and specified GCG at all 3 positions (Fig. 2).

The Strong Positional Dependence of GCA, GAT, GGT, GAA, and GCC Fingers—We used the code-based design QSGDLTR to specify the GCA triplet at three positions (F1, F2, and F3) and tested specificities of this design at 3 positions. To our surprise, the QSGDLTR worked for GCA triplet at only the F2 and F3 positions but not at the F1 position (Fig. 3*A*). We had tested this design at the F1 position in three different proteins, and each time it selected GC(T/N). Gln^{-1}-Gly^{+2}-Ser^{+3}-Arg^{+6} had been selected previously from a randomized F1 library using GCA as the target and had been shown to bind GCA well (8). The variant design, QSGSLTR with an $Asp^{+3} \rightarrow Ser^{+3}$ change, was then used at the F1 position in three different proteins. This $Asp^{+3} \rightarrow Ser^{+3}$ change did significantly enhance the overall F1 specificity, and the QSGSLTR specified all three bases of the GCA triplet at F1 (Fig. 2). To see how the QSGSLTR design behaves at F2 position, we did a side-by-side comparison experiment with two ZFPs harboring these two designs, QSGDLTR and QSGSLTR at F2; F1 and F3 were the same for these two ZFPs. The data showed that when used at F2 position, QSGSLTR specified GTA instead of GCA (data not shown). Thus it seemed that to specify a GCA triplet, the design QSGSLTR can only be used at the F1 position, whereas the QSGDLTR can only be used at F2 and F3 positions (Fig. 2). Ser^{+3} had been selected previously to bind middle base thymine at the F2 position by phage display (6). This is the first evidence that different fingers are needed to specify the same triplet while at different positions.

Positional effects also existed for the GAT and GTT fingers. The code-based design QSSNLAR for GAT only selected GAT at the F1 position but not at the F2 and F3 positions (Fig. 3*B*). Instead, QSSNLAR selected GAA at the F2 and F3 positions. When a previously selected finger, TSGNLVR (18), was used at F2 and F3 positions, it did select GAT at both F2 and F3 (Fig. 3*B*). Similar to the GAT situation, QSSHLTR, a code-based design for GGT, selected GGT at the F1 position but not at F2 and F3; At F2 and F3 positions, it specified GGA instead (Fig. 3*C*). For GGT at F2 and F3, we used TSGHLVR, another finger selected from the F2 libraries (18). Finger TSGHLVR was shown to be highly specific for GGT at the F2 and F3 positions. However, the TSGHLVR was not specific for the GGT at F1 (Fig. 3*C*). These results indicate that many fingers are position-dependent, and a finger selected from an F2 library may only work at the F2 or F3 position. There is evidence suggesting that fingers for GAA and GCC triplets were also position-dependent (Fig. 2).

Evaluation of ZFPs by in Vivo Functional Assays—To ask whether there is a correlation between the functionality and the affinity of a ZFP, we used cell-based transient reporter gene assays to analyze the functionality of one of the designed ZFPs. EP2C binds to a target sequence, GCGGTGGCT, with a 2 nM dissociation constant. Site selection results indicated that the overall specificity for this protein is reasonably good, except finger 3 of EP2C, a natural zif268 F3, recognizes both GCG and GTG (Fig. 4*A*). To confirm the site selection observation, we measured the binding affinities of EP2C to its cognate and the related target variants by standard gel shift analyses. To better compare the binding affinities of the EP2C to their targets, we also measured the binding affinities of Sp1 and zif268 to their respective targets under the same buffer condition (Fig. 5). The K_d values confirmed that the F3 of EP2C bound GTG equally well as GCG but bound GAG with a 2-fold lower affinity. Finger 2 was very specific for the GTG triplet; it bound 15-fold less tightly to a GGG triplet. Finger 1 was also very specific for the GCT triplet; it bound with a 4-fold lower affinity for the GAT triplet and a 2-fold lower affinity for the GCG triplet. Target 3B was a nonspecific control and was used only in the transient transfection assays.

