Toward rules relating zinc finger protein sequences and DNA binding site preferences

(transcription factors/Sp1/molecular recognition)

JOHN R. DESJARLAIS* AND JEREMY M. BERG†

*Thomas C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, MD 20218; and †Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205

ABSTRACT Zinc finger proteins of the Cys_2-His_2 type consist of tandem arrays of domains, where each domain appears to contact three adjacent base pairs of DNA through three key residues. We have designed and prepared a series of variants of the central zinc finger within the DNA binding domain of Sp1 by using information from an analysis of a large data base of zinc finger protein sequences. Through systematic variations at two of the three contact positions (underlined), relatively specific recognition of sequences of the form 5'-GGGGN(G or T)GGG-3' has been achieved. These results provide the basis for rules that may develop into a code that will allow the design of zinc finger proteins with preselected DNA site specificity.

Relationships between macromolecules are central to biology. While some such interactions are highly complex and unpredictable, certain classes are describable in terms of relatively simple rules. For example, the interaction between one strand of DNA and its antiparallel partner is governed by the Watson–Crick base-pairing scheme (1). Since the discovery of sequence-specific DNA binding proteins, there has been speculation about the potential for codes that control these interactions (2–4). The prospect for the existence of such codes has been improved by the discovery of superfamilies of DNA binding proteins. The zinc finger proteins typified by the human transcription factor Sp1 (5) represent such a family. These proteins contain tandem arrays of domains, each of which approaches the sequence Pro-(Tyr or Phe)-Xaa-Cys-Xaa-Xaa-Cys-Xaa-Xaa-Xaa-Phe-Xaa-Xaa[13]-Xaa-Xaa-Xaa[16]-Leu-Xaa-Xaa[19]-His-Xaa-Xaa-Xaa-His-Thr-Gly-Glu-Lys (6–8). Modeling (9–11), mutagenesis (11–13), and crystallographic studies (14) have revealed general features of the complexes formed between proteins of this class and DNA. Each domain is associated with 3 base pairs of DNA with the triplet subsites from adjacent domains being contiguous but nonoverlapping. The predominant sequence-specific contacts are made by residues Xaa-13, Xaa-16, and Xaa-19 in a region with helical structure. The complexes are asymmetric, with one DNA strand being much more extensively contacted than the other. The protein sequence, in the N to C direction, runs antiparallel to the more heavily contacted DNA strand in the 5' to 3' direction.

Given this basic structural information, we would like to achieve an understanding of the interactions between each domain and its DNA triplet that determine binding specificity. Previously, we took initial steps toward determining rules that relate the identity of amino acids in the contact positions to DNA binding site selectivity. Thus, analysis of a large data base of zinc finger domain sequences (8) allowed us to convert the DNA binding domain of Sp1 from selectivity for 5'-GGGGCGGGG-3' to GGGGCTGGG via three amino acid

changes in the second zinc finger domain (12, 13). The contact residue in position Xaa-13 was changed from arginine (which contacts G) to glutamine (which presumably contacts T), but additional changes from aspartic acid to serine in position Xaa-15 and glutamic acid to aspartic acid in position Xaa-16 were also necessary. These initial results encouraged us to design new variants based on other dominant correlations in the data base. Here we demonstrate that it is possible to obtain additional changes in DNA binding specificity by following these correlations. Based on the results reported here and previously (12, 13), we also begin the development of a code between zinc finger amino acid sequences and the DNA triplet sequences to which they bind.

MATERIALS AND METHODS

Production of Peptide Variants and DNA Probes. Peptide variants and DNA probes (79-base-pair restriction fragments) were constructed and prepared as described (12, 13). Approximate peptide quantities and relative concentrations were estimated visually by Coomassie staining of partially pure samples run on SDS/polyacrylamide gels. When necessary, proteins were further purified using a heparin Sepharose column (Pharmacia). After elution with KCl, the peptides were concentrated with a Centricon-10 concentrator (Amicon).

