Comprehensive DNA Recognition through Concerted Interactions from Adjacent Zinc Fingers

Mark Isalan, Aaron Klug, and Yen Choo*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

Received June 9, 1998; Revised Manuscript Received July 21, 1998

ABSTRACT: Zinc fingers are small DNA-binding modules noted for their occurrence in a large number of eukaryotic transcription factors, and their use in protein engineering. Although it was expected that zinc fingers can bind to a wide diversity of DNA sequences, previous studies using model zinc finger domains from Zif268 (and Sp1) have revealed a potential limitation to the DNA-binding specificity. For example, phage display selection of individual zinc fingers to recognize trinucleotide DNA subsites returned fingers that bound specifically only to triplets of the form GNN, i.e., triplets with guanine at the 5′ end. Following our recently reported work [Isalan, M., Choo, Y., and Klug, A. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94,* 5617−5621], we now show that this limitation can be overcome by the concerted randomization of certain amino acid positions in adjacent zinc fingers that specify overlapping DNA subsites. This illustrates an important mechanism underlying DNA recognition by arrays of zinc fingers, and points the way to improved strategies for the design of highly specific zinc finger proteins that bind any given nucleotide sequence.

The zinc fingers whose mode of DNA binding is best understood, and which are consequently used in protein engineering, derive from a subfamily that binds to guanine-rich sequences (*1*, *2*). An important member of this subfamily is the three-finger DNA-binding domain of the mouse transcription factor Zif268, which recognizes the 9 bp[1] sequence 5′-GCG-TGG-GCG-3′ (*3*, *4*). The first crystal structure of the DNA-binding domain in complex with the above binding site (*5*) indicated that each zinc finger bound a 3 bp subsite (Figure 1a). This modular mode of interaction suggested that zinc finger−DNA recognition could be studied by phage display (*6*). Phage display libraries of the Zif268 DNA-binding domain were constructed with residue randomizations in one of the three fingers, and then screened with a Zif268-like DNA-binding site in which the corresponding trinucleotide subsite was altered to a given sequence (*7*−*11*). In the most comprehensive selection experiment to date (*9*), we created a Zif268 library containing a randomized middle finger and screened this using 32 different DNA sequences containing variant middle triplet subsites. We found that only 17 of these triplets resulted in the selection of sequence-specific zinc fingers, while the remaining sequences always selected a set of 2 nonspecific zinc fingers. Other research groups also noted that certain triplets did not result in selection of sequence-specific zinc fingers when using phage display libraries of the first finger of Zif268 (*7*, *8*).

An analysis of the collective data from these experiments showed that zinc finger phage selections were successful only when the 5′ base of the triplet subsite was fixed as guanine or thymine, but consistently failed when the 5′ base was either adenine or cytosine. Of those zinc fingers which were selected to bind DNA triplets containing 5′-guanine or thymine, only a subset that recognized guanine showed true base-specificity. An exhaustive specificity check of the numerous fingers selected from our phage display library also showed that none were capable of binding adenine or cytosine at the 5′ end of a DNA triplet (*12*). Thus, the initial phage display experiments suggested that individual zinc fingers derived from Zif268 (a protein that binds to guanine-rich DNA sites) were inherently restricted to specifying a small repertoire of DNA triplet sequences of the form GNN. This raised interesting questions about the mechanism of DNA recognition employed by naturally occuring zinc fingers which are capable of binding DNA sites rich in adenine or cytosine. The restriction in sequence-specificity of the Zif268 fingers also appeared to be a potentially serious limitation to the design of zinc finger DNA-binding domains that recognize predetermined sequences.

We believe that the failure to select Zif268 fingers which recognize some of the DNA triplets tested so far was related to the context of the protein−DNA interaction, rather than being an inherent limitation of the zinc finger structure. We recently demonstrated that in contrast to the prevailing view that the Zif268 zinc finger modules interact with independent 3 bp subsites (Figure 1a), these fingers in fact specify overlapping 4 bp subsites (*13*) (Figure 1b). Further support for this mode of DNA recognition has emerged from the X-ray crystal structures of Zif268 and variant proteins determined by the Pabo laboratory (*14*, *15*). According to this refined model of DNA recognition, specificity for the base pair at the boundary between the two DNA subsites potentially arises from synergy of binding by amino acids from two adjacent zinc fingers. For instance, recognition of guanine in the 5′ position of a nominal triplet is often accomplished by concerted contacts from (i) Arg in position +6 of one zinc finger and (ii) a cross-strand interaction to

[1] Abbreviations: bp, base pair(s); F1, finger 1; F2, finger 2; F3, finger 3; PBS, phosphate-buffered saline; PCR, polymerase chain reaction.

