Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis

Jeannette Nardelli, Toby Gibson¹ and Patrick Charnay*

Laboratoire de Génétique Moléculaire, CNRS D 1302, Ecole Normale Supérieure, 46, rue d'Ulm, F-75230 Paris Cedex 05, France and ¹European Laboratory of Molecular Biology, Postfach 102209, Meyerhofstrasse 1, D-6900 Heidelberg, Germany

Received June 26, 1992; Accepted July 21, 1992

ABSTRACT

Zinc fingers of the Cys₂/His₂ class are conserved 28–30 amino acid motifs that constitute an important and widespread family of eukaryotic DNA-binding domains. It is therefore of great interest to understand the rules that govern specific recognition of DNA by zinc fingers. The DNA-binding domain of the transcription factor Krox-20 consists of three zinc fingers, each of them making its primary contacts with a three-base pair subsite. We have performed a data base-guided site-directed mutagenesis analysis of Krox-20: nine derivatives were generated, in which one to three amino acid changes had been introduced within finger 2, at positions which were likely to affect the specificity of DNA recognition. The affinities of the different proteins for a panel of potential DNA binding sites were estimated by gel retardation assay. Six of the derivatives bound specific targets with affinities comparable to that of wild type Krox-20 for its consensus binding site. However, the specificity of recognition was dramatically modified at the expected bases, in a manner that could be explained by examining the newly introduced amino acids within the context of the overall finger/triplet interaction. These data provide new insights into the details of zinc finger-DNA interactions and, combined with the modular nature of zinc fingers, illustrate both the potential and the difficulties of utilising these motifs for designing DNA-binding proteins with novel specificities.

INTRODUCTION

Zinc fingers (1, 2) constitute structural motifs folded around bound zinc ions that have appeared as major elements of eukaryotic protein-nucleic acid interactions. They are present in particular in many transcription factors (3–5) and can be grouped in several classes. The Cys₂/His₂ class consists of conserved domains of 28–30 amino acids which are usually present as tandem repeats. The Cys₂/His₂ motif was first detected within the amino acid sequence of the *Xenopus* transcription factor IIIA (TFIIIA) (1, 2), but, since then, vertebrate genomes have been

shown to encode hundreds (probably thousands) of these zinc finger domains (6). The structure of TFIIIA-like zinc fingers has been established both by NMR studies and by crystallography and shown to consist of an antiparallel β -hairpin packing onto an α -helix (7–9). Details of the interaction with DNA have been uncovered in the cases of two closely related proteins, Krox-24 and Krox-20: Krox-24 (also known as Zif268, Egr-1, NGFI-A and TIS8 (10–14)) is the product of an immediate-early serum response gene and may be involved in various signal transduction pathways; Krox-20 is also the product of an immediate early gene (15, 16) and is likely to be involved in pattern formation during hindbrain development (17, 18). Both proteins have almost identical DNA-binding domains consisting of three zinc fingers (10, 15), bind to the same 9 base pair DNA sequence (19, 20) and act as transcriptional activators (19, 20). The solution of the crystal structure of a complex containing the fingers from Krox-24/Zif268 bound to an oligonucleotide with the consensus target sequence (9) and a site-directed mutagenesis study of specificity determinants in the fingers of Krox-20 (21) have come to close conclusions: each finger has a similar relation to DNA and makes primary contacts in a three base pair subsite; 3 residue positions have been identified on the exposed face of the α -helix that make specific contacts with bases. In the crystal structure, base contacts were found to be limited to just one of the two DNA strands.

Based on the knowledge of the crystal structure, it is now possible to investigate in greater detail the general basis for specificity of DNA recognition by zinc fingers. We have introduced amino acid changes in the second finger of Krox-20, at positions involved in contacts with bases in the wild-type protein, to examine in particular whether these amino acids can still participate in DNA-binding, what is their specificity and whether the context is important for making particular changes. In an attempt to ensure that the changes would still allow DNA binding, combinations of amino acids were chosen that already occur in known zinc fingers. The analysis of the different Krox-20 derivatives indicates that most of them have kept the capacity of binding oligonucleotide sequences with high affinity, while their specificities of DNA recognition have been modified in ways that were not well predicted in advance but that can be understood

* To whom correspondence should be addressed

Table 1. Quantification of the gel retardation experiments. An estimate of the relative affinities of Krox-20 derivatives for different target sequences was obtained in the following way: the proportion of oligonucleotide retarded due to complex formation was estimated by densitometer scanning of the autoradiograms (Figure 3 and data not shown). This value was then compared with the value corresponding to the binding of wild type Krox-20 to its consensus target site. Only the central triplet of each oligonucleotide is shown, the surrounding nucleotides being identical. The mutated positions are shown in bold characters. + + +: proportion of complex similar to that of Krox-20 with the sequence 5'GCGGGGCG3', within a 2.5-fold range; + +: proportion 2.5- to 5-fold lower than the control; +: proportion 5- to 20-fold lower than the control; -: proportion at least 20-fold lower than the control. The result of the interaction of Q¹⁵S¹⁷R²¹ with the triplet 5'CCC3' is shown in parenthesis because in this case the interaction is likely to occur with the other strand of the DNA.