Gel Mobility-Shift Assay. Peptide preparations (0.2–4 μl; $\approx$0.1 μg of peptide) were added to labeled DNA probes (1–10 ng) so the final reaction conditions in 10-μl volumes were as follows: 35 mM Tris·HCl, pH 8.0/60 mM KCl/90 μM $ZnCl_2$/3 mM dithiothreitol/300 μg of bovine serum albumin (BSA) per ml/20 μg of poly(dI-dC)/10% (vol/vol) glycerol. Reaction mixtures were incubated at room temperature for 10 min and electrophoresed on a 1.8% SeaPlaque (FMC) agarose gel. The gels were dried and autoradiographed.
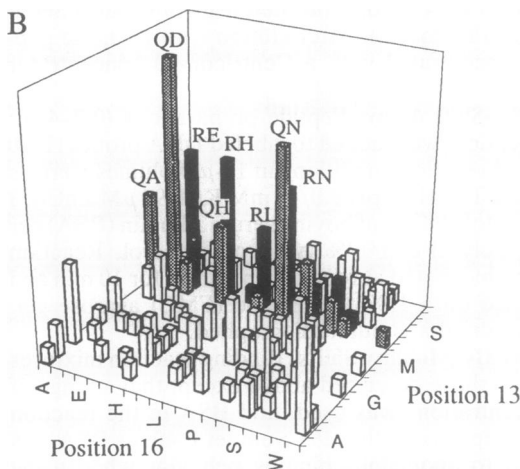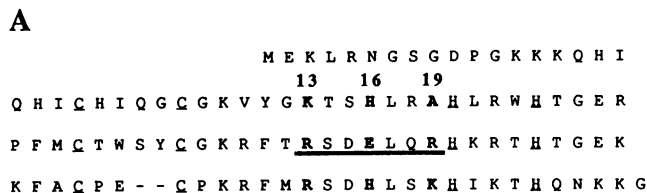
Quantitative Gel Mobility Shifts. Binding reaction mixtures were similar to those described above except that a range of peptide concentrations was used, and BSA in the reaction mixtures was replaced with 0.1% Nonidet P-40, as BSA was noted to lead to anomalous binding behavior when using purified protein. Quantitation of bound and free DNA was done on a Phosphorimager (Molecular Dynamics, Sunnyvale, CA). Binding isotherms were fit to the Langmuir function $Y = [P]/(K_d + [P])$, using the program KALEIDAGRAPH (Synergy Software, Reading, PA).
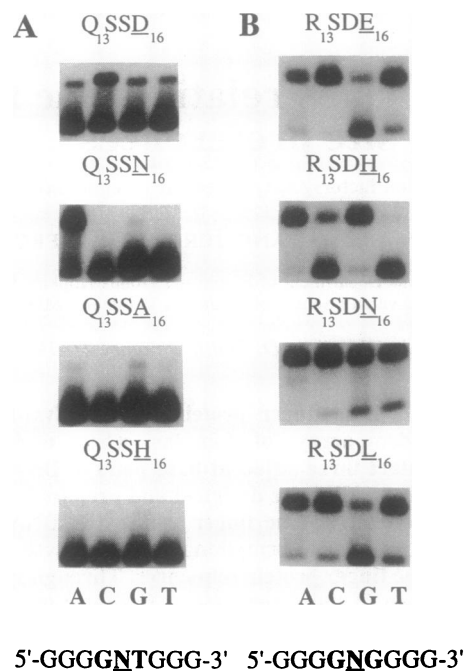
RESULTS

