The EMBO Journal vol.13 no.15 pp.3551-3560, 1994

The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins

Stephen C.Ekker, Donald G.Jackson, Doris P.von Kessler, Benjamin I.Sun, Keith E.Young and Philip A.Beachy

The Johns Hopkins School of Medicine, Howard Hughes Medical Institute, Baltimore, MD, USA

Communicated by M.Bienz

The homeodomain has been implicated as a major determinant of biological specificity for the homeotic selector (HOM) genes. We compare here the DNA sequence preferences of homeodomains encoded by four of the eight *Drosophila* HOM proteins. One of the four, *Abdominal-B*, binds preferentially to a sequence with an unusual 5′-T-T-A-T-3′ core, whereas the other three prefer 5′-T-A-A-T-3′. Of these latter three, the *Ultrabithorax* and *Antennapedia* homeodomains display indistinguishable preferences outside the core while *Deformed* differs. Thus, with three distinct binding classes defined by four HOM proteins, differences in individual site recognition may account for some but not all of HOM protein functional specificity. We further show that amino acid residues within the N-terminal arm are responsible for the sequence specificity differences between the *Ultrabithorax* and *Abdominal-B* homeodomains. Similarities and differences at the corresponding positions within the N-terminal arms are conserved in the vertebrate *Abdominal-B*-like HOM proteins, which play critical roles in limb specifications as well as in regional specification along the anterior–posterior axis. This and other patterns of residue conservation suggest that differential DNA sequence recognition may play a role in HOM protein function in a wide range of organisms.

Key words: development/DNA sequence recognition/ *Drosophila*/homeodomain

Introduction

Homeotic selector (HOM) genes are responsible for specifying the identity of spatial units along the anterior–posterior body axis of multicellular organisms, from insects to mammals (see McGinnis and Krumlauf, 1992 for review). In *Drosophila*, these spatial units correspond to segments or parasegments, and HOM mutations produce transformations of segment identity (homeotic transformations). HOM genes are proposed to specify positional identities through transcriptional modulation of specific sets of downstream genes (reviewed in McGinnis and Krumlauf, 1992). Functional studies in *Drosophila* have implicated the homeodomain, a highly conserved 61 amino acid region found in each of these proteins, as an important determinant of the specificity of this modulation (see Hayashi and Scott, 1990 for review).

The homeodomain is an autonomous, sequence-specific DNA binding domain (Mihara and Kaiser, 1988; Affolter *et al.*, 1990; Percival-Smith *et al.*, 1990; Ekker *et al.*, 1991, 1992; Florence *et al.*, 1991) with a helix–turn–helix motif that resembles the structure of some prokaryotic sequence-specific DNA binding positions (Otting *et al.*, 1988; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991; Klemm *et al.*, 1994). Early characterization of DNA sequence recognition indicated a high degree of promiscuity in the binding properties of homeodomain proteins (e.g. Desplan *et al.*, 1988; Hoey and Levine, 1988), which was probably due to a 4 bp DNA sequence element (the 'TAAT-core') present in many homeodomain binding sites (see Hayashi and Scott, 1990 for review). More recent studies have begun to resolve distinctions in the binding properties of homeodomain proteins and have suggested that differential DNA sequence recognition plays some role in the determination of biological specificity (Dessain *et al.*, 1992; Ekker *et al.*, 1992; Jones and McGinnis, 1993).

A comparison of the *Drosophila* HOM proteins encoded by *Ultrabithorax* (*Ubx*) and *Deformed* (*Dfd*) indicated that differential DNA sequence recognition involves base pairs flanking a common 5′-TAAT-3′ core recognition sequence (Dessain *et al.*, 1992; Ekker *et al.*, 1992). Simple target sequences in yeast were transactivated differently by these proteins, and the differences correlated well with the observed differences in DNA sequence recognition *in vitro* (Ekker *et al.*, 1992). More extensive mapping studies using proteins chimeric for regions within the homeodomains of *Ubx* and *Dfd* showed that determinants of differential binding specificity *in vitro* (Ekker et al., 1992) correspond to the same region of the homeodomain responsible for determining specificity in *Drosophila* embryos (Lin and McGinnis, 1992).

We present here the systematic characterization of DNA sequence recognition properties for two additional *Drosophila* HOM members, *Abdominal-B* (*Abd-B*) and *Antennapedia* (*Antp*), and a direct comparison of these properties with those of *Dfd* and *Ubx*. *Abd-B* is unusual in this group in binding preferentially to sequences including a 5′-TTAT-3′ core, where the other three prefer a 5′-TAAT-3′ core. Among these latter three, the *Ubx* and *Antp* homeodomains display nearly identical preferences outside the core while the *Dfd* homeodomain differs. Amino acid residue differences within the N-terminal arm are responsible for the major differences in sequence recognition between *Abd-B* and *Ubx*, as demonstrated by the conversion of *Ubx* sequence specificity when three N-terminal residues are altered. Two additional *Drosophila* HOMs also carry contact residue differences and thus appear likely to show distinct DNA sequence preferences. We discuss how such differences in DNA sequence recognition might contribute to the biological specificity of *Drosophila* HOM proteins and we consider alternative mechanisms that may distinguish the biological activities of *Ubx* and *Antp*, which display similar DNA sequence recognition properties.

A

```
             N-terminal                                                  C-terminal
               arm     HELIX 1           HELIX 2           HELIX 3       extension
             1 * * *   10         22   28       37   42   *  **  *   58             71
UBX HD       RRRGRQTYT RYQTLELEKEFHT NHYLT RRRRIEMAHA LCLT ERQIKIWFQNRRMKLKK EIQAIKELNEQEK

DFD HD       PK-Q-TA-- -H-I--------Y -R--- ------I--T -V-S --------------W-- DNKLPNTK-VRK-

ANTP HD      -K------- ------------F -R--- ------I--- ---- --------------W-- -NKTKG-PGSGGE

ABD-B HD     V-KK-KP-S KF---------LF -A-VS KQK-W-L-RN -Q-- ---V----------N-- NS-RQANQQNNNN

UBX K3       --K------ ------------- ----- ---------- ---- ----------------- -------------

UBX K3/K6/P7 --K--KP-- ------------- ----- ---------- ---- ----------------- -------------
```

B

```
Selection
 oligo-
nucleotide   Site                        Insert sequence                              Site     Strand

    I     (HindIII) 5'- N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  -3'     (EcoRI)     +
                    3'- N' N' N' N' N' N' N' N' N' N' N' N' N' N' N' N' N' N' -5'                 -

    II      (NotI)  5'- N  N  N  N  N  N  N  T  A  A  T  N  N  N  N  N  N  N  -3'     (BamHI)     +
                    3'- N' N' N' N' N' N' N' A  T  T  A  N' N' N' N' N' N' N' -5'                 -

    III     (NotI)  5'- N  N  N  N  N  N  N  T  T  A  T  N  N  N  N  N  N  N  N  -3'  (BamHI)     +
                    3'- N' N' N' N' N' N' N' A  A  T  A  N' N' N' N' N' N' N' N' -5'              -

    IV      (NotI)  5'- N  N  N  N  N  N  T  T  N  A  T  G  N  N  N  N  N  N  N  -3'  (BamHI)     +
                    3'- N' N' N' N' N' N' A  A  N' T  A  C  N' N' N' N' N' N' N' -5'              -

 Position                    -5 -4 -3 -2 -1  1  2  3  4  5  6  7  8  9  10 11 12
```

Fig. 1. Proteins and selection oligonucleotides used in this study. (**A**) Homeodomain peptides used in this study. The amino acid sequences of each homeodomain are shown in relation to UBX HD (Ekker *et al.*, 1991), with identity indicated by a hyphen. The numbering scheme and the positions of α-helices correspond to those of the *engrailed* homeodomain (Kissinger *et al.*, 1990). An asterisk denotes a potential sequence-specific contact residue. Because Edman degradation and sequence analysis of the first 10 amino acid residues of ABD-B HD detected no N-terminal methionine, we do not include this residue in any of the protein sequences. (**B**) The insert sequences of the selection oligonucleotides are shown with respect to their flanking cloning sites; other flanking sequences are given in Materials and methods. A random base within the insert region is indicated by an N for the (+) strand or by an N′ for the complementary (−) strand. The bottom three oligonucleotides are aligned according to the constant bases they contain, with the numbering scheme shown below.