	GGG	<u>CGG</u>	<u>AGG</u>	<u>IGG</u>	<u>GCG</u>	<u>GAG</u>	GIG	<u>GGC</u>	<u>GGA</u>	<u>GGI</u>	<u>CCG</u>	<u>CGC</u>	<u>GCC</u>	<u>CCC</u>
Krox-20	+++	+++	+++	+++	-	+++	+	+	+	+	-	-	-	-
E ¹⁸ R ²¹	+	-	-	+	+++	+++	+++	-	-	-	-	-	-	-
R ²¹	+++	-	+	++	+	++	-	-	-	-	-	-	-	-
E ¹⁸	-	-	-	++	-	++	++	-	-	-	-	-	-	-
Q ¹⁸ R ²¹	+++	-	-	+	+++	+++	+++	-	-	-	-	-	-	-
T ¹⁷ N ¹⁸ R ²¹	+++	-	-	-	++	+++	++	-	-	-	-	-	-	-
Q ¹⁵ S ¹⁷ R ²¹	+++	-	-	-	-	++	-	++	+++	+++	-	-	-	(+++)
Q ¹⁵ S ¹⁷	-	-	-	+	-	-	-	-	-	-	-	-	-	-
E ¹⁵ S ¹⁷	-	-	-	+	-	-	-	-	+	-	-	-	-	-
L ¹⁵ S ¹⁷	-	-	-	-	-	-	-	-	-	-	-	-	-	-

by examining the substitutions within the context of the crystal structure. This work illustrates the potential of zinc fingers for designing novel DNA-binding specificities and contributes to the growing catalogue of 3 base pair modules recognised by zinc fingers.

MATERIALS AND METHODS

DNA constructions and site directed mutagenesis

All the bacterial expression plasmids were derived from pET-Krox-20 (19), a plasmid in which the wild-type Krox-20 coding sequence has been inserted into the expression vector pET3a (22, 23). The new Krox-20 derivatives were constructed as described previously (21). Briefly, mutations were introduced into the Krox-20 cDNA or its derivative R²¹ by polymerase chain reaction (PCR) amplification using oligonucleotide primers containing the relevant mutations. The mutating primers covered positions 1038 to 1072 (position 1075 for T¹⁷N¹⁸R²¹) of the gene, as numbered in reference 24, and included the following mutations: T¹⁷N¹⁸R²¹, codon changes GAC to ACC and CAC to AAC; Q¹⁵S¹⁷R²¹, codon changes CGA to CAA and GAC to AGC; E¹⁵S¹⁷, codon changes CGA to GAA and GAC to AGC; Q¹⁵S¹⁷, codon changes CGA to CAA and GAC to AGC; L¹⁵S¹⁷, codon changes CGA to CTA and GAC to AGC. After recloning, the PCR amplified fragments were entirely sequenced using a Pharmacia sequencing kit.

Data base analysis

Zinc finger sequences were selected from the SWISSPROT data base, searching with the pattern C-X₁₋₅-C-X₁₂-H-X₁₋₅-H using the program SCRUTINEER (25). The selected sequences were then analyzed by eye to identify those similar to Krox-20 finger 2 between positions 13 and 22.

Bacterial protein extracts and Western blotting

The expression system and protocols of Studier and collaborators (22, 23) were used to produce the different derivatives in *E. coli* as described previously (19). Protein extracts were prepared according to Kadonaga and collaborators (4) and dialysed against buffer B containing 20% (v/v) glycerol, 20 mM Tris-HCl pH 7.5, 50 mM KCl, 10 mM MgCl₂, 1 mM EDTA, 10 μM ZnSO₄, 1 mM DTT, 0.2 mM PMSF (phenylmethyl-sulfonyl fluoride, Sigma) and 1 mM NaHSO₃.

For immunoblotting, 2 and 10 μl of bacterial protein extract were separated by SDS-PAGE. Blotting and immunodetection were carried out according to standard procedures (26). The 539A antiserum and the peroxidase conjugated anti-rabbit antibody (Jackson Immunoresearch Laboratories) were diluted 2000- and 2500-fold respectively.