Zinc Finger Peptide Middle Finger Variants Derived from Data Base Correlations. Residue Xaa-16 in the zinc finger sequence Pro-(Tyr or Phe)-Xaa-Cys-Xaa-Xaa-Cys-Xaa-Xaa-Xaa-Phe-Xaa-Xaa[13]-Xaa-Xaa-Xaa[16]-Leu-Xaa-Xaa[19]-His-Xaa-Xaa-Xaa-His-Thr-Gly-Glu-Lys appears to recognize the central base pair in the DNA triplet to which one zinc finger binds, whereas residue Xaa-13 recognizes the third base pair

in that triplet. However, the recognition of base pairs by these residues is not independent (12, 13). Correlations between residues Xaa-13 and Xaa-16 are apparent in the amino acid codistributions, derived from a data base of zinc finger sequences (8), shown in Fig. 1. The most dominant subset is that in which residue 13 is glutamine and residue 16 is aspartic acid, asparagine, alanine, or histidine (listed in order of frequency of occurrence). The other prominent subset of distributions is that in which residue 13 is arginine and residue 16 is glutamic acid, histidine, asparagine, or leucine. These distributions are especially intriguing since simple bonding arguments suggest that each of the four nucleotides might be recognized by one member of each set. We constructed variants of an overexpressed Sp1 zinc finger peptide in which pairs of contact residues in the recognition helix of the second zinc finger are replaced by residues representing each of these distributions. Thus, these helices have the sequences Gln[13]-Ser-Ser-(Asp, Asn, His, or Ala)[16]-Leu-Gln-Arg[19] and Arg[13]-Ser-Asp-(Glu, His, Leu, or Asn)[16]-Leu-Gln-Arg[19], which include the wild-type sequence Arg[13]-Ser-Asp-Glu[16]-Leu-Gln-Arg[19]. Binding of each peptide from the Gln-13 subset to DNA probes containing the sequence 5'-GGGGNTGGG-3' and of each peptide from the Arg-13 subset to probes with the sequence 5'-GGGGNGGGG-3', where N is A, C, G, or T, was assayed by gel mobility shift as shown in Fig. 2.

A

```
                        M E K L R N G S G D P G K K K Q H I
                        13    16    19
Q H I C H I Q G C G K V Y G K T S H L R A H L R W H T G E R

P F M C T W S Y C G K R F T R S D E L Q R H K R T H T G E K

K F A C P E - - C P K R F M R S D H L S K H I K T H Q N K K G
```

B

FIG. 1. Primary structure of Sp1 zinc finger peptide and contact residue codistributions. (A) Primary structure of the wild-type Sp1 peptide. Contact residues are boldface and are numbered. Peptide variants contain amino acid changes within the underlined region of the second zinc finger domain. (B) Amino acid codistributions derived from a data base of zinc finger sequences (8). The axes labeled Position 13 and Position 16 represent the amino acids found in these positions in zinc fingers listed in the data base, with each axis running alphabetically by the single-letter amino acid code, as indicated with representative amino acids. The height of each column reflects the relative frequency of occurrence. For example, the column labeled QD corresponds to the 15 (of 239) sequences that contain a glutamine in position 13 and an aspartic acid in position 16. The two dominant series of codistributions, Gln[13]-(Ala, Asp, His, or Asn)[16] and Arg[13]-(Glu, His, Leu, or Asn)[16], are shown as shaded and solid bars, respectively.

A Q SSD 16   **B** R SDE 16
      13   13

Q SSN 16   R SDH 16
  13   13

Q SSA 16   R SDN 16
  13   13

Q SSH 16   R SDL 16
  13   13

        A C G T   A C G T

5'-GGGGNTGGG-3' 5'-GGGGNGGGG-3'

FIG. 2. Mobility-shift assays of DNA binding. (A) Binding of the Gln[13]-Ser-Ser-(Asp, Asn, Ala, or His)[16] second finger variants to DNA probes containing the sequences GGGGNTGGG, where N is A, C, G, or T, as shown below the lanes. Upper and lower bands in each square indicate bound and free DNA, respectively. (B) Binding of Arg[13]-Ser-Asp-(Glu, His, Asn, or Leu)[16] second finger variants to DNA probes containing the sequences GGGGNGGGG.

