Development 124, 2007-2014 (1997)
Printed in Great Britain © The Company of Biologists Limited 1997
DEV7565

Switching the in vivo specificity of a minimal Hox-responsive element

Siu-Kwong Chan[1,*], Hyung-Don Ryoo[1], Alex Gould[2], Robb Krumlauf[2] and Richard S. Mann[1,†]

[1]Department of Biochemistry and Molecular Biophysics, Center for Neurobiology and Behavior, Columbia University, 630 West 168th Street, New York, NY 10032, USA
[2]MRC, National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, UK

*Present address: Howard Hughes Medical Institute, Department of Genetics and Development, Columbia University, 701 West 168th Street, New York, NY 10032, USA
†Author for correspondence (e-mail: rsm10@columbia.edu)

SUMMARY

The homeodomain proteins encoded by the Hox complex genes do not bind DNA with high specificity. In vitro, Hox specificity can be increased by binding to DNA cooperatively with the homeodomain protein extradenticle or its vertebrate homologs, the pbx proteins (together, the PBC family). Here we show that a two basepair change in a Hox-PBC binding site switches the Hox-dependent expression pattern generated in vivo, from *labial* to *Deformed*. The change in vivo correlates with an altered Hox binding specificity in vitro. Further, we identify similar Deformed-PBC binding sites in the *Deformed* and *Hoxb-4* genes and show that they generate *Deformed* or *Hoxb-4* expression patterns in *Drosophila* and mouse embryos, respectively. These results suggest a model in which Hox-PBC binding sites play an instructive role in Hox specificity by promoting the formation of different Hox-PBC heterodimers in vivo. Thus, the choice of Hox partner, and therefore Hox target genes, depends on subtle differences between Hox-PBC binding sites.

Key words: homeotic genes, cooperative binding, extradenticle, pbx, mouse, *Drosophila*

INTRODUCTION

In *Drosophila*, morphological differences along the anteroposterior axis are controlled by the homeotic (Hox) genes, which have close relatives throughout the animal kingdom (McGinnis and Krumlauf, 1992). All Hox genes encode homeodomain proteins and are thought to act by regulating distinct sets of downstream target genes (Garcia-Bellido, 1975; Botas, 1993). Yet, despite the precise morphologies they generate, Hox proteins do not bind DNA with high specificity (Desplan et al., 1988; Hoey and Levine, 1988; Ekker et al., 1992; Kalionis and O'Farrell, 1993; Ekker et al., 1994; Mann, 1995). For example, an oligonucleotide containing the strong Hox binding site, 5′-TAATTG, is able to bind at least 18 highly divergent homeodomains (Kalionis and O'Farrell, 1993). Moreover, by itself this binding site does not generate a pattern of expression in *Drosophila* embryos that correlates with any of its DNA binding activities (Vincent et al., 1990). This illustrates the difficulty in correlating the in vitro specificities of simple homeodomain binding sites with their in vivo activities. Thus, a major challenge has been to identify genuine in vivo Hox target sites and understand the basis of their specificity.

Although very few in vivo target sites of Hox proteins are known, Hox autoregulatory elements have provided some insight into this problem. The best characterized examples of Hox autoregulation come from analysis of the *Drosophila* and vertebrate *labial* and *Deformed* (*Dfd*) group genes (Kuziora and McGinnis, 1988; Chouinard and Kaufman, 1991; Regulski et al., 1991; Pöpperl and Featherstone, 1992; Tremml and Bienz, 1992; Pöpperl et al., 1995; Gould et al., 1997). The characterization of these elements has indicated that Hox cofactors are important for their in vivo functions (Zeng et al., 1994; Pöpperl et al., 1995; Chan et al., 1996; Gross and McGinnis, 1996). One family of cofactors, the homeodomain proteins encoded by the *Drosophila extradenticle* (*exd*) and vertebrate *pbx* genes (together, the PBC genes), bind cooperatively to DNA with Hox proteins (Chan et al., 1994; van Dijk and Murre, 1994; Chang et al., 1995; Lu et al., 1995; Phelan et al., 1995) and are important for Hox function in vivo (Peifer and Wieschaus, 1990; Chan et al., 1994; Rauskolb and Wieschaus, 1994; Gonzalez-Crespo and Morata, 1995; Harding et al., 1995; Pöpperl et al., 1995; Rauskolb et al., 1995; Chan et al., 1996). One model suggested by these observations is that PBC proteins provide Hox proteins with additional DNA binding specificity in vivo. Consistent with this model, Hox-PBC heterodimers exhibit more sequence specificity in vitro than Hox monomers (Chan and Mann, 1996; Chang et al., 1996; Shen et al., 1996). For example, different Hox-PBC binding sites prefer to bind different Hox-PBC heterodimers in vitro (Chan and Mann, 1996; Chang et al., 1996; Mann and Chan, 1996). However, these studies did not determine whether this increase in in vitro specificity is relevant in vivo. In the work presented here, we demonstrate that subtly different Hox-PBC binding sites result in distinct Hox-dependent transcriptional responses in vivo.