S0006-2960(98)01358-0 CCC: $15.00
Published on Web 08/14/1998

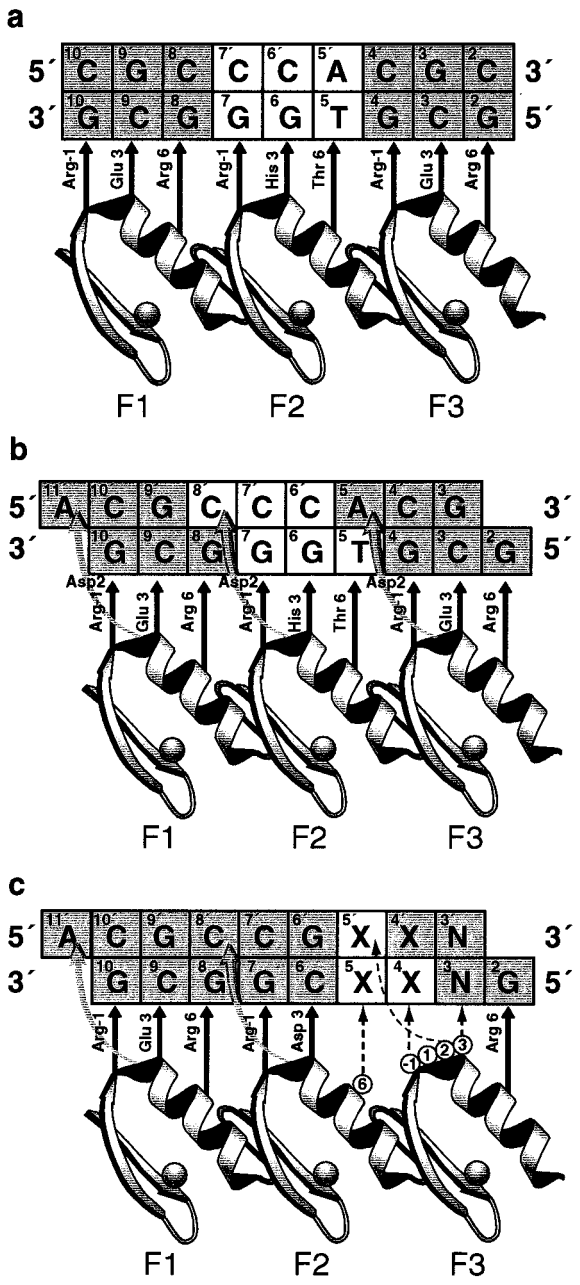

FIGURE 1: Schematic diagrams of zinc finger−DNA interactions discussed in the paper. (a) Modular interactions between individual zinc fingers of Zif268 (F1, F2, and F3) and the triplet subsites of an optimized DNA-binding site, as inferred from the X-ray crystal structure (*5*). Straight arrows indicate juxtapositioning of zinc finger residues (from helical positions −1, +3, and +6) and the bases on one DNA strand. (b) Model of synergistic Zif268−DNA interaction showing the potential for concerted contacts fom adjacent zinc fingers as discussed in detail in (*13*). Synergism between F2 and F3, for instance, could occur in specifying the ^{5}X-$^{5'}$X base pair that lies in the nominal triplet subsite of F2. In this case, a cross-strand contact from position +2 of F3 (shown by a curly arrow) is made to nucleotide $^{5'}$X on the complementary DNA strand. (c) Model of the large phage display library (LIB-A) designed to select zinc fingers that specify the base-step at the interface between the nominal triplet subsites of F2 and F3 (nucleotides ^{4}X and ^{5}X). Circled residue positions in F2 and F3 mark amino acids that were randomized in this library. DNA bases marked X were systematically varied in the 16 different binding sites used in selections. The DNA base pair marked ^{3}N $^{3'}$N was allowed to float in the binding sites used in selections, to allow for various pairings with randomized position +3 of F3.

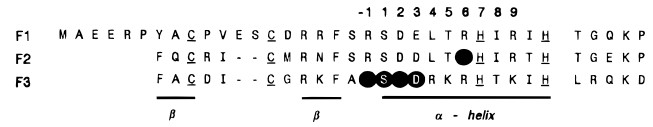

FIGURE 2: Amino acid sequence of the peptide comprising three zinc fingers (F1, F2, and F3) used in the phage display libraries for selections. The amino acid positions of each finger are aligned, residue positions −1 to +9 (relative to the first position of the α-helix, position +1) are numbered above the alignment, and elements of the zinc finger secondary structure are shown below. The conserved zinc-chelating residues of each zinc finger are underlined. The amino acid sequence is derived primarily from the DNA-binding domain of the transcription factor Zif268. Black circles show the five residue positions which were randomized in the larger library, LIB-A. The smaller library, LIB-B, had fixed amino acids in two of these positions (F3 positions +1 and +3 were fixed as Ser and Asp, respectively), and these residues are written in white letters over the black circles.

the complementary base from position +2 of the following finger (e.g., Asp in position +2 of F2 to cytosine at nucleotide position 8′, as shown in Figure 1b). This model accounts for the restricted sequence-specificity of zinc fingers that were selected from phage display libraries where only one finger was randomized at a time. In particular, the limitation in the original selections stemmed from the fact that the Asp residue present in position +2 of F3 was incompatible with cross-strand binding of certain nucleotides (*13*). The refined scheme of zinc finger−DNA recognition suggests that zinc fingers may achieve a more comprehensive repertoire of binding specificities through concerted interactions that straddle the DNA-binding interface between two synergistic zinc fingers.

In this paper, we constructed phage display libraries containing simultaneous randomizations of potentially synergistic base-contacting residues from two adjacent fingers, and then performed selections against a set of DNA-binding sites containing systematic variations in the base-step between nominal trinucleotide subsites. Zinc fingers selected from the new libraries bound sequence-specifically to a broader repertoire of DNA sequences than had been previously demonstrated. Together with previous data (*12*, *16*), the results illustrate that the zinc finger motif is capable of comprehensive nucleotide discrimination in any position of the nominal DNA triplet subsite. Thus, the strategy used to isolate these synergistic fingers by phage display may prove an improved method for the design of highly sequence-specific zinc fingers, allowing specific targeting to any DNA sequence.

