Nucleic Acids Research, 1994, Vol. 22, No. 16 3397–3405

Stereochemical basis of DNA recognition by Zn fingers

Masashi Suzuki*, Mark Gerstein[1] and Naoto Yagi[2]
MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK, [1]Beckman Laboratories for Structural Biology, Department of Cell Biology, Stanford Medical School, CA 94305-5400, USA and [2]School of Medicine, Tohoku University, Seiryo-machi, Sendai, 980-77, Japan

Received April 25, 1994; Revised and Accepted July 5, 1994

ABSTRACT

DNA-recognition rules for Zn fingers are discussed in terms of crystal structures. The rules can explain the DNA-binding characteristics of a number of Zn finger proteins for which there are no crystal structures. The rules have two parts: chemical rules, which list the possible pairings between the 4 DNA bases and the 20 amino acid residues, and stereochemical rules, which describe the specific base positions contacted by several amino acid positions in the Zn finger. It is discussed that to maintain the correct binding geometry, in which the N-terminus of the recognition helix is closer to the DNA than the C-terminus, the residues facing the DNA on the helix must be larger near the C-terminus, and that two different types of fingers (A and B) bind to DNA in distinctly different ways and cover different numbers of base pairs.

1. INTRODUCTION

The Zn finger motif was first proposed for TFIIIA (Miller *et al.*, 1985), and now many transcription factors are known to use the same motif for DNA-recognition. Crystal structures of DNA-finger complexes — Zif268 [Zif] (Pavletich and Pabo, 1991), GLI (Pavletich and Pabo, 1993), and tramtrack [TTK] (Fairall *et al.*, 1993) (Figures 1, 2a)— and several NMR structures of Zn fingers (Klevit *et al.*, 1989, Páraga *et al.*, 1990, Omichinski *et al.*, 1990, 1992, Neuhaus *et al.*, 1992, Lee *et al.*, 1989, Kochiyan *et al.*, 1991) have been determined. Many studies have been carried out towards understanding the DNA-recognition rules for Zn fingers (Fairall *et al.*, 1986, Gibson *et al.*, 1988, Nardelli *et al.*, 1991, Jacobs 1992, Desjarlais and Berg 1992, 1993, Rosenfeld and Margalit, 1993).

Although similarities in DNA-recognition by some fingers have been pointed out (Klevit 1991, Berg 1992), these did not provide rules which could explain DNA-binding specificity of Zn fingers generally. On the contrary, in a recent paper, Pavletich and Pabo (1993) expressed scepticism as to whether such rules exist. They based this discussion on the finding that not all fingers bind to DNA in the same way and that some fingers do not even bind to DNA strongly.

In this paper, we show that it is possible to deduce consistent rules for DNA-recognition by Zn fingers based on the known crystal structures.

2. ANALYSIS OF THE CRYSTAL STRUCTURES

We discuss in this section features found in the three crystal structures, that have not been discussed before.

2.1 'Good' and 'bad' fingers

The two fingers in TTK and the three fingers in Zif bind to DNA bases (Figure 1). GLI has five fingers: Finger 4 [F4] and finger 5 [F5] bind to DNA strongly. However, F1 and F3 have no residues contacting the DNA bases, and F2 has only one such residue (Figure 1). This is in one sense expected since some fingers are believed to bind to DNA only weakly (Zarkower and Hodgkin, 1992, Delwel *et al.*,1993, see also discussion on TFIIIA in **3.5**). It is therefore important to find out what makes a DNA–finger interaction specific.

Four amino acid positions are commonly used for base recognition among the fingers of TTK and Zif; 2, 3, and 6 in the recognition helix and −1 which is placed in a short linker connecting the helix to a β-sheet (Figure 1).

We argue that a 'good' or 'specific' finger has smaller residues at the N-terminus of the helix, and larger residues at the C-terminus. Such a configuration matches the binding-geometry, in which the N-terminus of the recognition helix is closer to the DNA than the C-terminus (Figure 2d). In the crystals large residues, such as Arg and Lys, at position 6, the position farthest from the DNA can reach a DNA base but small residues, Ala and Thr, cannot (Figure 1). Amino acid residues can be classified into four groups according to the shapes of their sidechains: small, medium, large, and aromatic (Suzuki,1994 and see Figure 3a of this paper). Aromatic residues have distinctive shapes but may often be included in the large group. The position −1 is not inside the helix and the larger the residue which occupies this position the better.

Therefore, we suggest that a 'very good' finger has a large residue at −1, small/medium at 2, medium/large at 3, and large

*To whom correspondence should be addressed

at 6. 'Bad' or 'non-specific' fingers have at least two wrong residues which do not meet this description.

2.2 Type A and B fingers

The fingers in TTK and Zif, and F5 of GLI all recognise DNA with a very similar geometry, while F4 of GLI binds to DNA in a very different geometry (Figures 2d, 4 and see **2.3**). We call the two DNA-binding modes: A (TTK, Zif, and GLI F5) and B (GLI F4).

A recognition helix in the A mode binds predominantly to bases on one DNA strand (the Watson strand) (Figure 2e), while that in the B mode binds to the other DNA strand (the Crick strand) (Figure 2f).

The A and B geometries seem to be fixed by placing phosphate-binding residues on different β-strands (Figure 1): If β-strand 2 is designed so that it binds to phosphates on the Watson strand and if β-strand 1 is designed so that it does not bind to phosphates on the Crick strand, the finger behaves as A (Figure 2b). Alternatively, if β-strand 1 binds to the Crick strand, while β-strand 2 does not bind to the Watson strand, the finger behaves as B (Figure 2c).

Four positions appear to be important for judging the mode of a finger: two positions on β-strand 2 (β21 and β23), one position on β-strand 1 (β11) and one position on the helix (position 4) (Figure 2). If one of these positions is occupied by a hydrogen-bond donor such as Lys/Arg or Tyr, it may be used for phosphate-binding, while if it is occupied by strictly hydrophobic residue such as Phe, it can not (Figure 1a).

Any finger can be placed somewhere between an 'ideal' A and an 'ideal' B finger (**3.4**) and act as the closest A/B type. The first His, which is required for Zn-binding and is always found at helix position 7, can be used to bind to a phosphate on the Watson strand but not to the Crick strand (Figures 1,2b) and therefore a finger seems to be biased to act as A.

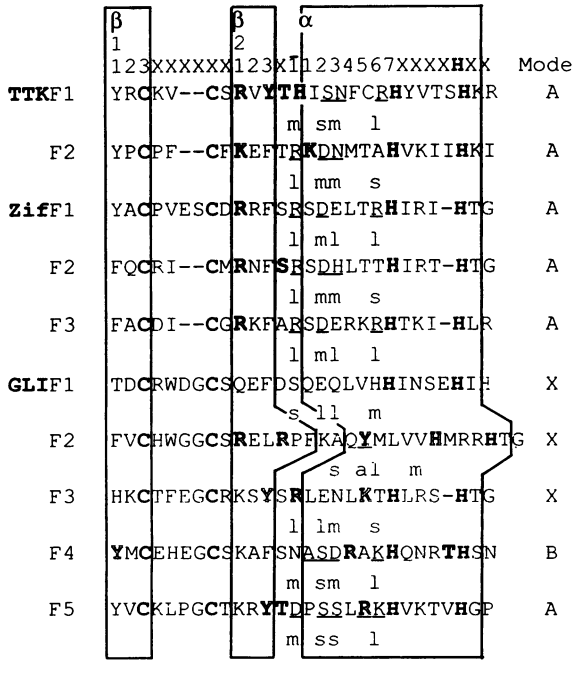


Figure 1. Sequences of the crystallised Zn fingers. The residues interacting with a DNA phosphate are outlined, while those interacting with a DNA base are underlined. Also shown are: the DNA-binding modes, A, B, and non-specific (X); the sizes of residues, small (s), medium (m), large (l), and aromatic (a); and the positions of the two β-strands and the recognition helix.

