The EMBO Journal vol.15 no.18 pp.4992–5000, 1996

A molecular code dictates sequence-specific DNA recognition by homeodomains

Giuseppe Damante[1], Lucia Pellizzari, Gennaro Esposito, Federico Fogolari, Paolo Viglino, Dora Fabbro, Gianluca Tell, Silvestro Formisano[2] and Roberto Di Lauro[3]

Dipartimento di Scienze e Tecnologie Biomediche, Università di Udine, Via Gervasutta 48, 33100 Udine, [2]Dipartimento di Biologia e Patologia Cellulare e Molecolare 'L.Califano', Via Pansini 5, 80121 Napoli and [3]Stazione Zoologica 'A.Dohrn', Villa Comunale 1, 80121 Napoli, Italy

[1]Corresponding author

Most homeodomains bind to DNA sequences containing the motif 5′-TAAT-3′. The homeodomain of thyroid transcription factor 1 (TTF-1HD) binds to sequences containing a 5′-CAAG-3′ core motif, delineating a new mechanism for differential DNA recognition by homeodomains. We investigated the molecular basis of the DNA binding specificity of TTF-1HD by both structural and functional approaches. As already suggested by the three-dimensional structure of TTF-1HD, the DNA binding specificities of the TTF-1, Antennapedia and Engrailed homeodomains, either wild-type or mutants, indicated that the amino acid residue in position 54 is involved in the recognition of the nucleotide at the 3′ end of the core motif 5′-NAAN-3′. The nucleotide at the 5′ position of this core sequence is recognized by the amino acids located in position 6, 7 and 8 of the TTF-1 and Antennapedia homeodomains. These data, together with previous suggestions on the role of amino acids in position 50, indicate that the DNA binding specificity of homeodomains can be determined by a combinatorial molecular code. We also show that some specific combinations of the key amino acid residues involved in DNA recognition do not follow a simple, additive rule.

Keywords: DNA binding/DNA recognition/ homeodomain

Introduction

Detailed knowledge of the molecular mechanisms used by transcription factors to recognize specific DNA sequences is important in understanding how gene expression is regulated. It has not been possible to derive a general protein–DNA recognition code (Pabo and Sauer, 1984; Pavletich and Pabo, 1993; Suzuki, 1994; Suzuki *et al.*, 1995), presumably because DNA binding domains have distinct structures and, therefore, different ways in which to contact DNA (Sauer *et al.*, 1982; Landshultz *et al.*, 1988; Pabo and Sauer, 1992; Miller *et al.*, 1995). Hence, each class of DNA binding domain might have a distinct recognition code. Recent data on zinc finger proteins support this notion as well as the possible practical outcome of this approach (Choo *et al.*, 1994).

The homeodomain (HD) is a 60 amino acid long DNA binding domain of a large family of transcriptional activators. HD-containing transcription factors control various cell fate decisions in a wide range of organisms including yeasts, insects and vertebrates (Gehring, 1987; Levine and Hoey, 1988; Kaufman *et al.*, 1990). The biological specificity of HD-containing proteins can, in part, be explained by differences in the DNA recognition properties of their homeodomains (Dessain *et al.*, 1992; Ekker *et al.*, 1992). It is, therefore, essential to rationalize the molecular basis of the HD–DNA recognition. To this end, it is necessary to integrate three types of information: (i) the structure of HDs; (ii) their DNA binding mode; and (iii) the identification of the amino acid–base pair contacts controlling the specificity of interaction. HDs show a conserved structure consisting of three helical regions (I, II and III) folded into a tight globular structure (Qian *et al.*, 1989, 1994; Tsao *et al.*, 1995). Helix I is preceded by an N-terminal arm and separated by a loose loop from helix II which, with helix III, forms a helix–turn–helix motif (HTH). The latter was described previously for several prokaryotic gene regulatory proteins (Pabo and Sauer, 1992). However, unlike prokaryotic HTH motifs that bind to DNA as dimers, HDs bind to DNA as monomers (Affolter *et al.*, 1990; Florence *et al.*, 1991).

The DNA binding mode of HDs is very well conserved (Kissinger *et al.*, 1990; Otting *et al.*, 1990; Wolberger *et al.*, 1991; Gehring *et al.*, 1994a; Hirsch and Aggarwal, 1995). Helix III (also called the recognition helix) lies in the major groove of the DNA and establishes specific contacts with bases. Although the HD has the same HTH protein fold as the prokaryotic HTH proteins, the residues used for contacting DNA bases in the recognition helix are shifted by one turn toward the C-terminus. Moreover, compared with classic prokaryotic HTH proteins, HDs have an extended C-terminus of the recognition helix with many basic residues that bind to DNA phosphates. As a consequence, it has been proposed that the HD recognition helix does not use the HTH-type DNA binding geometry, but the type called probe helix (PH) (Suzuki *et al.*, 1995), the same as that which occurs in basic zipper proteins. The HD makes additional specific contacts with bases, in the minor groove, with its N-terminal arm. The exact way in which the N-terminal arm contacts DNA can differ for the different HDs. In fact, the N-terminal arm in the Even-skipped (Eve)–DNA complex follows the minor groove instead of, as in the Engrailed (En)–DNA complex, reaching across it (Kissinger *et al.*, 1990; Hirsch and Aggarwal, 1995). In addition to the recognition helix and N-terminal arm, the loop between helices I and II interacts with the DNA backbone.

Most of the HDs bind with high affinity to sites which

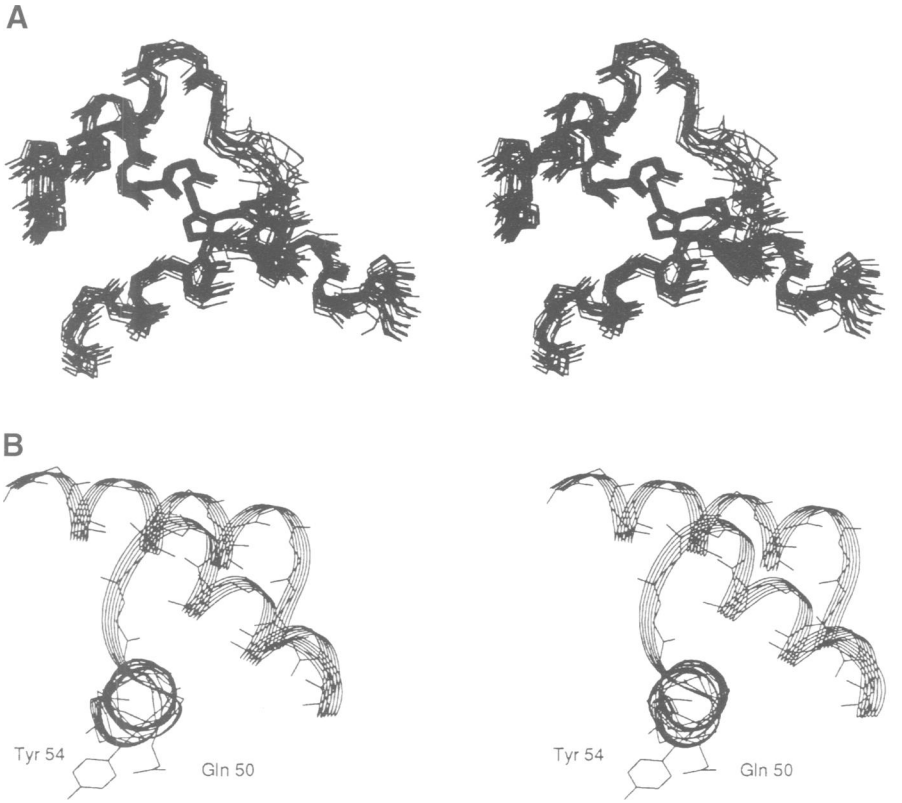

Fig. 1. Structure of TTF-1HD. (**A**) A stereoview of a superimposition of 20 AMBER-refined structures of TTF-1HD. Only the backbone atoms of residues 10–58 (superimposed on the best structure obtained from DIANA) are shown. (**B**) A stereoview of the ribbon structure of TTF-1HD with side chains of Gln50 and Tyr54.