Gel retardation assay

The DNA probes for the gel retardation assays (27-29) consisted of two complementary oligonucleotides, whose sequences were derived from those of the following oligonucleotides: 5'-CTCT-GTACGCGGGGCGGTTA-3' and 5'-CTCTAACCG-CCCCGCGGTACA-3', which include a Krox-20 consensus binding site (in bold characters). The derivatives were obtained by introducing changes within the central 5'GGG3' or 5'CCC3' triplets (underlined) and the different triplets tested are shown in Table 1. The oligonucleotides were annealed by incubation at 65°C for 10 min and slow cooling to room temperature and they were subsequently labeled by filling the recessed 3' ends with the Klenow enzyme and [α -³²P]-labeled nucleotides (30). The bacterial extracts (0.5 to 5 μl, corresponding to equivalent amounts of Krox-20 derivative) were first preincubated on ice for 10 min in 20 μl of buffer C (20 mM Tris-HCl pH 7.5, 50 mM KCl, 10 mM MgCl₂, 1 mM EDTA, 1 mM DTT) supple-

Figure 1. Krox-20 interactions with its cognate sequence. (A). Schematic representation of the finger 2 of Krox-20. The amino acid involved in base-specific contacts in finger 2 are shown in bold characters. The positions which have been substituted in some of the derivatives presented in this study are circled in grey. Numbering of the amino acids is performed according to reference 24. (B). Schematic representation of the specific interactions of Krox-20 amino acids with bases. The model is based on the structural data obtained in the case of Krox-24/Zif268 by Pavletich and Pabo (1991).

mented with 3 μ g of poly dIdC and 8% Ficoll. 0.5 ng of labeled oligonucleotide was subsequently added and the incubation was pursued for an additional 20 min on ice. The mixture was then loaded on a 5% polyacrylamide gel and electrophoresis was performed for 2 hours at 14.5 V/cm in 0.5 $\times$ TBE buffer (30). The gels were dried and autoradiographic exposures were performed at -70°C with an intensifying screen. The proportion of oligonucleotide retarded due to complex formation was estimated by densitometer scanning of the autoradiograms.

Molecular Graphics

Results were investigated by modelling the amino acid substitutions into the Zif268 crystal structure coordinates (9) using the program INSIGHT II (Biosym Technologies, San Diego) running on a Silicon Graphics IRIS computer. The drawings in Figure 4 were prepared from photographs of the VDU screen.

RESULTS

To analyze the basis of zinc finger-DNA specific recognition and to investigate the possibility of creating novel specificities, we have introduced changes in the second finger of Krox-20. The mutant proteins were then exposed to a series of oligonucleotides carrying mutated Krox-20 binding sites to investigate the specificity of the protein-DNA contacts. The Krox-20 protein and this particular finger have been chosen for the following reasons: i) the DNA-binding domain of Krox-20 is of limited

Figure 2. Western blotting analysis of the relative amount of Krox-20 derivatives present in the bacterial extracts. The different extracts were analyzed with the 539A antiserum to allow normalization of gel retardation assays. pET, control bacterial extract containing no Krox-20 protein.

size and complexity, consisting of only three homologous fingers (15); ii) the recognition specificity of the protein has already been partially studied (19, 21); iii) the crystal structure of the complex of a consensus DNA binding site and the three fingers of Krox-24/Zif268, which are almost identical to those of Krox-20, has been established (9); iv) finger 2 is essential for high affinity DNA-binding and specific recognition (21); v) gross conformational adaptations of finger 2 or the DNA can be ruled out due to the numerous and specific interactions anchoring protein and DNA on both sides.

Gel retardation analysis of Krox-20 derivatives produced in *E.coli*

Figure 1 shows a schematic representation of Krox-20 finger 2 (residues are numbered according to reference 24) and of the specific interactions of Krox-20 with the bases of a consensus binding site as deduced from the work of Pavletich and Pabo (9). We will now refer to mutant derivatives of Krox-20 as X^y, where X is the substituted residue, in the one-letter code (31), at position y of finger 2. Krox-20 and its derivatives were produced in *E. coli*, using the expression system of Studier and collaborators (19, 21–23). The relative concentrations of Krox-20 derivatives present in the different extracts were estimated by Western blotting (Figure 2), with a rabbit antiserum, 539A, directed against the N-terminal part of the protein (M.Zerial, P.Chavrier, P.Charnay and R.Bravo, unpublished result). Relative quantification was obtained by comparison of different dilutions of the extract (data not shown) and subsequently equivalent amounts of Krox-20 derivatives were used in the gel retardation assay (27–29). The extracts were tested with a panel of oligonucleotide probes. This approach allows an easy determination of the relative affinities of different derivatives for the various oligonucleotide sequences (21). The oligonucleotide probes used in these experiments (see Materials and Methods) contained a central 9 base pair sequence corresponding to the Krox-20 consensus binding site 5'GCGGGGGCG3' or to a mutated version of this binding site. According to the crystal structure of the Krox-24/Zif268-DNA complex and to our previous mutagenesis analysis (9, 21), finger 2 is likely to make specific contacts only within the three-base pair subsite central to the binding site. Therefore, the mutations introduced into the oligonucleotide probes were restricted to those three base pairs.