One of the best characterized Hox-PBC binding sites is present in a 20 bp oligonucleotide, repeat 3, which was identified in the 5′ promoter region of the mouse *Hoxb-1* gene (Pöpperl et al., 1995). Hoxb-1 protein or its *Drosophila* ortholog labial are both able to bind cooperatively with exd to

this binding site whereas other Hox proteins, such as Ultrabithorax (Ubx) or Hoxb-4, cannot (Pöpperl et al., 1995; Chan and Mann, 1996; Chan et al., 1996). Furthermore, when *lacZ* reporter constructs containing three tandem copies of repeat 3 were introduced into either mouse or *Drosophila* embryos, *lacZ* expression patterns resembling the endogenous *Hoxb-1* or *labial* expression patterns were generated, respectively (Pöpperl et al., 1995; Chan et al., 1996). In *Drosophila* expression driven by this reporter gene, *3Xrpt3-lacZ*, requires both *exd* and *labial* functions (Chan et al., 1996). Thus, in both mouse and *Drosophila* repeat 3 behaves as an *exd*-dependent *Hoxb-1*/*labial* autoregulatory enhancer element.

The highly specific characteristics of repeat 3 provide an excellent basis for assessing the role of individual basepairs in conferring in vivo Hox specificity. Biochemical studies have generated a model for how the Hox and PBC homeodomains bind to DNA (Chan and Mann, 1996; Chang et al., 1996; Lu and Kamps, 1996). A bipartite 10 bp Hox-PBC consensus binding site, 5′-*TGAT**NN***AT[g/t][g/a], has been defined in which the PBC and Hox half sites, indicated by italics and underlining, respectively, overlap (Fig. 1A) (Chan and Mann, 1996; Chang et al., 1996; Lu and Kamps, 1996). The central two basepairs (NN), which are predicted to contact the Hox N-terminal arm, have been shown to influence which Hox partner is incorporated in the heterodimer in vitro (Chan and Mann, 1996; Chang et al., 1996; Mann and Chan, 1996). In the repeat 3 binding site these two specificity-determining basepairs are GG (Fig. 1A). We report here that changing these central basepairs to TA results in an oligonucleotide that generates a *Deformed* expression pattern in vivo. Similar binding sites to these are present in the *Drosophila Deformed* gene and in the mouse *Deformed* ortholog, *Hoxb-4*. Remarkably, these binding sites are able to generate *Deformed* or *Hoxb-4* expression patterns in *Drosophila* or mouse embryos, respectively. These results demonstrate that PBC proteins cooperate with multiple Hox proteins in vivo. Further, they illustrate the importance of the two central basepairs in Hox-PBX binding sites in distinguishing between different Hox specificities in vivo.

MATERIALS AND METHODS

Fly stocks and transformants

All *Drosophila* reporter genes were generated by cloning three copies of 20 bp oligonucleotides (sequences are listed in Fig. 1A) in the same orientation into a blunted *Sph*I site of the *nuclear-lacZ*-encoding P element vector CPLZ as described by Chan et al. (1996). Transformant lines were generated by standard procedures. For each reporter construct the *lacZ* expression pattern was identical in multiple independent transformant lines with the exception that one line generated with repeat 3[TA] (#6) had, in addition to a *Dfd* expression pattern, ectopic expression in the central nervous system and in more posterior regions of the head. This line was not used in further studies. In addition, in occasional (~10%) *repeat 3[TA]-lacZ* embryos weak expression in endodermal cells was observed. Expression in endodermal cells, which is a position of labial expression, may reflect the fact that repeat 3[TA] retains some affinity for labial-exd complexes (see Fig. 3). Because all strong *lacZ* expression required *Dfd* function (Fig. 2), we suggest that the few Dfd+, β-gal– or Dfd–, β-gal+ nuclei (Figs. 1J and 4E) might be due to differences in the rates of accumulation or stabilities of the Dfd and β-gal proteins. The anti-β-gal antibody was a rabbit polyclonal (Cappell) and the anti-Dfd antibody was a guinea pig polyclonal provided by W. McGinnis. Secondary antibodies (conjugated with alkaline phosphatase, horseradish peroxidase, Texas red, or FITC) were from Jackson Labs.