bear the sequence 5′-TAAT-3′ as a core (Hoey and Levine, 1988). However, the preference for DNA sequences outside this core appears to differ considerably (Hanes and Brent, 1991; Ekker *et al.*, 1992; Catron *et al.*, 1993). In addition, some HDs show a preference for sites with core sequences other than the 5′-TAAT-3′ motif (Damante and Di Lauro, 1991; Ekker *et al.*, 1994; Tang and Breitman, 1995).

The amino acids which control the DNA binding specificity of HDs are located mostly in the recognition helix and in the N-terminal arm (Treisman *et al.*, 1989; Percival-Smith *et al.*, 1990; Hanes and Brent, 1991; Furukubo-Tokunaga *et al.*, 1992; Lin and McGinnis, 1992; Ades and Sauer, 1994; Damante *et al.*, 1994; Hanes *et al.*, 1994; Phelan *et al.*, 1994; Pomerantz and Sharp, 1994; Vershon *et al.*, 1995). However, the relative importance of each residue appears to be different, depending on the HD. The only amino acid that appears to have a general role in controlling DNA binding specificity is located at position 50. In fact, using genetic approaches, it has been demonstrated that the nature of the amino acid at position 50 dictates the preference for the dinucleotide immediately 3′ to the core sequence (5′-TAAT-3′ or others) (Treisman *et al.*, 1989; Percival-Smith *et al.*, 1990; Hanes and Brent, 1991; Ades and Sauer, 1994; Damante *et al.*, 1994; Hanes *et al.*, 1994; Phelan *et al.*, 1994; Pannese *et al.*, 1995). The structure of HD–DNA complexes has, however, not completely elucidated the nature of the specificity of these interactions, thus stressing the need for both genetic and structural approaches in order to gain better insights into the rules of DNA sequence selection by HD-containing proteins.

We have chosen the HD of the thyroid transcription factor 1 (TTF-1HD) as a model to explore the molecular mechanisms which form the basis of the DNA binding specificity of HDs. At variance from many other HDs, TTF-1HD recognizes sequences containing a 5′-CAAG-3′ core motif (Damante *et al.*, 1994). ^{1}H NMR studies revealed that TTF-1HD folds into the typical HD structural motif, namely three helices spanning the fragments 10–22, 28–39 and 42–59 of the sequence (Figure 1A, Brookhaven Protein Data Bank, id. code: 1FTT; G.Esposito, F.Fogolari, G.Damante, S.Formisano, G.Tell, A.Leonardi, R.Di Lauro and P.Viglino, unpublished data). Here, we show that residues located in the N-terminal arm and in the recognition helix control the binding specificity of TTF-1 and of other HDs. Moreover, we identify the nucleotide(s) that are differentially recognized in the cognate sequences and propose that the DNA binding specificity of HDs is dictated by a combinatorial code.

Results

The amino acid at position 54 controls the DNA binding specificity of TTF-1HD

The structure of TTF-1HD was useful in gaining insights into the residues which are in the appropriate position to contribute to the DNA binding specificity of this protein. The spatial arrangement of TTF-1HD shows that Tyr54 is in an excellent position to contact DNA (Figure 1B), as previously predicted by a modelling study (Fogolari *et al.*, 1993). Interestingly, a tyrosine residue at this position is present in all members of the NK class of HDs (Kim and Nirenberg, 1989; Gehring *et al.*, 1994b; Evans

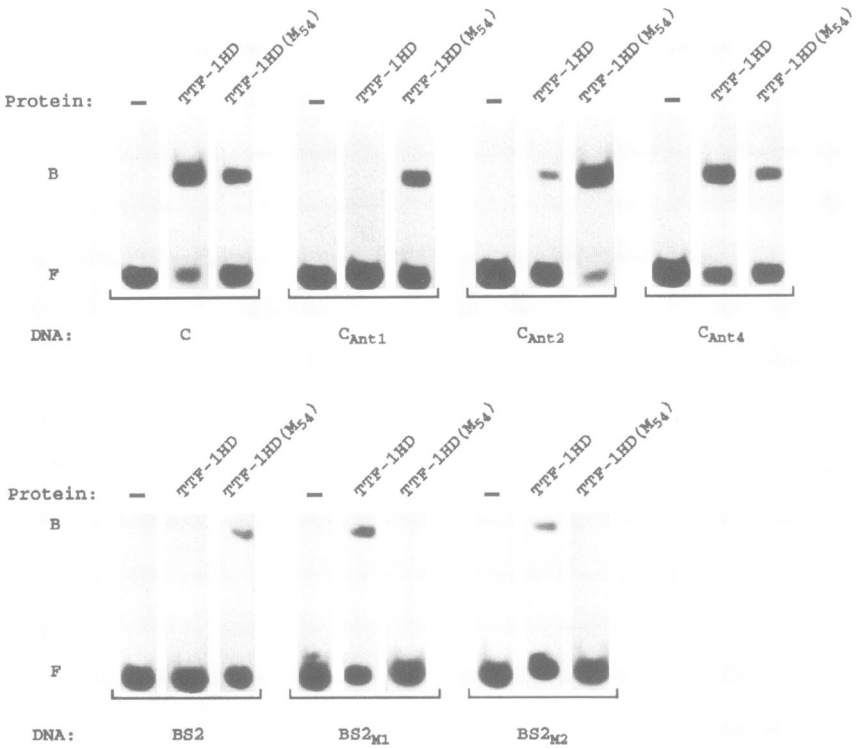

Fig. 2. The amino acid at position 54 controls the DNA binding specificity of TTF-1HD. The proteins used are indicated above each autoradiogram. The sequences of the oligonucleotides are shown in Table I. B: bound DNA; F: free DNA. Gel retardation assays were performed as described in Materials and methods.