Results

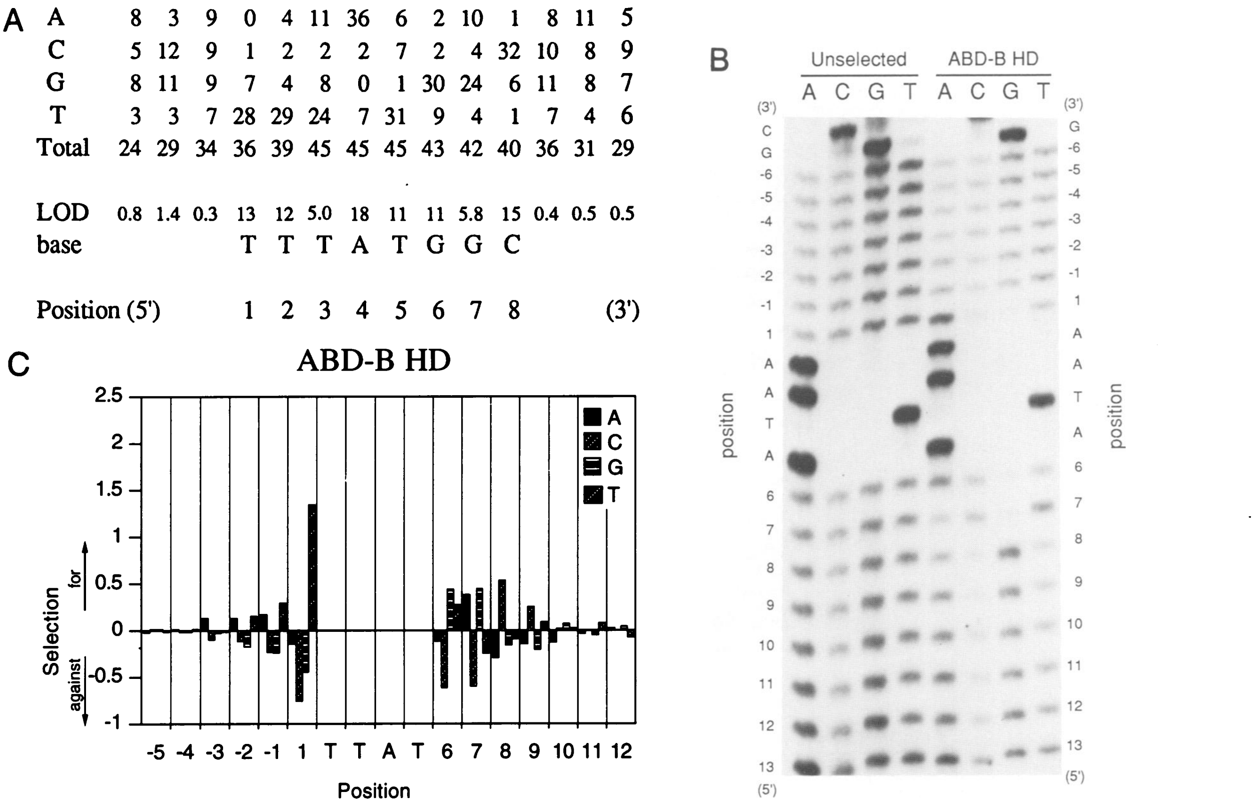
Homeodomain peptides

Earlier studies with *Ubx* demonstrated that sequence preferences of the full length protein are accurately reflected by the preferences of the more readily purified homeodomain peptide. Our binding studies of the *Ubx*, *Dfd*, *Abd-B* and *Antp* proteins therefore utilized homeodomain peptides (UBX HD, DFD HD, ABD-B HD and ANTP HD, respectively; see Figure 1A) purified to near homogeneity as described in Materials and methods. All four of these proteins contain the 61 amino acids of the homeodomain and an additional 10 C-terminal residues. This extension includes regions of extended inter-species homology for *Ubx* (Wysocka-Diller *et al.*, 1989) and *Dfd* (Regulski *et al.*, 1987), and encompasses the regions shown to be sufficient for altering the specificity of *Dfd* to that of either *Ubx* or *Abd-B* (Kuziora and McGinnis, 1989, 1991) in *Drosophila* embryos. Two additional homeodomain proteins (UBX K3 and UBX K3/K6/P7; see Figure 1A) were tested to identify residues responsible for differences in DNA sequence recognition between *Ubx* and *Abd-B*.

The consensus binding site sequence for ABD-B HD incorporates a non-TAAT core

The experiments presented here use sequence selection to identify optimal binding sites for the proteins of interest. This method involves identification of high affinity binding sites by protein-dependent selection from a pool of random sequence DNA, followed by alignment of these sequences to generate a consensus DNA binding site. Our initial experiments with ABD-B HD used selection oligonucleotide I (see Figure 1B), which contained an 18 bp stretch of random sequence flanked by specific sequences for cloning and amplification by PCR. Two rounds of selection for high-affinity binding sites were performed using a matrix containing covalently bound ABD-B HD and selection oligonucleotide I as described in Ekker *et al.* (1991). The sequences of 45 individual clones were aligned optimally using an automated algorithm (E.D.Perez-Albuerne, unpublished) that was first tested on data previously reported for the *Ubx* homeodomain (Ekker *et al.*, 1991). The *Ubx* consensus sequence derived from this automated alignment method (5′-TTAATGG-3′) matches the optimal sequence (5′-TTAATGGCC-3′; Ekker *et al.*, 1992) at the seven most highly constrained positions within the optimal site. Application of this alignment algorithm to the ABD-B HD data identified an eight base working consensus of 5′-TTTATGGC-3′ (Figure 2A).

When aligned with the *Ubx* and *Dfd* optimal sites, the *Abd-B* consensus contains a TTAT sequence at positions corresponding to the *Ubx* and *Dfd* TAAT core. We tested this difference in dissociation rate experiments (Table I,



A	A	8	3	9	0	4	11	36	6	2	10	1	8	11	5
	C	5	12	9	1	2	2	2	7	2	4	32	10	8	9
	G	8	11	9	7	4	8	0	1	30	24	6	11	8	7
	T	3	3	7	28	29	24	7	31	9	4	1	7	4	6
	Total	24	29	34	36	39	45	45	45	43	42	40	36	31	29
	LOD	0.8	1.4	0.3	13	12	5.0	18	11	11	5.8	15	0.4	0.5	0.5
	base				T	T	T	A	T	G	G	C			
	Position (5')				1	2	3	4	5	6	7	8			(3')

Fig. 2. ABD-B HD DNA sequence preferences. (**A**) Identification of a 5′-TTAT-3′ core consensus binding site for ABD-B HD. Two rounds of selection were performed using oligonucleotide I (see Figure 1B), and the sequences of 45 individual clones were aligned and tabulated. Alignment was done with the computer-based algorithm of E.D.Perez-Albuerne (unpublished). Only the random core of each sequence was included in this tabulation. At each position, the number of occurrences of the four bases is shown along with an estimate of the degree of skewing from random expectation. This estimate was derived as described (Ekker *et al.*, 1991) and is given as a probability in the form of a LOD score. (A LOD of 3 indicates a probability of one in 10^3.) The derived consensus binding site sequence is shown. (**B**) Base preferences at positions flanking the core. Sequence analysis after three rounds of enrichment for ABD-B HD and for unselected DNA using selection oligonucleotide III (see Figure 1B). The identities of the fixed bases are indicated, with numbers identifying the random sequence positions of the oligonucleotide. The (−) strand was sequenced. (**C**) Sequence preference histogram. Preferences for and against A, C, G and T are indicated by bars extending above or below zero respectively. The data are presented in relation to the (+) strand and were obtained by quantitative analysis of (B) as described in Ekker *et al.* (1992) and in Materials and methods.

sequences a−d; see Materials and methods), with the results indicating a T base preference by ABD-B HD at the second position within the core (position 3 in our numbering scheme). In addition, selection experiments using oligonucleotide IV (see Figure 1B) independently confirmed this preference for a non-TAAT core (Figure 3; described more fully below).