exd maternal$^-$ zygotic$^-$ embryos were generated as described by Chan et al. (1996) using the FLP recombination system and the null allele *exd*^{XP11}. Two independent *3Xrpt3[TA]-lacZ* insertions (#1 and #2) were crossed into a *Dfd*^{r11} background (provided by W. McGinnis) and *Dfd*$^-$ embryos were identified by the absence of staining with an anti-Dfd antibody. For both lines, although most of the expression was *Dfd*-dependent, the *Dfd*$^-$ embryos had a few cells in the head and endoderm that weakly expressed β-gal. A *HS:Dfd* transgene (provided by W. McGinnis) was crossed into *3Xrpt3[TA]-lacZ* line #3 and both P elements were homozygous in the final stock. Embryos were heat shocked for 45 minutes at 37°C between 3 and 6 hours of embryogenesis. Interestingly, heat shock-*Dfd*-dependent expression of repeat 3[TA] in the midgut endoderm is reminiscent of the *labial*-dependent expression driven by repeat 3 (Chan et al., 1996) and suggests that these cells are competent to express repeat 3 variants when the appropriate Hox protein (Dfd or labial) is present.

In the 2.7 kb *Dfd* EAE the sequence 5′-TGAT**TA**ATGA (site 1) begins at nucleotide 719 of the bottom strand and 5′-AGAT**TA**ATTA (site 2) begins at nucleotide 539 of the bottom strand. In the 600 bp *Dfd* NAE the sequence 5′-TGAT**TA**ATTA (site 1) begins at nucleotide 518 of the top strand and 5′-AGAT**TA**ATGA (site 2) begins at nucleotide 119 of the bottom strand. Strikingly, in both of these elements the only matches to the general consensus, 5′-TGAT**NN**AT[g/t][g/a] are EAE site-1 and NAE site-1, which both have TA in the specificity-determining positions.

Electrophoretic mobility shift assays (EMSA)

Electrophoretic mobility shift assays (band shifts) were carried out as described by Chan et al. (1996) except that the exd protein was nearly full length (from amino acid 1 to 323) fused to the His tag at its N terminus in the pQE vector (Qiagen). Dfd was full length, also fused to the His tag at its N terminus in the pET14b vector (Novagen) and labial was from amino acid 158 to its C terminus (Chan et al., 1996). All proteins were expressed in *E. coli* and purified by Ni^{2+} affinity chromatography (Chan et al., 1996). In the protein preparations used in Fig. 3, approximately 100% of the exd, 10% of the Dfd and 20% of the labial was full length. In the competition experiments the cold competitor oligonucleotide was present at time 0, before the addition of any protein.

Mouse transgenes

Three copies of the sequence 5′-GAACCATTAATCACTTCTTTT-CTTTAAATACG (the Hoxb-4/PBC binding site is underlined), present in CR3 of *Hoxb-4*, were cloned into the filled *Spe*I site of a minimal promoter-*lacZ* construct and transgenic embryos generated and analyzed as previously described (Pöpperl et al., 1995). Of 11 expressing 10.5 dpc embryos, 7 showed neural expression with an anterior limit at the r6/7 junction.

RESULTS

Generating a *Deformed* autoregulatory element from a *labial* autoregulatory element

We tested the effects of changing the central two basepairs in the labial-exd site present in repeat 3, 5′-TGATGGATGG, by constructing *lacZ* reporter genes with variants of the repeat 3 sequence. Transgenic fly stocks were generated and *lacZ* expression patterns examined during embryogenesis in several transformant lines (data not shown). Amongst those tested a variant containing the change from GG to TA (*3Xrpt3[TA]-lacZ*; Fig. 1A) displayed a dramatic change in expression. Unlike *3Xrpt3-lacZ*, which generated a *labial* expression

pattern in *Drosophila* (Fig. 1B,E), *3Xrpt3[TA]-lacZ* generated a pattern that was very similar to the *Dfd* expression pattern (Fig. 1C,D,F,G). Double label experiments using anti-Dfd and anti-β-galactosidase (β-gal) antibodies confirmed that *3Xrpt3[TA]-lacZ* and *Dfd* are expressed in overlapping patterns (Fig. 1H-J). Thus, whereas repeat 3 directs a *labial* expression pattern in vivo, repeat 3[TA] generates a *Deformed* expression pattern in vivo.