Specificity of Gln[13]-Ser-Ser-(Asp, Asn, His, or Ala)[16]-Leu-Gln-Arg[19] Peptides. Peptides in the Gln-13 subset have marked differences in specificity and levels of binding affinity for the four GGGGNTGGG binding sites. The Gln[13]-Ser-Ser-Asp[16] peptide binds with high affinity to the sequence GGGGCTGGG but also fairly well to GGGGATGGG and GGGGGTGGG. To establish the reliability of the gel-shift data for all of the mutant proteins, we performed more quantitative experiments with purified Gln[13]-Ser-Ser-Asp[16] peptide on all of the GGGGNTGGG binding sites. Gel shifts were performed at a range of peptide concentrations, with subsequent quantitation of bound and free DNA. The resulting binding isotherms are shown in Fig. 3. The results were also confirmed by quantitative DNase I footprinting (data not shown), which showed good agreement with the gel-shift data. The data show that the Gln[13]-Ser-Ser-Asp[16] peptide binds to the GGGGNTGGG probes in the order C > G > A > T with relative dissociation constants of ≈1:2:3:14. The preferences suggest that the aspartic acid in position 16 may be making a direct contact with cytosine, possibly accepting a hydrogen bond from the cytosine N-4 amino group as shown in Fig. 4A. Alternatively, the aspartic acid may not be contributing a direct contact to the DNA, by analogy with glutamic acids in equivalent positions as observed in the Zif268 cocrystal structure (14). The similarity between the quantitative results and the results at one protein concentration validate the use of the single point data in estimating relative affinities. The Gln[13]-Ser-Ser-Asn[16] peptide binds with high affinity only to the site GGGGATGGG. The asparagine may be interacting with adenine in the expected fashion (2), possibly donating a hydrogen bond from the asparagine amide group to the adenine N-7 and accepting a hydrogen bond from the adenine N-6 as in Fig. 4B. The Gln[13]-Ser-Ser-Ala[16] peptide binds approximately equally, with overall lesser affinity, to all four of the GGGGNTGGG binding sites. The Gln[13]-Ser-Ser-His[16] peptide shows no high-affinity binding to any of the four GGGGNTGGG bind-
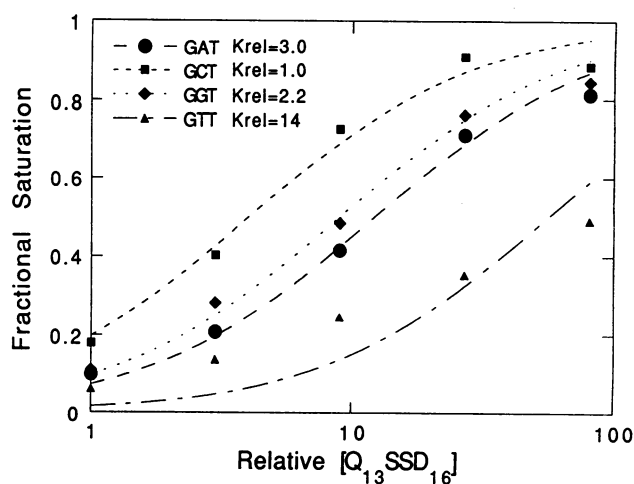
Biophysics: Desjarlais and Berg

Proc. Natl. Acad. Sci. USA 89 (1992) 7347

FIG. 3. Equilibrium binding isotherms for the Gln^{13}-Ser-Ser-Asp^{16} peptide interactions with the GGGG$\underline{N}$TGGG sites. Isotherms indicate that the binding data are concentration dependent and that the relative binding affinities for the different sites are preserved through a wide range of peptide concentrations. The relative dissociation constants (Krel) from the fitted curves are reported for each interaction. DNase I footprinting experiments (data not shown) performed with the same peptide and binding sites give consistent results and imply that the gel-shift data are thermodynamically valid.

ing sites, but it shows some slight binding to either GGG-G$\underline{A}$TGGG or GGGG$\underline{G}$TGGG. Histidine may hydrogen bond with a purine N-7, but because of its relatively large size, it may have a sterically unfavorable contribution to binding in the context of glutamine in position 13.