ABD-B HD preferences at positions flanking the core

Oligonucleotide III (5′-N_7TTATN_8-3′; see Figure 1B) was used in selection experiments with ABD-B HD to confirm and resolve base preferences flanking the 5′-TTAT-3′ core sequence. These experiments were performed essentially as those described for *Ubx* and *Dfd* proteins (Ekker *et al.*, 1992). After three rounds of selection, the pool was sequenced and quantified in relation to unselected control DNA to identify base preferences outside the defined core (see Figure 2B and Materials and methods). We noted strong sequence preferences at positions 1, 6, 7 and 8. These data yielded an 8 bp consensus sequence of 5′-T-T-T-A-T-G>T-G>A-C-3′ for ABD-B HD which was in good agreement with the initial consensus sequence determination of Figure 2A. The selection preferences at position 6 were further confirmed by measurements of complex stability with an oligonucleotide series that contained each of the four bases at position 6 (sequences e−h, Table I). The weak preferences at positions −1 and 9 were confirmed by further selection experiments (see Figure 3 below).

Table I. Dissociation rates of ABD-B HD complexes with various DNA sequences

	Position											k_d × 1000	$t_{1/2}$	$t_{1/2}$
	(5′) −1	1	2	3	4	5	6	7	8	9	(3)	(min^{-1})	(min)	(rel)
a	T	T	T	T	A	T	G	G	C	C		1.4 ± 0.1	495	1.00
b	–	–	–	A	–	–	–	–	–	–		4.4 ± 0.5	158	0.32
c	–	–	–	G	–	–	–	–	–	–		21.3 ± 0.7	33	0.07
d	–	–	–	C	–	–	–	–	–	–		27 ± 2	26	0.05
e	T	T	T	A	A	T	G	G	C	T		4.0 ± 0.4	173	0.35
f	–	–	–	–	–	–	T	–	–	–		6.4 ± 0.2	108	0.22
g	–	–	–	–	–	–	A	–	–	–		7 ± 2	99	0.20
h	–	–	–	–	–	–	C	–	–	–		16 ± 5	43	0.09

Dissociation rate constants (k_d) and half-lives ($t_{1/2}$) were determined as described in Ekker *et al.* (1992). The k_d values are given as an average of two independent determinations ± the standard error.

A scheme for direct comparison of four homeotic homeodomains

The sequence 5′-TTNATG-3′ is common to the preferred homeodomain binding sites of *Abd-B* (this work), *Ubx*

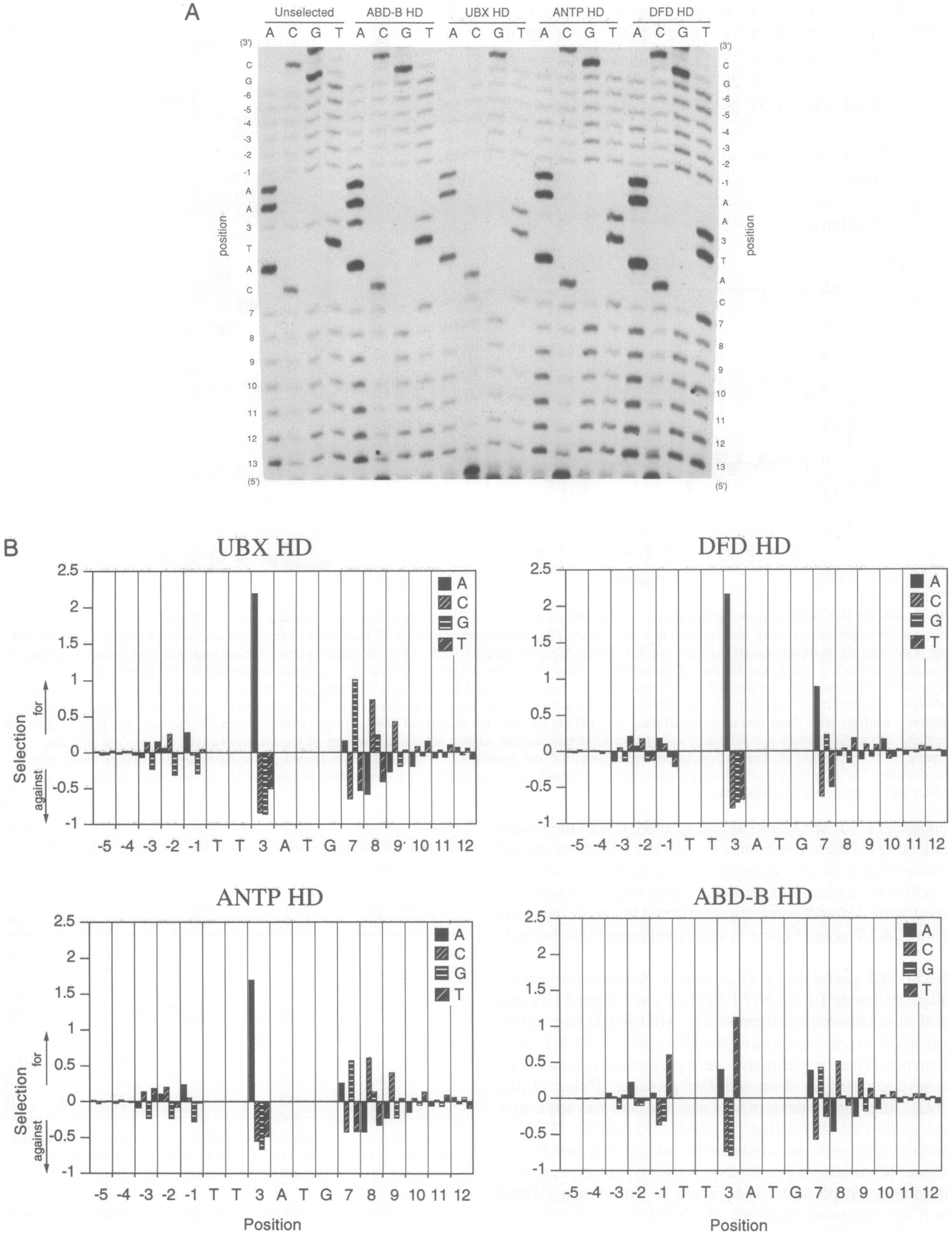


Fig. 3. Sequence preferences of four homeodomains encoded by the homeotic genes *Abd-B*, *Ubx*, *Antp* and *Dfd*. (**A**) Sequence analysis of selected oligonucleotide pools. Selection oligonucleotide IV (Figure 1B) was subjected to three rounds of selection with each of the homeodomain peptides ABD-B HD, UBX HD, ANTP HD and DFD HD. Fixed bases are identified and random bases are numbered by position. The (−) strand was sequenced. (**B**) Sequence preference histograms. Preferences are as described in the legend to Figure 2C and are shown with reference to the (+) strand (fixed bases 5′-TTNATG-3′).

(Ekker *et al.*, 1991, 1992) and *Dfd* (Ekker *et al.*, 1992). In addition, we examined a purified peptide containing the homeodomain of *Antp* (ANTP HD; see Figure 1A). Oligonucleotide IV, which contains this sequence (Figure 1B), was used in sequence selection experiments to test explicitly for differences in DNA sequence preference by these four homeodomains. As above, the oligonucleotide pools were sequenced after three rounds of selection or, as a control, after amplification without selection (Figure 3A). Sequence preference histograms derived from a quantification of these data are shown in Figure 3B.