2.3 Amino acid and base positions contacted

The amino acid and DNA base positions contacted are well conserved among the A fingers (Figure 4). Rules which describe the positions and sizes of residues used for each of the contacts are a consequence of the conserved binding-geometry (Figure 2d) and can be summarised in the form of a chart (Figure 3b). Briefly the stereochemical rules are (see Figure 3b for the numbering of bases):

a large residue at position −1 contacts base W1,
a small/medium at 2, C0 or W1,
a medium/large at 3, W2, and
a large at 6, W3.

In the crystal structure Glu (large) at position 3 of Zif F1 and F3 are very close to C[W2] but they do not make a hydrogen bond. However, as will be shown later in **3.2**, Glu occurs at position 3 almost always with its specific partner C or A at W2 (see **3.1**) and therefore it seems reasonable to assume that their interaction contributes to specificity.

The protein−DNA contacts found for the single B finger (GLI F4, Figure 4g) are not sufficient for generalising the stereochemical rules of B fingers. However some other fingers appear to use a very similar binding-geometry (Figure 4), which are summarised in Figure 3c.

GLI F5 is essentially an A finger (Figure 4g). However its binding-geometry is not that of a standard A finger. The slight differences may be caused by the connection towards the preceding B finger; thus, the binding-geometry of a finger may be affected by its connections with neighboring fingers (see **2.4**).

2.4 Spacing between fingers along the DNA

A recognition helix in the A mode is more radial with respect to the DNA axis than one in the B mode (Figure 2d). As a consequence, the number of base pairs covered by a B finger is larger than that by an A finger. Also a connection between B and A fingers covers a few more base pairs than is needed between two A fingers (Figure 5). Thus, the spacing between two neighboring fingers along the DNA is dependent on the types of the fingers.

An A finger covers four base pairs (base pairs 0 to 3) and is positioned every three base pairs along DNA by the sharing of one base pair between two neighboring fingers; that is, base pair 3 for one finger is simultaneously base pair 0 for the following finger (Figures 5a,b). In contrast, a B finger covers five base pairs (base pairs −1 to 3) (Figure 5c) and is likely to be positioned every four base pairs along DNA (Figure 5d, see also **3.4**).

3. EXAMINATION OF OTHER Zn FINGERS

There are number of other Zn finger proteins well characterised by biochemical techniques, such as foot-printing and PCR. In this section, we show that the DNA recognition of these proteins can be understood by the same principles described in section 2. We use the following strategy: (1) identify a 'good' finger, (2) determine its A/B mode, (3) predict its DNA-binding sequence, (4) compare the predicted binding sequence with

experimental data, and (5), if necessary, consider the binding specificity of less specific fingers neighboring the good finger.

3.1 Chemical code

It is essential to understand the possible specific contacts between amino acid sidechains and DNA bases (Suzuki, 1994). A contact between a sidechain and a base is achieved by either a hydrophobic interaction or a hydrogen-bond (Seeman *et al.*, 1976, Suzuki, 1994). A list of possible contacts, which we call the chemical code, is shown in Figure 3a.

Some sidechains can bind to only one or two of the four bases and thus such a contact is very specific. For instance, Ala has

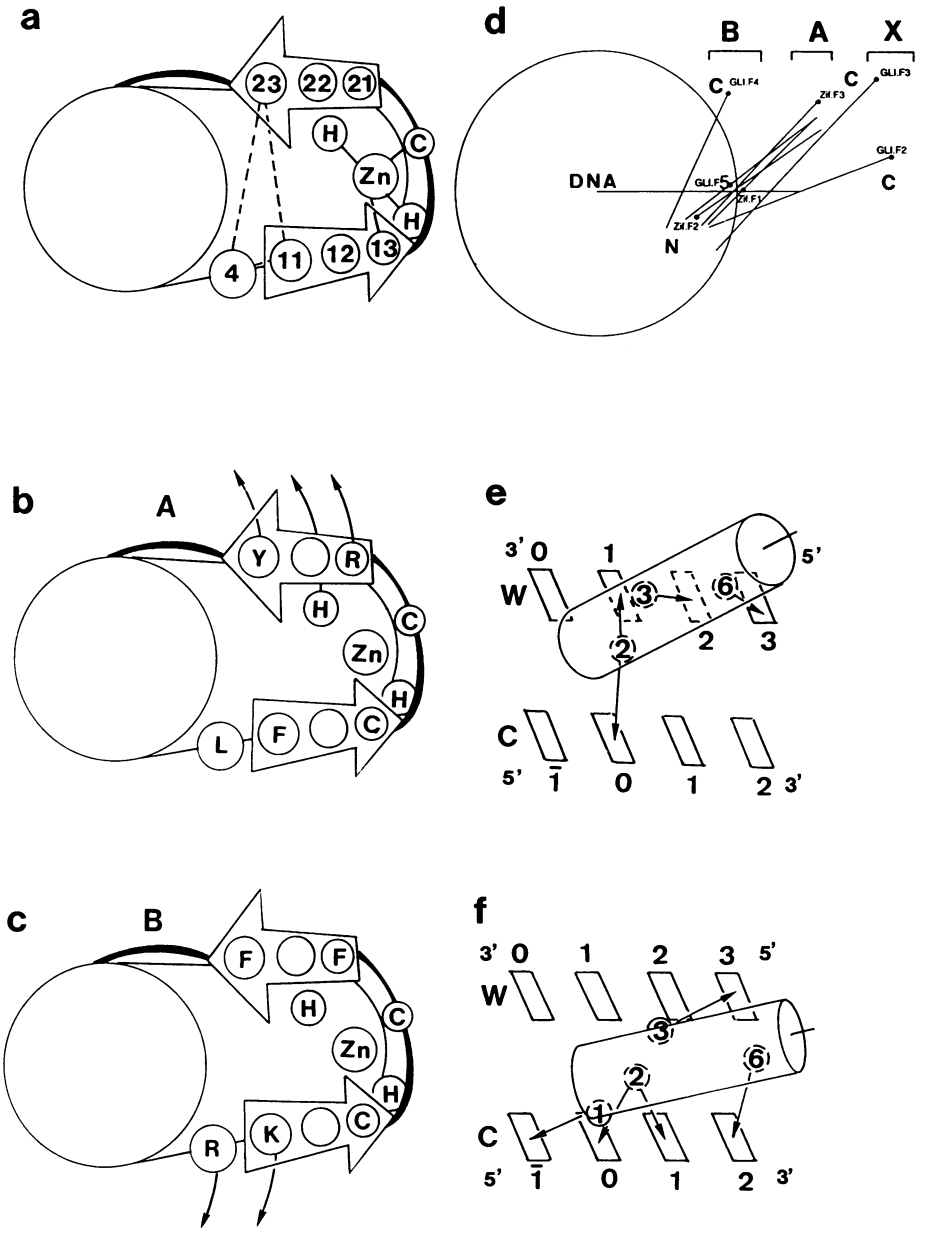


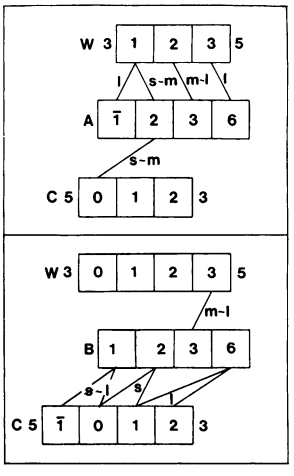
Figure 2. Orientation of recognition helix in the A (b,e) and B (c,f) modes. (**a**) shows the structure of a Zn finger schematically. An anti-parallel β-sheet packs against the recognition helix through the hydrophobic interaction between β11, β23 and α4 and by binding to the same Zn ion through two His and two Cys residues (one of the Cys residues is positioned at β13). (**b**) and (c) show ideal A and B fingers, respectively. If β-strand 1 has basic residues and β-strand 2 has hydrophobic residues, the finger binds to the Watson strand in the A mode. In contrast, if basic residues are positioned on β-strand 2 and at α4, the finger binds to the Crick strand in the B mode. The arrows show the direction towards the phosphates contacted. (**d**) shows the orientation of the recognition helices in the crystal structures in the A (Zif F1,F2,F3, GLI F5), B (GLI F4) and non-specific (GLI F2,F3) modes, viewed down the helix axis of DNA. These were calculated using Cα of position 4 to define the centre of the helix. The DNA axis is defined locally by the phosphate groups of the closest three to four base pairs. 'N' indicates the N-termini of the recognition helices. The apparent differences in the length of helices are due to differences in the angle of the helices from the plane on which they are projected; always thirteen residues (see Fig. 1) are used to draw the helices. The names of the fingers are shown near the filled circles on the corresponding helices. (**e**) and (**f**) show the positions of the contacting protein residues and DNA bases in the A and B(GLI F4) modes, respectively.

