Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 11168–11172, November 1994
Biochemistry

Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions

(zinc finger–DNA interaction/recognition code/binding-site signature/protein design/DNA–protein interaction)

YEN CHOO AND AARON KLUG

Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, United Kingdom

Contributed by Aaron Klug, June 6, 1994

ABSTRACT **In the preceding paper [Choo, Y. & Klug, A. (1994) Proc. Natl. Acad. Sci. USA 91, 11163–11167], we showed how selections from a library of zinc fingers displayed on phage yielded fingers able to bind to a number of DNA triplets. Here, we describe a technique to deal efficiently with the converse problem—namely, the selection of a DNA binding site for a given zinc finger. This is done by screening against libraries of DNA triplet binding sites randomized in two positions but having one base fixed in the third position. The technique is applied here to determine the specificity of fingers previously selected by phage display. We find that some of these fingers are able to specify a unique base in each position of the cognate triplet. This is further illustrated by examples of fingers which can discriminate between closely related triplets as measured by their respective equilibrium dissociation constants. Comparing the amino acid sequences of fingers which specify a particular base in a triplet, we infer that in most instances, sequence-specific binding of zinc fingers to DNA can be achieved by using a small set of amino acid–nucleotide base contacts amenable to a code.**

In principle, rules governing protein–DNA interactions can be deduced from a large database of correlations between the amino acid sequences of the proteins and the nucleotide sequences of their optimal binding sites. To this end, we have shown in the preceding paper (1) that functionally equivalent zinc fingers which bind to a given DNA sequence can be selected from a phage display library. However, determination of the optimal binding site for these fingers is still required, as a safeguard against spurious selections. One can determine the optimal binding sites of these (and other) proteins, by selection from libraries of randomized DNA. This approach, the principle of which is essentially the converse of zinc finger phage display, would provide an equally informative database from which the same rules can be independently deduced. However, until now the favored method for binding-site determination, involving iterative selection and amplification of target DNA followed by sequencing, has been a laborious process not conveniently applicable to the analysis of a large database (2, 3).

We present here a convenient and rapid method which can reveal the optimal binding site(s) of a DNA-binding protein by single-step selection from small libraries, and use this to check the binding-site preferences of those zinc fingers selected previously by phage display (1). For this application, we use 12 different minilibraries of the binding site for transcription factor Zif268, each one with the central triplet having one position defined with a particular base pair and the other two positions randomized. Each library therefore comprises 16 oligonucleotides and offers a number of potential binding sites to the middle finger, provided that the latter can tolerate the defined base pair. Each zinc finger phage is screened against all 12 libraries individually immobilized in wells of a microtiter plate, and binding is detected by an enzyme immunoassay. Thus, a pattern of acceptable bases at each position is disclosed, which we call a binding-site signature. The information contained in a binding-site signature encompasses the repertoire of binding sites recognized by a zinc finger.

The binding-site signatures obtained by using zinc finger phage selected as described in the preceding paper (1) reveal that the selection has yielded some highly sequence-specific zinc fingers which discriminate at all three positions of a triplet. From measurements of equilibrium dissociation constants, we find that these fingers bind tightly to the triplets indicated in their signatures and discriminate against closely related sites usually by at least a factor of 10. The binding-site signatures allow us to infer rules for a specificity code for the interactions of zinc fingers with DNA.