As seen in the top two histograms of Figure 3B, *Ubx* and *Dfd* proteins prefer an A at position 3 (the second position within the TAAT core), thus confirming the high TAAT core specificity reported earlier for these two proteins (Ekker *et al.*, 1991, 1992). Due to the fixed T at position 1, the experiments in Figure 3 could not have detected the differences reporter earlier at this position for *Dfd* and *Ubx* (Ekker *et al.*, 1992). To the 3′ side of the core we noted that UBX HD prefers a 5′-G>A-C-3′ dinucleotide at positions 7 and 8 while DFD HD exhibits a highly constrained A>G sequence preference at position 7 with no clear preference at position 8. No clear sequence preference for DFD HD was resolved at position 9, probably due to less stringent selection conditions imposed in this as compared with previous experiments (Ekker *et al.*, 1992). The results of our selections with UBX HD and DFD HD are consistent with our previous studies, and they validate direct comparisons based on selections using this oligonucleotide (see below).

The optimal Abd-B binding site

The ABD-B HD preference histogram (bottom right, Figure 3B) indicates a 10-base region of sequence preference, with clearly resolved preferences at positions −1, 3, 7, 8 and 9. Note in particular how ABD-B HD prefers a T at positions −1 and 3; these preferences are distinct from those of *Ubx* and *Dfd* (Figure 3B). In contrast, we observed that the sequence preferences of *Abd-B* at positions 7, 8 and 9 are similar to those of *Ubx* and *Antp*.

Selection experiments (Figures 2B and C and 3) and binding measurements (Table I) confirm and extend the working consensus derived earlier from ABD-B HD (Figure 2A). The order of base preference at position 3 is T>A>G>C in both binding measurements (sequences a−d in Table I) and selection experiments (Figure 3B). The binding measurements (Table I, sequences e−h) and selection data (Figure 3) are also identical with respect to the order of base preference at position 6. These selection experiments also extend the binding site sequence to include bases at positions −1 and 9. We therefore conclude that the 10 bp sequence (5′-T-T-T-T-A-T-G-G-C-C-3′) is the optimal DNA binding site for the *Abd-B* homeodomain.

Antp and Ubx homeodomains share the same DNA sequence preferences

Specific binding to the DNA sequence 5′-TAATG-3′ had been observed with *Antp* homeodomain proteins in biochemical (Affolter *et al.*, 1990) and structural (Otting *et al.*, 1990) studies. We confirmed and extended this sequence preference data using selection oligonucleotides II and IV (see Figure 1). In each of these experiments the ANTP HD showed sequence preferences very similar to those of UBX HD (see Figure 3; data for oligonucleotide II not shown). Some apparent quantititative differences can be observed between these proteins: compare, for example, the extent of *Ubx* and *Antp* preference for an A at position 3 (Figure 3B). Such differences may be due to the varying degrees of binding site enrichment obtained for these two homeodomains, a parameter that is difficult to control and measure. Conclusions from selection experiments such as these must therefore be restricted to the qualitative DNA sequence preferences of binding proteins, while quantitative interpretations must rely on binding measurements with individual sequences. At a qualitative level, the *Ubx* and *Antp* homeodomains display indistinguishable sequence preferences and appear to share the 9 bp optimal DNA binding site of 5′-T-T-A-T-G>T-G>A-C-C-3′. These results are not surprising given that these two proteins carry identical residues at positions responsible for DNA sequence recognition (see Figures 1A and 5).

The N-terminal arm is responsible for differences in DNA sequence recognition between Ubx and Abd-B

The binding of even the most diverged homeodomains to DNA involves an invariant interaction between N51 and an A:T base pair (Wolberger *et al.*, 1991). By this and other criteria the homeodomain binding sites of HOM proteins can be aligned with the structural models of closely related (Kissinger *et al.*, 1990; Otting *et al.*, 1990) as well as more diverged (Wolberger *et al.*, 1991; Klemm *et al.*, 1994) homeodomains. These alignments indicate that the N-terminal arm of the homeodomain contacts DNA at the 5′ side of our binding sites. For example, residues 3 and 5 (Kissinger *et al.*, 1990) and 7 (Wolberger *et al.*, 1991) appear to make base-specific contacts in the minor groove, while residue 6 contacts the sugar−phosphate backbone (Kissinger *et al.*, 1990; Klemm *et al.*, 1994). The details of these contacts are not well resolved; these studies nevertheless provide a general model applicable to most homeodomains.

Among the homeodomains studied here, residue 5 remains constant while residues 3, 6 and 7 differ (Figure 1A). To determine the roles of these residues in specifying *Ubx* versus *Abd-B* sequence preferences, we analyzed two *Ubx* homeodomains with residue switches at position 3 alone or at positions 3, 6 and 7 (UBX K3 and UBX K3/K6/P7; Figure 1A). The binding preferences shown in Figure 4 were determined by parallel selection experiments on oligonucleotide IV (Figure 1B) with UBX HD, ABD-B HD and the two modified *Ubx* homeodomains. This analysis indicated that UBX K3 retains *Ubx*-like binding preferences (note the strong preference for an A at position 3 and the lack of a clear preference at position −1) while the UBX K3/K6/P7 binding preferences resemble those of *Abd-B* (note the strong preference for a T at position −1 and the dual preference for a T or an A at position 3). Residues 6 and 7 thus appear to play a decisive role in sequence specificity of the N-terminal arm (see Discussion).

Discussion

The use of chimeric *Drosophila* genes (Kuziora and McGinnis, 1989, 1991; Gibson *et al.*, 1990; Furukubo-Tokunaga *et al.*, 1993; Zeng *et al.*, 1993) or vertebrate homologs (McGinnis *et al.*, 1990; Malicki *et al.*, 1990; Zhao *et al.*, 1993) in ectopic expression studies has demonstrated that the primary determinant of segmental specificity is the homeodomain plus several flanking residues