DNA recognition pattern of Krox-20 finger 2

In a first series of experiments, we have evaluated the relative affinities of the wild-type Krox-20 molecule for the different target

oligonucleotides. The panel included in particular all possible single mutations in the central triplet from the starting sequence 5'GGG3' which is efficiently recognized by finger 2 (21). Figure 3 shows examples of the results obtained with four of the oligonucleotide probes. The relative affinities of the protein for each target were estimated by the ratio of complexed versus free oligonucleotide. The complete results are summarized in Table 1. Krox-20 is able to bind with relatively high affinity to a number of different sequences. It accepts any nucleotide at the 5' position of the triplet, with a slight preference for T. This is consistent with the crystal structure which indicates that only the middle and the 3' nucleotide are specifically contacted (Figure 1B and reference 9). Only A is permitted instead of G at the middle position. This is again consistent with the crystal structure which indicates that His¹⁸ donates an hydrogen bond to the N₇ of the G and suggests that a similar interaction can occur with an A. Finally, only G is permitted at the 3' position, where it accepts two hydrogen bonds from Arg¹⁵ (9). In conclusion, this study defines the triplet recognition pattern of Krox-20 finger 2 as the sequences 5'N(G/A)G3', where N can be A, T, G or C. This pattern is consistent with the structure of the protein-DNA complex (9) and with previous studies on Krox-20 and Krox-24 (19–21).

Mutagenesis based on the comparison of Krox-20 zinc fingers

To analyze the specificity of DNA recognition by zinc fingers, a further step was to introduce mutations at the finger positions which have been defined by the crystal structure as being capable of involvement in base-specific contacts (9). There are three such positions: 15, 18 and 21 (Figure 1) but, in addition, Asp¹⁷ was shown to be involved in orienting Arg¹⁵ (9) and this position is so close to the DNA that it may be directly involved in base-specificity in other proteins. In a previous study (21), we constructed three Krox-20 derivatives mutated in finger 2 at positions 18 and/or 21. These positions were selected because they showed variations between the three fingers of Krox-20 within a strongly conserved region (residues 15–22). We could demonstrate that when Glu and Arg were introduced respectively at positions 18 and 21, as in finger 1, the resulting protein, E¹⁸R²¹ (formerly referred to as Krox-20/2ER), did not bind to the Krox-20 consensus binding sequence, 5'GCGGGGCG3', but formed an abundant complex with an oligonucleotide carrying the sequence 5'GCGGCGCG3', indicating that the specificity of finger 2 had been modified and that, like finger 1, it could recognize the triplet 5'GCG3' but not 5'GGG3'. Analysis of the other mutant proteins R²¹ and E¹⁸ (originally referred to as Krox-20/2HR and Krox-20/2ET respectively) indicated that positions 18 and 21 both contributed to binding affinity while position 18 was the strongest determinant of specificity for the central base pair. We have now extended this analysis to our larger panel of oligonucleotide probes. Figure 3 shows examples of the results obtained with four of the oligonucleotide probes. In addition to the major retarded bands corresponding to the complexes including the complete Krox-20 derivatives (Figure 3, complexes B), bands of variable relative intensities and corresponding to faster migrating complexes were observed. They are likely to correspond to partially degraded Krox-20 proteins. Quantification was performed only when more than 80% of the Krox-20 derivative was intact. The complete data are summarized in Table 1. In the case of the E¹⁸R²¹ derivative, Arg²¹ is in position to donate two hydrogen bonds to the 5' nucleotide of the triplet. Accordingly, a G is allowed at this position but not

a C. Extrapolating from the crystal structure, Glu¹⁸ is not able to provide a hydrogen bonding ligand to the middle base, although it is in Van der Waals contact distance. Consistent with the lack of a ligand, discrimination is observed to be weak, with the exception of G, which is not accepted. This is likely to be due to solvent exclusion (and possibly electrostatic repulsion) caused by proximity of the Glu¹⁸ residue. Interestingly, 5'GAG3' is the only triplet bound efficiently both by E¹⁸R²¹ and by the wild-type protein. The analysis of the single mutants, E¹⁸ and R²¹ is consistent with the above observations. It reaffirms that the main determinant of discrimination of the middle base is the amino acid at position 18.