MATERIALS AND METHODS

Binding-Site Signatures. Flexible flat-bottomed 96-well plates (Falcon) were coated overnight at 4°C with streptavidin (0.1 mg/ml in 0.1 M $NaHCO_3$, pH 8.6/0.03% NaN_3). Wells were blocked by incubation for 1 hr with PBS/Zn (phosphate-buffered saline plus 50 μM zinc acetate) containing 2% (wt/vol) fat-free dried milk (Marvel) and were washed three times with PBS/Zn containing 0.1% Tween and three times with PBS/Zn. The "bound" strand of each oligonucleotide library was made synthetically and the other strand was extended from a 5′-biotinylated universal primer by DNA polymerase I (Klenow fragment). Products of fill-in reactions were added to wells (0.8 pmol of DNA library in each) in PBS/Zn for 15 min and then washed once with PBS/Zn containing 0.1% Tween and once with PBS/Zn. Overnight bacterial cultures each containing a selected zinc finger phage (1) were grown at 30°C in 2×TY medium containing 50 μM zinc acetate and 15 μg of tetracycline per ml (2×TY/Zn/Tet). Culture supernatants containing phage were diluted 10-fold by addition of PBS/Zn containing 2% (wt/vol) fat-free dried milk, 1% (vol/vol) Tween 20 and 20 μg of sonicated salmon sperm DNA per ml. Diluted phage solutions (50 μl) were applied to wells and binding was allowed to proceed for 1 hr at 20°C. Unbound phage were removed by washing five times with PBS/Zn containing 1% Tween and then three times with PBS/Zn. Bound phage were detected as described (4) or by using horseradish peroxidase-conjugated anti-M13 IgG (Pharmacia) and quantitated with SOFTMAX 2.32 (Molecular Devices).

Determination of Apparent Equilibrium Dissociation Constants (K_d Values). Overnight bacterial cultures were grown in 2×TY/Zn/Tet at 30°C. Culture supernatants containing phage were diluted 2-fold by the addition of PBS/Zn containing 4% fat-free dried milk, 2% Tween 20, and 40 μg of sonicated salmon sperm DNA per ml. Binding reaction mixtures containing appropriate concentrations of specific

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

5′-biotinylated DNA and equal volumes of zinc-finger phage solution were allowed to equilibrate for 1 hr at 20°C. All DNA was captured on streptavidin-coated paramagnetic beads (500 μg per well), which were subsequently washed six times with PBS/Zn containing 1% Tween and then three times with PBS/Zn. Bound phage were detected with horseradish peroxidase-conjugated anti-M13 IgG (Pharmacia) and developed as described (4). Optical densities were quantitated with SOFTMAX 2.32 (Molecular Devices).

K_d values were estimated by fitting to the equation K_d = [DNA][protein]/[DNA–protein] with the program KALEIDAGRAPH version 2.0 (Synergy Software, Reading, PA). Owing to the sensitivity of the ELISA used to detect protein–DNA complexes, we can use zinc finger phage concentrations far below those of the DNA, as is required for accurate calculations of K_d. The technique we use has the advantage that while the concentration of DNA (variable) must be known accurately, that of the zinc fingers (constant) need not be known (5). This circumvents the problem of calculating the number of zinc finger peptides expressed on the tip of each phage, although since only 10–20% of the gene III protein (pIII) carries such peptides, we would expect on average less than one copy per phage. Binding is performed in solution to prevent any effects caused by the avidity (6) of phage for DNA immobilized on a surface. Moreover, in this case measurements of K_d by ELISA are made possible because