a methyl group and bind only to T, the single base which has a hydrophobic group in the major group. On the other hand, the binding specificities of other residues, such as Ser, are weak and therefore less important for the following discussion.

a

aa	base	specificity
Ala(s)	T	++
Val(m) Leu(l) Ile(m) Phe(a)	T/C	+
Met(l)	T/C≥A	±
Trp(a) Tyr(a)	T/C>A/G	±
Asp(m) Glu(l)	C≥A	+
Asn(m) Gln(l)	A>T/G/C	+
Cys(s) Ser(s) Thr(s)	A/T/G/C	
Arg(l) Lys(l)	G>>T≥A	++
His(m)	G≥T≥A≥C	±

b



c

Figure 3. Chemical and stereochemical rules. (**a**) Chemical rules that list the DNA-binding specificity of each amino acid (see Suzuki, 1994, for details). The sizes of the residues are also shown: small(s), medium(m), large(l) and aromatic(a). (**b**) and (**c**) Stereochemical rules which describe the residue and base positions contacted in the A and B modes. The sizes of the residues used for the contacts are also shown.

3.2 'A' fingers using several Arg residues

Some finger proteins, which have two or three A fingers, are predicted to use several Arg or Lys residues for base-recognition (Figure 6a). The DNA-binding specificity of these proteins is easy to understand, since Arg residue most likely binds to the G base (Klevit, 1991, Berg, 1992). However, for understanding the interactions fully, consideration of other types of contacts is necessary. For example, Glu (aa3) of WT1, SP1 and Krox20 binds to C or A at W2.

Some proteins have more than three fingers but not all of them are 'very good'. MAZ has five fingers, but three of them (F1–F3) appear to be sufficient to explain the binding-specificity of the protein (Figure 6a).

3.3 'A' fingers not using many Arg residues

Some fingers do not possess Arg at a base-contacting position. By using the chemical code table, it is possible to discuss the amino acid-DNA base contacts for these proteins (Figure 7a).

For instance, hunchback binds to A/T-rich sequences and a theory is needed to explain such specificity (Berg, 1992). We account for its specificity in terms of the binding specificity of fingers 2 to 4, which bind in the A mode and make specific contacts to T bases from Tyr, Phe, Met, and Leu (Figures 6b, 4n–p).

3.4 'B' fingers

F4 of GLI (Kinzler and Vogelstein, 1990) and a related protein Tra1 (Zarkower and Hodgkin, 1991,1993) behave as B fingers.

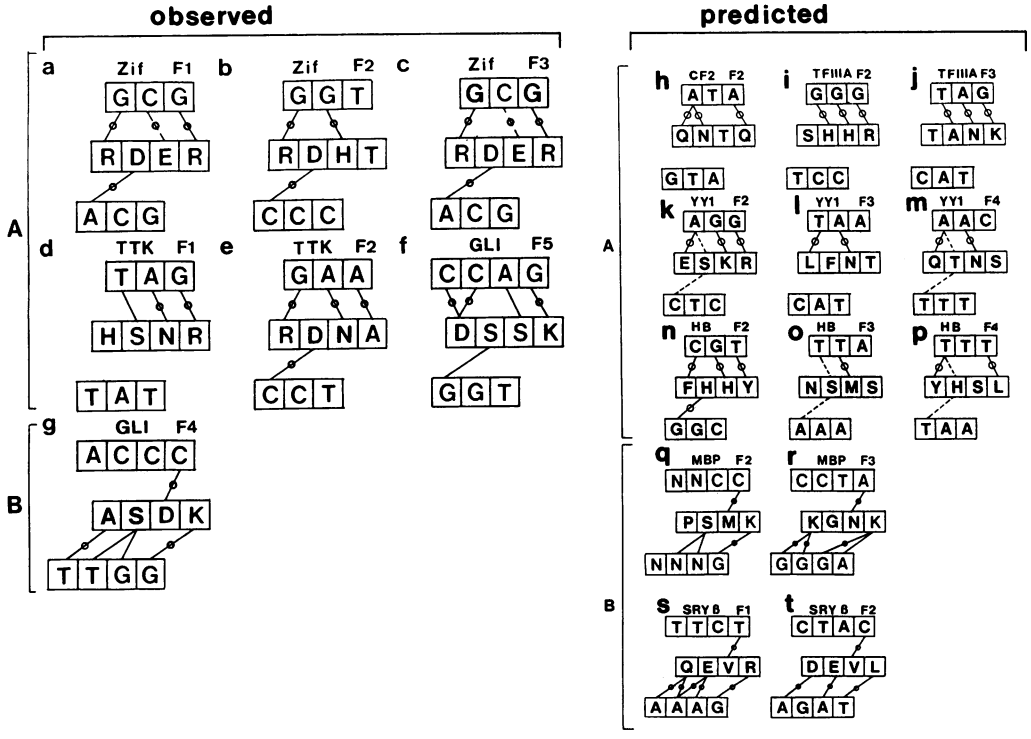


Figure 4. Contacts observed (a–g) and predicted (h–t) between amino acid residues and DNA bases. Those in A and B modes are shown. The charts are drawn in the same way as in Figures 3b,c. The lines show contacts, and the circles indicate specific contacts (see Figure 3a). In some of the predictions an ambiguity is seen as aa2 can bind to two base positions (shown with broken lines). See following references for the MBP family; MBP1 (Baldwin, 1990, Maekawa *et al.*, 1989), MBP2 (Van 'T Veer *et al.*, 1992), PRDII (Fan and Maniatis, 1990, Nakamura *et al.*, 1990), KBF1 (Henseling *et al.*, 1990, Rustgi *et al.*,1990), Rc (Wu *et al.*, 1993), AGIE-BP (Ron *et al.*,1991), ATBP (Mitchelmore *et al.*, 1990).

Likewise, MBP1 and related proteins have two fingers and bind to essentially the same DNA sequences (see references in caption to Figure 4). One of the two fingers has two Phe residues on strand 2 (at positions β21 and β23), the other has Ile and Cys. This may indicate that these fingers are of the B type. Indeed, the binding-specificity of the fingers is not explained by the contacting profiles of the A mode but by those similar to that of GLI F4 (Figures 4q,r).

The binding specificity of Sryβ (Vincent *et al.*, 1985, Payre and Vincent, 1991) can be explained by that of two of the five fingers, F1 and F2. These two fingers again appear to be in the B mode (Figures 4s,t) as F1 has Glu and Phe at β21 and β23, respectively, and F2, Leu and Val. The Glu in F1, which is repulsive to DNA phosphates, at the position which is usually occupied by Lys or Arg in the A mode, making it particularly unlikely that F1 binds in A.

In figure 7 we list all the finger sequences discussed in this paper from those very 'A-like' to those very 'B-like.' By comparing a new finger sequence with those in the list, the mode of finger may be determined. There is a gray area in the middle of the list (marked AB), in which the A and B characters are not clear cut.

Sryβ F2, for instance, has many hydrophobic residues on its β-strands and therefore may not bind to either of the DNA strands strongly. The behavior of such an ambivalent finger may well depend on fingers neighboring it or by the features of the linker connecting two fingers (Choo and Klug, 1993).