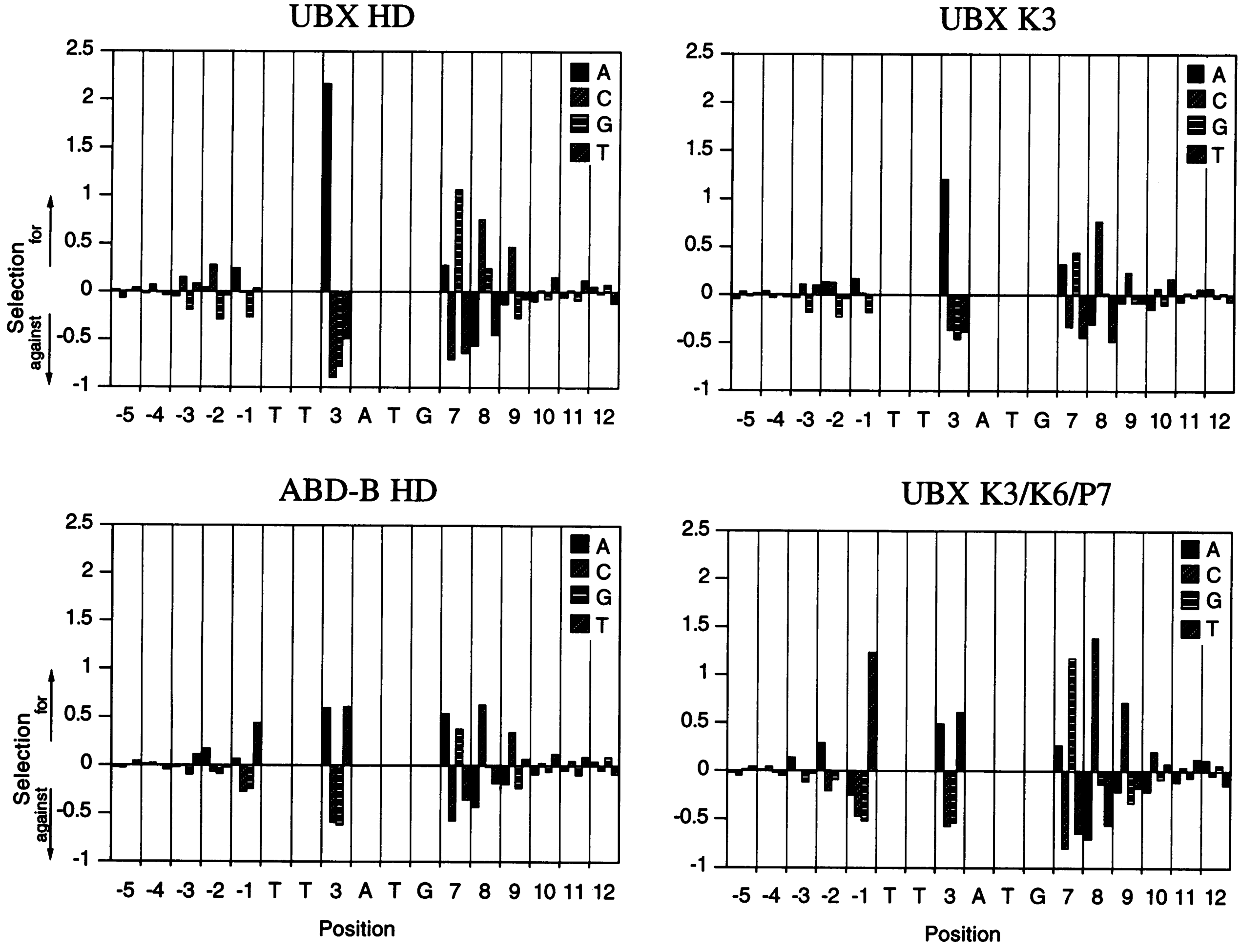


Fig. 4. Sequence preferences of ABD-B HD, UBX HD and two modified *Ubx* homeodomains. Amino acid residues from position 3 (UBX K3) or positions 3, 6 annd 7 (UBX K3/K6/P7) of *Ubx* were changed to those of *Abd-B*. Selections were done with oligonucleotide IV (Figure 1B) as described in Figure 3 and Materials and methods, and the resulting sequence preference histograms are shown. Preferences are as described in the legend to Figure 2C and are shown with reference to the (+) strand (fixed bases 5′-TTNATG-3′).

(reviewed by Hayashi and Scott, 1991). These experiments showed, for example, that when the homeodomain and five C-terminal amino acid residues of *Dfd* were replaced by the corresponding regions of *Ubx* (Kuziora and McGinnis, 1989) or *Abd-B* (Kuziora and McGinnis, 1991) the resulting chimeric proteins lost the ability to activate an endogenous *Dfd* target gene and instead activated *Ubx*- or *Abd-B*-specific target gene promoters. The homeodomains of *Dfd*, *Ubx* and *Abd-B* thus have distinct biological properties. In the case of *Ubx* and *Dfd*, more extensive mapping studies using proteins chimeric for regions within the homeodomains showed that determinants of DNA binding specificity *in vitro* (Ekker *et al.*, 1992) correspond to the determinants of segmental specificity *in vivo* (Lin and McGinnis, 1992). We concluded from these studies that differential DNA sequence recognition by *Dfd* and *Ubx* homeodomains may contribute to the specificity of homeotic gene action. Our current work shows that *Abd-B* homeodomain has distinct DNA sequence recognition properties and that differential DNA sequence recognition therefore could play a role in *Abd-B* biological specificity.

Structural basis for the distinctive Abd-B specificity

The base preferences of the modified *Ubx* homeodomains permit certain conclusions regarding the physical basis of the distinctive *Abd-B* specificity. The strong base preference of *Abd-B* at position −1 appears to depend on residues 6 and 7, since it is observed in the triple but not the single switch mutant; this observation is consistent with structural models in which residue 7 contacts bases −1 and 1 (Wolberger *et al.*, 1991). Structural models also indicate that base 3 (base 2 within the core) is contacted by residue 3 (Kissinger *et al.*, 1990). Our results demonstrate, however, that the base preference at position 3 does not depend simply on the identity of residue 3, since either K3 or R3 in a *Ubx* context prefers an A, while K3 in conjunction with K6 and P7 (the *Abd-B* residues) displays the dual preference for A and T characteristic of *Abd-B*. Residues 6 and 7 thus appear to influence base preferences throughout the region contacted by the N-terminal arm, in part perhaps by specifying the conformation of the arm within the minor groove. Such a role is consistent with the backbone contacts made by residue 6 in the *en* and Oct-1 complexes (Kissinger *et al.*, 1990; Klemm *et al.*, 1994). The importance of N-terminal arm conformation is further suggested by a relaxed preference of the Oct-1 homeodomain for an A or a T base at position 2 (position 1 of the core; Verrijzer *et al.*, 1992), while other homeodomains that share the R5 base-contact residue with Oct-1 display a strict preference for a T. A more precise understanding of N-terminal arm confirmation, how it is specified, and how sequence specificity is affected must await higher resolution studies of the Oct-1 and other homeodomain DNA complexes.

A non-TAAT core preference for Hox genes involved in limb specification

Within the ancestral vertebrate HOM cluster, the *Abd-B*-like gene appears to have duplicated to five copies prior to duplication of the entire cluster (Schubert *et al.*, 1993). Thus, in the mouse, the current arrangement of HOM genes in four clusters includes 15 *Abd-B*-like genes. Among these 15 genes are *Hox C9–C13* and *Hos D9–D13* (nomenclature according to Scott, 1992), which have been implicated in specification of positional information in vertebrate limb development along the proximal–distal and anterior–posterior axes, respectively (see Tabin, 1992 for review). Remarkably, the residues present at positions 3 and 7 in the *Drosophila Abd-B* homeodomain are conserved in all of the vertebrate *Abd-B*-like genes. For two of these, *Hox D9* and *D10*, preliminary binding studies have identified binding sites containing TTAT (Arcioni *et al.*, 1992). The vertebrate *Abd-B*-like genes, including those important in limb specification, thus appear likely to bind preferentially to non-TAAT core sequences.

Identical DNA recognition properties for some Drosophila HOM proteins

The similarity in sequence preference between the *Ubx* and *Antp* homeodomains is not surprising given their overall similarity in amino acid sequence and their identity at all positions presumed to make base-specific contacts in the major or minor grooves. The homeodomain of one other *Drosophila* HOM, *abdominal-A* (*abd-A*), is also similar in overall amino acid sequence and identical at residues making base-specific contacts; we therefore expect that it will display a DNA sequence preference similar to that of *Ubx* and *Antp*. What then accounts for the differences in segmental specificity of these three proteins?

The recent work of Chan and Mann (1993) with ectopically expressed chimeric proteins suggests that sequences C-terminal to the homeodomain may play a role in the segmental specificity of *Ubx* and *Antp* function. An interesting feature of these proteins is that C-terminal sequences of the *Ubx* and *abd-A* but not the *Antp* proteins are predicted to form coiled coil structures (Lupas *et al.*, 1991; Beachy *et al.*, 1993). In addition, full length *Ubx* protein binds cooperatively to multiple individual DNA sites in a fashion dependent upon the presence of sequences outside the homeodomain, including the C-terminus (Beachy *et al.*, 1993). One possible source of segmental specificity for HOM proteins with identical DNA recognition properties thus might be in protein interaction surfaces encoded at positions outside the homeodomain. The presence or absence of such surfaces might influence the binding behavior of HOM proteins to multiple sites, or possibly their interactions with other non-HOM proteins. We note that the coiled coil predicted for *Ubx* is 43 residues in length, slightly more than half the size of that for *abd-A*, which is predicted to encompass 71 residues (Beachy *et al.*, 1993). Differences in biological specificity of these two proteins therefore might rely on the size difference or on other properties of the predicted coiled coils.