To test the specificities of EP2C to these targets inside cells, we cloned three copies of each of these target sequences into the pGL-3 promoter vector (Promega) as reporter constructs. These luciferase reporters were transiently transfected into a stable Tet-regulated EP2C-VP16 expression stable cell line (21). After the induction of EP2C by doxycycline, the luciferase activities of these reporters were measured and normalized to the co-transfected β-galactosidase activities (Fig. 5*B*). The reporter data showed that overall luciferase activities correlated very well to the actual binding affinities of EP2C to these various targets. The selected consensus sequence gave a luciferase value of 66. The highest affinity targets, 2C0 and 2C2, gave the highest activities, 66 and 62, respectively (Fig. 5*B*). The 2-fold lower affinity targets, 2C1 and 2C5 (4 nM), gave roughly half of the luciferase activity of the target 2C0. The lowest affinity target, 2C3, gave the lowest luciferase activity, and the next lowest affinity target 2C4 gave the second lowest luciferase activity. Target 3B, a negative control, gave a luciferase activity of 0.81, almost the same as the no-target control. The tight correlation of affinity with luciferase activity strongly supports the idea that the *in vivo* functionality of ZFPs is closely related to their DNA binding specificities.

GAG	F1: RSDNLAR			F2: RSDNLAR			F3: RSDNLTR			GAT	F1: QSSNLAR			F2: TSGNLVR			F3: TSANLSR		
5'-3'	**G**	**A**	**G**	**G**	**A**	**G**	**G**	**A**	**G**	5'-3'	**G**	**A**	**T**	**G**	**A**	**T**	**G**	**A**	**T**
	G22	A22	G21	G24	A23	G24	G16	A15	G16		G14	A10	T9	G14	A12	T12	G18	A18	T18
			A1		T1			G1				G2	G2		G1	G2			
												C2	A2		T1				
													C1						
GGG	F1: RSDHLAR			F2: RSDHLTR			F3: RSDHLSR			GAC	F1: DRSNLTR			F2: DRSNLTR			F3: DRSNLTR		
5'-3'	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	5'-3'	**G**	**A**	**C**	**G**	**A**	**C**	**G**	**A**	**C**
	G14	G11	G12	G20	G17	G21	G15	G14	G14		G12	A10	C11	G15	A15	C12	G16	A11	C11
		A2	A1	T1	A4			T1	C1		T1	G2	G2			T2	T2	C4	A6
		C1	C1								C1	T1	A1			G1	C1	G3	G1
												C1						T1	T1
GTG	F1: RSDALTR			F2: RSDALSR			F3: RSDALTR			GGA	F1: QSGHLAR			F2: QSGHLQR			F3: QSGHLQR		