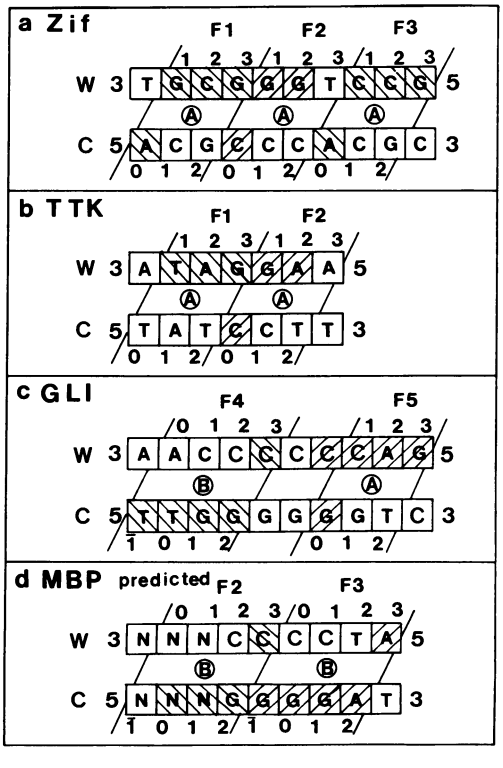


Figure 5. Spacing between Zn fingers. The spacing between between A fingers (**a** and **b**) and that between A and B fingers (**c**) observed in the crystal structures are shown. The spacing between B fingers, which is predicted for MBP1, is shown in (**d**). The bases shadowed are those contacted by amino acid sidechains.

3.5 DNA-binding of multiple finger proteins

Some proteins have many fingers. Following the strategy of the previous sections, the DNA-binding specificity of these proteins can be explained by that of a small number of fingers (Figure 8b).

The DNA-binding of TFIIIA, extensively studied by footprinting experiments (Fairall *et al.*, 1986, Churchill *et al.*, 1990, Christensen *et al.*, 1991, Fairall *et al.*, 1992, Liao *et al.*, 1992, Hayes and Tullius, 1992, Hansen *et al.*, 1993), can be explained very well by the chemical and stereochemical rules (see legend for Figure 9). It is predicted here that with one helical turn of nine fingers, TFIIIA wraps around 3 turns of DNA. F2, F3 and F5 bind to DNA in a specific fashion and are most important for the overall recognition.

4. PREDICTION OF Zn-FINGER BINDING SITES IN REGULATORY SEQUENCES

We have designed a computer program that can predict contacts between amino acid-DNA base contacts according to the chemical and stereochemical rules summarised in Figure 3 (Suzuki and Yagi, 1994b, see also Suzuki and Chothia, 1994, Suzuki and Yagi, 1994a). In figure 9 we show examples of such calculation for TTK, TFIIIA, and Hunchback.

The predicted sites are consistent well with the experimentally identified binding sites. DNA-recognition by a Zn-finger can be understood in terms of the chemical and stereochemical rules.

ACKNOWLEDGEMENTS

We thank Prof. C.O.Pabo for the co-ordinates of GLI. We thank Drs A.Klug, J.Finch, and C.Chothia for their critical reading of the manuscript and for useful comments.

REFERENCES

1. Baldwin Jr, A.S., LeClair, K.P., Singh, H., and Sharp, P.A., 1990, Mol. Cell Biol. 10, 1406–1414.
2. Berg, J.M., 1992, Proc. Natl. Acad. Sci. USA 89, 11109–11110.
3. Blumberg, H., Eisen, A., Sledziewski, A., Bader, D., and Young, E.T., 1987, Nature, 328, 443–445.
4. Bossone, S.A., Asselin, C., Patel, A.J., and Marcu, K.B., 1992, Proc. Natl Acad. Sci. USA 89, 7452–7456.
5. Call, K.M., Glaser, T., Ito, C.Y., Buckler, A.J., Pelletier, J., Haber, D.A.,Rose, E.A., Kral, A., Yeger, H., Lewis, W.H., Jones, C., and Housman, D.E., 1990, Cell 60, 509–520.
6. Choo, Y., and Klug, A., 1993, Nucl. Acids Res. 21, 3341–3346.
7. Christensen, J.H., Hansen, P.K., Lillelund, O., Thøgersen, H.C., 1991, FEBS Letters 281, 181–184.
8. Christy, B. and Nathans, D., 1989, Proc. Natl. Acad. Sci. USA 86, 8737–8741.
9. Churchill, M.E.A., Tullius, T.D., and Klug, A., 1990, Proc. Natl. Acad. Sci. USA 87, 5528–5532.
10. Delwel, R., Funabaki, T., Jreider, B.L., Morishita, K., and Ihle, J.N., 1993, Mol. Cell Biol. 13, 4291–4300.
11. Desjarlais, J.R., and Berg, J.M., 1992, Proteins 12, 101–104.
12. Desjarlais, J.R., and Berg, J.M., 1993, Proc. Natl. Acad. Sci. USA 90, 2256–2260.
13. Ebright, R.H., Cossart, P., Gicquel-Sanzey, B., and Beckwith, J., 1984, Proc. Natl. Acad. Sci. USA, 81, 7274–7278.
14. Fairall. L., and Rhodes, D., 1992, Nucleic Acids Res. 20, 4727–4731.
15. Fairall, L., Harrison, S.D., Travers, A.A. and Rhodes, D., 1992, J.Mol. Biol 226, 349–366.
16. Fairall, L., Rhodes, D., and Klug, A., 1986, J. Mol. Biol. 192, 577–591.
17. Fairall, L., Schwabe, J.W.R., Chapman, L., Finch, J., and Rhodes, D., 1993, Nature 366, 483–487.

a

Name		$\bar{1}$ 23 6	Cl	Predicted	Observed	Reference
MIG1	F1	R EH R		GGG	GGG/GGG	Nehlin and Ronne,1990
	F2	R DE R	O	GC(A)G	GCG/GCC	Nehlin et al., 1991
TTK						Fairall et al.,1992a,1993
	F1	H SN R		XAG	TAG/AAG/CAG/AGG/AAG/GCG/AAG	
	F2	R DN A		GAX	GAA/GAT/GAC/GAT/GAC/GAT/GAT	
ADR1	F1	R EH R		GGG	GAG	Blumberg et al.,1987
	F2	R DL R	C	GT(C)G	GTT	Thukral et al.,1991
Krox20						
	F1	R DE R	O	GC(A)G	GTG/GAG/GAG	Nardelli et al.,1991
	F2	R DH T		GGX	GGG/GGG/GGG	Nardelli et al.,1992
	F3	R DE R	O	GC(A)G	GCG/GAG/GCG	Sham et al.,1993
Zif	F1	R DE R	O	GC(A)G	GCG/GCG/GCG	Christy & Nathans,1989
	F2	R DH T		GGX	GGT/GGG/GGC	Lemaire et al.,1990
	F3	R DE R	O	GC(A)G	GCG/GCG/GCG	Hung et al.,1993
Sp1						Letovsky & Dynan,1989 Kriwacki et al.,1992
	F1	K SH A		GGX	GGG/GGG/GAG/GAG/GGT/AGG/GGT	
	F2	R DE R	O	GC(A)G	GCG/GCG/GCG/GCG/GAG/GCG/GCG	
	F3	R DH K	O	GGG	GGG/GGT/GGT/GGG/GGG/GGT/GAG	
EKLF	F1	K SH A		GGX	GGT	Miller & Bieker, 1993
	F2	R DE R	O	GC(A)G	GTG	
	F3	R SR D		GGX	GGA	
SWI5	F1	R YN S	X	GAX	GGT/TTG	Nagai et al,1988
	F2	R HD R	O	GC(A)G	GCG/TCG	Tebb et al.,1993
	F3	R DA V	X	GXX	GTA/TAC	
WT1	F1	K SH M	C	GGT		Call et al.,1990
	F2	R DQ R	C	GAG		Rauscher et al.,1990
	F3	R NH T		GGX		Madden et al.,1991
	F4	R DE R	C	GC(A)G		Pelletier et al.,1991
observed						
1 GGG/GGG/GGG/GGT/GGG/GGG or						
2 GCG/GAG/GAG/GAG/GAG/GTC 2 GCG/CTG/GGC/GTG/ AC/GGA						
3 3 GGG/GGG/GGG/GGT/GGG/GGG						
4 4 GCG/GAG/GAG/GAG/GAG/GTC						
MAZ	F1	D YH R	X	XGG	GGG/AGG/CGG	Bossone et al.,1992
	F2	R DR Y	C	GGT(C)	GGA/GGA/GAC	Pyrc et al.,1992
	F3	R DH S		GGX	GGG/GGG/GGG	Moberg et al.,1992
	F4	T DR A	X			
	F5	S AY D	X			