et al., 1995). The relevance of the residue at position 54 for DNA binding specificity has been proposed already on the basis of structural considerations (Laughon, 1991); furthermore, when MATα2 HD is mutagenized so that the naturally occurring Arg54 is changed to Met, a dramatic change occurs in its '*in vivo*' activity (Vershon *et al.*, 1995). In order to test the relevance of Tyr54 for the recognition of the motif 5′-CAAG-3′ by TTF-1 HD, the mutant TTF-1HD(M_{54}) was constructed, with a Met residue replacing the wild-type Tyr54. Met54 is conserved in many HDs, such as those of the Antennapedia (Antp) class, that preferentially bind to DNA sequences containing the 5′-TAAT-3′ core motif (Damante *et al.*, 1994). The DNA binding specificity of wild-type and TTF-1HD(M_{54}) was evaluated by using either a TTF-1 (oligonucleotide C) or an Antp binding site (oligonucleotide BS2), and mutants thereof. Results are shown in Figure 2 and, in a quantitative form, in Table I. For the sake of clarity, we will refer to the crucial bases in these two oligonucleotides using the numbering of Table I. Hence, the core sequences, 5′-CAAG-3′ or 5′-TAAT-3′, in oligonucleotides C and BS2, respectively, correspond to nucleotides 2 (C or T), 3 (A), 4 (A) and 5 (G or T). The data shown in Figure 2 and Table I demonstrate that, independently of whether the flanking DNA sequence is that of oligonucleotide C or BS2, the wild-type TTF-1HD prefers G at position 5 ($C>C_{Ant2}$, $C_{Ant4}>C_{Ant1}$ and $BS2_{M2}>BS2$). In contrast, TTF-1HD(M_{54}) prefers oligonucleotides bearing T in place of G at position 5 ($C_{Ant2}>C$, $C_{Ant1}>C_{Ant4}$ and $BS2>BS2_{M2}$). These findings demonstrate that the nature of the residue at position 54 of TTF-1HD determines a preference for specific nucleotides at position 5 in the cognate binding sites and hence contributes to the overall DNA binding specificity of this HD.

The amino acid at position 54 controls the DNA binding specificity of Antp and En HDs

To verify whether the nature of residue 54 is a determinant for the DNA binding specificity of other HDs, we first measured the binding of Antp and En HDs, which bear respectively Met and Ala at position 54, to the oligonucleotides C and BS2 and the mutants thereof already described above. Both Antp and En HDs bind to 5′-TAAT-3′-containing sequences, irrespective of whether the sequence context is that of the C or of the BS2 oligonucleotide (Table I). This result is exactly the reverse of that obtained with TTF-1HD, showing that the only major difference between the recognized sequences of Antp and En HDs, on one hand, and of TTF-1 HD, on the other, is in the core motif 5′-YAAK-3′, where Y and K must be T for the former and C and G for the latter. These results also suggest that the presence of either Ala or Met in position 54 does not affect the DNA binding specificity, as we could not detect any difference in DNA sequence selection between Antp or En HDs. Both Antp and En HDs were mutagenized to give the HDs Antp(Y_{54}) and En(Y_{54}) which bear Tyr at position 54. As predicted, both mutant HDs switch their preference toward sequences that contain the core motif 5′-TAAG-3′, such as C_{Ant4} or $BS2_{M2}$ (Table I). We conclude that the amino acid residue in position 54 of HDs is of relevance in determining the DNA sequence recognized. In addition to Tyr (TTF-1), Met (Antp) and Ala (En), several other residues can be found in this position in different HDs (Gehring *et al.*,

Table I. DNA binding specificity of TTF-1HD, AntpHD, EnHD and their mutants

Oligos	Position No.	Proteins									
	0 1 2 3 4 5 6 7 8	TTF-1HD	TTF-1HD (M_{54})	AntpHD	AntpHD (Y_{54})	EnHD	EnHD (Y_{54})	TTF-1HD (QTY)	AntpHD (VLF)	TTF-1HD (QTY-M_{54})	AntpHD (VLF-Y_{54})
C	G T C A A G T G T	100	43 ± 6	2 ± 1	37 ± 4	3 ± 1	32 ± 3	47 ± 4	23 ± 3	4 ± 2	49 ± 4
C_{Ant1}	G T **T** A A T T G T	5 ± 1	85 ± 2	88 ± 3	24 ± 4	131 ± 11	21 ± 2	15 ± 3	85 ± 7	95 ± 7	13 ± 2
C_{Ant2}	G T C A A **T** T G T	20 ± 3	105 ± 6	12 ± 5	7 ± 2	29 ± 8	4 ± 1	6 ± 2	118 ± 14	n.d.	n.d.
C_{Ant4}	G T **T** A A G T G T	65 ± 5	30 ± 2	18 ± 3	94 ± 9	17 ± 5	117 ± 15	89 ± 6	22 ± 3	n.d.	n.d.
BS2	T C T A A T G G C	3 ± 1	30 ± 6	80 ± 4	18 ± 3	79 ± 14	14 ± 3	8 ± 3	77 ± 6	34 ± 5	11 ± 2
$BS2_{M1}$	T C **C** A A **G** G G C	78 ± 7	1 ± 1	8 ± 1	32 ± 4	4 ± 2	26 ± 4	82 ± 8	7 ± 2	1 ± 1	46 ± 4
$BS2_{M2}$	T C T A A **G** G G C	43 ± 3	1 ± 1	15 ± 2	60 ± 6	9 ± 3	53 ± 7	n.d.	n.d.	n.d.	n.d.

The proteins used are indicated at the top of the table. The names and sequences (partial) of the oligonucleotides are indicated on the left. Only one strand is reported, in the 5′→3′ direction. Bases shown in bold at positions 2 or 5 indicate mutations introduced with respect to the wild-type C and BS2 oligonucleotides. Values indicate the DNA binding activity of each protein with the corresponding oligonucleotide. The DNA binding activity was determined by measuring the protein-bound/free DNA ratio in gel retardation assays. Results are expressed as a fraction of the value obtained for the TTF-1–C complex, considered arbitrarily as 100. Each value represents the mean ± SD of at least three independent determinations. n.d.: not done.

1994b). We suggest that such a difference in residues might contribute to dictating the *in vivo* target genes for the corresponding HDs.

Residues at the N-terminus control the DNA binding specificity of TTF-1 and Antp HDs

Even though residue 54 is important for DNA binding specificity, some other amino acid residue must be responsible for differential recognition of sequences containing either a C or a T in position 2 of the recognized site. In fact, regardless of whether the residue at position 54 is Tyr or Met, TTF-1HD prefers sequences containing a C in that position (in Table I: C>C_{Ant4}, C_{Ant2}>C_{Ant1}, $BS2_{M1}$>$BS2_{M2}$). Conversely, both Antp and En, regardless of the residue present in position 54, preferentially bind to sequences containing T in position 2 of the recognized sequence (in Table I: C_{Ant4}>C, C_{Ant1}>C_{Ant2}, $BS2_{M2}$> $BS2_{M1}$). Since it has been demonstrated that TTF-1HD DNA binding specificity is also conferred by residues located in the first 28 positions of the HD (Damante and Di Lauro, 1991), several TTF-1HD mutants were tested, in which amino acids of this region were changed to those of AntpHD (data not shown). A striking change in the binding properties was observed with the mutant TTF-1HD(QTY) in which Val6, Leu7 and Phe8 were changed to Gln, Thr and Tyr, respectively (these amino acids are present at position 6, 7 and 8 in AntpHD). TTF-1HD(QTY), in contrast to the wild-type HD, prefers the 5′-TAAG-3′-containing sequence (C_{Ant4}>C, etc.), even though it still binds efficiently to 5′-CAAG-3′-containing sequences (Table I)

The effect of changing residues at positions 6, 7 and 8 was also tested in AntpHD. The corresponding amino acids of TTF-1HD were introduced into the wild-type protein to produce AntpHD(VLF). As predicted, this mutant binds preferentially to the 5′-CAAT-3′-containing sequence (C_{Ant2}>C_{Ant1}) (Table I).