Specificity of Arg^{13}-Ser-Asp-(Glu, His, Leu, or Asn)^{16}-Leu-Gln-Arg^{19} Peptides. The Arg-13 subset of peptides exhibits a significant but less dramatic variation in binding activity and specificity for the Arg-13-specific GGGG$\underline{N}$GGGG binding sites. The Arg^{13}-Ser-Asp-Glu^{16} peptide, representing the wild-type Sp1 sequence, binds optimally to GGGG$\underline{C}$GGGG, but also with slightly lower affinity to GGGG$\underline{A}$GGGG and GGGG$\underline{T}$GGGG, consistent with earlier results using native Sp1 that showed binding with C > A > T > G in a 1:3:6:30 ratio of dissociation constants (15). The similarity between our results and these observations with regard to affinity order and apparent affinity validate both the use of the zinc finger region alone and the gel-shift assay for relative binding constant determination. The modest selectivity of the Arg^{13}-Ser-Asp-

Glu^{16} peptide is similar to the small discrimination for cytosine by the Gln^{13}-Ser-Ser-Asp^{16} peptide. The lack of strong discrimination by glutamic acid is explained by the Zif268–DNA cocrystal structure, which showed that glutamic acids in equivalent positions do not interact directly with the DNA (14). The Arg^{13}-Ser-Asp-His^{16} peptide binds with high affinity to either GGGG$\underline{A}$GGGG or GGGG$\underline{G}$GGGG. These affinities are approximately equal, as determined with subsaturating amounts of the Arg^{13}-Ser-Asp-His^{16} peptide (data not shown). The most likely interaction is the donation of a hydrogen bond from histidine to a purine N-7, consistent with the favored interpretation of the Zif268–DNA cocrystal data (14) and as shown in Fig. 4C. The Arg^{13}-Ser-Asp-Asn^{16} peptide binds with high affinity to all four of the GGGG$\underline{N}$GGGG binding sites, although it has a moderate preference for GGGG$\underline{A}$GGGG. The Arg^{13}-Ser-Asp-Leu^{16} peptide binds equally, with high affinity, to GGGG$\underline{A}$GGGG, GGGG$\underline{C}$GGGG, or GGGG$\underline{T}$GGGG, and so its binding behavior is similar to that of the Arg^{13}-Ser-Asp-Glu^{16} peptide.

Specificity Due to Position 19. Finally, we examined the effects of changes in the final contact residue in position 19. In the context of Gln^{13}-Ser-Ser-Asp^{16} in the second domain, genes encoding proteins with lysine and threonine in position 19 in addition to the arginine were produced. Binding experiments with the parent Arg-19 protein revealed the expected specificity for G in the 5′ base of the central triplet (data not shown). Examination of the binding preferences of the other two proteins was thwarted by low levels of expression. Extension of the rules to include position 19 will depend on overcoming these technical difficulties.