5'-3'	**G**	**T**	**G**	**G**	**T**	**G**	**G**	**T**	**G**	5'-3'	**G**	**G**	**A**	**G**	**G**	**A**	**G**	**G**	**A**
	G15	T14	G15	G15	T12	G15	G14	T13	G12		G15	G15	A10	G17	G13	A13	G14	G14	A14
		G1			G2			G1	T2				G3		A2	G3			
					A1								T2		C2	T1			
GCG	F1: RSDDLTR			F2: RSDDLQR			F3: RSDDLTR			GGT	F1: QSSHLTR			F2: TSGHLVR			F3: TSGHLVR		
5'-3'	**G**	**C**	**G**	**G**	**C**	**G**	**G**	**C**	**G**	5'-3'	**G**	**G**	**T**	**G**	**G**	**T**	**G**	**G**	**T**
	G13	C11	G14	G21	C18	G21	G18	C16	G16		G17	G16	T12	G17	G17	T16	G16	G15	T14
	T1	T3			G2			G2	T1			A1	C4			C1	T1	T2	G4
					T1				C1				A1				C1	A1	
GCA	F1: QSGSLTR			F2: QSGDLTR			F3: QSGDLTR			GGC	F1: DRSHLTR			F2: DRSHLAR			F3: DRSHLAR		
5'-3'	**G**	**C**	**A**	**G**	**C**	**A**	**G**	**C**	**A**	5'-3'	**G**	**G**	**C**	**G**	**G**	**C**	**G**	**G**	**C**
	G17	C12	A12	G19	C18	A18	G7	C7	A6		G15	G12	C9	G13	G10	C7	G13	G14	T6
		G3	T2		G1	C1	A2	G1	T3			A2	T4		A3	T3	A1	T1	C3
		T2	G2					T1				T1	G2			A2			A3
			C1													C1			G3
GCT	F1: QSSDLTR			F2: QSSDLTR			F3: QSSDLQR			GTA	F1: QSGALTR			F2: QSGALAR			F3: QRASLTR		
5'-3'	**G**	**C**	**T**	**G**	**C**	**T**	**G**	**C**	**T**	5'-3'	**G**	**T**	**A**	**G**	**T**	**A**	**G**	**T**	**A**
	G19	C19	T16	G15	C15	T15	G10	C17	T10		G11	T7	A7	G10	T6	A9	G5	G4	A5
			G3				A9	T1	G7			A2	G2		A3	G1	A4	T3	G3
								G1	A2			G1	T2		G1			A1	T1
												C1						C1	
GCC	F1: ERGTLAR			F2: DRSDLTR			F3: DRSDLTR			GTT	F1: TTSALTR			F2: TSGALTR			F3: QSSALTR		
5'-3'	**G**	**C**	**C**	**G**	**C**	**C**	**G**	**C**	**C**	5'-3'	**G**	**T**	**T**	**G**	**T**	**T**	**G**	**T**	**T**
	G19	C12	C11	G20	C20	C13	G17	C13	C13		G21	T16	G8	G9	T8	G5	G5	T2	T3
		T4	G3			T7	A1	A2	T3			G3	T6		A1	T4	A1	G2	A3
		A3	C3					G2	A2			A1	C4					A1	
			T2					T1				C1	A3					C1	
GAA	F1: QRSNLVR			F2: QSGNLAR			F3: QSGNLAR			GTC	F1: DRSALAR			F2: DRSALAR			F3: DRSALAR		
5'-3'	**G**	**A**	**A**	**G**	**A**	**A**	**G**	**A**	**A**	5'-3'	**G**	**T**	**C**	**G**	**T**	**C**	**G**	**T**	**C**
	G19	A19	A10	G11	A10	A10	G8	A8	A6		G14	T7	C9	G15	T11	C9	G13	T7	C6
			T6		T1	G1	A1	G1	G3			A5	T2	T1	A3	T6		G5	G6
			G3									G2	A2		G1	A1		C1	T1
													G1		C1				