Combinatorial effects of mutations at positions 6, 7, 8 and 54

Previously, we have demonstrated a role for the amino acid at position 50 in controlling TTF-1HD binding specificity (Damante *et al.*, 1994). Here it is shown that the peculiar DNA binding specificity of TTF-1HD is contributed by two additional determinants, one located in the N-terminal arm (residues 6, 7 and 8) and the other in the recognition helix (residue 54). By constructing double mutant HDs in which residues at the N-terminus and at position 54 were changed simultaneously, we tested whether these determinants act in a combinatorial fashion. The specificity conferred by these two determinants appears to be additive both in TTF-1HD and in AntpHD (Table I). The TTF-1HD double-mutant (QTY-M_{54}) prefers sequences that contain 5′-TAAT-3′ (C_{Ant1} and BS2) and therefore its DNA binding specificity is very similar to that observed for AntpHD. Similarly, the Antp double mutant (VLF-Y_{54}), binds most efficiently to 5′-CAAG-3′-containing sequences (C and $BS2_{M1}$). From data listed in Table I, it can be observed that there must be additional contributions to DNA binding preferences, since there are quantitative changes among the different mutants that cannot be explained entirely by the mutations introduced.

Interference between determinants of the DNA binding specificity

Figure 3 presents a model summarizing the DNA–protein contacts responsible for the binding specificity of TTF-1HD. The structural analogies between TTF-1 and other HDs indicate that this domain can be oriented and aligned to the cognate sequence with respect to an invariant interaction between residue Asn51 and the adenine partner of the A:T base pair corresponding to the third base of the 5′-YAAK-3′ core (Ekker *et al.*, 1992). Besides the Asn51 contact, the interactions described in the model are based on functional data presented here and elsewhere (Damante *et al.*, 1994). We have demonstrated previously that TTF-1HD(K_{50}), in which lysine replaces glutamine at position 50, recognizes the sequence 5′-CAAGCC-3′, as opposed to the wild-type protein that prefers the sequence 5′-CAAGTG-3′ (Damante *et al.*, 1994), indicating that, as in most other HDs, a functional interaction occurs between amino acid 50 of TTF-1HD and the dinucleotide attached to the 3′ end of the 5′-YAAK-3′ core motif. The model shown in Figure 3 envisages the possibility of delineating a code for HD-specific DNA recognition: once all relevant residues in HDs and their preferred nucleotide in the recognition site are known, it should be possible to predict the sequence of a binding site from the specific combination of amino acid residues at the key positions in the HDs. To test further the validity

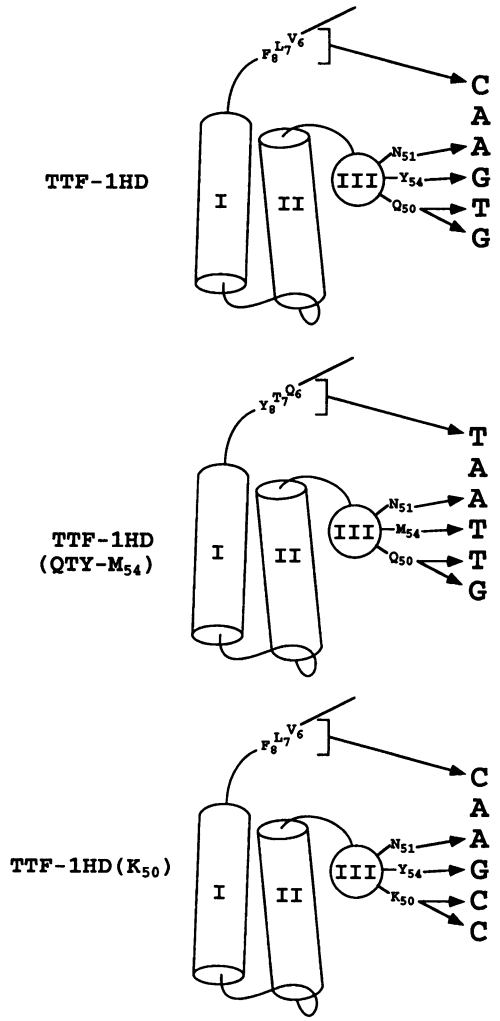

Fig. 3. Amino acids controlling DNA binding specificity of TTF-1HD. Three proteins are shown: TTF-1HD, TTF-1HD(VLF-M_{54}) and TTF-1HD(K_{50}), each flanked by the optimal cognate DNA sequence. I, II and III indicate the first, second and third helix of HD, respectively. Amino acids are indicated by the single letter code. The arrows indicate functional interactions between amino acids and base pairs. In the case of the N_{51}–A interaction, the arrow indicates a direct structural contact. The data supporting the scheme of the TTF-1HD(K_{50})–DNA interaction are from Damante *et al.* (1994).

of the recognition code summarized in Figure 3, we used single and double mutants of TTF-1HD, in which Gln and Lys at position 50, and Tyr and Met at position 54, were present in all combinations. We focused on amino acids 50 and 54 because of their relative proximity in the DNA-contacting surface of the recognition helix. The binding activity of wild-type, single and double mutant proteins was tested on a set of C-based oligonucleotides in which all combinations of the bases preferred by amino acids at position 50 and 54 were present. The results of this approach are shown in Figure 4. The combinations Gln50–Tyr54, Gln50–Met54 and Lys50–Tyr54 follow the code, i.e. each protein prefers the DNA sequence containing the appropriate combination of nucleotides. However, the DNA sequence preference of the combination Lys50–Met54 does not correspond to that predicted and, furthermore, TTF-1HD(K_{50}–M_{54}) shows a poor DNA binding activity. From these data, we deduce that there are restrictions to the code due to functional or structural interferences between specific amino acid residues. In this respect, it is of great interest to point out that in the naturally occurring HDs the nature of the amino acid at position 54 appears to be constrained by the presence of either Gln or Lys at position 50 (Table II). Gln50 allows a wide spectrum of different residues at position 54, Met being the most abundant. In contrast, Met54 is never found when position 50 is occupied by Lys. This observation correlates well with the findings that in the context of TTF-1 the combination Lys50–Met54 does not show optimal DNA binding properties. The strong association Cys50–Gln54 (this combination is found in all POU-HD proteins) further supports the existence of constraints between amino acids at positions 50 and 54.

Discussion

We present evidence indicating that amino acids in the N-terminal arm, as well as the amino acid at position 54, control the DNA binding specificity of several HDs. Therefore, it appears that the DNA binding specificity of a single HD is conferred by several determinants. Evidence supporting our conclusion has been obtained recently by studying the differential DNA binding specificity existing among Msx-1 and HoxA3 HDs (Ebu Isaac *et al.*, 1995). The control of the DNA binding specificity is not restricted to base-contacting residues, since backbone-contacting amino acids could also play a significant role (Furukubo-Tokunaga *et al.*, 1992).