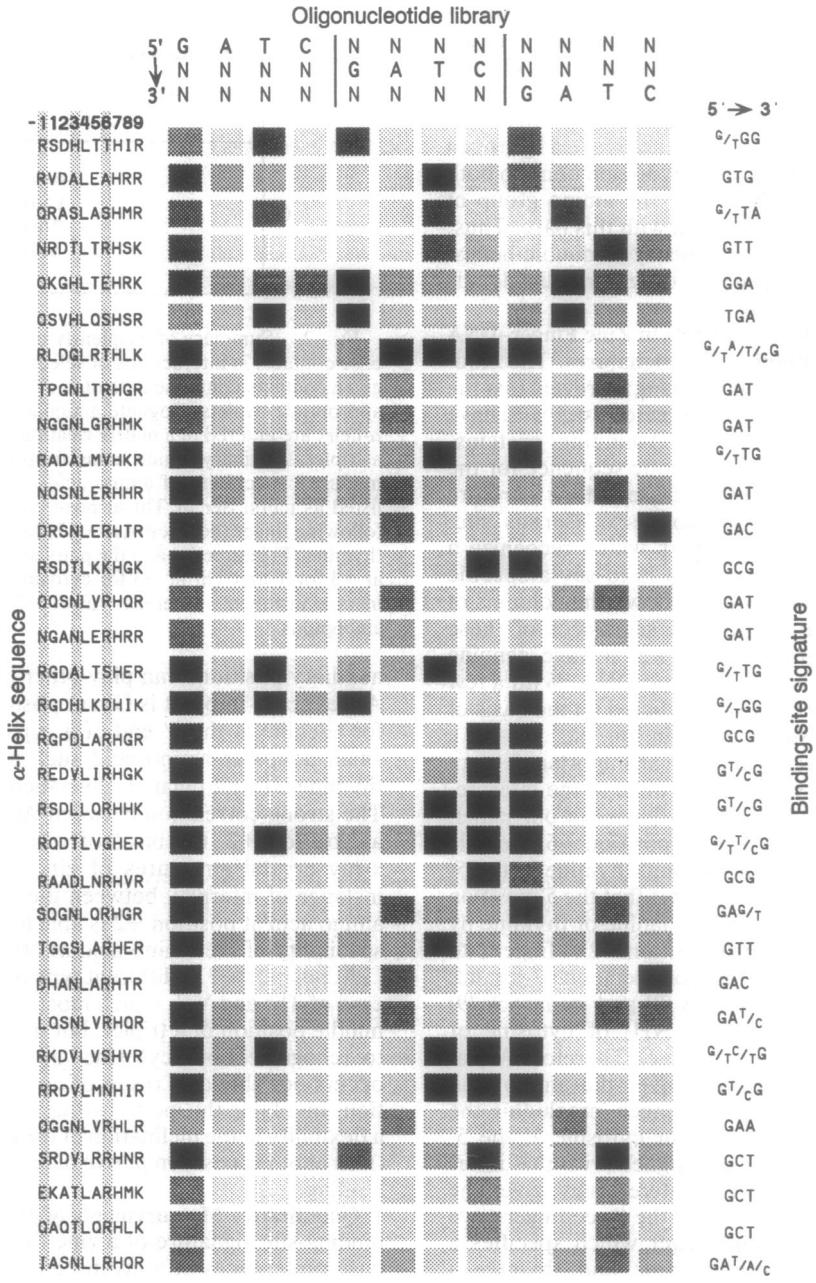

FIG. 1. Binding-site signatures of individual zinc finger phage. The diagram is of raw data and represents binding of zinc finger phage to randomized DNA immobilized in the wells of microtiter plates. To test each zinc finger phage against each oligonucleotide library (see text), DNA libraries are applied to columns of wells (down the plate), while rows of wells (across the plate) contain equal volumes of a solution of a zinc finger phage. The identity of each library is given as the middle triplet of the "bound" strand of Zif268 operator, where N represents a mixture of all four nucleotides. The zinc finger phage is specified by the sequence of the variable region of the middle finger, numbered relative to the first helical residue (position +1), and the three primary recognition positions are highlighted. Bound phage are detected by an enzyme immunoassay. The approximate strength of binding is indicated by a gray scale proportional to the enzyme activity. From the pattern of binding to DNA libraries, called the signature of each clone, one or a small number of binding sites can be read off; these are shown at right.

equilibrium is reached in solution prior to capture on the solid phase.