Data base-guided mutagenesis

The mutations analyzed above affected only two positions of finger 2. In addition, modifications introduced at these positions corresponded to residues present in the highly homologous finger 1 of Krox-20. We wanted to test the effect of substitutions affecting the two other positions possibly involved in base sequence recognition: in Krox-24/Zif268, Arg¹⁵ was shown to hydrogen bond to the 3' G of the triplet and Asp¹⁷ to stabilize this interaction (9). In addition, these positions have been shown to be essential for DNA-binding in an alanine scanning site-directed mutagenesis of the zinc fingers of yeast transcription factor ADR1 (32) and in a mutational analysis of Krox-24/NGFI-A (33). In order to modify these particular amino acids without creating unexpected distortions due to radical changes in the context, we analyzed zinc finger amino acid sequences available in data bases, looking for sequences similar to that of the finger 2 of Krox-20 between positions 13 and 22, but eventually differing at positions 15, 17, 18 or 21. We selected four such sequences: finger 2 of WT1 (34) (Phe-Ser-Arg-Ser-Asp-Glu-Leu-Lys-Arg-His) which differs only at position 18, 20 and 21, finger 9 of Mkr2 (35) (Phe-Ser-Arg-Ser-Thr-Asn-Leu-Thr-Arg-His) which differs only at positions 17, 18 and 21; finger 1 of Kox8 (36) (Phe-Ile-Leu-Ser-Ser-His-Leu-Thr-Thr-His) which differs at positions 14, 15 and 17; and finger 7 of ZNF8 (37) (Phe-Ile-Gln-Ser-Ser-His-Leu-Ile-Arg-His) which differs at positions 14, 15, 17, 20 and 21. On the basis of these naturally occurring zinc finger sequences, we have generated the following Krox-20 derivatives: Q¹⁸R²¹, T¹⁷N¹⁸R²¹, L¹⁵S¹⁷ and Q¹⁵S¹⁷R²¹. In addition, two other derivatives, Q¹⁵S¹⁷ and E¹⁵S¹⁷, were constructed to evaluate the importance of Arg²¹ in Q¹⁵S¹⁷R²¹ and to test the effect of different amino acids at position 15. Of special interest was the possibility that Gln and Asn would be found to show a preference for an A base with which the possibility exists for double-liganded interaction, while only a single bond is permitted with the other bases.

The derivative Q¹⁸R²¹ shows a less specific pattern of recognition than E¹⁸R²¹, since any base is fully accepted at the central position of the triplet. Although this is consistent with the capacity of Gln¹⁸ to act as both hydrogen bond donor or acceptor, the crystal structure (9) indicates that the explanation does not involve a direct ligand to the base (see Discussion). The derivative T¹⁷N¹⁸R²¹ also allows any base at the central position, analogous to Q¹⁸R²¹ (Figure 3 and Table 1). This suggests that the Asn residue at position 18 may be oriented to provide a single contact to any base at the middle position of the triplet. The change of Asp¹⁷, which according to the crystal structure is involved in positioning Arg¹⁵, into a Thr residue does not seem to affect the selectivity of Arg¹⁵ for G at the 3' position.

The derivative Q¹⁵S¹⁷R²¹ also presents a change in specificity.

Figure 3. Gel retardation analysis of the interactions between Krox-20 derivatives and four potential target sequences. The nucleotide sequence of the central part of each oligonucleotide is indicated: A, 5'GCGGGGGCG3' and 5'GCGGGCGCG3'; B, 5'GCGAGGGCG3' and 5'GCGGTGGCG3'. The bacterial extracts contained the indicated Krox-20 derivatives. pET: control bacterial extract containing no Krox-20 derivative. Krox-20: wild type Krox-20. F and B: free and bound oligonucleotide respectively.

It shows little discrimination for the 3' base of the triplet, in contrast to both wild type and the R²¹ derivative. Again, this might be consistent with an orientation of Gln15 allowing only one ligand to base. However, molecular graphics analysis suggested another explanation (see Discussion). Surprisingly, the derivative Q¹⁵S¹⁷ did not show any high affinity binding to the proposed target sequences, although Gln¹⁵ and His¹⁸ had been expected to provide a total of two or three hydrogen-bonds to bases. This may question the existence or the quality of contacts established by the Gln residue at position 15 (see Discussion). As with Q¹⁵S¹⁷, the other position 15 modifications E¹⁵S¹⁷ and L¹⁵S¹⁷ did not bind efficiently to any of the oligonucleotides.

DISCUSSION

Most of the Krox-20 derivatives tested in this study have the ability for high affinity binding to oligonucleotide sequences but show specificities that are different from that of the original protein. This is true for the derivatives originally designed by comparison of the three homologous fingers of Krox-20 as well as for those obtained from the analysis of zinc finger sequences present in data bases. These results indicate that the zinc fingers offer a great deal of flexibility at the level of the recognition positions since an amino acid residue contacting a base can be substituted by another one to alter specificity for the predicted base pair. Although there is so far no evidence that neighbouring fingers can modify a finger specificity, the possibility cannot be formally discounted. With this caveat, the results for database fingers may be extrapolated to these fingers in their natural sequence context. The results show that for all the new cases analysed here, there will be redundancy of triplet recognition. Redundancy has already been observed for the Krox-20 and Sp1 sites and may be quite a common feature, since the sites recognised by multifinger proteins will be much larger than for other types of DNA binding proteins and have more specificity determinants than are needed.