A large body of evidence indicates that residues at the N-terminal arm of HD control the functional specificity of HD-containing proteins (Kuziora and McGinnis, 1989; Gibson *et al.*, 1990; Mann and Hogness, 1990; Lin and McGinnis, 1992; Zeng *et al.*, 1993; Mann, 1995). The difference in the functional specificity between Antp and Scr is determined by four amino acids of the N-terminal arm (Furukubo-Tokunaga *et al.*, 1993). Also the difference in functional specificity between Dfd and Ubx resides in the N-terminal arm of their HD (Mann and Hogness, 1990; Lin and McGinnis, 1992; Chan and Mann, 1993). A direct influence of residues of the N-terminal arm on DNA binding specificity has been observed: amino acids at positions 6 and 7 play a decisive role in the different sequence specificity displayed by Ubx and Abd-B HDs (Ekker *et al.*, 1994). Consistently, the ability of the N-terminal arm to discriminate base pairs in the minor groove has been demonstrated (Ades and Sauer, 1995; Dragonescu *et al.*, 1995). Residues at positions 6 and 7 may operate as a pair in conferring sequence specificity. They diverge widely among all HDs, but exhibit a strong pairwise conservation within related groups (Laughon, 1991). Besides NK class HDs, the combination Val6, Leu7 and Phe8 is present in the *Drosophila* cutHD. Interestingly, by using a genetic screen, it has been demonstrated that cutHD binds to a TAA repeat sequence (therefore more similar to 5′-TAAT-3′ than to 5′-CAAT-3′ sequences used in this study) with an affinity 10-fold lower with respect to Antp class HDs (Kalionis and O'Farrel, 1993). In several of the resolved HD–DNA complexes, residues 6, 7 and 8 are involved in a variety of contacts mainly to the sugar-phosphate backbone at the interface between the major and minor grooves. In the Antp–DNA complex, Gln6 should be able to establish an interaction with the

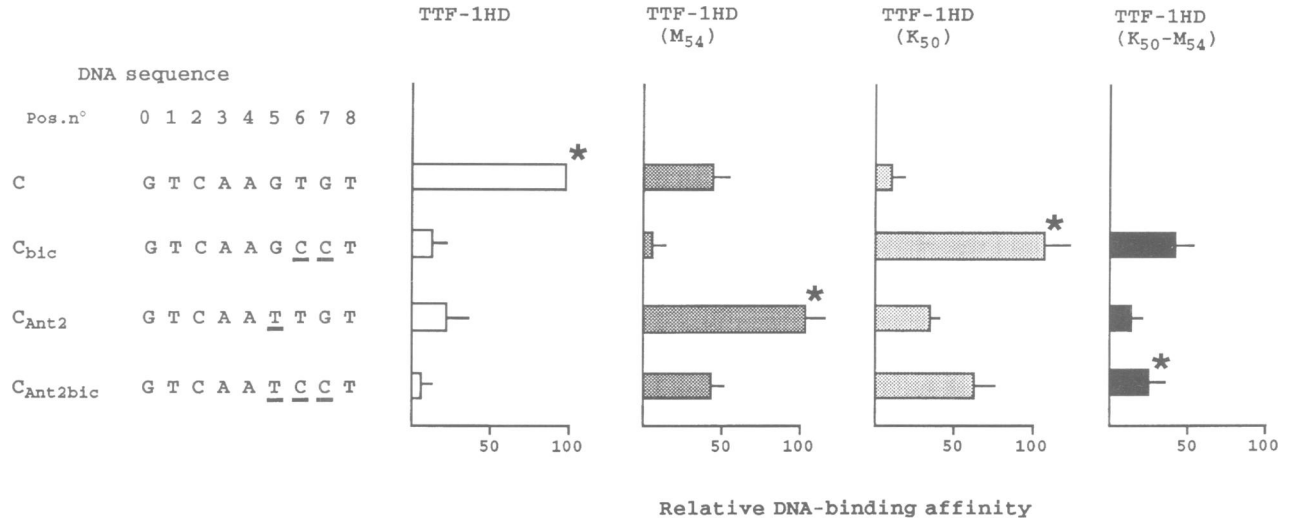

Fig. 4. DNA binding affinity of mutants of TTF-1HD with single or double mutations at positions 50 and 54. On the left, the names and sequences of the oligonucleotides are indicated. Underlined bases indicate mutations with respect to the wild-type C sequence. The relative DNA binding affinity was measured as described in Materials and methods, and in both panels is expressed as the percentage of the binding affinity detected for the TTF-1HD–C complex considered arbitrarily as 100. Asterisked columns indicate the expected optimal sequence for each protein, in the case of a combinatorial effect of amino acids at positions 50 and 54. Each bar indicates the mean ± SD of at least three measurements performed in separate gels.

Table II. Number of homeodomains possessing the various combinations of amino acids at positions 50 and 54

Amino acid at position 50	Total (385)	Amino acid at position 54 (No.)									
		Met (183)	Ala (86)	Gln (37)	Thr (22)	Tyr (13)	Ser (13)	Val (9)	Asn (6)	Arg (6)	Others (12)
Gln	324	183	71	2	22	13	11	9	6	4	11
Lys	9	–	7	–	–	–	–	–	–	2	–
Cys	35	–	–	35	–	–	–	–	–	–	–
Others	17	–	8	–	–	–	2	–	–	–	1

Data are from Gehring *et al.* (1994b). The homeodomains of the POU class are included, all containing Cys and Gln at positions 50 and 54, respectively.

phosphodiester backbone between base pairs 2 and 3 of the BS2 sequence which, in turn, may facilitate the contacts of Tyr8 with the deoxyribose moiety of A2. This concerted interaction pattern would be lost with a valine at position 6, resulting in a different anchoring of the protein at the interface between the major and minor groove, which may determine a packing shift in the remainder of the complex.

It is of interest to observe the DNA binding effects of either methionine or tyrosine at position 54 in the light of the structural data on protein–DNA complexes. The reduction of the binding activity of both Antp and En HDs with sequences bearing the 5′-TAAG-3′ motif in place of 5′-TAAT-3′ should occur because of the loss of the hydrophobic contact made by Ile47 with the methyl group of the thymine at the 3′ extremity of the 5′-TAAT-3′ sequence. EnHD(Y_{54}) binds preferentially to 5′-TAAG-3′-bearing sequences because of the favourable interactions between Tyr54 and the base pair at the 3′ end of the core. The inability of the same protein to bind 5′-TAAT-3′-bearing sequences efficiently could be due to the rigid and bulky aromatic ring of tyrosine that would not allow the Ile_{47}–DNA interaction to be accommodated. A similar model could be invoked to explain the results obtained with the AntpHD(Y_{54}) mutant where the contacts of both Ile47 and Gln50 would be disrupted by Tyr54. It is very important to note that, in the AntpHD–DNA complex, residue 54 interacts with the base pairs at position 6 and 7, i.e. outside the 5′-TAAT-3′ core (Billeter *et al.*, 1993). Therefore, the functional interaction between Met54 and the thymine at the 3′ end of the 5′-YAAT-3′ core should not be due to a direct interaction but to a lack of steric hindrance of this amino acid in the contact between Ile47 and thymine at the 3′ end of the 5′-YAAT-3′ core. A direct interaction involving Met54 and the base pair 3′ to the conserved Asn51–adenine (5′-YAAT-3′) contact has been observed recently in the crystal of the Eve–DNA complex (Hirsch and Aggarwal, 1995). In that case, Met54 establishes a van der Waals contact with the *N*7 of the adenine pairing thymine at the 3′ end of the 5′-TAAT-3′ core motif. The mutation 5′-TAAT-3′→5′-TAAG-3′ introduces a cytosine in place of the Met54-contacted adenine, which removes the possibility of establishing the appropriate hydrophobic interaction (Seeman *et al.*, 1976).