MATERIALS AND METHODS

Construction of Phage Display Libraries. The first library to be constructed, LIB-A, contained randomizations at F2 residue position +6 and F3 residue positions −1, 1, 2, and 3 (Figures 1c and 2), and was sorted using the DNA sequence 5′-GNX-XCG-GCG-3′, where X-X denotes a known combination of the two bases at DNA positions ^{4}X and ^{5}X, and N denotes an equal probability of any of the four bases at DNA position 3 (Figure 1c). The second library, LIB-B, contained randomizations at F2 residue position +6 and F3 residue positions −1 and 2 (Figure 2), and was sorted using the DNA sequence 5′-GCX-XCG-GCG-3′, where DNA position 3 was fixed as cytosine as shown, and X-X denotes a known combination of the two bases at DNA positions ^{4}X and ^{5}X.

The genes for the two different zinc finger phage display libraries were assembled from four synthetic DNA oligonucleotides by directional end-to-end ligation using three short complementary DNA linkers. The oligonucleotides contained selectively randomized codons (of sequence NNS; N = A/C/G/T, S = G/C) in the appropriate amino acid positions of fingers 2 and 3. The constructs were amplified by PCR using primers containing *Not*I and *Sfi*I restriction sites, digested with the above endonucleases to produce cloning overhangs, and ligated into vector Fd-Tet-SN. Electrocompetent *E. coli* TG1 cells were transformed with the recombinant vector and plated onto TYE medium [1.5% (w/v) agar, 1% (w/v) Bactotryptone, 0.5% (w/v) Bactoyeast extract, 0.8% (w/v) NaCl] containing 15 μg/mL tetracycline.

Phage Selections. Tetracycline resistant colonies were transferred from plates into 2×TY medium (16 g/L Bactotryptone, 10 g/L Bactoyeast extract, 5 g/L NaCl) containing 50 μM $ZnCl_2$ and 15 μg/mL tetracycline, and cultured overnight at 30 °C in a shaking incubator. Cleared culture supernatant containing phage particles was obtained by centrifuging at 300*g* for 5 min.

Biotinylated DNA target sites (1 pmol) were bound to streptavidin-coated tubes (Boehringer Mannheim). Phage supernatant solutions were diluted 1:10 in PBS selection buffer [PBS containing 50 μM $ZnCl_2$, 2% (w/v) fat-free dried milk (Marvel), 1% (v/v) Tween, 20 μg/mL sonicated salmon sperm DNA, 10 pmol/mL each of the 15 remaining DNA sites (unbiotinylated)], and 1 mL was applied to each tube for 1 h at 20 °C. After this time, the tubes were emptied and washed 20 times with PBS containing 50 μM $ZnCl_2$, 2% (w/v) fat-free dried milk (Marvel), and 1% (v/v) Tween. Retained phage were eluted in 0.1 mL of 0.1 M triethylamine and neutralized with an equal volume of 1 M Tris (pH 7.4). Logarithmic-phase *E. coli* TG1 (0.5 mL) were infected with eluted phage (50 μL), and used to prepare phage supernatants for subsequent rounds of selection. After 3 rounds of selection, *E. coli* infected with selected phage were plated, and individual colonies were picked and used to grow phage for binding site signature assays and DNA sequencing.

Binding Site Signatures. Procedures were as described previously (*12*). Briefly, oligonucleotides were chemically synthesized, and double-stranded DNA libraries were made by (biotinylated or unbiotinylated) primer extension. DNA libraries (1.2 pmol/well for LIB-A and 0.4 pmol/well for LIB-B) were added to streptavidin-coated ELISA wells (Boehringer-Mannheim) in PBS containing 50 μM $ZnCl_2$ (PBS/Zn). Phage solution [overnight bacterial culture supernatant diluted 1:10 in PBS/Zn containing 2% (w/v) fat-free dried milk (Marvel), 1% (v/v) Tween, and 20 μg/mL sonicated salmon sperm DNA] was applied to each well (50 μL/well). Binding was allowed to proceed for 1 h at 20 °C. Unbound phage were removed by washing 6 times with PBS/Zn containing 1% (v/v) Tween, and then 3 times with PBS/Zn. Bound phage were detected by ELISA using horseradish peroxidase-conjugated anti-M13 IgG (Pharmacia Biotech), and the colorimetric signal was quantitated using SOFTMAX 2.32 (Molecular Devices).

DNA Sequence Analysis. The coding sequence of individual zinc finger clones was amplified by PCR using external primers complementary to phage sequence. These PCR products were then sequenced manually using Thermo Sequenase cycle sequencing (Amersham Life Science).

RESULTS

Design of the Zinc Finger Phage Display Library. Zinc finger−DNA recognition at the interface between adjacent DNA subsites was studied using a zinc finger phage display library. This library was based on the three-finger DNA-binding domain of Zif268, but contained randomizations of amino acids from both F2 and F3, at residue positions which could form a network of contacts across the interface of their DNA subsites. The detailed design of the library is shown in Figure 1c together with the generic DNA-binding site used in selections. Briefly, the library contained amino acid randomizations at F2 residue position +6 (hereafter denoted F2[+6]) and F3 residue positions −1, +1, +2, and +3 (hereafter denoted F3[−1], F3[+2], etc.).