However, the decreased specificity at the discrimination site, rather than a strong specificity for a new base, was not anticipated for the database fingers which we chose. This result was particularly unexpected for Gln at position 15 because, as this is the most common sidechain at this position, it is expected to be highly base-specific (with adenine as the natural candidate). Likewise, Asn is frequently found at position 18 and greater specificity by it had also been anticipated. To generalise from this aspect of our results to conclude that Gln¹⁵ and Asn¹⁸ are always unspecific has an awkward corollary: that very many known finger sequences, which have the preferred amino acids at the base contacting positions, have poor triplet discrimination. This seems difficult to accept. Therefore it seemed important to look at the structural context of these substitutions, which could be done by modelling them into the Krox24/Zif268 crystal structure on a graphics computer.

Proposed explanations for the specificity changes

Using molecular graphics, three sidechains were substituted into finger 2 of the crystal structure to produce derivative Q¹⁵S¹⁷R²¹. Arg²¹ was modelled in the finger 1 conformation, contacting the first base of the triplet, while the second base contact by His¹⁸ remained unchanged. Gln15 was then extended as far as possible toward the hydrogen bonding atoms of the G in the 3rd position. As shown in Figure 4A, the orientation of the Gln is good but the ~5 Å distance between it and the base is too great to allow

Figure 4. Molecular graphics modelling of zinc finger-DNA interactions. (A) Model of the recognition α -helix of Krox-20 Q¹⁵S¹⁷R²¹ and specified triplet of basepairs, adapted from the crystal structure coordinates. The recognition α -helix is represented as a cylinder, the DNA backbone has thin lines while DNA bases and protein sidechains have thick lines. His¹⁸ and Arg²¹ can make hydrogen bonds to bases at the correct ~3 Å distance. The Gln¹⁵ amide group determinants however can only reach to 4.7 and 5.0 Å which is not close enough to allow hydrogen bonding, thereby accounting for the lack of specificity and affinity with this amino acid. (B) Model of the environment of Asn¹⁸ with Arg at positions 15 and 21. The Asn can reach to make a single contact, in this case donating a hydrogen bond to the purine N₇, but cannot provide the second ligand necessary to specify an A base.

stable hydrogen bonding (distance 2.9–3.5 Å) and water molecules would be expected to intrude. The lack of direct contact by Gln is consistent both with the absence of third base discrimination by Q¹⁵S¹⁷R²¹ as well as with the poor affinity and failure to bind by Q¹⁵S¹⁷, which would have only the single ligand from His¹⁸ to the middle base N₇. Similarly the failure of the variants E¹⁵S¹⁷ and L¹⁵S¹⁷ to bind stably can be explained by a failure to provide a complementary surface to the DNA with no contact to the 1st and 3rd base of the triplet.

Table 2. Catalogue of zinc finger specificities. Specificities of recognition of different zinc fingers have been deduced from this work (#) or from published reports. The identity of the amino acids at the specificity positions has been indicated. So far position 17 has not been shown to make direct contact to a base. In the recognition triplets, lowercase indicates partial or uncertain specificity for the base; N is for any base and R is for purine.

Name of Finger	Specificity Positions				Recognition Triplet (5'-3')	References
	21	18	17	15		
Krox-20 F1,F3; Sp1 F2; Krox-20 F2 E ¹⁸ R ²¹	R	E	D	R	GcG	9, 21, 42, #
Krox-20 F2	T	H	D	R	NRG	9, 21, #
Krox-20 F2 R ²¹	R	H	D	R	GRG	#
Krox-20 F2 Q ¹⁸ R ²¹	R	Q	D	R	GNG	#
Krox-20 F2 T ¹⁷ N ¹⁸ R ²¹	R	N	T	R	GNG	#
Krox-20 F2 Q ¹⁵ S ¹⁷ R ²¹	R	H	S	Q	GRN	#
Sp1 F1	A	H	S	K	grx	42
Sp1 F3	K	H	D	R	ggg	42
SP1 F2 QSDD	R	D	S	Q	GcT	38, 39
ADR1 F1	R	H	E	R	GRG	40
ADR1 F1 R115Q	R	H	E	Q	GRA	40
ADR1 F1 H118T	R	T	E	R	GcG	40
ADR1 F1 R115Q-H118T	R	T	E	Q	Gca	40
ADR1 F2	R	L	D	R	ttG	40

The Asn¹⁸ substitution was examined by computer graphics in the context of Arginines, at positions 15 and 21, binding to the 1st and 3rd base pair guanines. As shown in Figure 4B, Asn can provide one ligand to the purine N₇ position. However it cannot reach to make the double ligands necessary to confer specificity for adenine while, not only may the amide group act as both donor and acceptor for a base ligand, but there is also space available for the sidechain to move away from the base to avoid unfavourable contact, which might be the case for T. The observed lack of discrimination by the variant T¹⁷N¹⁸R²¹ for the middle base pair of the triplet is therefore consistent with the context of the Asn sidechain provided by the modified finger in the crystal structure.