RESULTS AND DISCUSSION

Binding-Site Signature of the Second Zinc Finger of Zif268. The top row of Fig. 1 shows the signature of the second finger of wild-type Zif268. From the pattern of strong signals indicating binding to oligonucleotide libraries having GNN, TNN, NGN, and NNG as the middle triplet, it emerges that the optimal binding site for this finger is (T/G)GG, in accord with the published consensus sequence (7). This has implications for the interpretation of the x-ray crystal structure of Zif268 solved in complex with a consensus operator having TGG as the middle triplet (8). For instance, His at position +3 of the middle finger was modeled as donating a hydrogen bond to N7 of guanine, suggesting an equivalent contact to be possible with N7 of adenine, but from the binding-site signature we can see that there is discrimination against adenine. This implies that the His may prefer to make a hydrogen bond to O6 of guanine or a bifurcated hydrogen bond to both O6 and N7 or that a steric clash with the amino group of adenine may prevent a tight interaction with this base. Thus, from consideration of the stereochemistry of double-helical DNA, binding-site signatures can give insight into the details of zinc finger–DNA interactions.

Amino Acid–Nucleotide Base Contacts in Zinc Finger–DNA Complexes Deduced from Binding-Site Signatures. The binding-site signatures of other zinc fingers (Fig. 1) reveal that the phage selections we performed in our previous study (1) have yielded highly sequence-specific DNA-binding proteins. Some of these are able to specify a unique sequence for the middle triplet of a variant Zif268 binding site and are therefore more specific than is Zif268 itself for its consensus site. Moreover, one can identify the fingers which recognize a particular oligonucleotide library—that is to say a specific base at a defined position—by looking down the columns of Fig. 1. By comparing the amino acid sequences of these fingers we can identify any residues which have genuine preferences for particular bases on bound DNA. With a few exceptions, these are as previously predicted on the basis of phage display (1) and are summarized in Fig. 2.

The binding-site signatures also reveal an important feature of our phage display library which is crucial to the interpretation of our selection results. All the fingers in our panel, regardless of the amino acid present at position +6, are able to recognize guanine or both guanine and thymine at the 5′ end of a triplet. Our explanation for this is that the 5′ position of the middle triplet is fixed as either guanine or thymine by a contact from the invariant Asp at position +2 of finger 3 to the partner of either base on the complementary strand, analogous to those seen in the Zif268 (8) and tramtrack (9) crystal structures (a contact to the NH_2 of cytosine or adenine, respectively, in the major groove). Therefore Asp at position +2 of finger 3 is dominant over the amino acid present at position +6 of the middle finger, precluding the possibility of recognition of adenine or cytosine at the 5′ position. Future libraries must be designed with this interaction omitted or the position varied. Interestingly, given the framework of the conserved regions of the three fingers, we can identify a rule in the second finger which specifies a frequent interaction with both guanine and thymine—namely, the occurrence of Ser or Thr at position +6, which may donate a hydrogen bond to either base.

Modulation of Base Recognition by Auxiliary Positions. As we have noted above, position +2 is able to specify the base directly 3′ of the "cognate triplet" and can thus work in conjunction with position +6 of the preceding finger. The binding-site signatures, while pointing to amino acid–base contacts from the three primary positions, indicate that