If, as is the case for repeat 3, repeat 3[TA] is an *exd*-dependent Hox autoregulatory element, its activity in vivo should depend on both *exd* and *Dfd* functions. To test the requirement for *exd*, which is a highly expressed maternal gene (Rauskolb et al., 1993; Mann and Abu-Shaar, 1996), we generated females with mosaic *exd*$^{-}$ germlines. When crossed to males containing the *3Xrpt3[TA]-lacZ* reporter gene, embryos that had no maternal or zygotic *exd* function did not show any detectable *lacZ* expression (Fig. 2D). Similarly, nearly all *lacZ* expression was eliminated when the *3Xrpt3[TA]-lacZ* reporter gene was crossed into a *Dfd*$^{-}$ background (Fig. 2B). Thus, the *3Xrpt3[TA]-lacZ* reporter gene is expressed in a *Dfd*-like pattern and requires both *exd* and *Dfd* functions in vivo.

If *3Xrpt3[TA]-lacZ* is activated by Dfd in vivo then ectopic expression of Dfd might result in the ectopic activation of this reporter gene. To test this we expressed ubiquitous and high levels of Dfd from a heat shock-inducible transgene, *HS:Dfd* (Kuziora and McGinnis, 1988), in embryos containing the *3Xrpt3[TA]-lacZ* reporter gene. Ubiquitous Dfd expression resulted in ectopic *3Xrpt3[TA]-lacZ* expression (Fig. 2C). Activation was especially high throughout the head and in the midgut endoderm. Interestingly, the reporter gene was not uniformly activated, suggesting that other factors limit the activity of this enhancer.

Repeat 3[TA] requires the exd half-site

The requirement for *exd* for expression of *3Xrpt3[TA]-lacZ* could be indirect; for example, *exd* could be required for the synthesis of another necessary cofactor. To obtain evidence that exd is directly regulating this reporter gene we mutated the G at position 2 of the Hox-exd binding site within repeat 3[TA] to generate repeat 3[TA],G>A (Fig. 1A). Previously, methylation interference and hydrazine modification studies of repeat

A

	PBC HOX
consensus	....TGATNNATGG/TA.....
repeat 3	ggggTGAT**GG**ATGGgcgctg
repeat 3[TA]	ggggTGAT**TA**ATGGgcgctg
repeat 3[TA],G>A	ggggT**a**AT**TA**ATGGgcgctg
Deformed	
EAE-1	ataaTGAT**TA**ATGAacgcgc
EAE-2	gttaAGAT**TA**ATTAgcgata
NAE-1	atttTGAT**TA**ATTAttcagt
NAE-2	acttAGAT**TA**ATGAtatgtg
Hoxb-4	
mouse	...gTGAT**TA**ATGGt.....
chicken	...gTGAT**TA**ATGAt.....
pufferfish	...gTGAT**TA**ATGAt.....