Library selections were carried out using DNA oligonucleotides that resembled a Zif268-binding site (Figure 1a,b), but which contained systematic combinations of bases in the DNA doublet which forms the base-step between the DNA subsites of F2 and F3 (nucleotides ^{4}X and ^{5}X in Figure 1c). More specifically, DNA-binding sites were of the generic form 5′-GNX-XCG-GCG-3′, where X-X denotes a given combination of the bases at the interface between the DNA subsites, and N denotes that the four bases are equally represented at DNA position 3. Thus, the interaction between F3[+3] and nucleotide position ^{3}N is allowed complete freedom in this experiment. This feature of the library allowed selection of a large family (or database) of related zinc fingers that bound a given combination of bases at nucleotide positions ^{4}X and ^{5}X, but which were nonidentical owing to different interactions with the middle base in the nominal triplet subsite of F3.

Allowing this contextual freedom to some protein−DNA interactions that are not being studied is a useful strategy to increase the diversity of clones which can be obtained from any one selection experiment. However, at the same time, it is important to limit the nature and number of contacts that are allowed contextual freedom at any one time; otherwise there is a danger that a subset of particularly strong intermolecular interactions will dominate the selections, as was apparent in (*17*). Anticipating this possibility, we also created a smaller sublibrary that contained randomized residues only in positions F2[+6] and F3[−1 and +2], with F3[+1 and +3] fixed. This smaller library was screened with a DNA-binding site in which nucleotide position 3 was also fixed, and therefore did not allow for any contextual freedom in selections. Clones selected from this library are marked with an asterisk when they are discussed in this paper.

Experimental Strategy. Phage selections from the two zinc finger libraries were performed separately in order to determine the diversity of DNA sequences which could be bound specifically by members of each library. Sixteen selections were performed on each library, using the different DNA-binding sites that correspond to all 16 possible combinations of bases at nucleotide positions ^{4}X and ^{5}X (Figure 1c). The DNA-binding site used to select phage was immobilized on a solid surface, while a 10-fold excess of each of the remaining 15 DNA sites was present in solution as a specific competitor.

After three rounds of phage selection against a particular DNA-binding site, individual zinc finger clones were recovered, and the DNA-binding specificity of a few clones was determined by the binding site signature method (*12*). This involved screening each zinc finger phage for binding to eight different libraries of the DNA-binding site, designed such that each library contained one fixed base and one randomized base at either position ^{4}X or position ^{5}X (i.e., libraries GN, AN, TN, CN, and NG, NA, NT, NC). In this system, each of the 16 DNA-binding sites used in selection experiments was present in a unique combination of two libraries—for example, the DNA-binding site containing ^{4}G^{5}G was present in only 2 of the 8 libraries in which the relevant doublet had 1 nucleotide randomized and the other nucleotide fixed as guanine, i.e., libraries ^{4}G^{5}N and ^{4}N^{5}G. The eight DNA libraries used in binding site signatures were arrayed across a microtiter plate, and zinc finger phage binding to these was detected by phage ELISA (*12*). Thus, the pattern of phage binding to the eight DNA libraries revealed the DNA-sequence specificity (or preference) of each phage clone, and those clones found to give a high degree of specificity were subsequently sequenced to reveal the identity of the amino acids selected in the randomized residue positions of the zinc fingers.

Analysis of Phage-Selected Zinc Fingers. Figure 3 shows the binding site signatures of numerous zinc finger phages selected from both libraries, using binding sites that contain the 16 different DNA doublets which form the base-step between the DNA subsites of fingers 2 and 3. The results show that zinc finger clones were selected which bound specifically to almost all subsites, including those triplets whose 5′ position (nucleotide ^{5}X in Figure 1c) was fixed as a base other than guanine. Overall, the selections showed that any of the four bases can be bound specifically in both the 5′ and 3′ positions of a nominal triplet subsite.

Selections from the smaller sublibrary yielded fingers that could bind specifically to only 8 of the 16 doublets, whereas members of the larger library yielded fingers that recognized 15 out of the 16 doublets. It is not known whether this difference in efficacy originated from the inclusion of more randomized positions in the larger library, or the conformational flexibility afforded by the contextual freedom designed into the larger library, or both. The only base-step that did not yield specific zinc fingers was ^{4}G^{5}A. Again, it is not known whether this dinucleotide induced an unfavorable DNA deformation in the context of the DNA-binding sites used for selection, or whether there is a more general problem with binding of this base-step at the interface between zinc finger binding sites. Specific zinc finger binding to the GA dinucleotide has been confirmed a number of times in other sequence contexts when it occurs within a zinc finger subsite (*7*, *12*), but not when present at the subsite interface.

DISCUSSION

Rules That Describe Zinc Finger-DNA Recognition. The unique geometry of interaction of Zif268 with the DNA double helix allows a series of one-to-one interactions between amino acids on consecutive turns of the zinc finger α-helix and nucleotides from consecutive base pairs on DNA (*5*). Consequently, many aspects of the DNA-binding specificity of Zif268-like zinc fingers can be described by a set of rules which relates the amino acids in these base-recognition positions to the nucleotides that they contact in the major groove of DNA (*18*). This recognition code is based strictly on the mode of intermolecular interaction observed in zinc finger−DNA complexes such Zif268 (*5*) and Tramtrack (*19*), and therefore does not account for DNA recognition by all zinc finger proteins, nor by other proteins that contain heterologous DNA-binding motifs. Despite this limitation, thinking about zinc finger−DNA recognition in terms of a code is important because it provides a paradigm of specific protein−DNA recognition that can be rationalized by reference to the stereochemistry of a protein−DNA complex. Further elucidation of the recognition code will also allow the formulation of more effective experimental strategies to select zinc finger DNA-binding domains, and may yet permit the rational design of such molecules to bind any given DNA sequence.