DISCUSSION

Nature of the Code. The rules that relate zinc finger protein sequence to preferred DNA binding site sequence may be combined to form a code. Since discrimination between different bases is rarely absolute (and may be somewhat dependent on assay conditions), some degeneracy of any such code is expected. A code might exist with various degrees of complexity. The simplest code would involve a correspondence between the identity of the amino acid in a contact position and the base pair recognized that is context independent. If the code is not this simple, the next level of complexity would involve intradomain context. Thus, the amino acid in a given contact position necessary to bind a particular base pair would depend on the amino acids in the other contact positions within the same zinc finger domain.
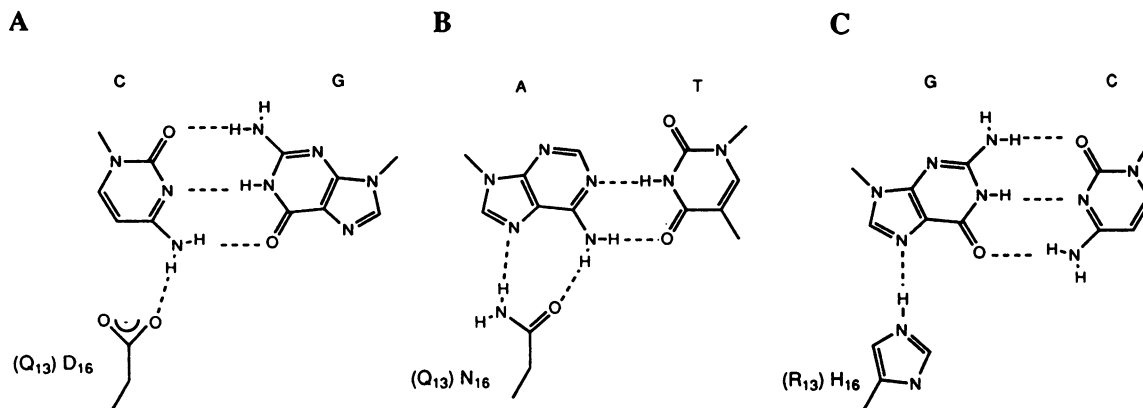

FIG. 4. Proposed contacts between amino acids and nucleotides. (*A*) A possible hydrogen bonding contact between Asp-16, in the context of Gln-13, and cytosine. This structure may explain the preference of aspartic acid for cytosine, as shown in Fig. 2*A*. (*B*) Hydrogen bonding contact proposed for the interaction between adenine and Asn-16, in the context of Gln-13, suggested by the strong discrimination by asparagine for adenine, as shown in Fig. 2*A*. (*C*) Proposed contact between His-16, in the context of Arg-13, and adenine or guanine, consistent with the observation of equal discrimination by histidine for adenine or guanine, as shown in Fig. 2*B*. This structure is also consistent with observations from the Zif268–DNA cocrystal structure (14).

If the zinc finger domains act as completely independent units, then no further complexity is expected. However, since the 3-base-pair sites contacted by adjacent zinc fingers are contiguous, the interdomain context may also play a role. In other words, is the extent of modularity of zinc fingers such that any set of zinc fingers of defined specificities can be linked together to recognize similarly linked base-pair triplets such that the specificities and affinities are preserved? Finally, the interactions between zinc finger proteins and DNA may be sufficiently subtle, in terms of the effects of noncontact amino acids, sequence-dependent DNA structure, and so on, such that no useful code may be easily conceived.

Intradomain Context Effects. It is apparent from the Zif268-DNA cocrystal structure (14) and from the results described here and earlier (11-13) that each base in the base-pair triplet recognized by a zinc finger is recognized by a single amino acid, at least in the representative peptides studied. However, the data presented here and elsewhere indicate the importance of intradomain context. The amino acids required at the contact positions are not independent; the residue required at one position for a given specificity may be different depending on the residues at the other positions. The most striking instances observed here are the pronounced differences in specificities of asparagine and histidine in position 16 when in the context of arginine versus glutamine in position 13. With glutamine in position 13, asparagine in position 16 was highly specific for A, whereas with arginine in position 13, the asparagine in position 16 showed only modest discrimination. Different effects were observed with histidine in position 16. In the Arg-13 context, a peptide with histidine in position 16 was found to bind well to sequences with A or G in the central position. However, the peptide with glutamine in position 13 and histidine in position 16 did not bind any of the oligonucleotides with high affinity under the conditions used. Thus, intradomain context can dramatically affect both specificity and overall affinity. The underlying structural basis for these effects appears to involve side-chain length; arginine is longer than glutamine and therefore favors larger residues in position 16 in order to make simultaneous contacts with DNA.