The structural observation made in the Antp–DNA complex, where Gln50 and Met54 contact the same nucleotides (3′ to the 5′-TAAT-3′ core) could provide the explanation as to why TTF-1HD (K_{50}–M_{54}) shows an overall weak DNA binding activity that is not predictable by a combinatorial code. The side chain of Gln is relatively short and can work simultaneously as an acceptor and

donor of hydrogen bonds. Therefore, this residue appears ideally suited to establishing the network of short-lived fluctuating interactions, described in the Antp–DNA complex (Billeter *et al.*, 1993). By contrast, Lys, with a longer and positively charged side chain, appears to be less versatile than Gln. Therefore, when Gln50 is present, an optimal fluctuating network of hydrogen bonds may occur, allowing Met54 to establish its correct contacts with DNA. Lys50, on the other hand, could disrupt Met54–DNA contacts. Interestingly, Table II shows that most of the HDs with Lys at position 50 contain an alanine at position 54, i.e. a residue with a reduced side chain hindrance that, therefore, would not interfere with the Lys50–DNA contacts. Structural studies on Lys50-containing HD–DNA complexes are required to test this prediction.

In conclusion, this study provides an important refinement of the molecular mechanisms used by HDs in recognizing DNA sequences. It has been demonstrated recently that HDs require protein–protein interactions to select target genes (Smith and Johnson, 1992; Beachy *et al.*, 1993; Chan *et al.*, 1994; Zappavigna *et al.*, 1994; Chang *et al.*, 1995; Johnson *et al.*, 1995; Li *et al.*, 1995; Mann, 1995; Smith *et al.*, 1995; Wilson *et al.*, 1995; Copeland *et al.*, 1996; Shen *et al.*, 1996; Wolberger, 1996). Moreover, some HDs could recognize RNA (Dubnau and Struhl, 1996). Nevertheless, the understanding of the molecular basis of HD DNA binding specificity must still be considered as a fundamental issue in the refinement of models of gene expression regulation by HDs. We have confirmed that residues at the N-terminal arm control the DNA binding specificity of these proteins. Moreover, we have demonstrated the role of the amino acid at position 54 and identified the base pair responsible for the underlying functional interaction. In exploring the possibility of a combinatorial HD–DNA recognition code, we provide an explanation of why only certain combinations of contacting amino acids have been selected during evolution.

Materials and methods

Plasmids and construction of mutants

Plasmids encoding TTF-1, Antp and En HDs have already been described (Muller *et al.*, 1988; Damante and Di Lauro, 1991; Wilson *et al.*, 1993). In all of these plasmids the transcription of the HD-coding sequence is driven by the T7 RNA polymerase promoter. Mutants were constructed by PCR (Ho *et al.*, 1989), cloned in the bacterial expression vector pT7.7 and verified by nucleotide sequencing.

Protein expression and purification

Proteins were expressed using the *Escherichia coli* strain BL21 (DE3) (Studier *et al.*, 1991). Briefly, BL21 cells were transformed with expression plasmids, and cultures were grown in LB at 37°C overnight with shaking. Then, 2–4 l of LB were inoculated with the overnight cultures and shaken at 37°C until the OD_{600} reached 0.6. At this point, the expression of the T7 RNA polymerase was induced by 0.5 mM of IPTG and, 3 h after induction, bacteria were pelleted and frozen at –80°C. TTF-1HD protein to be used in NMR analysis was purified essentially as described (Muller *et al.*, 1988) with two chromatographic steps using Biorex and Mono S ion-exchange matrices. At the end of the procedure, the purity of the proteins was checked by SDS–PAGE and reverse phase chromatography and was shown to be at least 99%. The protein concentration was determined spectrophotometrically. The HD proteins used for DNA binding assays were partially purified using Econo Pac S cartridges (Bio-Rad). A purity of 40–80% was achieved, enough for the proteinsto be detected by SDS–PAGE. The concentration of the active protein was measured by oligonucleotide saturation assay. A gel retardation assay (see below) was performed without calf thymus DNA and with increasing amounts of oligonucleotides (from 0.3 to 50 nM), then the protein-bound and free oligonucleotide concentration values were subjected to a Scatchard plot analysis. value The percentage of active protein was usually 20–40%

Gel retardation assay and quantitation of the binding activity

Double-stranded oligodeoxynucleotides, labelled at the 5′ end with ^{32}P, were used as probes in the gel retardation assays. The C site is a 24mer whose upper strand sequence is 5′-CACTGCCCAGTCAAGTGTTCTTGA-3′. The BS2 site is an 18mer whose upper strand sequence is 5′-GAGAAAAAGCCATTAGAG-3′. Mutants of C and BS2 described in the Results section were based on wild-type oligodeoxynucleotides; thus, C and BS2 mutants were 24mers and 18mers respectively. The gel retardation assay was performed by incubating protein and DNA in a buffer containing 20 mM Tris–HCl pH 7.6, 75 mM KCl, 0.25 mg/ml bovine serum albumin (BSA), 5 mM dithiothreitol (DTT), 5 mg/ml calf thymus DNA, 10% glycerol for 30 min at room temperature. Protein-bound DNA and free DNA were separated on a native 7.5% polyacrylamide gel run in 0.5× TBE for 1.5 h at 4°C. Gels were dried, exposed to X-ray films and bands were quantitated by densitometric scanning of the autoradiogram using an LKB laser densitometer. The DNA binding activities of HDs were determined by calculating the ratio of protein-bound and free DNA signals using the same amount of active protein. Results are expressed as the percentage of the binding activity of the TTF-1HD–C complex, arbitrarily considered as 100.

Acknowledgements

We thank M.Muller and D.Wilson for providing the expression vectors encoding Antp and En HDs. L.Minichiello was of great help in providing suggestions and support. We acknowledge L.Cattarossi for computational and digital communication support and G.Superti Furga and S.Guerra for help in the revision of the manuscript. This work is supported by a grant from the Associazione Italiana Ricerca sul Cancro (A.I.R.C.). G.D.D.F. is supported by an A.I.R.C. fellowship. L.P. is supported by the Ph.D. programme 'Diagnosi e Terapie Cellulari e Molecolari' of Udine University.

References

Ades,S.E. and Sauer,R.T. (1994) Differential DNA-binding specificity of the Engrailed homeodomain: the role of residue 50. *Biochemistry*, **33**, 9187–9194.

Ades,S.E. and Sauer,R.T. (1995) Specificity of minor groove and major groove interaction in a homeodomain–DNA complex. *Biochemistry*, **34**, 14601–14608.

Affolter,M., Percival-Smith,A., Muller,M., Leupin,W. and Gehring,W.J. (1990) DNA binding properties of the purified Antennapedia homeodomain. *Proc. Natl Acad. Sci. USA*, **87**, 4093–4097.