FIG. 2. **Summary of the 16 GNN finger designs.** Tested designs are presented for each of the 16 GNN triplets at each position. The α-helix sequence of each finger is presented from the −1 to +6 position, and target triplets are listed in 5′ to 3′ order underneath each finger. The selected consensus sequences are aligned beneath their target sequences.

DISCUSSION

Recently several laboratories have stated that the there is no universal recognition code for directing zinc finger-DNA binding (11, 22). It has been suggested that the combination of selection and rational design is needed to produce specific fingers (12). The design codes do not address the potential positional dependence of fingers and unanticipated protein-DNA interactions. There are several selection schemes that have been developed to select three-finger ZFPs. The first one is the simple parallel pre-selection of individual zinc fingers and the subsequent stitching together of these selected fingers to produce a three-finger ZFP (12, 18, 23). The second scheme

A

Design	F1: QSGDLTR	F2: QSGDLTR	F3: QSGDLTR
Target	**G C A**	**G C A**	**G C A**
5'-3'	G19 C16 T13	G19 C18 A18	G7 C7 A6
	G2 G2	G1 C1	A2 G1 T3
	T1 A2		T1
	C2		

B

Design	F1: QSSNLAR	F2: QSSNLAR	F3: QQSNLAR
Target	**G A T**	**G A T**	**G A T**
5'-3'	G14 A10 T9	G17 A17 A10	G21 A18 A11
	G2 G2	T5	G2 T9
	C2 A2	G2	C1 G1
	C1		
Design	F1: TSGNLVR	F2: TSGNLVR	F3: TSANLSR
Target	**G A T**	**G A T**	**G A T**
5'-3'	G7 A5 T4	G14 A12 T12	G18 A18 T18
	A3 C4 G3	G1 G2	
	T1 T2 C1	T1	
	C1		

C

Design	F1: QSSHLTR	F2: QSSHLAR	F3: QSSHLAR
Target	**G G T**	**G G T**	**G G T**
5'-3'	G17 G16 T12	G18 G18 A16	G13 G13 A7
	A1 C4	T2	A2 A1 T4
	A1		T1 G4
Design	F1: TSGHLVR	F2: TSGHLVR	F3: TSGHLVR
Target	**G G T**	**G G T**	**G G T**
5'-3'	G18 G13 A9	G17 G17 T16	G16 G15 T14
	T2 T4 T5	C1	T1 T2 G4
	A2 C5		C1 A1
	C1 G1		

FIG. 3. **Positional dependence of GCA, GAT, and GGT fingers.** The design of each finger is presented above each target triplet. The amino acid residues from −1 to +6 of each α-helix are listed from *left* to *right*, and the target sequence is listed in 5′ to 3′ order. The selected sequences for each base position are listed in descending order. *A*, positional effects of GCA fingers. The *shaded area* indicates that *F1: QSGDLTR* selected GCT instead of the intended target, GCA. *B*, positional effects of GAT fingers. The *shaded area* indicates that *F2: QSSNLAR* and *F3: QQSNLAR* selected GAA most often and not the intended target sequence, GAT; and *F1: TSGNLVR* selected GNN. *C*, positional effects of GGT fingers. The *shaded area* indicates that *F2: QSSHLAR* and *F3: QSSHLAR* both selected GGA and not the intended target sequence, GGT; and *F1: TSGHLVR* selected GGN.

is the serial zinc finger selection, selecting one finger at a time (17). The third is the recently published parallel pre-selection of two halves of a three-finger ZFP followed by further selection of the recombined three-finger ZFP (24). Although these selection schemes have generated many good fingers, some of the selections also produced nonspecific fingers, for example, the 16 F2 GNN finger selections. Not all of the 16 GNN targets produced specific fingers. The selection did not produce any specific finger for the GTG and GCG triplets (18). Selection can be slow, typically taking 2 weeks to finish the whole process from pre-made libraries. These factors argue for the building of a directory or archive which then could allow the assembly of ZFPs to any target site.

Alternative methods can be used to measure the specificities of ZFPs, such as binding site signatures (7) and the recently published DNA microarrays (25). To validate and further improve our designs, we chose the site selection method as the way to evaluate our finger designs. By selecting DNA fragments from a completely randomized oligonucleotide pool, the DNA binding profiles of the ZFPs can be reflected by the occurrence of the selected nucleotides at each base position. Gel shift experiments supported the selection results, showing that the occurrence of nucleotides at each target position represents the binding affinities of these bases to the tested ZFPs (Fig. 1). Depending on how the site selection experiments are performed (number of selection cycles, stringency, etc.), site selection results can include the consensus sequence and many less optimal interactions; for example, ZFP1 selected the ninth base of the target with C_{11}, G_2, and A_1, and the gel shift confirmed that a C to G change did decrease the binding affinity to ZFP1 by more than 8-fold (Fig. 1, ZFP1).

In contrast to the previous findings, we found that RSDHLTR is very specific for the GGG triplet, although it does have some cross-recognition to the GAG triplets but to a much lesser degree than the reported data based on enzyme-linked immunosorbent assays (18). We used Ala^{+3} to specify the middle base thymine in the GTG triplet and found that Ala^{+3} was more specific than Ser^{+3} in specifying thymine. Ser^{+3} had difficulties in discriminating the middle thymine from cytosine (12, 18). Our designs for the four GNG triplets stand out as very specific for their targets, as they can specify all three bases of their target triplets (Fig. 2). The GNG designs are the most tested designs, with each design listed in Fig. 2 having been tested at least 5 times while in different proteins. In addition to the interchangeability at the +5 position for many amino acid residues, we noticed that the charged residue, Arg^{+5}, could decrease the specificities of the involved fingers.