Interdomain Effects. Zinc finger domains appear to exist as relatively independent structural units. The Zif268 cocrystal structure confirmed this point although some interdomain interactions were observed. In addition, since the triplets contacted by adjacent domains are directly abutted, the relationship between amino acids in position 19 in one domain and position 13 in the next domain is similar to that between amino acids in positions 13 and 16 and in positions 16 and 19 within a single domain. Thus, *a priori*, it is not clear if the DNA binding specificity of each domain will be independent of the nature of the surrounding domains. It is reasonable to assume that finger to finger context may, indeed, be an additional factor in determining DNA binding activity in some situations. In light of this, the presence of alanine, presumably a noncontacting residue, in position 19 of the first domain of the Sp1 peptide, may have permitted the successful replacement of Arg-13 with Gln-13 in the second domain with the change in binding specificity that followed (12, 13).

Developing a Single Zinc Finger-DNA Recognition Code. Two kinds of codes, a design code and a prediction code, might exist for zinc finger-DNA recognition. The simpler of these, the design code, would allow for the construction of a zinc finger protein that would bind with appropriate specificity to a given DNA sequence. In contrast, a prediction code, translating information in the other direction, enables the prediction of the DNA sequence to which a given zinc finger protein would optimally bind. Beginning the development of a design code, the results discussed in this report can be combined in the form of a set of rules that correlate the

sequence of amino acids in the recognition region of a zinc finger domain with the triplet of base pairs to which it is capable of binding. Because of the 3-base-pair per zinc finger domain relationship, these rules can be tabulated as shown in Fig. 5. It is important to note here that these specificities may be somewhat dependent on the interdomain context of the zinc fingers in which they arise. Thus, this set of rules serves only as a starting point for a more general set of rules, which would incorporate context-dependent information as well.

Notably, there are different levels of specificity for the zinc fingers studied here. For instance, as discussed above, the Gln[13]-Ser-Ser-Asn[16] peptide binds with at least 100-fold higher affinity to GGGGATGGG than to any other of the GGGGNTGGG sites and, thus, at an appropriate protein concentration, would readily discriminate against these sites in a competitive binding situation. The Arg[13]-Ser-Asp-Glu[16] peptide, however, should be incapable of discriminating strongly between the GGGG(A, C, or T)GGGG sites. The *in vivo* Sp1 binding sequences previously described (5) suggest that the C in the consensus binding site GGGG$\underline{C}$GGGG is strongly conserved, implying that the modest discrimination may be sufficient, or that the C is conserved for other reasons. For purposes of design, some level of degeneracy in binding specificity may be tolerated, or perhaps even desired. Indeed, if some base-pair triplets cannot be recognized specifically, as may be the case for C- or T-rich triplets, then the zinc fingers exhibiting less specificity but reasonable affinity might prove to be essential. The rules described here, as well as additional ones that can be expected from other recurring patterns in the data base, and more complete knowledge of interfinger effects, should form a basis for the design of DNA binding proteins with desired site selectivities.

In addition, these rules provide some information concerning the prediction code. DNA binding sites are known for only a small fraction of the sequenced zinc finger proteins. For those proteins with known binding sites, some of the contact positions contain residues whose specificity has not yet been examined. An exception is yeast ADR1, which has two tandem zinc finger domains that have recognition helices Arg[13]-Gln-Glu-His[16]-Leu-Lys-Arg[19] and Arg[13]-Arg-Asp-Leu[16]-Ile-Arg[19]. From our rules, we would predict a binding site of the form 5'-G(A, C, or T)GG(A or G)G-3'. Footprinting experiments revealed binding to 5'-GT$\underline{TGGAGA}$-3' and 5'-A$\underline{GAGGAGA}$-3' (16), where we have underlined the bases that are consistent with our specificity rules. This suggests recognition of A by His-16 in the first domain and tolerance of several bases by Leu-16 in the second domain, as well as the canonical arginine-guanine interactions. The expected specificity of Arg-19 for G is not accounted for in the first example, but it may be suggestive of an interdomain context dependence of the specificity of this arginine, consistent with published results showing that this arginine is not required for ADR1 function (17). A similar prediction has been reported