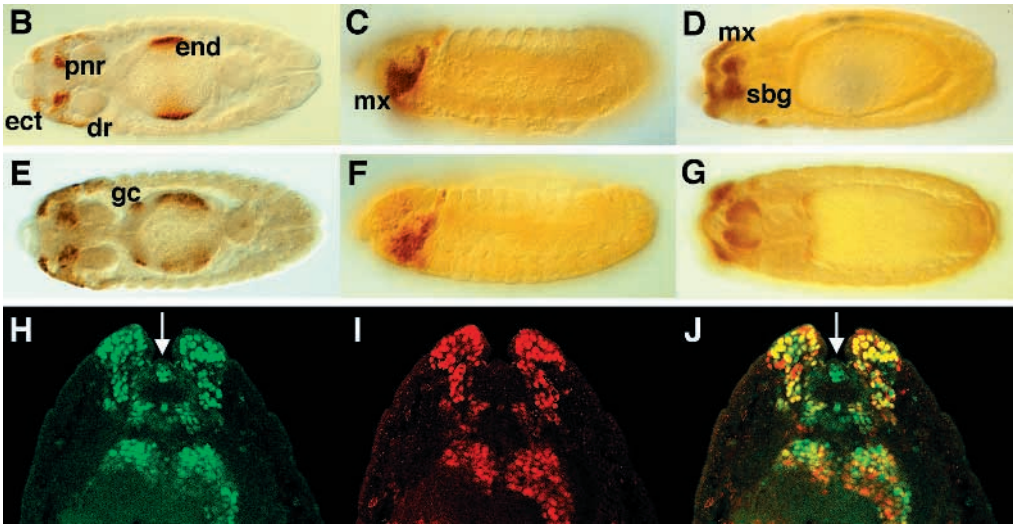

Fig. 1. Switching the in vivo specificity of repeat 3. (A) The bipartite Hox-PBC consensus sequence (Mann and Chan, 1996) is shown on top with the two specificity-determining basepairs (NN) in bold. The sequences of repeat 3, which is an evolutionarily conserved element from the mouse *Hoxb-1* gene (Pöpperl et al., 1995), and two repeat 3 variants are shown. The sequences of Dfd-exd binding sites found in the epidermal (EAE) or neural (NAE) autoregulatory elements of the *Drosophila Dfd* gene or in an evolutionarily conserved region (CR3) of the vertebrate *Hoxb-4* gene are listed. In all seven of these naturally occurring sequences the specificity-determining basepairs are TA. See text for further details. Sequences outside the Hox-PBC consensus sequence are in lower case. B-G are stage 13-14 wild-type (B-D), *3Xrpt3-lacZ* (E), or *3Xrpt3[TA]-lacZ* (F,G) embryos stained with anti-labial (B), anti-Dfd (C,D), or anti-β-gal (E-G) antibodies, detected by horseradish peroxidase-conjugated secondary antibodies. *labial* and *3Xrpt3-lacZ* are expressed in ectodermal (ect) and presumptive neural (pnr) cells of the head and endodermal cells of the midgut (end). *3Xrpt3-lacZ* is also expressed in the gastric caeca primordia (gc) (Chan et al., 1996). *Dfd* and *3Xrpt3[TA]-lacZ* are expressed in the maxillary segments (mx) and subesophogeal ganglion (sbg) of the central nervous system. C and F are lateral views; D and G are ventral views. (H-J) Confocal images of a stage 13 *3Xrpt3[TA]-lacZ* embryo showing nuclear-β-gal (green; H and J) or Dfd (red; I and J). Overlapping expression in J appears yellow. While β-gal and Dfd are detected in very similar patterns, a few β-gal-positive, Dfd-negative nuclei can be seen (arrows in H and J).

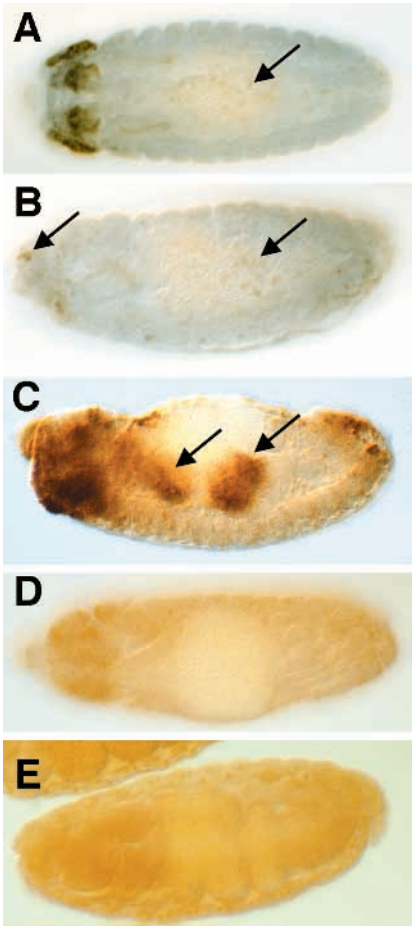

Fig. 2. *3Xrpt3[TA]-lacZ* is an *exd*-dependent *Dfd* autoregulatory element. A and B are sibling, stage 12, *3Xrpt3[TA]-lacZ* embryos doubly stained for Dfd (blue) and nuclear-β-gal (brown). A is wild type (blue and brown stains are visible) and B is *Dfd*$^{-}$ (a few β-gal-positive (brown) cells remain, arrows). In addition, weak β-gal expression can be observed in endoderm cells in A (arrow). (C) *HS:Dfd*; *3Xrpt3[TA]-lacZ* heat shocked, stage 12 embryo stained for β-gal. Ectopic expression of *3Xrpt3[TA]-lacZ* is observed throughout the head and in the endoderm (arrows). The endoderm appears as two darkly labeled regions because ubiquitous Dfd expression blocked fusion of the two midgut primordia. (D) A *3Xrpt3[TA]-lacZ* embryo devoid of all (maternal and zygotic) *exd* function stained for β-gal. No expression of the reporter gene is detected. (E) A *3Xrpt3[TA],G>A-lacZ* stage 15 embryo stained for β-gal. No expression of the reporter gene is detected.

3 demonstrated that both bases of this G:C basepair are important for exd binding, but not Hox binding (Chan and Mann, 1996). The analogous G to A mutation in repeat 3 abolished all detectable exd binding, greatly diminished the ability to form labial-exd complexes, and destroyed enhancer function in vivo (Chan and Mann, 1996). Similarly, a reporter gene made with the repeat 3[TA],G>A oligonucleotide generated no detectable *lacZ* expression in vivo (Fig. 2E). Thus, the exd half site is required for the in vivo function of repeat 3[TA]. This result suggests that exd is directly binding to repeat 3[TA] in vivo, consistent with the genetic requirement for *exd* function.