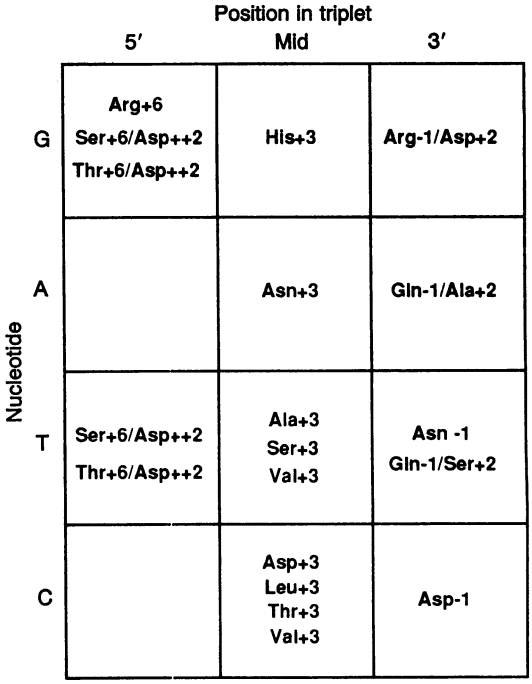

FIG. 2. Summary of frequently observed amino acid–nucleotide base contacts in interactions of selected zinc fingers with DNA. The given contacts comprise a "syllabic" recognition code (see text) for appropriate triplets. Cognate amino acids and their positions in the α-helix are entered in a matrix relating each base to each position of a triplet. Auxiliary amino acids from position +2 can enhance or modulate specificity of amino acids at position −1, and these are listed as pairs. Ser or Thr at position +6 permit Asp at +2 of the following finger (denoted Asp++2) to specify both guanine (G) and thymine (T) indirectly, and the pairs are listed. The specificity of Ser at +3 for T and Thr at +3 for cytosine (C) may be interchangeable in rare instances, whereas Val at +3 appears to be consistently ambiguous.

auxiliary positions can play other parts in base recognition. A clear case in point is Gln at position −1, which is specific for adenine at the 3′ end of a triplet when position +2 is a small nonpolar amino acid such as Ala but is specific for thymine when a polar residue such as Ser is at position +2. The strong correlation between Arg at position −1 and Asp at position +2, the basis of which is understood from the x-ray crystal structures of zinc fingers (8, 9), is another instance of interplay between these two positions. Thus the amino acid at position +2 is able to modulate or enhance the specificity of the amino acid at other positions.

At position +3, a different type of modulation is seen in the case of Thr and Val, which most often prefer cytosine in the middle position of a triplet, but in some zinc fingers are able to recognize both cytosine and thymine. This ambiguity occurs possibly as a result of different hydrophobic interactions involving the methyl groups of these residues, and here a flexibility in the inclination of the finger rather than an effect from another position *per se* may be the cause of ambiguous reading.

Quantitative Measurements of Dissociation Constants. The binding-site signature of a zinc finger reveals its differential base preferences at a given concentration of DNA. As the concentration of DNA is altered, one can expect the binding site signature of any clone to change, being more distinctive at low [DNA], and becoming less so at higher [DNA] as the K_d of less favorable sites is approached and further bases become acceptable at each position of the triplet. Further, because two base positions are randomly occupied in any one library of oligonucleotides, binding-site signatures are not formally able to exclude the possibility of context depen-

dence for some interactions. Therefore to supplement binding-site signatures, which are essentially comparative, quantitative determinations of the K_d values of each phage for different DNA binding sites are required. After phage display selection and binding-site signatures, this is the third and definitive stage in assessing the specificity of zinc fingers.

Examples of such studies presented in Fig. 3 reveal that zinc finger phages bind the operators indicated in their binding-site signatures with K_d values in the range of 10^{-8} to 10^{-9} M and can discriminate against closely related binding sites by factors greater than an order of magnitude. Indeed, Fig. 3 shows such differences in affinity for binding sites which differ in only one out of nine base pairs. Since the zinc fingers in our panel were selected from a library by noncompetitive affinity purification, there is the possibility that fingers which are even more discriminatory can be isolated by a competitive selection process.

Measurements of K_d allow different triplets to be ranked in order of preference according to the strength of binding. The examples here indicate that the contacts from either position -1 or $+3$ can contribute to discrimination. Also, the ambiguity in certain binding-site signatures referred to above can be shown to have a basis in the equal affinity of certain fingers for closely related triplets. This is demonstrated by the K_d values of the finger containing the amino acid sequence RGDALTSHER for the triplets TTG and GTG.