In this paper, we used libraries of zinc fingers to study a network of protein−DNA contacts that mediates specific DNA recognition of the base-step between two DNA subsites of adjacent zinc fingers. Phage display experiments were designed such that residues selected at F3[−1] recognized nucleotide ^{4}X, whereas residues selected at F2[+6] and F3-[+2] could cooperate in order to recognize the adjacent base pair, ^{5}X-$^{5'}$X (Figure 1c). In the model system used in our experiments, nucleotide ^{4}X represents the 3′ position of a nominal triplet subsite (in this case the subsite of F3), and nucleotide ^{5}X represents the 5′ position of the nominal triplet subsite bound by the preceding finger (in this case F2). While recognition rules have been elucidated to account for interactions between the 3′ base position and zinc finger position −1, the only rule discovered so far to describe a specific interaction at the 5′ base position is a pairing between guanine and Arg at zinc finger position +6 (*12*, *16*, *18*) (e.g., see Figure 1a).

The results of the phage display selections generally bear out the conclusions of previous studies regarding the code of DNA recognition from position −1 of the zinc finger α-helix (*12*). In particular, strong correlations were derived between the nucleotide specified at position ^{4}X and the amino acids selected at F3[−1] (Figure 4A), indicating that certain recurring contacts were largely responsible for intermolecular recognition.

A similar analysis, however, did not show any one-to-one correlation between the identity of the ^{5}X-$^{5'}$X base pair and the amino acids selected in F2[+6] or F3[+2], other than for the expected pairing of Arg and guanine (Figure 4b). In other cases, rather than reveal such simple recognition rules, the biases in interactions between ^{5}X:F2[+6] and $^{5'}$X:F3[+2] follow a particular pattern. In both these residue positions, certain amino acids (e.g., Glu) were selected to accept a hydrogen bond from either adenine or cytosine, while others (e.g., Lys) were selected to donate a hydrogen bond to either guanine or thymine (Figure 4b). It has been pointed out that these degeneracies arise because of the stereochemistry of the base pairs, but can be broken by two hydrogen bonds to the same base pair, hence, the importance (and prevalence in many naturally occurring zinc fingers) of Arg in position +6 for specific recognition of guanine (*20*). Curiously, no pairings between Gln and adenine were observed, even though these two can form analogous bidentate contacts (*20*). The absence of a simple coding