Positional dependence existed for many of the non-GNG fingers. For example, QSGDLTR could not specify the GCA triplet when presented at the F1 position (Fig. 3*A*). The Asp^{+3} could still specify the middle C, but it seems to have prevented the Gln^{-1} from making a specific interaction with the 3′ base adenine. To specify GCA at the F1 position, we replaced the Asp^{+3} with a smaller Ser^{+3} residue, as the Ser^{+3} had also been selected to bind base cytosine at F1 previously (8). We surmised that the smaller Ser^{+3} could also specify the middle base cytosine while at the same time not interfering with the $Gln^{-1} \rightarrow$ Ala interaction at the N-terminal tip of the ZFP. As expected, Ser^{+3} maintained the finger's specificity to the middle base cytosine, while in the meantime allowing the Gln^{-1} to specify the 3′ base adenine. In three different proteins, the site selection results consistently showed the design QSGSLTR specified GCA triplet at the F1 position. Another potential change we could make was to use Asn^{-1} to replace the Gln^{-1}, instead of making the Asp^{+3} to Ser^{+3} change, to try to improve the finger's specificity for GCA. Asn had been shown to make base-specific contact with the middle base adenine when at the +3 position (6). This is the first time it has been demonstrated that two different fingers are required to specify the same triplet while at different positions. The positional effects demonstrated for GCA, GAT, GGT, and possibly GAA and GCC fingers (Fig. 2) clearly indicate that many code-based designs are position-specific. Without testing a finger at each specific position, we do not know how to use these fingers properly. The finger's positional dependence shown here is in contrast to earlier published work (26), which suggested that zinc fingers behave qualitatively like modules and can be randomly stitched together. Other than the positional effects shown here, other factors such as the context of a finger with respect to the other two fingers of a three-finger protein and the context of the binding target DNA (27) can also influence the modularity of a finger. The building of a complete finger directory is a step toward the rapid construction of sequence-specific ZFPs.

A

EP2C	F3: RSDERKR	F2: RSDALSR	F1: QSSDLTR
Target	**G C G G T G G C T**		
K_d, 2 nM	G19 T10 G19 G19 T18 G19 G19 C19 T16		
	C9	C1	G3

B

Name	Target	K_d (nM)	Luc/β-gal
pGL3			0.4
2C0	GCG GTG GCT	2	66.1
2C1	G**A**G GTG GCT	4	31.8
2C2	G**T**G GTG GCT	2	62.4
2C3	GCG G**G**G GCT	31	3.3
2C4	GCG GTG G**A**T	7.8	14.4
2C5	GCG GTG GC**G**	4	48.6
3B	GGT GAG GAG		0.8

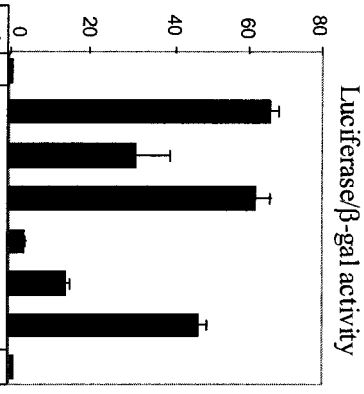

FIG. 4. **The properties of EP2C.** *A*, site selection results of EP2C. Fingers are presented in F3 to F1 order with each finger listed from −1 to + 6 residues. The target sequence is presented in the 5′ to 3′ direction. The selected consensus sequence is listed underneath the intended target sequence. *B*, specificities of EP2C are supported by both *in vitro* and *in vivo* assays. Gel shift assays were performed with a 2-fold dilution series of EP2C to its various targets. Single base pair mutants of the EP2C target are shown by *boldface, underlined letters*. Relative luciferase activities are presented from three individual samples with standard deviations marked. pGL3 is the empty vector with SV40 promoter-driven luciferase gene. 3B is a negative control for EP2C, with three copies of the 3B sequences cloned into pGL3 vector in the same fashion as the rest of EP2C target sequences.