b

Name	$\bar{1}$ 23 6	Cl	Predicted	Observed	Reference
YY1					Natesan & Gilman,1993
F1	D SA K	X			Shi et al.,1991
F2	E SK R	O	C(A)GG	AGG/AGG/CGG/CGG/CTG/CGG/AGG/TGG/AGG	
F3	L FN T	X	T(C)AX	TAA/TAG/TAG/TAA/TAA/TAG/TAT/TAG/TAG	
F4	Q TN S		AAX	AAC/AAC/AAC/AAC/AAC/AAC/AA /AGG/GTA	
CF2					Shea et al.,1990
F1	T GT M	X	XXT(C)	CAT	
F2	Q NT Q		AXA	ATA	
F3	V DY K	X	XT(C)G	TCC	
F4	Q SA V	X	AXX	AGT	
Glass					Moses et al.,1989
F1	R ST T	X	GXX	GTA	Moses and Rubin,1991
F2	Q AN A		AAX	AAG	
F3	Q SS T	X	AXX	TTC	
F4	D ST K	X	XXG	CAA	
F5	Q GN R	O	AAG	AAG	
Krüppel					Schuh et al.,1986,Zuo et al.,1991 Parkratz et al.,1989,Stanojevic et al.,1989 Treisman and Desplan,1989
F1	Y HV N	T	(C)T(C)X	TTG/TTG/TTG/TAG/TTG/TTG/TTG/TTG/TTG	
F2	R HH T		GGX	GGC/GGA/GGT/GGA/GGC/GGT/GGG/GCG/GGC	
F3	Q AN R	O	AAG	AAT/AAA/CA /AG /CA /TA /CA /AA /AG	
F4	D NQ S	X			
F5	R HH N		GGX		
PRDI					Keller and Maniatis, 1991/1992
F1	Q SN V		AAX	AAA	
F2	Q AH K	O	AGG	GTG	
F3	S SN T	X	XAX	AAG	
F4	Q VH L	O	AGT(C)	AGG	
F5	H CS V	X	XXX		
Snail					Tony et al.,1992 Kasai et al.,1992
F1	T MG K	X			
F2	T GA M	X			
F3	R WL G	X			
F4	D SN A	X	XAX	TAC/AAC/GAA/GAA/GAC/GAC/GAC/CAC/CAC/TCT/CAA	
F5	R SL K	O	GT(C)G	GTG/GTG/GTG/GTT/GTG/ATG/TTG/GTT/GTC/TTG/GTC	
Hunchback					Treisman and Desplan,1989 Zuo et al.,1991
F1	T VD A	X			
F2	F HH Y	O	T(C)GT(C)	CGT/CGG/CGG/TGA/CCT/CGA/CGG/TGT/TGC	
F3	N SM S	X	XT(C)X	TTA/AAC/TTT/TTT/TTT/ATT/GTT/TTT/GTT	
F4	Y HS L		T(C)XT(C)	TTT/TTT/TTT/TTT/TTT/TTT/TTT/TTA/TTG	
				F2 GGT/CGT/AGT/TGA/GTT/CGT/TGT/TGT/TTA	
				F3 TTT/TTT/ATT/TTT/ATT/TTT/TTC/TTT/ATT	
				F4 TTT/TTT/TTT/TTA/TTT/TTA/TT /TT /TTT	
F5	D VL I	O	XT(C)T(C)		
F6	G VG V	X			
SRYδ					Vincent et al.,1985
F1	S YQ K	X			Payre and Vincent,1991
F2	D EY L	X	XTX	GTA/GTA/GTT/GTA/GTA	
F3	R VN R	O	GAG	GAG/GAG/GAG/GAG/GAG	
F4	Q NL N		AT(C)X	ATC/ATT/ATA/AAA/ATT	
F5	S KT H	X			
F6	E YT M	X			
F7	N KE H	X			

Figure 6. Binding specificty predicted for Zn fingers, which use several Arg residues for base-recognition (**a**) and which do not (**b**). 'Very good' (O) and 'bad' (X) finger are indicated in the class column (Cl). Predicted DNA binding sequences are compared with sequences found in the experimentally identified binding regions. The DNA sequences shown are those of W1−W3 written from 3′ to 5′. The predictions are given by contacts from residue positions −1, 3, and 6 using the rules of the A mode in Fig. 3. Helix position 2 can contact two DNA positions, so there is some ambiguity. The DNA bases (observed) shown in bold are the bases same as predicted as contacted by specific partner residues. The bases underlined are those inconsistent with the prediction. The bases in plain are those predicted to be not contacted, contacted but not specified, or contacted and consistent with the binding specificty of the amino acid residue but not the most specific candidate. Foot-printing and base-modification experiments are useful to identify the approximate position of a protein on the DNA. However, these are not always specific enough to pin-point the base contacted by an amino acid sidechain. Interpretation of the results may not be always easy and the results may depend on the details of the experiment; for example, slight differences can be seen in the two independent methylation protection experiments of CAP binding (see Figure 1 of Ebright *et al.*, 1984 and compare the two profiles with the crystal structure in Schultz *et al.*, 1991). Also, if a whole protein is used instead of its fingers for the experiment, interpretation becomes more difficult as some additional contacts may occur from outside the fingers. Therefore, there might be slight differences between our prediction shown here and experimental observation. Also it might be dangerous to conclude a protein's binding-specificity from a single or a small number of identified binding sites; slight deviations are seen among the identified binding sites of the same protein (Figures 6,8). In this study we tried to include as many Zn finger proteins as possible. The DNA-binding specificity of most of them can be explained by the rules very well. However we find two notable exceptions (see the following) which require further study for their understanding. The binding-specificity of SP1 has been extensively studied and it matches well with the rules; SP1 binds to G-rich sequences by using Arg/Lys residues. However, Zhu *et al.* (1993) have reported that HF1B which has almost the same three fingers as those of SP1, binds to an entirely different DNA sequence which is rich in A/T bases. This is quite puzzling and is inconsistent with any discussion so far published. EPF1 (Takatsuji *et al.*, 1992) has two fingers separated from each other. The two fingers, however, have 'wrong' types of residues at the base-recognition positions, small at −1, large at 4, small at 5, small at 8. Therefore, its DNA-binding specificity cannot be explained by the rules discussed in this paper. The rest of the protein might be important for the DNA-binding, or the fingers adopt a third binding geometry which is unknown at present.

18. Fan, C.-M., and Maniatis, T., 1990, Genes Develop. 4, 29−42.
19. Gibson, T.J., Postma, J.P., Brown, R.S., and Argos, P., 1988, Protein Engineering 2, 209−218.
20. Gorgos, J.A., Hsu, T., Bolton, J., and Kafatos, F.C., 1992, Science 257, 1951−1955.
21. Hansen, P.K., Christensen, J.H., Nyborg, J., Lillelund, O., and Thøgersen, H.C., 1993, J. Mol. Biol. 233, 191−202.
22. Hartshorne, T.A., Blumberg, H., and Young, E.T., 1986, Nature 320, 283−287.
23. Hayes, J., and Tullius, T.D., 1992, J. Mol. Biol. 227, 407−417.