Repeat 3[TA] prefers to bind Dfd-exd complexes over labial-exd complexes

To test for a change in in vitro specificity we performed band shift experiments with labeled oligonucleotide probes and purified proteins. Repeat 3 efficiently formed complexes with labial plus exd, but not with Dfd plus exd (Fig. 3A, lanes 1-12). In contrast, repeat 3[TA] formed complexes with Dfd plus exd (lanes 16-18). Labial plus exd also formed complexes on repeat 3[TA], but 3- to 8-fold less efficiently than did Dfd plus exd (lanes 22-24). We used competition experiments to further assess the change in DNA-binding specificity. When formed on repeat 3[TA], Dfd-exd complexes were nearly eliminated by a 100-fold excess of unlabeled repeat 3[TA] but were largely unaffected by similar amounts of repeat 3 (Fig. 3B, lanes 21-24). Conversely, when formed on repeat 3, labial-exd complexes were more efficiently competed (by 2- to 3-fold) by unlabeled repeat 3 than by repeat 3[TA] (lanes 9-12). Thus, changing the two central basepairs in the Hox-exd binding site present in repeat 3 changed the Hox preference from labial to Dfd. Interestingly, we note that in some embryos *3Xrpt3[TA]-lacZ* is weakly expressed in *labial*-expressing cells of the endoderm (Fig. 2), consistent with repeat 3[TA]'s ability to weakly bind the labial-exd heterodimer. In addition to binding labial-exd heterodimers, repeat 3[TA] also promotes heterodimer formation between exd and additional Hox proteins, for example abdominal-A and Ultrabithorax (data not shown). Thus, although the GG to TA mutation results in a change in the in vitro binding specificity that correlates with the change of specificity in vivo, the in vitro properties of the repeat 3[TA] oligonucleotide do not account for the highly specific expression pattern generated by *3Xrpt3[TA]-lacZ* in vivo. In contrast, the in vitro binding properties of repeat 3, which is very specific for labial-exd heterodimers, is consistent with the highly specific expression pattern generated by *3Xrpt3-lacZ* in vivo.

Repeat 3[TA]-like binding sites in *Deformed*

In light of the *labial* to *Dfd* change in specificity resulting from the GG to TA mutation, we investigated if repeat 3[TA]-like binding sites might participate in the normal autoregulation of the *Dfd* group genes. In *Drosophila* the *Dfd* gene contains well characterized epidermal (EAE) and neural (NAE) autoregulatory elements (Bergson and McGinnis, 1990; Regulski et al., 1991; Zeng et al., 1994; Lou et al., 1995). One of these, module E from the EAE, generates several aspects of the Dfd pattern but does not contain a repeat 3[TA]-like binding site. Instead, the activity of module E requires a high affinity Dfd binding site and a binding site for a novel DNA binding protein, DEAF-1 (Zeng et al., 1994; Gross and McGinnis, 1996). However, we found two repeat 3[TA]-like binding sites elsewhere in the 2.7 kb EAE and two in the 600 bp NAE (Fig. 1A). One of the sites in the EAE is in a subfragment of the enhancer known as module C which is sufficient to generate a *Dfd* pattern of expression (Zeng et al., 1994). Both binding sites found in the NAE are conserved in *Drosophila hydei*, suggesting that they are important for the function of this enhancer (Lou et al., 1995).

We tested the in vivo activity of two of the sequences identified in the *Dfd* gene, EAE site 1 and NAE site 2 (Fig. 1A). As with repeat 3[TA], three tandem copies of 20 bp oligonucleotides encompassing these binding sites were cloned upstream of a minimal promoter driving *lacZ* to generate the reporter genes *Dfd-exd[EAE]-lacZ* and *Dfd-exd[NAE]-lacZ*. Strikingly, both of these elements generated a *Dfd* pattern of

expression in vivo (Fig. 4). As the only similarity between repeat 3[TA], Dfd-exd[EAE], and Dfd-exd[NAE] is the Hox-exd binding site (Fig. 1A), we infer that this 10 bp sequence is important, and may be sufficient, for generating a *Dfd* pattern of expression in vivo. We note, however, that the first position of this sequence can be either T (as in EAE-1) or A (as in NAE-2) (Fig. 1A). Further, the absence of a repeat 3[TA]-like binding site in module E of the EAE, which can also generate a *Dfd*-dependent expression pattern in vivo, suggests that there are additional ways for Dfd to activate transcription in vivo.