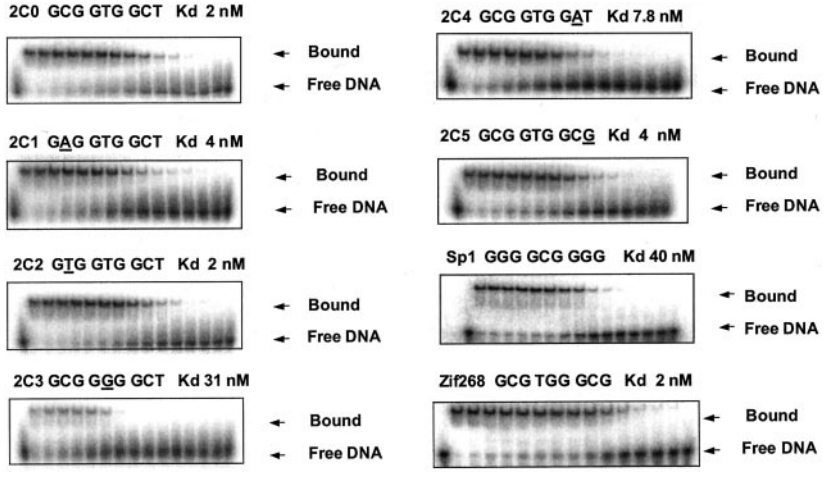

FIG. 5. **Gel shift assays of EP2C to its various target sequences.** The name of each target sequence is indicated at the *top left-hand corner* of each *gel panel*. Proteins were used to bind to their targets in 2-fold serial dilutions, with the highest protein concentration in the *second left-most lane* and the lowest concentration in the *right-most lane*. The *left-most lane* is a control lane containing radiolabeled DNA alone. The equilibrium dissociation constants for each target are indicated at the *top right-hand corner* of each *panel*. The binding affinities of Sp1 and Zif268 to their respective targets were also measured under the same binding conditions as the affinity references.

Although we came out with very specific fingers for most of our GNN triplets, there were still a few triplets we could not specify all three bases. Our designs for the GTT triplet could only specify GTN. Many different designs including the TSGSLTR, TSGALTR, TTSALTR, and QSSALTR were tried in order to improve the specificities for GTT triplet, but all gave similar or worse specificity profiles than those presented in Fig. 2. For these triplets, unconventional designs might be needed, and this might be better accomplished using selection methods.

The *in vivo* functionality of a ZFP is closely related to its *in vitro* DNA binding affinities. *In vivo*, there are a lot of factors that could influence a ZFP function, such as the ZFP concentration, location, and accessibility to its target DNA in chromatin. Here, we demonstrated that the specificity of these ZFPs, their binding affinity to various target sequences, is one of the main factors in deciding the function of a ZFP *in vivo*. The luciferase activities in Fig. 4*B* showed that inside cells the EP2C-VP16 activates the cognate 2C0, GCG target as well as the one-base change 2C2, GTG target to a similar level. This was in striking agreement with the gel shift data showing the EP2C bound these two targets with equal affinity. Overall, the luciferase activities closely mirrored the binding affinities of the various target sequences to the ZFP EP2C.

We have identified very specific zinc fingers for 14 of 16 GNN triplets through a design and site selection process. During this process, we found that many GNN fingers were position-de-

pendent, and different fingers were needed to specify the same target triplet at different finger positions. We also found that the existing design code was not applicable to all of the context situations. Many code-based designs could be used only at certain finger positions. Building a zinc finger directory brings us a step closer to solving these complex issues. Our future goal is to expand this finger directory to include specific fingers for the maximum number of all 64 triplets. The cell-based assays demonstrated that the ZFP *in vivo* functionality is highly dependent on its binding affinity to the target sequence. Highly specific ZFPs are naturally expected to function better inside of cells. The highly specific ZFPs will be useful tools in the study of gene functions and will find broad usage in human therapeutics and plant engineering.