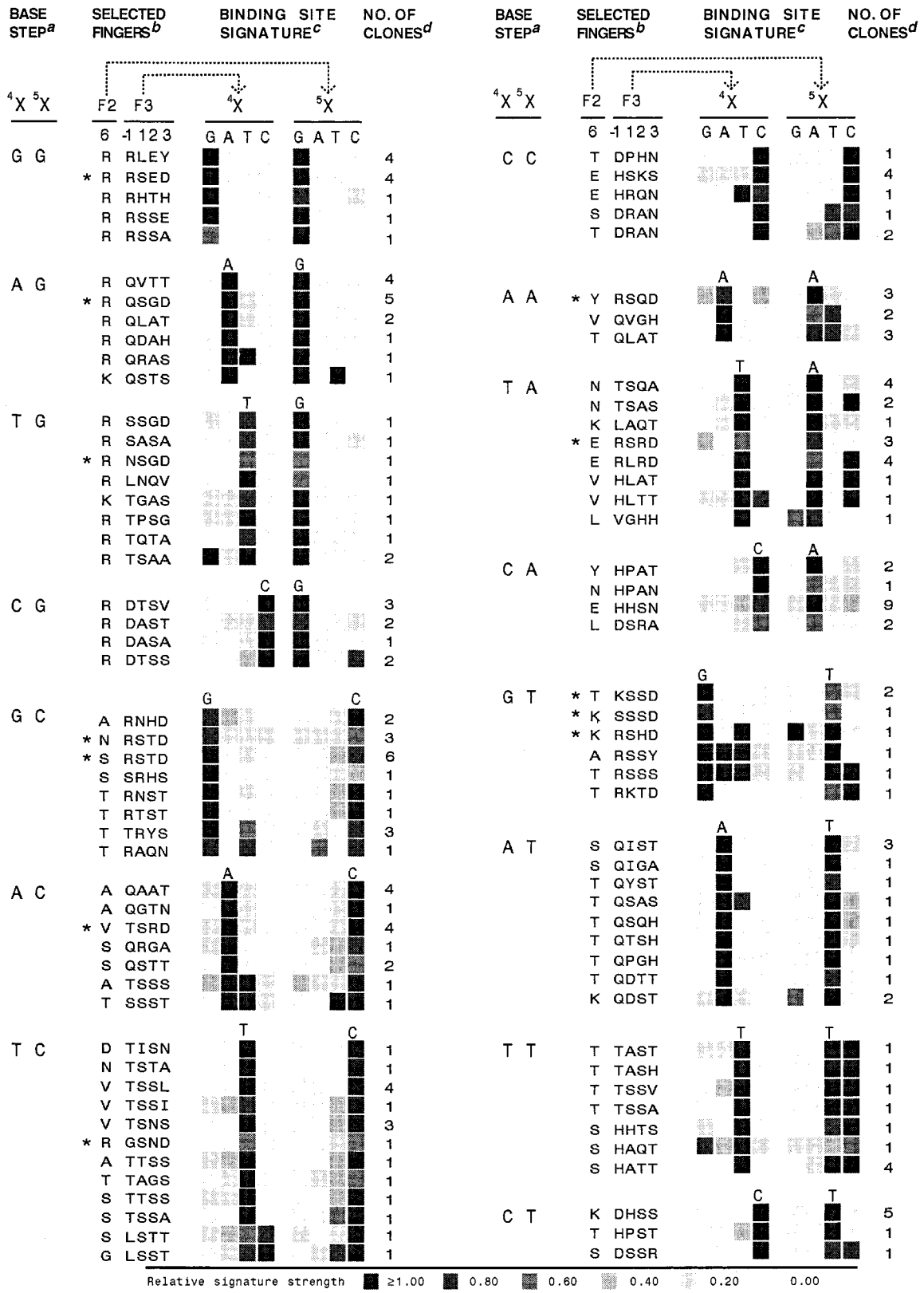

FIGURE 3: List of different sequence-specific zinc finger clones obtained from phage selections, together with their binding site signatures. Different zinc finger clones, selected using different base steps in position $^4X^5X$ ([a]), are denoted by the amino acids selected at the randomized residue positions of F2 and F3 ([b]), and their frequency among the total number of sequenced clones is indicated ([d]). The binding site signature of each clone (i.e., the base preference at positions ^4X and ^5X) is shown in a gray scale that is proportional to DNA binding as measured by phage ELISA ([c]). Arrows at the top of the diagram remind the reader that ^4X is in the nominal triplet subsite of F3, while ^5X is in the nominal triplet subsite of F2. Zinc finger clones selected from the smaller sublibrary (LIB-B), in which residue positions F3[+1] and F3[+3] were fixed as Ser and Asp, respectively, are labeled with an asterisk (*).

relationship that explains base-specificity at the 5′ position of the nominal DNA triplet subsite may arise as a result of the failure to randomize the relevant network of zinc finger residues (e.g., α-helical position +5, which may play a role

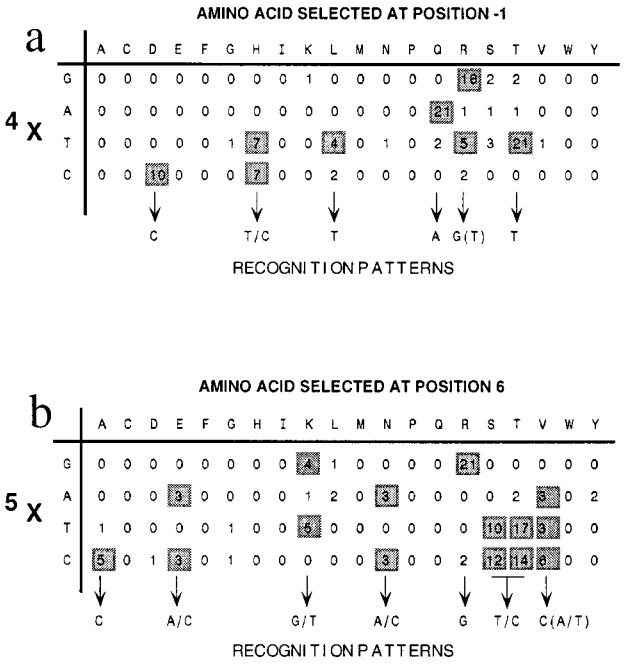

FIGURE 4: (a) Matrix comparison of amino acids present at F3[−1] and the base recognized in nucleotide position ^{4}X. The number plotted for each pair in the matrix is derived from the number of times that pair occurs in the binding site signature of Figure 3, and does not reflect the frequency of selection of particular clones. The four most common pairings are Gln:^{4}A, Arg:^{4}G, Asp:^{4}C, and Thr:^{4}T. The first three of these pairings are common interactions that form part of the zinc finger−DNA recognition rules (*18*). Among the new pairings to be observed was a strong correlation between ^{4}T and Thr, and a weaker one with Leu. Both of these were much more significant than was observed in a previous study where corresponding pairings between ^{4}T and Asn or Gln were selected more frequently (*9*, *12*). An additional interaction that was prevalent, and which had not been observed at all in the past, was between His in residue position −1 and either cytosine or thymine on DNA. The amino acids found to occur with the highest frequency at position −1 in a database of naturally occurring zinc fingers are Gln(25%), Arg (16%), His (7.9%), Asp (6.2%), and Ser (5.9%) (*24*). (b) Matrix comparison, again based on Figure 3, of amino acids present at F2[+6] and the base recognized in nucleotide position ^{5}X. Strong pairings were observed between Arg:^{5}G, as would be expected, and also Ala:^{5}C. Other significant correlations involved amino acids that bound bases with similar hydrogen bonding groups exposed in the DNA major groove−thus, Lys was grouped with ^{5}G or ^{5}T, while Glu and Asn were grouped with ^{5}A or ^{5}C. Additionally, Ser and Thr were frequently selected at F2-[+6], to bind opposite pyrimidines. Residues that occur with the highest frequency at position +6 in a database of naturally occurring zinc fingers are Arg (16%), Lys (15%), Val (10%), Thr (8.7%), Gln(8.6%), Ile (6.9%), Glu (6.0%), Ser (5.6%), and Leu (5.2%) (*24*). Ala, which occurs commonly in our selections, was present in the database at a relatively low frequency (1−5%). Conversely, Gln, which is highly represented in the database (8.6%), was selected only once in our experiments.

in DNA recognition, was not randomized). The lack of a simple code, however, could also be a reflection of the multiple ways in which one particular hydrogen bonding pattern can be reproduced by different combinations of amino acids from separate fingers. Moreover, inter-finger synergy would allow stereochemical complementarity at the protein−DNA interface to be modulated by the variation in docking arrangements between two adjacent zinc fingers. Although no substantial direct inter-finger contacts were seen in the X-ray cocrystal structure of wild-type Zif268 (*5*), there is no reason derivatives of these fingers could not interact more extensively, and adjacent zinc fingers have recently been found to interact in the NMR structure of a TFIIIA−DNA complex (*21*).