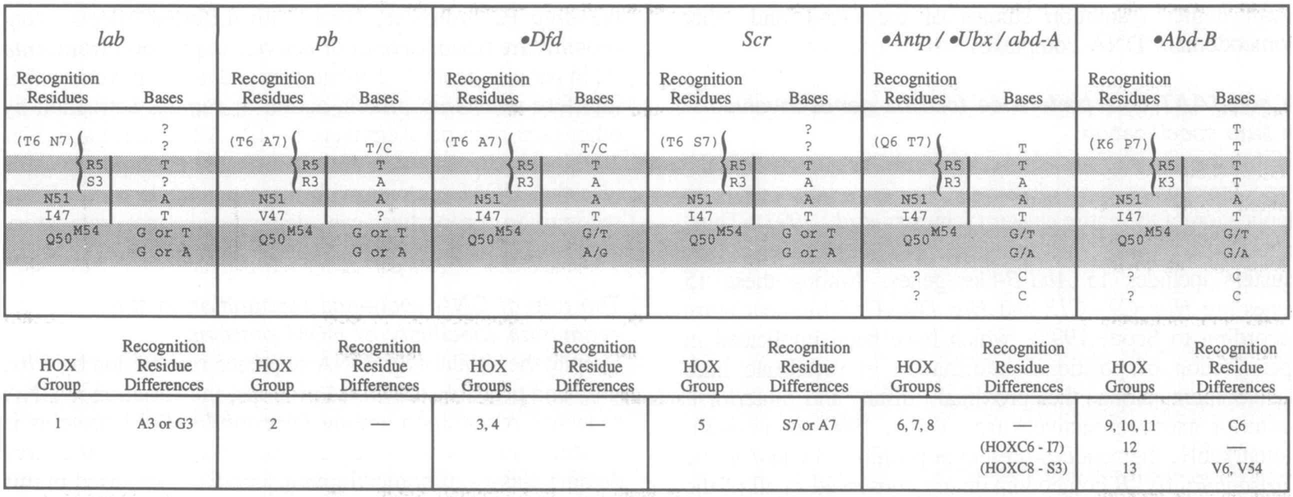
Genetic and biochemical evidence from Appel and Sakonju (1993) indeed demonstrates that all three of these proteins act upon 30 binding sites within a 2.3 kb DNA segment near the *Antp* P2 promoter, albeit with different effects. *Antp* proteins are permissive of or activate expression from *Antp* P2 in parasegments 3–5 while *Ubx* and *abd-A* proteins either interfere with *Antp* protein action or repress activation by other factors in parasegments 6–12. All of these activities are blocked by mutation of the binding sites, supporting the idea that HOM proteins with the same fundamental DNA recognition properties can generate distinct regulatory outcomes at a particular promoter.

The role of DNA sequence recognition in the segmental specificity of HOM proteins

Despite the similarity in DNA sequence recognition by *Ubx*, *Antp* and probably the *abd-A* proteins, the variation in DNA sequence recognition among *Drosophila* HOM proteins is greater than previously appreciated. In addition to the three distinct classes of homeodomain specificity reported in this work (*Dfd*, *Ubx*/*Antp* and *Abd-B*), three additional *Drosophila* HOM proteins differ within the homeodomain at positions corresponding to residues important for DNA sequence recognition: *proboscipedia* (*pb*) has a valine instead of the TAAT core-contacting isoleucine at position 47 (Cribbs *et al.*, 1992), *labial* (*lab*) has a unique set of residues at 3, 6 and 7 (S3, T6 and N7; Diederich *et al.*, 1989) and *Sex combs reduced* (*Scr*), while very similar to the *Ubx*/*Antp* group in overall sequence, differs at residues 6 and 7 (T6 and S7; LeMotte *et al.*, 1989). Consistent with a role for residues 6 and 7 in generating a novel specificity for *Scr*, we note that recent studies with chimeric homeodomains indicate that these residues may be important for biological specificity (Furukubo-Tokunaga *et al.*, 1993; Zeng *et al.*, 1993). We therefore suggest that *pb*, *lab* and *Scr* proteins may represent DNA sequence recognition classes distinct from the three DNA sequence recognition classes (*Dfd*, *Ubx*/*Antp* and *Abd-B*) identified biochemically for *Drosophila* HOM proteins.

All of the *Drosophila* HOM protein recognition classes are summarized in Figure 5 along with their actual or proposed optimal sites and the relevant sequence recognition residues. With the exception of residue 6, residues contacting the sugar–phosphate backbone are not included in Figure 5 because they are largely conserved; we note, however, that differences in backbone-contacting residues may alter specificity indirectly by influencing orientation or position of base-contacting residues (as seems likely for residue 6; see above) or by detecting base-dependent features of the DNA conformation. We also note that a difference in contact residue does not necessarily indicate a difference in base recognition since different amino acid side chains may resemble each other functionally, as appears to be the case for UBX K3 when residue three is changed from an arginine to a lysine (Figure 4).

Below the actual and proposed *Drosophila* specificity classes are shown the most closely related groups of vertebrate HOM genes. The group numbers are taken from the nomenclature of Scott (1992), and each designates a group of paralogous vertebrate HOM genes. Note that not all base-specific residues in vertebrate HOMs are identical to their *Drosophila* counterparts. The base-contacting residues of vertebrate group 12, for instance, are identical to those of *Abd-B*, while groups 9–11 contain C6 and group 13 contains V6 and V54. Groups 9–13 nevertheless all contain K3 and P7 which, along with other sequence features and



lab		*pb*		•*Dfd*		*Scr*		•*Antp* / •*Ubx* / *abd-A*		•*Abd-B*	
Recognition Residues	Bases	Recognition Residues	Bases	Recognition Residues	Bases	Recognition Residues	Bases	Recognition Residues	Bases	Recognition Residues	Bases
	?						?				T
(T6 N7)	?	(T6 A7)	T/C	(T6 A7)	T/C	(T6 S7)	?	(Q6 T7)	T	(K6 P7)	T
R5	T	R5	T	R5	T	R5	T	R5	T	R5	T
S3	?	R3	A	R3	A	R3	A	R3	A	K3	T
N51	A	N51	A	N51	A	N51	A	N51	A	N51	A
I47	T	V47	?	I47	T	I47	T	I47	T	I47	T
Q50^{M54}	G or T	Q50^{M54}	G or T	Q50^{M54}	G/T	Q50^{M54}	G or T	Q50^{M54}	G/T	Q50^{M54}	G/T
	G or A		G or A		A/G		G or A		G/A		G/A
								?	C	?	C
								?	C	?	C

HOX Group	Recognition Residue Differences	HOX Group	Recognition Residue Differences	HOX Groups	Recognition Residue Differences	HOX Group	Recognition Residue Differences	HOX Groups	Recognition Residue Differences	HOX Groups	Recognition Residue Differences
1	A3 or G3	2	—	3, 4	—	5	S7 or A7	6, 7, 8	— (HOXC6 - I7) (HOXC8 - S3)	9, 10, 11 12 13	C6 — V6, V54

Fig. 5. Proposed DNA sequence recognition classes for HOM proteins. Six proposed or actual DNA sequence recognition classes (see text) for *Drosophila* HOM proteins are given with their residues responsible for DNA sequence recognition listed alongside their optimal binding sites. The four proteins studied in this work are preceded by a filled circle. The vertical bar indicates the four base core sequence (e.g. 5′-TAAT-3′ for *Dfd*); shaded rows indicate recognition residues that are invariant in HOM proteins and actual or proposed bases within optimal sites. The residues at positions 6 and 7 may influence base preferences throughout the region contacted by the N-terminal arm, possibly by specifying the conformation of the arm within the minor groove of the DNA (see Discussion). Below each *Drosophila* recognition class are shown the most closely related groups of vertebrate HOM genes. Group numbers are from Scott (1992), and each number designates a group of paralogous vertebrate HOM genes. HOX group differences from *Drosophila* HOM proteins at recognition residues are specified; identities are indicated by a dash. Two unusual sequence differences for HOXC6 and HOXC8 are specifically indicated.

their positions in the clusters, identify them as *Abd-B* related. The differences at residues 6 and 54 may serve to refine or alter specificities of specific *Abd-B*-like groups and thus distinguish them from each other.