A Code for Zinc Finger–DNA Recognition. One would expect that the versatility of the zinc finger motif will have allowed evolution to develop various modes of binding to DNA (and even to RNA) which will be too diverse to fall under the scope of a single code. However, although a code may not apply to all zinc finger–DNA interactions, there is now convincing evidence that a code applies to a substantial subset. This code will fall short of being able to predict unfailingly the DNA binding-site preference of any given zinc finger from its amino acid sequence but may yet be sufficiently comprehensive to allow the design of zinc fingers with specificity for a given DNA sequence.

Using the selection methods of phage display (1) and of binding-site signatures, we find that in the case of Zif268-like zinc fingers, DNA recognition involves four fixed principal (three primary and one auxiliary) positions on the α-helix, from which a limited and specific set of amino acid–base contacts result in recognition of a variety of DNA triplets. In other words, a code can describe the interactions of zinc fingers with DNA. Toward this code, we can propose amino acid–base contacts for almost all the entries in a matrix relating each base to each position of a triplet (Fig. 2). Where there is overlap, our results complement those of Desjarlais and Berg (10, 11), who have derived similar rules by altering zinc finger specificity, using database-guided mutagenesis.

Combinatorial Use of the Coded Contacts. The individual base contacts listed in Fig. 2, though part of a code, may not always result in sequence-specific binding to the expected base triplet when used in any combination. First, we must be aware of the possibility that zinc fingers may not be able to recognize certain combinations of bases in some triplets by use of this code, or even at all. Otherwise, the majority of inconsistencies may be accounted for by considering variations in the inclination of the trident reading head of a zinc finger with respect to the triplet with which it is interacting. It appears that the identity of an amino acid at any one α-helical position is attuned to the identity of the residues at the other two positions to allow three base contacts to occur simultaneously. Therefore, for example, in order that Ala may pick out thymine in the triplet GTG, Arg must not be used to recognize guanine from position +6, since this would distance the Ala residue too far from the DNA (see for example the finger containing the amino acid sequence RGDALTSHER). Second, since the pitch of the α-helix is 3.6