Another result which was unclear was the tolerance of Gln¹⁸, but not Glu¹⁸ for G in the middle base, since, according to the cocrystal, neither of these residues are expected to make ligands to bases, while they occupy essentially the same space. Examination of the space around these residues indicates that the only conformation which is sterically allowed and which moves the sidechain away from the base, requires direct electrostatic contact to a phosphate. Gln, but not Glu, can make the hydrogen bond which will stabilise this position.

Importance of context for sidechain specificity

The use of the middle finger for substitutions should have minimised the ability of the finger and DNA to undergo major conformational adjustments. On the assumption that the conformations of all the variants are close to the cocrystal structure then the surprisingly low specificity of Gln¹⁵ and Asn¹⁸ are explainable. However, in all cases examined, an Arg or His original to finger 2 was present. These residues are large and therefore serve to push away the bases that they contact from the α -helical backbone. Examining fingers in the database, it is

found that when Gln¹⁵ is present, position 18 has a strong preference for small sidechains, with Asp the most abundant followed by Asn (38). The implication is that Gln, which is shorter than Arg, requires a smaller residue on the next helical turn if it is to reach to a base from position 15.

Evidence for just such a context dependence of Gln¹⁵ binding specificity has been provided recently by Desjarlais and Berg using finger 2 of Sp1 as the system (38, 39). Adopting the surmise that Gln¹⁵ would be A-specific, a construct having sidechains Q¹⁵S¹⁷E¹⁸R²¹ was made. On testing, this combination failed to bind to any DNA. When however, Asp¹⁸ was introduced into the construct to match the observed database preference, DNA binding was obtained but, against expectation, the Q showed a strong T preference and did not bind to the expected A base. The difference between Asp and Glu is size, strongly implying that the new context which allows Gln¹⁵ specificity is a closer approach of the finger to DNA. Thukral et al. (40) have also investigated Gln¹⁵, this time using ADR1 finger 1 as the system. A construct having the residues Q¹⁵E¹⁷H¹⁸R²¹ shows an A preference for Gln¹⁵, in contrast to the very similar Krox20 finger 2 modification tested here which shows no specificity. There is a likelihood of DNA conformational adjustment allowing Gln¹⁵ to contact the base for the ADR1 finger1, since it is an end finger. It will be most interesting to see which of A- or T-specificity proves the more general for fingers with Gln¹⁵.

Towards a recognition catalogue for zinc fingers

It is becoming clear from the results presented here and elsewhere (38–40) that there is no prospect of achieving a 'zinc finger recognition code' as hoped for by some authors. A genuine code implies a small number of rules which, combinatorially, have strong explanatory power. However, if the recognition by one amino acid is dependent on the amino acids at neighbouring positions, there will be no or, at best, very little combinatorial effect. In this respect, the zinc fingers are not different from other DNA-binding motifs which also cannot be described by recognition codes (41). Therefore the most realistic approach to utilising the specificity information which will be generated in the near future will be to collate the finger specificities in catalogue form.

Although the first finger specificities were identified only last year, change of specificity substitutions in Krox-20, Sp1 and ADR1 fingers have now been analysed. Given this burgeoning dataset, it is now appropriate to begin the cataloguing process and a preliminary recognition catalogue for the known finger specificities is presented in Table 2. Currently, all the source proteins represent a related subfamily of fingers with overt similarities to the Krox family, having in common G-rich binding sites and the use of Arg in G recognition. Therefore there is a strong bias to the catalogue which is not representative of the majority of finger proteins. There is as yet no understanding of how A and C and little of how T can be accurately specified. Nevertheless, the catalogue can be expected to grow rapidly as other interactions, especially those involving A/T specificity are analysed.

ACKNOWLEDGEMENTS

We thank Nikola Pavletich and Carl Pabo for providing the Zif268 coordinates. This work was supported by grants from the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Association pour

la Recherche sur la Cancer, the Ligue Nationale Française Contre le Cancer, the Ministère de la Recherche et de la Technologie, the Fondation pour la Recherche Médicale and the Association Française contre les Myopathies.