Are the demonstrated differences in DNA sequence recognition great enough to influence significantly the segmental specificity of HOM proteins? We do not wish to argue that DNA sequence recognition alone is the source of HOM protein specificity. Indeed, our demonstration of identical preferences for the *Ubx* and *Antp* homeodomains provides concrete evidence that other mechanisms must play some role. It is worth noting, however, that many of the differences between *Drosophila* HOMs at base-contacting residues are conserved in the paralogous vertebrate groups. The most striking examples are the *Abd-B*-like genes (groups 9–13), all of which carry K3 and P7 residues specific to *Abd-B* (Figure 4). With our current understanding of DNA sequence preferences and the amino acid residues responsible for differences between homeodomains, it should now be possible to design experiments with physiologically active enhancer elements that precisely address the importance of differential sequence recognition in the biological specificity of HOM proteins.

Vertebrate and *Drosophila* HOM proteins have also been shown to bind cooperatively to multiple individual sites (Galang and Hauser, 1992; Beachy *et al.*, 1993). In the case of *Ubx*, this cooperativity can lead to multi-protein complexes involving multiple clusters of sites located at some distance apart in the DNA (Beachy *et al.*, 1993). Binding cooperativity of this type suggests that functionally equivalent regulatory elements could be built from a few high affinity sites, from many lower affinity sites, or from sites of some intermediate number and affinity. The significance for specificity is that even small differences in binding of HOM proteins to individual sites could be summed to yield large overall differences in binding to multiple sites. For example, an enhancer element specific for a particular HOM protein could be constructed through the use of a large number of low affinity but highly discriminatory sites.

Consistent with these ideas, candidate elements for regulation by HOM proteins thus far isolated contain multiple individual sites distributed throughout sizable DNA regions (Regulski *et al.*, 1991; Gould and White, 1992; Vachon *et al.*, 1992; Appel and Sakonju,1993; Jones and McGinnis, 1993; Capovilla *et al.*, 1994). In addition, multiple individual sites, some of them weak, have been established as important for the action of other non-HOM homeodomain proteins (Driever *et al.*, 1989; Struhl *et al.*, 1989; Schier and Gehring, 1992, 1993; Small *et al.*, 1992). We wish to emphasize that a multi-site integrative model of HOM protein action such as the one described could operate autonomously or in conjunction with other specificity-enhancing mechanisms (e.g. HOM protein interactions with DNA binding partner proteins). In addition, this type of model indicates that optimal binding site determinations such as those reported here do not predict or constrain the individual binding site sequences used *in vivo*, although they should be helpful in analyzing such sites.

Materials and methods

Plasmid constructions

Plasmids pABD-B HD72, pANTP HD72, pUBX K3 and pUBXK3/K6/P7 were made by the insertion of PCR generated homeodomain expression cassettes into the expression vector pET3c (Rosenberg *et al.*, 1987) using standard procedures (Ausubel *et al.*, 1991). The ABD-B HD expression cassette was generated from primers ABD-B HD-A (5′-TATGGTCCGGAA-AAAGCGCAAG-3′) and ABD-B HD-B (5′-GCGTGGATCCTAGTTGTT-GTTGTTCTGCTG-3′) with 1 ng *Abd-B* P5 cDNA (Celniker *et al.*, 1989) as template. The ANTP HD expression cassette was generated using the primers ANTP HD-A (5′-ACGGCATATGCGCAAACGCCAAGG-3′) and ANTP HD72-B (5′-GATTGGATCCTATTCGCCTCCGGATCCCG-3′) with 1 μg *Drosophila* genomic DNA as template. The UBX K3 expression cassette was made using primers UBX K3-A (5′-CCACGGCATATGCGA-AGAAAGGGCCGACAGACATAC-3′) and UBX HD-D (Ekker *et al.*, 1991) with 1 ng *Ubx* cDNA template (p3712; Beachy *et al.*, 1985). The

UBX K3/K6/P7 expression cassette was made by recombinant circular PCR mutagenesis using primers UBX K3/K6/P7-sense (5′-CGAAGAAAAGGCCGAAAGCCATACACCCGCTACCAG-3′) and UBX K3/K6/P7-antisense (5′-GTGTATGGCTTTCGGCCTTTTCTTCGCA-3′) in standard protocols (Jones and Winistorfer, 1992) with a derivative of the pUHD expression construct (Ekker *et al.*, 1991) as template. The structures of all constructs were verified by double-strand sequence analysis using Sequenase 2.0 (US Biochemicals).

Purification of homeodomain proteins

UBX HD and DFD HD were purified as described previously (Ekker *et al.*, 1991, 1992). Plasmids pABD-B HD72 and pANTP HD72 were transformed into *Escherichia coli* strain BL21(DE3) pLysS (Rosenberg *et al.*, 1987). Induction, harvest and purification to homogeneity of ABD-B HD by chromatography were as described for DFD HD (Ekker *et al.*, 1992), with the identity of the purified product confirmed by Edman degradation of the first 10 N-terminal residues (Applied Biosystems Model 477A). UBX K3 and UBX K3/K6/P7 were partially purified using a single BioRex 70 column as described (Ekker *et al.*, 1991); the estimated purity of these proteins was ≥50% and ≥75% for UBX K3 and UBX K3/K6/P7, respectively, as judged by Coomassie Blue staining of an SDS−polyacrylamide gel (not shown). ANTP HD was purified to homogeneity as described (Ekker *et al.*, 1992), with an ammonium sulfate (85%) precipitation and subsequent dialysis (3.5 kDa size inclusion) step included before chromatography. The ANTP HD was identified by its chromatographic properties in relation to other homeodomain peptides and by its slightly anomalous (11−12 kDa) migration in SDS−polyacrylamide gels, as reported for a similar *Antennapedia* homeodomain peptide (Affolter *et al.*, 1990). Protein concentrations were measured using the absorbance at 280 and 205 nm (Scopes, 1987).

Structure, amplification and sequence analysis of the selection oligonucleotides

The sequences of the selection oligonucleotides given in Figure 1B were followed by the sequence 5′-ACTGGCCGTCGTTTTAC-3′ and preceded by either 5′-GTTTTCCCAGTCACGAC-3′ for oligonucleotide I or 5′-GTTTTCCCAGTCAG-3′ for oligonucleotides II, III and IV.

Three rounds of selection with each homeodomain peptide were performed as described in Ekker *et al.* (1992). Sequence analysis was performed with Sequenase 2.0 and $MnCl_2$ (US Biochemicals) using ^{32}P-labeled (Figures 2 and 3) or ^{33}P-labeled (Figure 4) primer on a 10% D600 (J.T.Baker), 7 M urea, 1 × TBE gel. The D600 matrix has approximately twice the resolving power of an equivalent acrylamide:bisacrylamide matrix in these experiments.

Quantitative sequence analysis of selected oligonucleotides

A PhosphorImager and storage phosphor screen (Molecular Dynamics) were used as described by Ekker *et al.* (1992) for analysis of selection gels, with the volume integration function used to yield values for the intensities of each band. Peak values at positions −5 and −4 were used to normalize the values in each lane to the corresponding unselected lanes for positions −3, −2, −1 and either 1 (Figure 2B) or 3 (Figures 3A and 4); positions 11 and 12 were similarly used for normalization of positions 10, 9, 8, 7 (Figures 3A and 4) and 6 (Figure 2B). Preference indices and the selection histograms were produced as described previously (Ekker *et al.*, 1992).

Dissociation rate constant measurements

DNA sequences used in the dissociation rate constant studies were 5′-AATTCAGATCT(N1−N10)ATGGATCCCTCGA-3′ where N1−N10 are the bases shown in Table I. Sequences b and e−h are from Ekker *et al.* (1992). Sequences a, c and d were synthetic oligonucleotides made double stranded by extension with the large fragment of DNA polymerase I of 34 base oligonucleotides annealed to a common primer (5′-TCGAGGGATCCATGGCC-3′). Double-stranded DNA was purified on a 20% polyacrylamide gel (19:1 bisacrylamide to acrylamide), eluted and further purified with a NACS column (Bethesda Research Labs). DNA labeling and dissociation rate quantification and analysis were performed as described (Ekker *et al.*, 1992) except binding reactions contained 2 nM ABD-B HD.