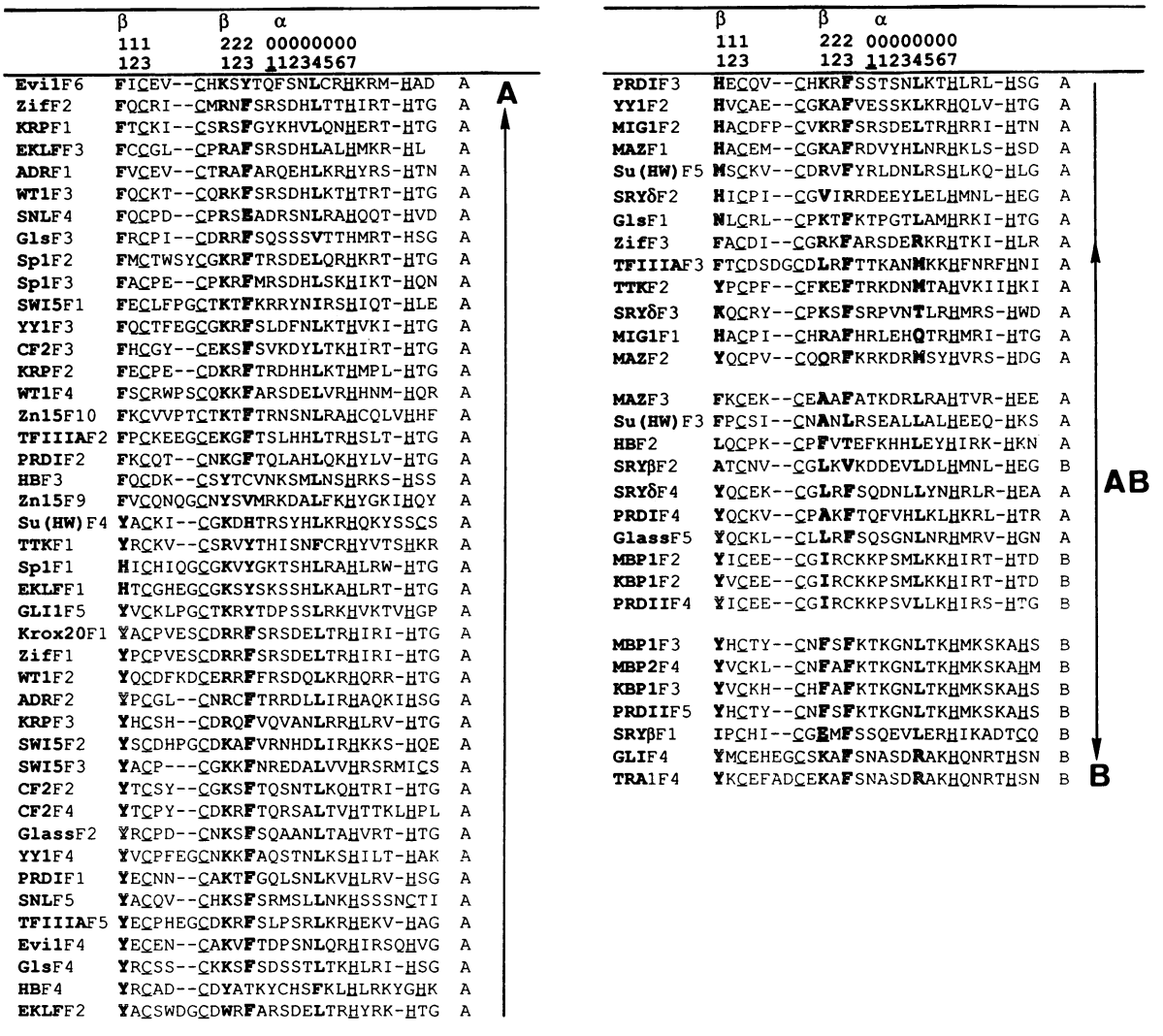


Figure 7. Sequences of Zn fingers arranged from A-like to B-like. The first group, Evi1F6-MAZF2, are predicted to be A, the third group, MBPF3-TRA1F4, B, while, the second group, MAZF3-PRDIIF4, a mixture of A and B (marked AB).

24. Henseling, U., Schnidt, W., Schöler, H.R., Gruss, P., and Hatzopoulos, A.K., 1990, Mol. Cell Biol. 10, 4100–4109.
25. Jacobs, G.H., 1992, EMBO J. 11, 4507–4517.
26. Kasai, Y., Nambu, J.R., Lieberman, P.M., and Crews, S.T., 1992, Proc. Natl. Acad. Sci. USA 89, 3414–3418.
27. Keller, A., and Maniatis, T., 1991, Genes Develop. 5, 868–879.
28. Keller, A., and Maniatis, T., 1992, Mol. Cell Biol. 12, 1940–1949.
29. Kinzler, K.W., and Vogelstein, B., 1990, Mol. Cell Biol. 10, 634–642.
30. Klevit, R.E., 1991, Science 253, 1367–1393.
31. Klevit, R.E., Herriot, L.R., and Horvath, S.J., 1990, Proteins 7, 215–226.
32. Kochoyan, M., Keutmann, H.T., and Weiss, M.A., 1991, Biochemistry 30, 7063–7072.
33. Kriwacki, R.W., Schultz, S.C., Steitz, T.A., and Caradonna, J.P., 1992, Proc. Natl. Acad. Sci. USA 89, 9759–9763.
34. Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A., and Wright, P.E., 1989, Science 245, 635–637.
35. Lemaire, P., Vesque, C., Schmitt, J., Stunnenberg, H., Frank, R., and Charnay, P., 1990, Mol. Cell Biol. 10, 3456–3467.
36. Letovsky, J., and Dynan, W.S., 1988, Nucleic Acids Res. 17, 2639–2653.
37. Liao, X., Clemens, K.R., Tennant, L., Wright, P.E., and Gottesfeld, J.M., 1992, J. Mol. Biol. 223, 857–871.
38. Lipkin, S.M., Näär, A.M., Kalla, K.A., Sack, R.A., and Rosenfeld, M.G., 1993, Genes Develop 7, 1674–1687.
39. Madden, S.L., Cook, D.M., Morris, J.F., Gashler, A., Sukhatme, V.P., and Rauscher III, F.J., 1991, Science 253, 1550–1553.
40. Maekawa, T., Sakura, H., Sudo, T., and Ishi, S., 1989, J. Biol. Chem. 264,14591–14593.
41. Miller, I.J., and Bieker, J. J., 1993, Mol. Cell Biol. 13, 2776–2786.
42. Miller, J., McLachlan, A.D., and Klug, A., 1985, EMBO J. 4, 1609–1614.
43. Mitchelmore, C., Traboni, C., and Cortese, R., 1990, Nucl. Acids Res. 19, 141–147.
44. Moberg, K.H., Logan, T.J., Tyndall. W.A., and Hall, D.J., 1992, Oncogene 7, 411–421.
45. Morishita, K., Parker, D.S., Mucenski, M.L., Jenkins, N.A., Copeland, N.G., and Ihle, J.N., 1988, Cell 54, 831–840.
46. Moses, K., Ellis, M.C., and Rubin, G.M., 1989, Nature 340, 531–536.
47. Moses, K., and Rubin, G.M., 1991, Genes Develop 5, 583–593.
48. Nagai, K., Nakaseko, Y., Nasmyth, k., and Rhodes, D., 1988, Nature 332, 284–286.
49. Nakamura, T., Donovan, D.M., Hamada, K., Norman, C.M.S.B., Flanagan, J.R., Ozato, K., Westphal, H., and Piatigorsky, J., 1990, Mol.Cell Biol. 10, 3700–3708.
50. Nardelli, J., Gibson, T.J., Vesque, C., and Charnsy, P., 1991, Nature 349, 175–178.
51. Nardelli, J., Gibson, T.J., Vesque, C., and Charnsy, P., 1992, Nucleic Acids Res. 20, 4137–4144.
52. Natesan, S. and Gilman, M.Z., 1993, Genes and Develop. 7, 2497–2509.
53. Nehrin,J.O., and Ronne, H., 1990, EMBO J. 9, 2891–2898.
54. Nehrin,J.O., Carlberg, M., and Ronne, H., 1991, EMBO J. 10, 3373–3377.
55. Nguyen, H.Q., Hoffman-Lieberman, B., and Liebermann, D,A., 1993, Cell