Acknowledgments—We thank Hong Zhong for technical assistance, S. Kaye Spratt and Yolanda Santiago for providing the EP2C expression stable cell line, Becky Vanderslice for providing protocols for site selection experiments, and Yen Choo for comments on the manuscript.

REFERENCES

1. Berg, J. M., and Shi, Y. (1996) *Science* **271,** 1081–1085
2. Pabo, C. O., Peisach, E., and Grant, R. A. (2001) *Annu. Rev. Biochem.* **70,** 313–340
3. Christy, B. A., Lau, L. F., and Nathans, D. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 7857–7861
4. Berg, J. M. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 11109–11110
5. Shi, Y., and Berg, J. M. (1995) *Chem. Biol.* **2,** 83–89
6. Choo, Y., and Klug, A. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 11163–11167
7. Choo, Y., and Klug, A. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 11168–11172
8. Rebar, E. J., and Pabo, C. O. (1994) *Science* **263,** 671–673
9. Jamieson, A. C., Kim, S. H., and Wells, J. A. (1994) *Biochemistry* **33,** 5689–5695
10. Wu, H., Yang, W. P., and Barbas, C. F., III (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92(2),** 344–348
11. Wolfe, S. A., Greisman, H. A., Ramm, E. I., and Pabo, C. O. (1999) *J. Mol. Biol.* **285,** 1917–1934
12. Dreier, B., Segal, D. J., and Barbas, C. F., III (2000) *J. Mol. Biol.* **303(4),** 489–502
13. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) *Structure* **4,** 1171–1180
14. Pavletich, N. P., and Pabo, C. O. (1991) *Science* **252,** 809–817
15. Isalan, M., Choo, Y., and Klug, A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 5617–5621
16. Isalan, M., Klug, A., and Choo, Y. (1998) *Biochemistry* **37,** 12026–12033
17. Greisman, H. A., and Pabo, C. O. (1997) *Science* **275,** 657–661
18. Segal, D. J., Dreier, B., Beerli, R. R., and Barbas, C. F., 3rd. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96(6),** 2758–2763
19. Swirnoff, A. H., and Milbrandt, J. (1995) *Mol. Cell. Biol.* **15,** 2275–2287
20. Kadonaga, J. T., Carner, K. R., Masiarz, F. R., and Tjian, R. (1987) *Cell* **51,** 1079–1090
21. Zhang, L., Spratt, S. K., Liu, Q., Johnstone, B., Qi, H., Raschke, E. E., Jamieson, A. C., Rebar, E. J., Wolffe, A. P., and Case, C. C. (2000) *J. Biol. Chem.* **275,** 33850–33860
22. Pabo, C. O., and Nekludova, L. (2000) *J. Mol. Biol.* **301,** 597–624
23. Choo, Y., Sanchez-Garcia, I., and Klug, A. (1994) *Nature* **372,** 642–645
24. Isalan, M., Klug, A., and Choo, Y. (2001) *Nat. Biotechnol.* **19,** 656–660
25. Bulyk, M. L., Huang, X., Choo, Y., and Church, G. M. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 7158–7163
26. Desjarlais, J. R., and Berg, J. M. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 2256–2260
27. Fairall, L., Schwabe, J. W., Chapman, L., Finch, J. T., and Rhodes, D. (1993) *Nature* **366,** 483–487

PROTEIN STRUCTURE AND FOLDING:

Validated Zinc Finger Protein Designs for All 16 GNN DNA Triplet Targets

Qiang Liu, ZhenQin Xia and Casey C. Case

J. Biol. Chem. 2002, 277:3850-3856.
doi: 10.1074/jbc.M110669200 originally published online November 28, 2001