	A	C	G	T	
5'-G	R N R 19 16 13	R E R 19 16 13	R H R 19 16 13	R L R 19 16 13	G-3'
	R N Q 19 16 13	R D Q 19 16 13	R H Q 19 16 13	R A Q 19 16 13	T-3'

(If R$_{13}$ then D$_{15}$; if Q$_{13}$ then S$_{15}$)

FIG. 5. A set of rules for zinc finger–DNA recognition. These rules are expected to be more useful as part of a design code than as part of a prediction code (see text). The three key zinc finger recognition residues that would be used to optimally recognize a given base-pair triplet is denoted in the box representing that triplet. The smaller lettering for two sets of residues indicates that these sets exhibit relatively low affinity when used in our Sp1-based system and so may not be as useful for design.

Biophysics: Desjarlais and Berg

Proc. Natl. Acad. Sci. USA 89 *(1992)*   7349

recently by Klevit (18) based largely on the arginine–guanine interactions. In addition, our rules allow the prediction of partial binding site sequences for some proteins for which no binding site data are yet available. Thus, for example, *Xenopus* Xfin contains 37 zinc finger domains divided into six sets (19). The fourth set has seven fingers with the recognition helices Gln^{13}, Ala^{16}, Lys^{19}–Gln^{13}, Asp^{16}, Lys^{19}–Gln^{13}, Asp^{16}, Lys^{19}–Gln^{13}, Ala^{16}, Lys^{19}–Gln^{13}, Asp^{16}, Lys^{19}–Glu^{13}, Ser^{16}, Lys^{19}–Gln^{13}, Asn^{16}, Lys^{19}. We do not yet know the binding preferences of Lys-19. We shall denote the base(s) recognized by this by N. Thus, a favorable binding site recognized by this set of fingers is predicted to be 5'-NAT-N??-NCT-NXT-NCT-NCT-NXT-3', where X denotes positions for which low specificity is expected and ? denotes unknown bases. The elucidation of binding sites for this and other zinc finger proteins will allow testing and extension of the rules developed here for binding site prediction.

1. Watson, J. D. & Crick, F. H. C. (1953) *Nature (London)* **171**, 737–738.
2. Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 804–808.
3. Matthews, B. W. (1988) *Nature (London)* **335**, 294–295.
4. Lehming, N., Sartorius, J., Kisters-Woike, B., von Wilcken-Bergmann, B. & Müller-Hill, B. (1990) *EMBO J.* **9**, 615–621.
5. Kadonaga, J. T., Jones, K. A. & Tjian, R. (1986) *Trends Biochem. Sci.* **11**, 20–23.
6. Rhodes, D. & Klug, A. (1986) *Cell* **46**, 123–132.
7. Evans, R. M. & Hollenberg, S. M. (1988) *Cell* **52**, 1–3.
8. Berg, J. M. (1990) *Annu. Rev. Biophys. Biophys. Chem.* **19**, 405–421.
9. Berg, J. M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 99–102.
10. Gibson, T. J., Postma, J. P. M., Brown, R. S. & Argos, P. (1988) *Protein Eng.* **2**, 209–218.
11. Nardelli, J., Gibson, T. J., Vesque, C. & Charnay, P. (1991) *Nature* **349**, 175–178.
12. Desjarlais, J. R. & Berg, J. M. (1992) *Proteins* **12**, 101–104.
13. Desjarlais, J. R. & Berg, J. M. (1992) *Proteins* **13**, 272.
14. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252**, 809–817.
15. Letovsky, J. & Dynan, W. S. (1989) *Nucleic Acids Res.* **17**, 2639–2653.
16. Eisen, A., Taylor, W. E., Blumberg, H. & Young, E. T. (1988) *Mol. Cell. Biol.* **8**, 4552–4556.
17. Thukral, S. K., Morrison, M. L. & Young, E. T. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9188–9192.
18. Klevit, R. E. (1991) *Science* **253**, 1367.
19. Altaba, A. R., Perry-O'Keefe, H., Melton, D. A. (1987) *EMBO J.* **6**, 3065–3070.