a

Name	-1 2 3 6	Cl	Predicted	Observed	Reference
TFIIIA					
F1	K WK A	X			(see text)
F2	S HH R		GGG	GGG	
F3	T AN K		TAG	TAG	
F4	K NQ V				
F5	L SR R	O	T(C)GG	CGG	
F6	T TL K				
F7	H DY D	X			
F8	T FN S	X			
F9	M KS R	X			
Evi1F1	S AE D	X			Morishita et al., 1988 Perkins et al.,1991
F2	D QS K	X			Delwel et al.,1993
F3	W SN R	O	TAG		
F4	D SN R		XAG	TAG/TAG/CAG/CAG	
F5	T SG Q	X	XXA	AA/ AA/ AA/ AA	
F6	Q SN R	O	AAG	TAG/CAG/TAG/CAG	
F7	T SS K	X			
F8	R AN R	O	GAG		
F9	I SN R	O	T(C)AG		
F10	Q TN R	O	AAG		
Su(HW)					
F1	R QS K	X			Parkhurst et al., 1988 Spana and Corces, 1990
F2	T TS R	X			
F3	S AL L		XT(C)T(C)	ATA/ATA/ATA/ATA/ATA	
F4	R YH R		CGG	CGT/CGT/CGT/CGT/CGT	
F5	R DN S		GAX	TAC/TGT/TAT/TGC/CGT	
F6	S PT I	X			
F7	A VA K	X			
F8	V EV R	X			
F9	R TQ T		GAX		
F10	T KQ R	X			
F11	T AL E		XT(C)C(A)		
F12	S DT R	X			
Zn15F1	L HR R	O	T(C)GG		Lipkin et al.,1993
F2	S ES P	X			
F3	Y KN A	X			
F4	S TH D	X			
F5	S AE S	X			
F6	S SE K		XC(A)G		
F7	S QS K	X			
F8	N RS G	X			
F9	R DA K		GXG	GAG/GGT	
F10	R SN A		GAX	GAC/GAC	
F11	I QN L				
F12	A TG Q	X			
F13	T LS V	X			
F14	T SN R		XAG		
F15	S SN R		XAG		

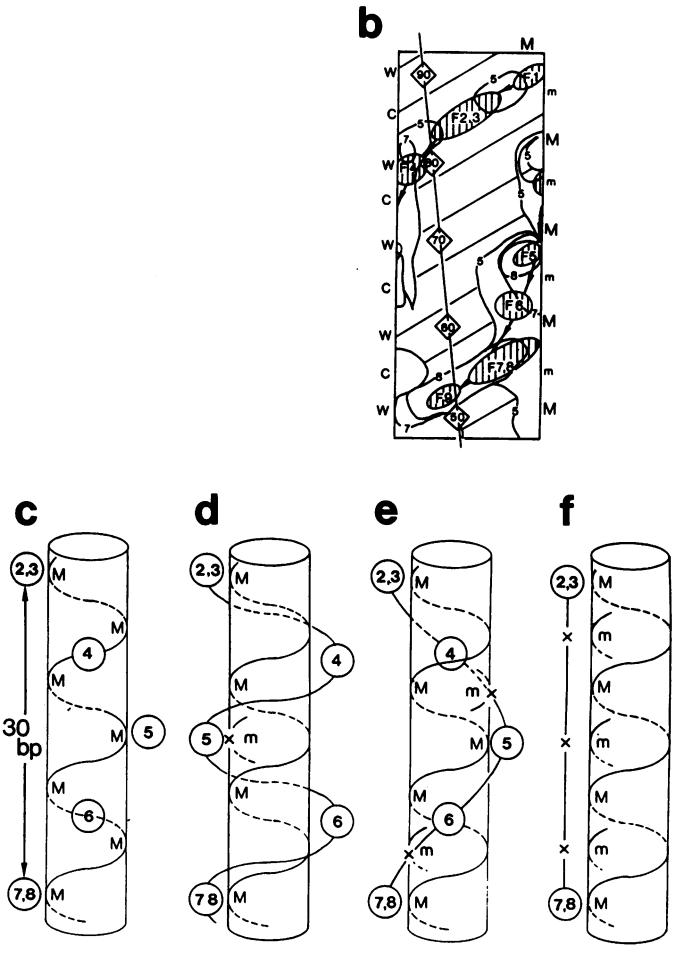


Figure 8. DNA-binding of multifinger proteins. (**a**) Binding specificity predicted for the fingers. The figure is drawn in the same way as Fig. 6. The space between lines indicates a break; Evi1, for example, has ten fingers grouped into three: F1, F2-F7, and F8-F10. (b)-(f) DNA-binding of TFIIIA. In (**b**) binding sites predicted for the fingers of TFIIIA (F2, F3 etc.) are projected onto the foot-printing contours (5–8) reported by Churchill *et al.* (1990). The base pair numbers are shown in diamonds. M and m: the major and the minor grooves, respectively. W and C: the Watson and the Crick strands, respectively. In (c)-(f) four different ways of wrapping TFIIIA around the DNA are shown. (**d**) shows the proposed mode. The DNA double helix makes three turns, while TFIIIA makes three turns in (b), two in (c), one in (d), and no turn in (e). The protein crosses the minor groove, not at all in (b), once in (c), twice in (d), and three times in (e) (shown with 'X'). It is predicted here that the N-terminal fingers, F2–F5, are better than fingers in the C-terminus, which agrees well with the experimental observation that a fragment containing F1–F3 is the core for DNA-binding (Liao *et al.*, 1992, Hansen *et al.*, 1993). Using the established N to C direction of TFIIIA on the DNA and why the foot-printing results (Miller *et al.*, 1985, Churchill *et al.*,1990, Liao *et al.*, 1992), the binding site of F5 is predicted as positioned about 15 base pairs from that of F2 and F3. A further 15bp from the F5-binding site another patche is protected in the experimental data probably in a less-specific way by the C-terminal fingers;

3′-A(90)GA<u>GGGTAG</u>GT(80)TCATGATTGG(70)TC<u>CGG</u>GCTGG(60)
F2,3 F5
GACGA<u>ACCGA(50)A</u>GGG-5′.
(F7-F9)

F6 and maybe F4 are used to cross over the minor groove twice. This DNA-binding mode of TFIIIA corresponds to (d) and is essentially consistent with previous proposals (Fairall and Rhodes, 1992, Hayes ad Tullius, 1992, Hansen *et al.*, 1993).

72, 197–209.
56. Neuhaus, D., Nakaseko, Y., Schwabe, J.W.R., and Klug, A., 1992, J. Mol., Biol. 228, 637–651.
57. Omichinski, J.G., Clore, M., Appella, E., Sakaguchi, K., and Gronenborn, A.M., 1990, Biochemistry 29, 9324–9334.
58. Omichinski, J.G., Clore, M., Robien, M., Sakaguchi, K., Appella, E., and Gronenborn, A.M., 1992, Biochemistry 31, 3907–3917.
59. Pankratz, M.J., Hoch, M., Seifert, E., and Jäckle, H., 1989, Nature 341, 337–339
60. Páraga, G., Horvath, S., Hood, L., Young, E.T., and Klevit, R.E., 1990, Proc. Natl. Acad. Sci. USA 87, 137–141.
61. Parkhurst , S.M., Harrison, D.A., Remington, M.P., Spana, C., Kelley, R.L., Coyne, R.S., and Corces, V.G., 1988, Genes Develop. 2, 1205–1215.
62. Pavletich, N.P., and Pabo, C.O., 1991, Science 252, 809–817.
63. Pavletich, N.P., and Pabo, C.O., 1993, Science 261, 1701–1707.
64. Payre, F., and Vincent, A., 1991, EMBO J. 10, 2533–2541.
65. Pelletier, J., Bruening, W., Kashtan, C.E., Mauer, S.M., Manivel, J.C., Striegel, J.E., Houghton, D.C., Junien, C., Habib, R., Fouser, L., Fine, R.N., Silverman, B.L., Haber, D.A., and Housman, D., 1991, Cell 67, 437–447.
66. Perkins, A.S., Fishel, R., Jenkins, N.A., and Copeland, N.G., 1991, Mol. Cell Biol. 11, 2665–2674.
67. Pyrc, J.J., Moberg, K.H., and Hall, D.J., 1992, Biochemistry 31, 4102–4110.
68. Rauscher III, F.J., Morris, J.F., Tournay, O.E., Cook, D.M., and Curran, T.,1990, Science 250, 1259–1262.
69. Ron, D., Brasier, A.R., and Habener, J.F., 1991, Mol. Cell Biol. 11, 2887–2895.