Beachy,P.A., Varkey,J., Young,K.E., von Kessler,D., Sun,B.I. and Ekker,S.C. (1993) Cooperative binding of an Ultrabithorax homeodomain to nearby and distant DNA sites. *Mol. Cell. Biol.*, **13**, 6941–6956.

Billeter,M., Qian,Y.Q., Otting,G., Muller,M., Gehring,W.J. and Wutrich,K. (1993) Determination of the three-dimensional structure of the antennapedia homeodomain from *Drosophila* in solution by H nuclear magnetic resonance spectroscopy. *J. Mol. Biol.*, **234**, 1084–1097.

Catron,K.M., Iler,N. and Abate,C. (1993) Nucleotides flanking conserved TAAT core dictate the DNA binding specificity of three murine homeodomain proteins. *Mol. Cell. Biol.*, **13**, 2354–2365.

Chan,S.K. and Mann,R.S. (1993) The segment identity functions of Ultrabithorax are contained within its homeodomain and carboxyterminal sequences. *Genes Dev.*, **7**, 796–811.

Chan,S., Jaffe,L., Capovilla,M., Botas,J. and Mann,R.S. (1994) The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell*, **78**, 603–615.

Chang,C., Shen,W., Rozenfeld,S., Lawrence,H.J., Largman,C. and Cleary,M.L. (1995) Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev.*, **9**, 663–674.

Choo,Y., Sanchez-Garcia,I. and Klug,A. (1994) *In vivo* repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature*, **372**, 642–645.

Copeland,J.W., Nasiadka,A., Dietrich,B.H. and Krause,H.M. (1996) Patterning of the *Drosophila* embryo by a homeodomain-deleted ftz polypeptide. *Nature*, **379**, 162–165.

Damante,G. and Di Lauro,R. (1991) Several regions of antennapedia and thyroid transcription factor 1 homeodomains contribute to DNA binding specificity. *Proc. Natl Acad. Sci. USA*, **88**, 5388–5392.

Damante,G. *et al.* (1994) Sequence-specific DNA recognition by the thyroid transcription factor-1 homeodomain. *Nucleic Acids Res.*, **22**, 3075–3083.

Dessain,S., Gross,C.T., Kuziora,M.A. and McGinnis,W. (1992) Antp-type homeodomains have distinct DNA-binding specificities that correlate with their different regulatory functions in embryos. *EMBO J.*, **11**, 991–1002.

Dragonescu,A., Levin,J.R. and Tullius,T.D. (1995) Homeodomain protein: what governs their ability to recognize specific DNA sequences? *J. Mol. Biol.*, **250**, 595–608.

Dubnau,J. and Struhl,G. (1996) RNA recognition and translational regulation by a homeodomain protein. *Nature*, **379**, 694–699.

Ebu Isaac,V., Sciavolino,P. and Abate,C. (1995) Multiple amino acids determine the DNA binding specificity of the Msx-1 homeodomain. *Biochemistry*, **34**, 7127–7134.

Ekker,S.C., von Kessler,D.P. and Beachy,P.A. (1992) Differential DNA sequence recognition is a determinant of specificity in homeotic gene action. *EMBO J.*, **11**, 4059–4072.

Ekker,S.C., Jackson,D.G., von Kessler,D.P., Sun,B.I., Young,K.E. and Beachy,P.A. (1994) The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. *EMBO J.*, **13**, 3551–3560.

Evans,S.M., Yan,W., Murillo,M.P., Ponce,J. and Papalopulu,N. (1995) tinman, a *Drosophila* homeobox gene required for heart and visceral mesoderm specification, may be represented by a family of genes in vertebrates: XNkx-2.3, a second vertebrate homologue of tinman. *Development*, **121**, 3889–3899.

Florence,B., Handrow,R. and Laughon,A. (1991) DNA-binding specificity of the fushi-tarazu homeodomain. *Mol. Cell. Biol.*, **11**, 3613–3623.

Fogolari,F., Esposito,G., Viglino,P., Damante,G. and Pastore,A. (1993) Homology model building of the thyroid transcription factor 1 homeodomain. *Prot. Engng*, **6**, 513–519.

Furukubo-Tokunaga,K., Muller,M., Affolter,M., Pick,L., Kloter,U. and Gehring,W.J. (1992) *In vivo* analysis of the helix–turn–helix motif of the fushi tarazu homeo domain of *Drosophila melanogaster*. *Genes Dev.*, **6**, 1082–1096.

Gehring,W.J. (1987) Homeo boxes in the study of development. *Science*, **236**, 1245–1252.

Gehring,W.J., Qian,Y.Q., Billeter,M., Furukubo-Tokunaga,K., Schier, A.F., Resendez-Perez,D., Affolter,M., Otting,G. and Wuthrich,K. (1994a) Homeodomain–DNA recognition. *Cell*, **78**, 211–223.

Gehring,W.J., Affolter,M. and Burglin,T. (1994b) Homedomain proteins. *Annu. Rev. Biochem.*, **63**, 487–526.

Gibson,G., Schier,A., LeMotte,P. and Gehring,W.J. (1990) The specificities of sex combs reduced and antennapedia are defined by a distinct portion of each protein that includes the homeodomain. *Cell*, **62**, 1087–1103.

Hanes,S.D. and Brent,R. (1991) A genetic model for interaction of the homeodomain recognition helix with DNA. *Science*, **251**, 426–429.

Hanes,S.D., Riddihough,G., Ish-Horowicz,D. and Brent,R. (1994) Specificity, DNA recognition and intersite spacing are critical for action af the bicoid morphogen. *Mol. Cell. Biol.*, **14**, 3364–3375.

Hirsch,J.A. and Aggarwal,A.K. (1995) Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. *EMBO J.*, **14**, 6280–6291.

Ho,S.N., Hunt,H.D., Horton,R.M., Pullen,J.K. and Pease,L.R. (1989) Site-directed mutagenesis by overlap extension using polymerase chain reaction. *Gene*, **77**, 51–60.

Hoey,T. and Levine,M. (1988) Divergent homeo box proteins recognize similar DNA sequences in *Drosophila*. *Nature*, **332**, 858–861.

Johnson,F.B., Parker,E. and Krasnow,M.A. (1995) Extradenticle protein is a selective cofactor for the *Drosophila* homeotics: role of the homeodomain and YPWM amino acid motif in the interaction. *Proc. Natl Acad. Sci. USA*, **92**, 739–743.

Kalionis,B. and O'Farrel,P.H. (1993) A universal target sequence is bound *in vitro* by diverse homeodomains. *Mech. Dev.*, **43**, 57–70.

Kaufman,T.C., Seeger,M.A. and Olsen,G. (1990) Molecular and genetic organization of the Antennapedia gene complex of *Drosophila melanogaster*. *Adv. Genet.*, **27**, 309–362.

Kim,Y. and Nirenberg,M. (1989) *Drosophila* NK-homeobox genes. *Proc. Natl Acad. Sci. USA*, **86**, 7716–7720.

Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) Crystal structure of an engrailed homeodomain–DNA complex at 2.8 Å resolution: a framework for understanding homeodomain–DNA interactions. *Cell*, **63**, 579–590.