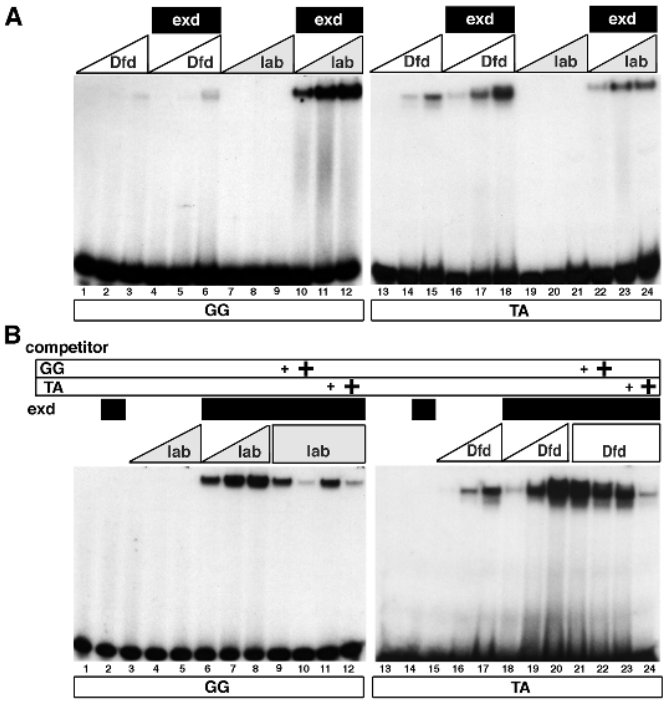

Fig. 3. Repeat 3 and repeat 3[TA] have different Hox binding preferences in vitro. (A) Band shift assays with labeled repeat 3 (GG, left panel) or labeled repeat 3[TA] (TA, right panel) bound to Dfd, exd, or labial as indicated. Repeat 3 is highly specific for labial plus exd (lanes 10-12) and repeat 3[TA] has a preference for Dfd plus exd (lanes 16-18) over labial plus exd (lanes 22-24). In this and other experiments (not shown) the mobility of the complex generated with Dfd plus exd (lanes 16-18) is slightly slower than that generated by Dfd, alone (lanes 13-15), suggesting that Dfd plus exd heterodimers are formed. (B) competition experiments with cold oligonucleotides. Labial plus exd complexes formed on repeat 3 (left panel) or Dfd plus exd complexes formed on repeat 3[TA] (right panel) were competed with unlabeled repeat 3 (GG) or unlabeled repeat 3[TA] (TA) as indicated. Amounts of protein used: (A) Dfd: 15 ng (lanes 1, 4, 13, 16), 60 ng (lanes 2, 5, 14, 17), 250 ng (lanes 3, 6, 15, 18); labial (lab): 12 ng (lanes 7, 10, 19, 22), 50 ng (lanes 8, 11, 20, 23), 200 ng (lanes 9, 12, 21, 24); exd: 100 ng. (B) Dfd: 30 ng (lanes 15, 18), 125 ng (lanes 16, 19), 500 ng (lanes 17, 20-24); labial (lab): 12 ng (lanes 3, 6), 50 ng (lanes 4, 7), 200 ng (lanes 5, 8-12); exd: 100 ng. Cold repeat 3 competitor: 10-fold excess (lanes 9 and 21), 50-fold excess (lane 10), 100-fold excess (lane 22); cold repeat 3[TA] competitor: 10-fold excess (lanes 11 and 23), 50-fold excess (lane 12) and 100-fold excess (lane 24). In these reactions (A,B), 100 ng exd ≈ 2.8 pM full length; 500 ng Dfd ≈ 0.75 pM full length; 200 ng labial ≈1.1 pM full length.

Repeat 3[TA]-like binding site in *Hoxb-4*

To extend this analysis to a vertebrate *Dfd* ortholog we relied on previous studies that identified regulatory sequences conserved among the *Hoxb-4* genes from mouse, chicken, and pufferfish (Aparicio et al., 1995; Morrison et al., 1995; Gould et al., 1997). Strikingly, in conserved region 3 (CR3) we found a match to the Dfd-exd binding site in repeat 3[TA] (Fig. 1A). To test if this CR3 sequence is sufficient to generate a *Hoxb-4* expression pattern in the mouse we analyzed embryos transgenic for a minimal promoter-*lacZ* construct with three copies of a 32 bp oligonucleotide containing the Dfd-exd[Hoxb-4] binding site (Fig. 5A,B). At 10.5 days post coitum (dpc), *lacZ* expression was observed in the anterior spinal cord and posterior hindbrain in a pattern resembling the CNS expression of the *Hoxb-4* gene. Like *Hoxb-4* expression (Wilkinson et al., 1989; Gould et al., 1997), Dfd-exd[Hoxb-4]-dependent staining has a rostral limit just posterior to the otic vesicle at the junction between rhombomeres 6 and 7 (r6/7). Thus, like the EAE and NAE binding sites, the binding site from the *Hoxb-4* gene mimics the expression pattern of an endogenous *Dfd* group gene. This underscores the importance of the central TA basepairs in specifying *Dfd*-like patterns of expression in both vertebrates and insects.