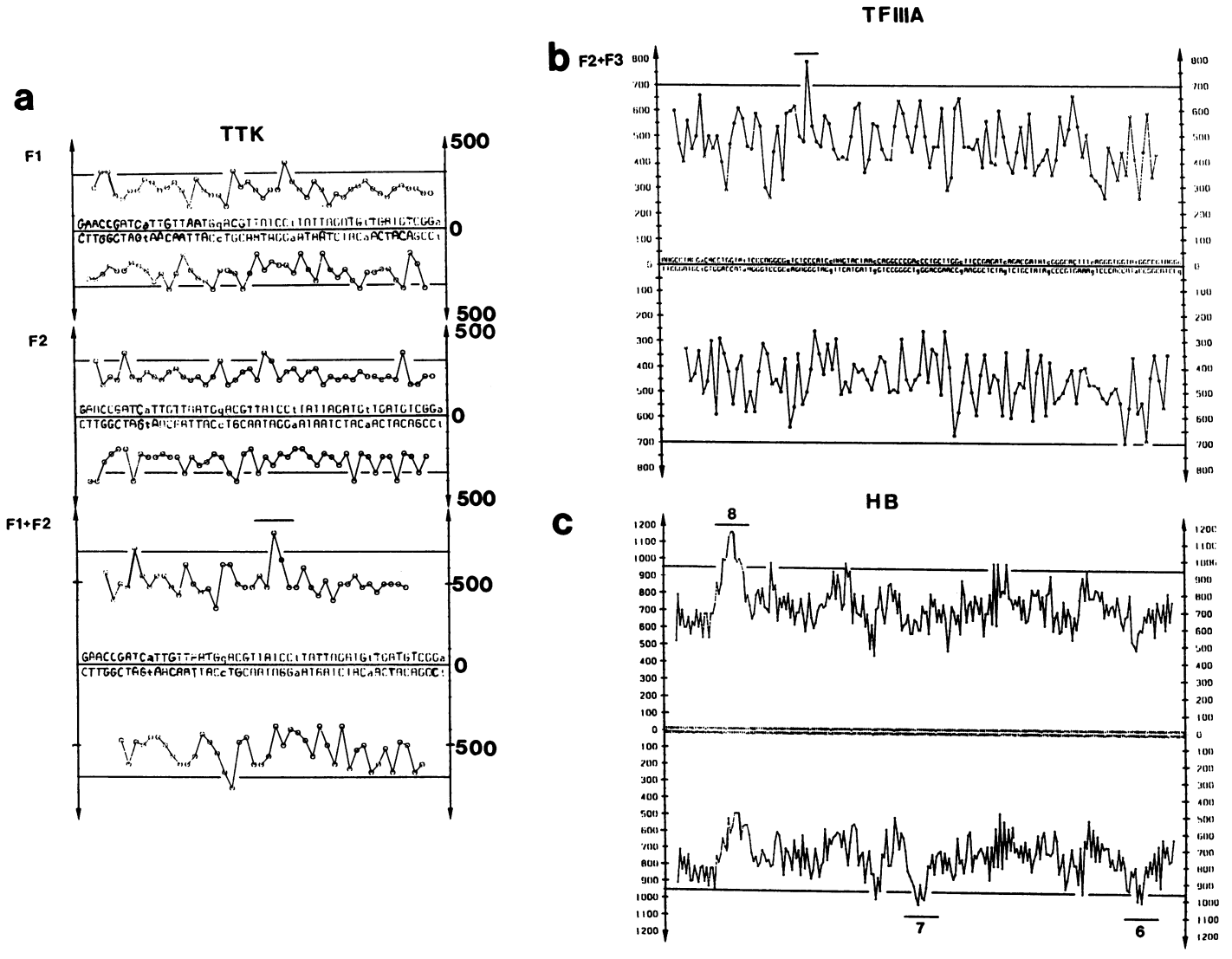


Figure 9. Prediction of finger-binding sites in regulatory DNA elements. That of F1 and F2 of TTK in the *fushi tarazu* promoter (**a**), F2 and F3 of TFIIIA in the 5SRNA promoter (**b**), and F2-F4 of Hunchback in the *hunchback* distal promoter (**c**) are shown. The binding sites identified experimentally (Fairall *et al.*, 1992, Liao *et al.*, 1992, Treisman and Desplan, 1989) are shown with bars. In (**a**) the prediction using F1 only and that using F2 only are also shown for comparison. Note that neither F1 nor F2 of TTK on their own are sufficient to specify the TTK binding site. However the two fingers in combination are sufficient.

70. Rosenfeld, R., and Margalit. H., 1993, J. Biomol. Struc. Dynamic. 11, 557–570.
71. Rustgi, A.K., van 'T Veer, L., and Bernards, R., 1990, Proc.Natl.Acad.Sci.USA 87, 8707–8710.
72. Schuh, R., Aicher, W., Gaul, U., Coté, S., Preiss, A., Maier, D., Seifert, E., Nauber, U., Schröder, C., Kemler, R., and Jäckle, H., 1986, Cell 47, 1025–1032.
73. Schultz, S.C., Schields, G.C., and Steitz, T.A., 1991, Science 253, 1001–1007.
74. Seeman, N.C., Rosenberg, J.M., amd Rich, A., 1976, Proc. Natl. Acad. Sci. USA 73, 803–808.
75. Sham, M.H., Vesque, C., Nonchev, S., Marshall, H., Frain, M., Gupta, R.D., Whiting, J., Wikinson, D., Charnay, P., and Krumlauf, R., 1993, Cell 72, 183–196.
76. Shea, M.J., King, D.L., Conboy, M.J., Mariani, B.D., and Kafatos, F.C., 1990, Genes Develop. 4, 1128–1140.
77. Shi, Y., Seto, E., Chang, L.-S., and Shenk, T., 1991, Cell 67, 377–388.
78. Spana, C., and Corces, V.G., 1990, Genes Develop. 4, 1505–1515.
79. Stanojevic, D., Hoey, T., and Levine, M., 1989, Nature 341, 331–335.
80. Suzuki, M., 1994, Structure 2, 317–327.
81. Suzuki, M., 1994, Proc. Japan Acad. B70, in press.
82. Suzuki, M., and Chothia, L., 1994, Proc. Japan Acad. B70, 58–61.
83. Suzuki, M., and Yagi, N., 1994a, Proc. Japan Acad. B70, 62–66.
84. Suzuki, M., and Yagi, N., 1994b, Proc. Natl. Acad. Sci. USA, in press.
85. Takatsugi, H., Mori, M., Benefey, P.N., Ren, L., and Chua, N.-H., 1992, EMBO J. 11, 241–249.
86. Tebb, G., Moll. T., Dowzer, C., and Nasmyth, K., 1993, Genes Develop. 7, 517–528.
87. Thukral, S.K., Eisen, A., and Young, E.T., 1991, Mol. Cell Biol. 11, 1566–1577.
88. Tony Ip, Y., Park, R.E., Kosman, D., Bier, E., and Levine, M., 1992, Genes and Develop. 6, 1728–1739.
89. Treisman, J., and Desplan, C., 1989, Nature 341, 335–337.
90. Van 'T Veer, L., Lutz, P.M., Isselbacher, K.J., and Bernarda, R., 1992, Proc.Natl.Acad.Sci.USA 89, 8971–8975.
91. Vincent, A., Colot, H.V., and Rosbash, M., 1985, J. Mol. Biol. 186, 149–166.
92. Wu, C., Mak, C.-H., Dear, N., Boehm, T., Foroni, L., and Rabbits, T.H., 1993, Nucleic Acids Res. 21, 5067–5073.
93. Zarkower, D., and Hodgkin, J., 1992, Cell 70, 237–249.
94. Zarkower, D., and Hodgkin, J., 1993, Nucl Acids Res. 21, 3691–3698.
95. Zhu, H., Nguyen, V.T.B., Brown, A.B., Pourhosseini, A., Garcia, A.V., Van Bilsen, M., and Chien, K.R., 1993, Mol. Cell Biol. 13, 4432–4444.
96. Zuo, P., Stanojevic, D., Colgan, J., Han, K., Levine, M., and Manley, J.L., 1991, Genes Develop. 5, 254–264.