Possible Zinc Finger−DNA Interactions Mediated by Zinc Finger α*-Helical Position +2.* The amino acid present in α-helical position +2 of a zinc finger can help determine the specificity for the base pair at the interface of two overlapping DNA subsites (Figure 1b). An Asp residue present in F3[+2] of wild-type Zif268 was shown to play a role in DNA recognition though a cross-strand contact (*13*), and further examples were generated by the current phage display experiments (e.g., Figure 5a). In these, a small subset of amino acids selected in F3[+2] showed significant correlations to the identity of the base pair in position $^{5'}$X (Figure 5b), suggesting that cross-strand interactions may be a general mechanism of DNA recognition. Most of these correlations could be rationalized as pairings between hydrogen bond donors in F3[+2] and guanine or thymine in DNA position $^{5'}$X, in accordance with the framework of the Zif268 model (Figure 1b). In contrast to amino acids that were never selected in position +2, or those that were selected but which showed no significant correlations, the amino acids which consistently appear to play a role in DNA recognition from this position have side chains with multiple hydrogen bonding groups. It is possible that these residues can play a role in base recognition because they achieve greater specificity by participating in buttressing networks [e.g., intramolecular networks analogous to the pairing of Arg-1 and Asp+2 of each finger in wild-type Zif268 (*5*)]. Interestingly, the X-ray crystal structures of Zif268 variants show that position +2 of a zinc finger can also make cross-strand contacts to the DNA backbone or to a base position which is further along the complementary strand in the 5′ direction (*15*).

The Role of Zinc Finger Synergy. To what extent is zinc finger synergy required for specific recognition of the 5′ position of a nominal triplet subsite? First, selection of individual zinc finger modules by the classical phage display strategies (*7*, *8*−*10*, *22*) has failed to produce zinc finger domains that bind specifically to diverse DNA sequences. Second, previous phage display experiments using a variant of our original Zif268 library (*9*), in which the library of randomized middle fingers was modified by mutating F3-[+2] to Ala in order to remove the constraint imposed by a cross-strand interaction, did not result in selection of fingers that could discriminate adenine or cytosine in the 5′ position of the middle triplet subsite (unpublished work). Although this may indicate a requirement for synergistic fingers, a negative result of course does not necessarily indicate that individual zinc fingers are incapable of recognizing any base triplet. The failure to bind subsites containing ^{5}A, for instance, could have derived from the effect of a ^{4}G^{5}A base-step which would arise as a result of the experimental design (the binding sites used in selections were of the form 5′-GCG-XXX-GCG-3′, where the underlined base is nucleotide position 5).

A novel zinc finger phage selection strategy based on serial selection of C-terminal zinc finger modules was recently reported to produce fingers which bound to diverse DNA sequences containing various bases at the 5′ position of triplet subsites (*22*). However, the sequence-specificity of these selected fingers was not assessed in detail, and therefore the