Kuziora,M.A. and McGinnis,W. (1989) A homeodomain substitution changes the regulatory specificity of the deformed protein in *Drosophila* embryos. *Cell*, **59**, 563–571.

Landshultz,W.H., Johnson,P.F. and McKnight,S.L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*, **24**, 1759–1754.

Laughon,A. (1991) DNA binding specificity of homeodomains. *Biochemistry*, **30**, 11357–11367.

Levine,M. and Hoey,T. (1988) Homeobox proteins as sequence-specific transcription factors. *Cell*, **55**, 537–540.

Li,T., Stark,M.R., Johnson,A.D. and Wolberger,C. (1995) Crystal structure of the Mata1/Matα2 homeodomain heterodimer bound to DNA. *Science*, **270**, 262–269.

Lin,L. and McGinnis,W. (1992) Mapping functional specificity in the Dfd and Ubx homeodomains. *Genes Dev.*, **6**, 1071–1081.

Mann,R.S. and Hogness,D.S. (1990) Functional dissection of Ultrabithorax protein in *D.melanogaster*. *Cell*, **60**, 597–610.

Mann,R.S. (1995) The specificity of homeotic gene function. *Bioassay*, **17**, 855–863.

Miller,J., McLachlan,A.D. and Klug,A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.*, **4**, 1609–1614.

Muller,M., Affolter,M., Leupin,W., Otting,G., Wuthrich,K. and Gehring,W.J. (1988) Isolation and sequence specific DNA binding of the Antennapedia homeodomain. *EMBO J.*, **7**, 4299–4304.

Otting,G., Qian,Y.Q., Billeter,M., Muller,M., Affolter,M., Gehring,W.J. and Wuthrich,K. (1990) Protein–DNA contacts in the structure of a homeodomain–DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.*, **9**, 3085–3092.

Pabo,C.O. and Sauer,R.C. (1984) Protein DNA recognition. *Annu. Rev. Biochem.*, **53**, 293–321.

Pabo,C.O. and Sauer,R.C. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.*, **61**, 1053–1095.

Pannese,M., Palo,G., Andreazzoli,M., Vignali,R., Kablar,B., Barsacchi,G. and Boncinelli,E. (1995) The *Xenopus* homologue of Otx2 is a maternal homeobox gene that demarcates and specifies anterior body region. *Development*, **121**, 707–720.

Pavletich,N.P. and Pabo,C.O. (1993) Crystal structure of a five-finger GLI–DNA complex: new perspectives on zinc fingers. *Science*, **261**, 1701–1707.

Percival-Smith,A., Muller,M., Affolter,M. and Gehring,W.J. (1990) The interaction with DNA of wild-type and mutant fushi-tarazu homeodomains. *EMBO J.*, **9**, 3967–3974.

Phelan,M.L., Sadoul,R. and Featherstone,M. (1994) Fuctional differences between HOX proteins conferred by two residues in the homeodomain N-terminal arm. *Mol. Cell. Biol.*, **14**, 5066–5075.

Pomerantz,J.L. and Sharp,P.A. (1994) Homeodomain determinants of major groove recognition. *Biochemistry*, **33**, 10851–10858.

Qian,Y.Q., Billeter,M., Otting,G., Muller,M., Gehring,W.J. and Wuthrich,K. (1989) The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell*, **59**, 573–580.

Qian,Y.K., Furukubo-Tokunaga,K., Muller,M., Resendez-Perez,D., Gehring,W.J. and Wuthrich,K. (1994) Nuclear magnetic resonance solution structure of the fushi tarazu homedomain from *Drosophila* and comparison with the antennapedia homeodomain. *J. Mol. Biol.*, **238**, 333–345.

Sauer,R.C., Yocum,R.R., Doolittle,R.F., Lewis,M. and Pabo,C.O. (1982) Homology among DNA-binding proteins suggests use of conserved super-secondary structure. *Nature*, **298**, 447–451.

Seeman,N.C., Rosenberg,J.M. and Rich,A. (1976) Sequence specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.

Shen,W.F., Chang,C.P., Rozenfeld,S., Sauvageau,G., Humphries,R.K., Lu,M., Lawrence,H.J., Cleary,M.L. and Lragman,C. (1996) Hox homeodomain proteins exhibit selective complex stabilities with Pbx and DNA. *Nucleic Acids Res.*, **24**, 898–906.

Smith,L.D. and Johnson,A.D. (1992) A molecular mechanism for combinational control in yeast: MCM 1 protein sets the spacing and

orientation of the homeodomains of an alpha 2 dimer. *Cell*, **68**, 133–142.

Smith,R.L., Redd,M.J. and Johnson,A.D. (1995) The tetratricopeptide repeats of Ssn6 interact with the homeo domain of α2. *Genes Dev.*, **9**, 2903–2910.

Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendroff,J.W. (1991) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.

Suzuki,M. (1994) A framework for the DNA–protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. *Structure*, **2**, 317–326.

Suzuki,M., Brenner,S.E., Gerstein,M. and Yagi,N. (1995) DNA recognition code of transcription factors. *Prot. Engng*, **8**, 319–328.

Tang,S. and Breitman,M.L. (1995) The optimal binding sequence of the HOX 11 protein contains a predicted recognition core motif. *Nucleic Acids Res.*, **23**, 1928–1935.

Treisman,J., Gonczy,P., Vashinshtha,M., Harris,E. and Desplan,C. (1989) A single amino acid can determine the DNA binding specificity of homeodomain proteins. *Cell*, **59**, 553–562.

Tsao,D.H.H., Gruschus,J.M., Wang,L., Nirenberg,M. and Ferretti,J.A. (1995) The three dimensional solution structure of the NK-2 homeodomain from *Drosophila*. *J. Mol. Biol.*, **251**, 297–307.

Vershon,A.K., Jin,Y. and Johnson,A. (1995) A homomolecular protein lacking specific side chains of helix 3 can still bind DNA and direct trascriptional repression. *Genes Dev.*, **9**, 182–192.

Wilson,D.S., Sheng,G., Lecuit,T., Dostatni,N. and Desplan,C. (1993) Cooperative dimerization of paired class homeo domains on DNA. *Genes Dev.*, **7**, 2120–2134.

Wilson,D.S., Guenther,B., Desplan,C. and Kuriyan,J. (1995) High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell*, **82**, 709–719.

Wolberger,C. (1996) Homeodomain interactions. *Curr. Opin. Struct. Biol.*, **6**, 62–68.

Wolberger,C., Vershon,A.K., Liu,B., Johnson,A.D. and Pabo,C.O. (1991) Crystal structure of a MATα2 homeodomain–operator complex suggests a general model for homeodomain–DNA interactions. *Cell*, **67**, 517–528.

Zappavigna,V., Sartori,D. and Mavilio,F. (1994) Specificity of HOX protein function depends on DNA–protein and protein–protein interactions both mediated by the homeodomain. *Genes Dev.*, **8**, 732–744.

Zeng,W., Andrew,W., Mathies,L.D., Horner,N.A. and Scott,M.P. (1993) Ectopic expression and function of the Antp and Scr homeotic genes: the N-terminus of the homeodomain is critical to functional specificity. *Development*, **118**, 339–352.

Received on April 24, 1996; revised on May 29, 1996