REFERENCES

1. Miller, J., McLachlan, A.D. and Klug, A. (1985) *EMBO J.*, **4**, 1609–1614.
2. Brown, R.S., Sander, C. and Argos, P. (1985) *FEBS Lett.*, **186**, 271–274.
3. Hartshorne, T.A., Blumberg, H. and Young, E.T. (1986) *Nature*, **320**, 283–287.
4. Kadonaga, J., Carner, K.R., Maslarz, F.R. and Tjian, R. (1987) *Cell*, **51**, 1079–1090.
5. Vrana, K.E., Churchill, M.E.A., Tullius, T.D. and Brown, D.D. (1988) *Mol. Cell. Biol.*, **8**, 1684–1696.
6. Jacobs, G. and Michaels, G. (1990) *New Biol.*, **2**, 583–584.
7. Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. and Wright, P.E. (1989) *Science*, **245**, 635–637.
8. Omichinski, J. G., Clore, G. M., Appella, E., Sakaguchi, K. and Gronenborn, A. M. (1990) *Biochemistry*, **29**, 9324–9334.
9. Pavletich, N.P. and Pabo, C.O. (1991) *Science*, **252**, 809–817.
10. Lemaire, P., Revelant, O., Bravo, R. and Charnay, P. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4691–4695.
11. Christy, B.A. and Nathans, D. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7857–7861.
12. Sukhatme, V.P., Cao, X., Chang, L.C., Tsai-Morris, C.-H., Stamenkovich, D., Ferreira, P.C.P., Cohen, D.R., Edwards, S.A., Shows, T.B., Curran, T., Le Beau, M.M., and Adamson, E.D. (1988) *Cell*, **53**, 37–43.
13. Milbrandt, J. (1987) *Science*, **238**, 797–799.
14. Lim, R.W., Vanum, B.C. and Herschman, H.R. (1987) *Oncogene*, **1**, 263–270.
15. Chavrier, P., Zerial, M., Lemaire, P., Almendral, J., Bravo, R. and Charnay, P. (1988) *EMBO J.*, **7**, 29–35.
16. Chavrier, P., Janssen-Timmen, U., Mattéi, M.-G., Zerial, M., Bravo, R. and Charnay, P. (1989) *Mol. Cell. Biol.*, **9**, 787–797.
17. Wilkinson, D., Bhatt, S., Chavrier, P., Bravo, R. and Charnay, P. (1989) *Nature*, **337**, 461–464.
18. Gilardi, P., Schneider-Maunoury, S. and Charnay, P. (1991) *Biochimie*, **73**, 85–91.
19. Chavrier, P., Vesque, C., Galliot, B., Vigneron, M., Dollé, P., Duboule, D. and Charnay, P. (1990) *EMBO J.*, **9**, 1209–1218.
20. Lemaire, P., Vesque, C., Schmitt, J., Stunnenberg, H., Frank, R., and Charnay, P. (1990) *Mol. Cell. Biol.*, **10**, 3456–3467.
21. Nardelli, J., Gibson, T., Vesque, C. and Charnay, P. (1991) *Nature*, **349**, 175–178.
22. Studier, F.W. and Moffatt, B.A. (1986) *J. Mol. Biol.*, **189**, 113–130.
23. Rosenberg, A.H., Lade, B.N., Chui, D.S., Lin, S.W., Dunn, J.J. and Studier, F.W. (1987) *Gene*, **56**, 125–135.
24. Gibson, T.J., Postma, J.P.M., Brown, R.S. and Argos, P. (1988) *Protein Engineering*, **2**, 209–218.
25. Sibbald, P. R. and Argos, P. (1990) *Comp. Applic. Biosci.*, **6**, 279–288.
26. Towbin, H., Staehelin, T. and Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4350–4354.
27. Fried, A. and Crothers, D.M. (1981) *Nucl. Acids Res.*, **9**, 6505–6525.
28. Garner, M.M. and Revzin, A. (1981) *Nucl. Acids Res.*, **9**, 3047–3059.
29. Strauss, F. and Varshavsky, A. (1984) *Cell*, **37**, 889–901.
30. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *In Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).
31. Dayhoff, M. O., ed. Atlas of protein sequence and structure (Natl. Biomed. Res. Found., Washington, DC), **5**, suppl. 3 (1978).
32. Thukral, S. K., Morrison, M. L. and Young, E. T. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9188–9192.
33. Wilson, T. E., Day, M. L., Pexton, T., Padgett, K.A., Johnston, M. and Milbrandt, J. (1992) *J. Biol. Chem.* **267**, 3718–3724.
34. Gessler, M., Poustka, A., Cavenee, W., Neve, R. L., Orkin, S.T. and Bruns, G.A.P. (1990) *Nature*, **343**, 774–778.
35. Chowdhury, K., Deutsch, U. and Gruss, P. (1987) *Cell* **48**, 771–778.
36. Thiesen, H.-J (1990) *The New Biologist* **2**, 363–374.
37. Lania, L., Donti, E., Pannuti, A., Pascucci, A., Pengue, G., Feliciello, I., La Mantia, G., LanFrancone, L. and Pelicci, P. G. (1990) *Genomics* **6**, 333–340.
38. Desjarlais, J. R. and Berg, J. M. (1992) *Proteins* **12**, 101–104.
39. Desjarlais, J. R. and Berg, J. M. (1992) *Proteins*, in press.
40. Thukral, S. K., Morrison, M. L. and Young, E. T. (1992) *Mol. Cell Biol.* **12**, in press.
41. Matthews, B. W. (1988) *Nature*, **335**, 294–295.
42. Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) *Trends Biochem.*, **11**, 20–23.