Acknowledgements

The authors wish to thank E.Perez-Albuerne for the development and use of his sequence alignment program, V.Vogel and C.Riley for protein sequencing, J.Ptak and C.Davenport for oligonucleotide synthesis, and S.Celniker for *Abd-B* cDNA. S.C.E. was supported by a predoctoral fellowship from the March of Dimes Birth Defects Foundation, and D.G.J. and B.I.S. by a NIH training grant.

References

Affolter,M., Percival-Smith,A., Müller,M., Leupin,W. and Gehring,W.J. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 4093−4097.

Appel,B. and Sakonju,S. (1993) *EMBO J.*, **12**, 1099−1109.

Arcioni,L., Simeone,A., Guazzi,S., Zappavigna,V., Boncinelli,E. and Mavilio,F. (1992) *EMBO J.*, **11**, 265−277.

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1991) *Current Protocols in Molecular Biology.* Greene Publishing Associates and Wiley-Interscience, New York.

Beachy,P.A., Helfand,S.L. and Hogness,D.S. (1985) *Nature*, **313**, 545−551.

Beachy,P.A., Varkey,J., Young,K.E., von Kessler,D.P., Sun,B.I. and Ekker,S.C. (1993) *Mol. Cell. Biol.*, **13**, 6941−6956.

Capovilla,M., Brandt,M. and Botas,J. (1994) *Cell*, **76**, 461−475.

Celniker,S.E., Keelan,D.J. and Lewis,E.B. (1989) *Genes Dev.*, **3**, 1425−1437.

Chan,S.-K. and Mann,R.S. (1993) *Genes Dev.*, **7**, 796−811.

Cribbs,D.L., Pultz,M.A., Johnson,D., Mazzulla,M. and Kaufman,T.C. (1992) *EMBO J.*, **11**, 1437−1449.

Desplan,C., Theis,J. and O'Farrell,P.H. (1988) *Cell*, **54**, 1081−1090.

Dessain,S., Gross,C.T., Kuziora,M.A. and McGinnis,W. (1992) *EMBO J.*, **11**, 991−1002.

Diederich,R.J., Merrill,V.K.L., Pultz,M.A. and Kaufman,T.C. (1989) *Genes Dev.*, **3**, 399−414.

Driever,W., Thoma,G. and Nusslein-Volhard,C. (1989) *Nature*, **340**, 363−367.

Ekker,S.C., von Kessler,D.P. and Beachy,P.A. (1992) *EMBO J.*, **11**, 4059−4072.

Ekker,S.C., Young,K.E., von Kessler,D.P. and Beachy,P.A. (1991) *EMBO J.*, **10**, 1179−1186.

Florence,B., Handrow,R. and Laughon,A. (1991) *Mol. Cell. Biol.*, **11**, 3613−3623.

Furukubo-Tokunaga,K., Flister,S. and Gehring,W.J. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6360−6364.

Galang,C.K. and Hauser,C.A. (1992) *New Biol.*, **4**, 558−568.

Gibson,G., Schier,A., Lemotte,P. and Gehring,W. (1990) *Cell*, **62**, 1087−1103.

Gould,A.P. and White,R.A.H. (1992) *Development*, **116**, 1163−1174.

Hayashi,S. and Scott,M.P. (1990) *Cell*, **63**, 883−894.

Hoey,T. and Levine,M. (1988) *Nature*, **332**, 858−861.

Jones,B. and McGinnis,W. (1993) *Genes Dev.*, **7**, 229−240.

Jones,D.H. and Winistorfer,S.C. (1992) *BioTechniques*, **12**, 528−534.

Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579−590.

Klemm,J.D., Rould,M.A., Aurora,R., Herr,W. and Pabo,C.O. (1994) *Cell*, **77**, 21−32.

Kuziora,M.A. and McGinnis,W. (1989) *Cell*, **59**, 563−571.

Kuziora,M.A. and McGinnis,W. (1991) *Mech. Dev.*, **33**, 83−94.

LeMotte,P., Kuroiwa,A., Fessler,L. and Gehring,W.J. (1989) *Proc. Natl Acad. Sci. USA*, **90**, 6360−6364.

Lin,L. and McGinnis,W. (1992) *Genes Dev.*, **6**, 1071−1081.

Lupas,A., VanDyke,M. and Stock,J. (1991) *Science*, **252**, 1162−1164.

Malicki,J., Schughart,K. and McGinnis,W. (1990) *Cell*, **63**, 961−967.

McGinnis,N., Kuziora,M.A. and McGinnis,W. (1990) *Cell*, **63**, 969−976.

McGinnis,W. and Krumlauf,R. (1992) *Cell*, **68**, 283−302.

Mihara,H. and Kaiser,E.T. (1988) *Science*, **242**, 925−927.

Mlodzik,M., Fjose,A. and Gehring,W.J. (1985) *EMBO J.*, **4**, 2961−2969.

Otting,G., Qian,Y.Q., Muller,M., Affolter,M., Gehring,W. and Wuthrich,K. (1988) *EMBO J.*, **7**, 4305−4309.

Otting,G., Qian,Y.Q., Billeter,M., Müller,M., Affolter,M., Gehring,W. and Wüthrich,K. (1990) *EMBO J.*, **9**, 3085−3092.

Percival-Smith,A., Müller,M., Affolter,M. and Gehring,W.J. (1990) *EMBO J.*, **9**, 3967−3974.

Regulski,M., McGinnis,N., Chadwick,R. and McGinnis,W. (1987) *EMBO J.*, **6**, 767−777.

Regulski,M., Dessain,S., McGinnis,N. and McGinnis,W. (1991) *Genes Dev.*, **5**, 278−286.

Rosenberg,A.H., Lade,B.N., Chui,D.-s., Lin,S.-W., Dunn,J.J. and Studier,F.W. (1987) *Gene*, **56**, 125−135.

Schier,A.F. and Gehring,W.J. (1992) *Nature*, **356**, 804−807.

Schier,A.F. and Gehring,W.J. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 1450−1454.

Schubert,F.R., Nieselt-Struwe,K. and Gruss,P. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 143−147.

Scopes,R.K. (1987) *Protein Purification: Principles and Practice.* Springer-Verlag, New York.

Scott,M.P. (1992) *Cell,* **71**, 551–553.

Small,S., Blair,A. and Levine,M. (1992) *EMBO J.*, **11**, 4047–4057.

Struhl,G., Struhl,K. and Macdonald,P.M. (1989) *Cell,* **57**, 1269–1273.

Tabin,C. (1992) *Development,* **116**, 289–296.

Vachon,G., Cohen,B., Pfeifle,C., McGuffin,M.E., Botas,J. and Cohen,S. (1992) *Cell,* **71**, 437–450.

Verrijzer,C.P., Alkema,M.J., van Weperen,W.W., van Leewen,H.C., Strating,M.J.J. and van der Vliet,P.C. (1992) *EMBO J.*, **11**, 4993–5003.

Wolberger,C., Vershon,A.K., Liu,B., Johnson,A.D. and Pabo,C.O. (1991) *Cell,* **67**, 517–528.

Wysocka-Diller,J.W., Aisemberg,G.O., Baumgarten,M., Levine,M. and Macagno,E.R. (1989) *Nature,* **341**, 760–763.

Zeng,W., Andrew,D.J., Mathies,L.D., Horner,M.A. and Scott,M.P. (1993) *Development,* **118**, 339–352.

Zhao,J., Lazzarini,R.A. and Pick,L. (1993) *Genes Dev.*, **7**, 343–354.

Received on August 25, 1993; revised on May 9, 1994