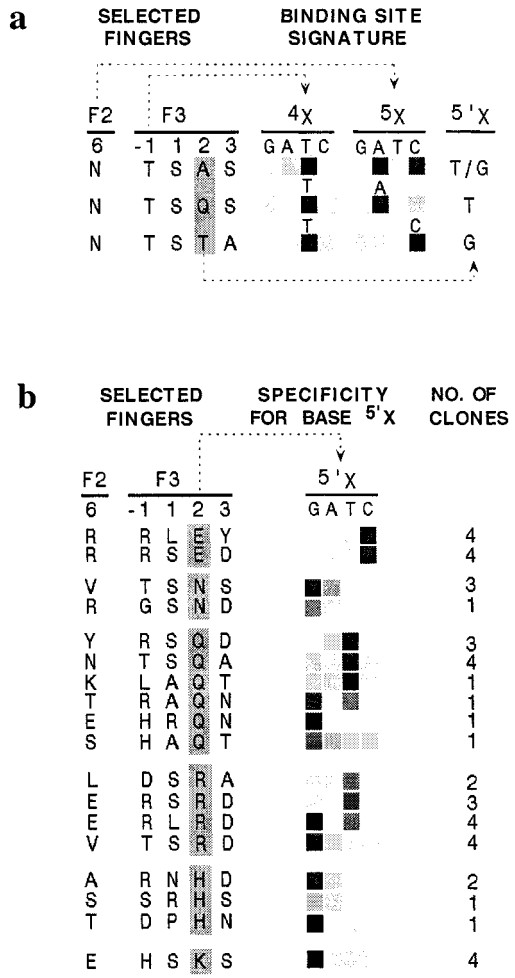

FIGURE 5: (a) Example of related zinc finger clones showing the effect of α-helical position +2 on DNA-binding specificity. Data are shown in a similar format to Figure 3. This family of zinc fingers was derived from selections using DNA-binding sites containing $^{4}T^{5}A$ or $^{4}T^{5}C$ subsite interfaces. The base preference for the ^{5}X-$^{5'}X$ base pair appears to be determined by the amino acid present at F3[+2], probably by the formation of cross-strand contacts, as indicated by the arrow at the bottom of the figure. (b) Examples of correlations between certain amino acids selected at F3[+2] and the identity of the base present at position $^{5'}X$. Possible cross-strand interactions are marked by the arrow. Selections revealed the possibility of DNA contacts from five amino acids (Asn, Gln, Arg, Lys, and His) which were all capable of donating a H-bond to the exocyclic oxygen atom of either guanine (O_6) or thymine (O_4) in nucleotide position $^{5'}X$. The majority of clones isolated with these amino acids at F3[+2] are listed in this diagram together with the binding site signature showing the base preference at position $^{5'}X$. Note that a few clones were isolated which do not follow this pattern, and that these are omitted from the diagram. Overall, Ser dominated the selections in position +2 with an occurrence of 38%, in accord with its presence in that position in over half of all known zinc fingers. Threonine, Ala, and Gly occurred frequently in our selections (15%, 15%, and 9%, respectively) but did not show any discernible patterns of discrimination. Certain amino acids (Cys, Asp, Phe, Ile, Leu, Met, Pro, Val, and Trp) were never selected in position +2 in our experiments, and it emerges that these are the rarest residues found at that position in naturally occurring fingers (*24*). As always, these phage selection results should be interpreted with caution, however, as they do not necessarily imply a lack of function for these residues in all fingers. For instance, if position −1 is fixed as Arg (to recognize guanine), position +2 is selected as Asp when phage selections are performed within F2, while Glu is preferred when selections are performed in the context of F3 (unpublished experiments).

extent of nucleotide discrimination, for example, at the 5′ position of each triplet subsite, is not known.

In this paper, we have described phage display libraries containing concerted variations in amino acids from adjacent fingers, created in order to exploit the potentially enhanced specificity arising from the overlap between abutting DNA subsites. In contrast to previous studies, we have now demonstrated that it is possible to select synergistic finger pairs that specifically recognize any base at the 5′ position of a nominal triplet subsite. In addition to defining a mechanism that is likely widely used by naturally occurring zinc fingers, the results prove that simultaneous randomization of the DNA-binding residues of adjacent zinc fingers is a means of selecting fingers that bind specifically to a more diverse set of DNA sequences.

Design of Zinc Finger Domains To Bind Any Given DNA Sequence. Selections from the synergistic zinc finger libraries described in this paper have demonstrated that zinc fingers are capable of base-specific discrimination in the 5′ position (and also the 3′ position) of their nominal triplet subsite, while other phage display and site-directed mutagenesis experiments have previously shown base discrimination in the middle and 3′ positions of the DNA triplet. Hence, an important principle that finally emerges from these studies is that the zinc finger module is capable of comprehensive DNA sequence recognition, justifying the use of this framework in the design of DNA-binding proteins. Our current results additionally show that two adjacent zinc fingers can specify virtually any combination of bases at the interface between their respective DNA subsites. Extrapolating these findings to zinc finger design using the Zif268 framework implies that arrays of zinc fingers can be selected to specify practically any given sequence of nucleotides. It was recently shown that arrays of up to six finger modules can be made to bind very tightly to 18 bp long sequences (*23*). In effect, it should thus be possible to engineer zinc finger DNA-binding proteins that bind to any unique region of a large genome, for applications in medicine, biotechnology, and research.

REFERENCES

1. Klug, A., and Rhodes, D. (1987) *Trends Biochem. Sci. 12*, 464−469.
2. Berg, J. M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 11109−11110.
3. Christy, B. A., Lau, L. F., and Nathans, D. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7857−7861.
4. Christy, B., and Nathans, D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 8737−8741.
5. Pavletich, N. P., and Pabo, C. O. (1991) *Science 252*, 809−817.
6. Smith, G. P. (1985) *Science 228*, 1315−1317.
7. Rebar, E. J., and Pabo, C. O. (1994) *Science 263*, 671−673.
8. Jamieson, A. C., Kim, S.-H., and Wells, J. A. (1994) *Biochemistry 33*, 5689−5695.
9. Choo, Y., and Klug, A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11163−11167.
10. Wu, H., Yang, W.-P., and Barbas, C. F., III (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 344−348.
11. Choo, Y., and Klug, A. (1995) *Curr. Opin. Biotechnol. 6*, 431−436.
12. Choo, Y., and Klug, A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11168−11172.
13. Isalan, M. D., Choo, Y., and Klug, A. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 5617−5621.

14. Elrod-Erickson, M., Rould, M. A., Nekludova, L., and Pabo, C. O. (1996) *Structure 4*, 1171−1180.
15. Elrod-Erickson, M., Benson, T. E., and Pabo, C. O. (1998) *Structure 6*, 451−464.
16. Desjarlais, J. R., and Berg, J. M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7345−7349.
17. Jamieson, A. C., Wang, H., and Kim, S.-H. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 12834−12839.
18. Choo, Y., and Klug, A. (1997) *Curr. Opin. Struct. Biol. 7*, 117−125.
19. Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T., and Rhodes, D. (1993) *Nature 366*, 483−487.
20. Seeman, N. D., Rosenberg, J. M., and Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804−808.
21. Foster, M. P., Wuttke, D. S., Radhakrishnan, I., Case, D. A., Gottesfeld, J. M., and Wright, P. E. (1997) *Nat. Struct. Biol. 4*, 605−608.
22. Greisman, H. A., and Pabo, C. O. (1997) *Science 275*, 657−661.
23. Kim, J.-S., and Pabo, C. O. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 2812−2817.
24. Jacobs, G. (1993) Ph.D. Thesis, Cambridge University, Cambridge, U.K.

BI981358Z