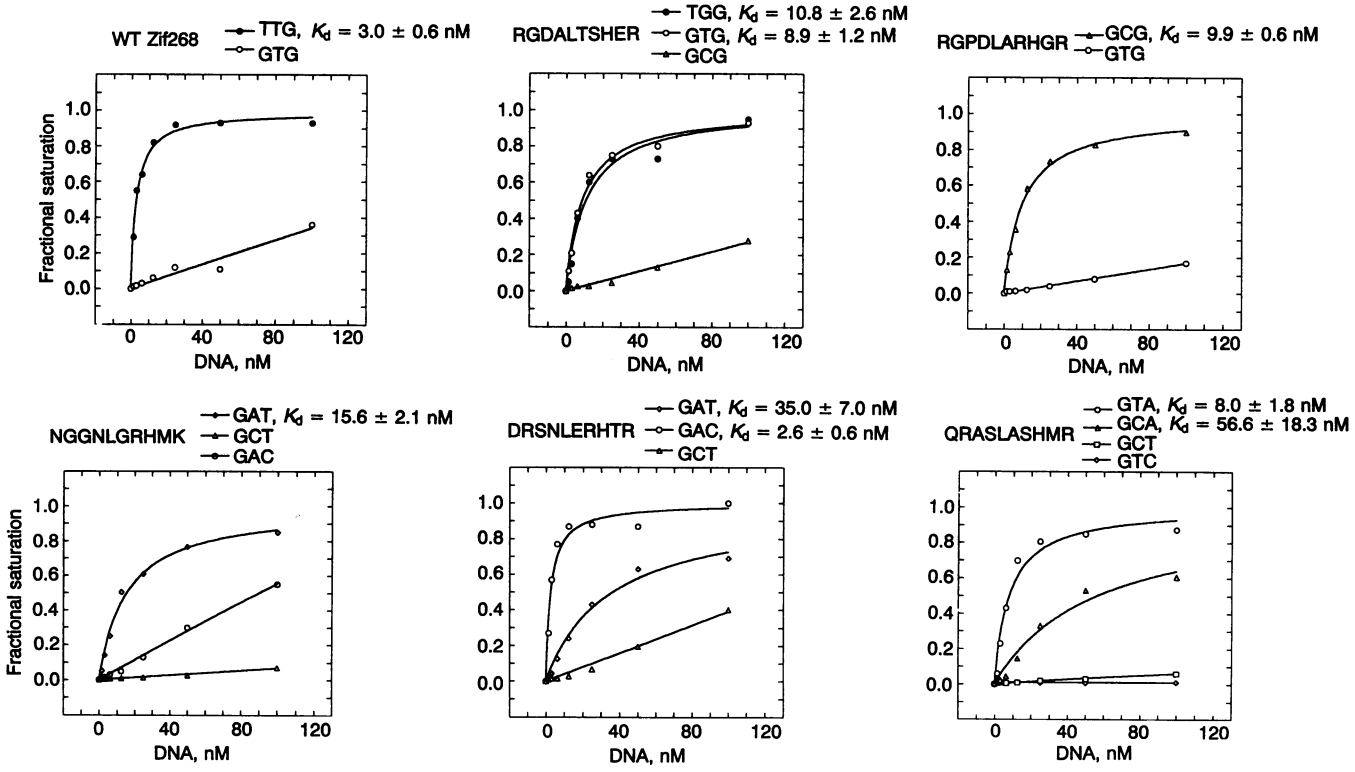

FIG. 3. Determination of apparent equilibrium dissociation constants of zinc finger phage for variants of the Zif268 binding site, showing discrimination of closely related triplets by the middle finger, usually by factors of >10. The two outer fingers carry the native sequence, as do the two cognate outer DNA triplets. The sequence of amino acids occupying helical positions -1 to +9 of the varied middle finger is shown in each case. WT, wild type (RSDHLTTHIR).

amino acids per turn, positions −1, +3, and +6 are not an integral number of turns apart, so that position +3 is nearer to the DNA than is −1 or +6. Hence, for example, short amino acids such as His and Asn, rather than the longer Arg and Gln, are used for the recognition of purines in the middle position of a triplet.

As a consequence of these distance effects, we might say that the code is not really "alphabetic" (always identical amino acid–base contact) but rather "syllabic" (use of a small repertoire of amino acid–base contacts). An alphabetic code would involve only four rules, but syllabicity adds an additional level of complexity, since systematic combinations of rules comprise the code. Nevertheless, the recognition of each triplet is still best described by a code of syllables, rather than a catalogue of "logograms" (idiosyncratic amino acid–base contact depending on triplet).

Conclusions. The syllabic code of interactions with DNA is made possible by the versatile framework of the zinc finger: this allows an adaptability at the interface with DNA by slight changes of orientation, which in turn maintains a stoichiometry of one coplanar amino acid per base pair in many different complexes. Given this mode of interaction between amino acids and bases, it is to be expected that recognition of guanine and adenine by Arg and Asn/Gln, respectively, is an important feature of the code; but remarkably, other interactions can be more discriminatory than was anticipated (12). Conversely, it is clear that degeneracy can be programmed in the zinc fingers in varying degrees, allowing for intricate interactions with different regulatory DNA sequences (7, 13). One can see how this principle makes possible the regulation of differential gene expression by a limited set of transcription factors.

As we have noted, the versatility of the finger motif will most likely allow other modes of binding to DNA. Similarly, we must take into account the malleability of nucleic acids, such as was observed in ref. 9, where a deformation of the double helix at a flexible base step allows a direct contact from Ser at position +2 of finger 1 to a thymine at the 3′ position of the cognate triplet. Even in our selections there are instances of fingers whose binding mode is obscure and may require structural analyses for clarification. Thus, water may be seen to play an important role, for example, where short side chains such as those of Asp, Asn, or Ser interact with bases from position −1 (14, 15).

Eventually, it might be possible to develop a number of codes describing zinc finger binding to DNA, which could predict the binding-site preferences of some zinc fingers from their amino acid sequences. The functional amino acids selected in this study at positions −1, +3, and, to some extent, +6 are very frequently observed at the same positions in naturally occurring fingers (e.g., see figure 4 of ref. 16), supporting the existence of coded contacts from these three positions. However, the lack of definitive predictive methods is not a serious practical limitation, as current laboratory techniques (this paper and refs. 2 and 3) will allow the identification of binding sites for a given DNA-binding protein. Rather, we can apply phage selection and a knowledge of the recognition rules to the converse problem, the design of proteins to bind predetermined DNA sites.

Prospects for the Design of DNA-Binding Proteins. The ability to manipulate the sequence specificity of zinc fingers implies that we are on the eve of designing DNA-binding proteins with desired specificity for applications in medicine and research (11, 17). This is possible because of the modular nature of the zinc finger, by contrast to all other DNA-binding motifs, since DNA sites can be recognized by appropriate combinations of independently acting fingers linked in tandem.

The coded interactions of zinc fingers with DNA can be used to model the specificity of individual zinc fingers *de novo* or, more likely, in conjunction with phage display selection of suitable candidates. In this way, according to requirements, one could modulate the affinity for a given binding site or even engineer an appropriate degree of indiscrimination at particular base positions. Moreover, the additive effect of multiply repeated domains offers the opportunity to bind specifically and tightly to extended, and hence very rare, genomic loci. Thus, zinc finger proteins might well be a good alternative to the use of antisense nucleic acids in suppressing or modifying the action of a given gene, whether normal or mutant. To this end, extra functions could be introduced into these DNA-binding domains by appending suitable natural or synthetic effectors.

We thank L. Fairall, A. Griffiths, D. Rhodes, and J. Schwabe for critical reading of the manuscript. Y.C. thanks the Medical Research Council and the British Council for funding.

1. Choo, Y. & Klug, A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11163–11167.
2. Thiesen, H. J. & Bach, C. (1990) *Nucleic Acids Res.* **18,** 3203–3209.
3. Pollock, R. & Treisman, R. (1990) *Nucleic Acids Res.* **18,** 6197–6204.
4. Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I., Waterhouse, P., Crosby, W. L., Kontermann, R. E., Jones, P. T., Low, N. M., Allison, T. J., Prospero, T. D., Hoogenboom, H. R., Nissim, A., Cox, J. P. L., Harrison, J. L., Zaccolo, M., Gherardi, E. & Winter, G. (1994) *EMBO J.* **13,** 3245–3260.
5. Choo, Y. & Klug, A. (1993) *Nucleic Acids Res.* **21,** 3341–3346.
6. Marks, J. D., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1992) *J. Biol. Chem.* **267,** 16007–16010.
7. Christy, B. & Nathans, D. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8737–8741.
8. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252,** 809–817.
9. Fairall, L., Schwabe, J. W. R., Chapman, L., Finch, J. T. & Rhodes, D. (1993) *Nature (London)* **366,** 483–487.
10. Desjarlais, J. R. & Berg, J. M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7345–7349.
11. Desjarlais, J. R. & Berg, J. M. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 2256–2260.
12. Seeman, N. D., Rosenberg, J. M. & Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 804–808.
13. Harrison, S. D. & Travers, A. A. (1990) *EMBO J.* **9,** 207–216.
14. Qian, Y. Q., Otting, G. & Wuthrich, K. (1993) *J. Am. Chem. Soc.* **115,** 1189–1190.
15. Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A. & Sigler, P. B. (1994) *Nature (London)* **368,** 469–473.
16. Desjarlais, J. R. & Berg, J. M. (1992) *Proteins* **12,** 101–104.
17. Rebar, E. J. & Pabo, C. O. (1994) *Science* **263,** 671–673.