DISCUSSION

PBC proteins provide specificity to multiple Hox proteins in vivo

Despite the unique functions that Hox genes carry out in vivo, they all encode proteins with similar homeodomains and DNA binding specificities. Therefore, an important problem in Hox biology has been to understand how they achieve their unique specificities in vivo. The results presented here, together with previous studies (reviewed by Mann and Chan, 1996), demonstrate that at least some of this in vivo specificity, in both vertebrates and insects, is due to the ability of Hox proteins to form heterodimers with PBC homeodomain proteins. In addition, our ability to switch between two different Hox-dependent transcriptional responses in vivo provides strong evidence that the choice of Hox partner, and thus Hox specificity, depends, at least in part, on subtle differences between Hox-PBC binding sites.

The formation of heterodimers between homeodomain proteins is an important and general mechanism for generating specificity in vivo. For example, a heterodimer of the homeodomain proteins UNC-86 and MEC-3 is important for the differentiation of a specific set of touch-sensitive neurons in *C. elegans* (Xue et al., 1992, 1993). Similarly, in the yeast *S. cerevisiae*, a heterodimer between the homeodomain proteins Mat**a**1 and Matα2 is critical for the execution of mating type-specific pathways (Goutte and Johnson, 1993, 1994). In both examples, these heterodimers have been shown to be important for DNA binding specificity in vitro and target gene selection in vivo.

We suggest that there is an important distinction between these examples and the interaction between Hox and PBC proteins. Specifically, these yeast or *C. elegans* proteins are likely to interact with very few or perhaps only a single homeodomain protein in vivo. In contrast, extradenticle, and PBC proteins in general, has the capacity to interact with most, if not all, members of the Hox protein family (Chan et

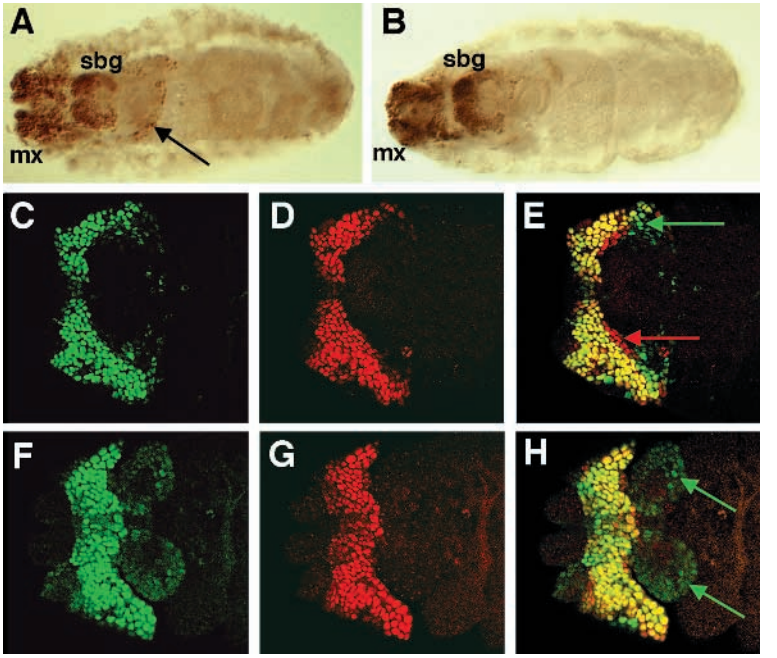

Fig. 4. Dfd-PBC binding sites from the *Dfd* gene function in fly embryos. A and B are *Dfd-exd[EAE]-lacZ* and *Dfd-exd[NAE]-lacZ* stage 15 embryos, respectively, stained for β-gal. For both reporter genes expression in the maxillary segment (mx) and the subesophogeal ganglion (sbg) is observed. In addition, weak expression in gastric caeca cells is seen in A (arrow). C-H are confocal micrographs of *Dfd-exd[EAE]-lacZ* (C-E) and *Dfd-exd[NAE]-lacZ* (F-H) stage 12 embryos doubly stained for β-gal (C,F; in green), Dfd (D,G; in red), or both (E,H; overlapping expression appears yellow). Although *Dfd* and *lacZ* expression are mostly coincident, in E a few Dfd-positive, β-gal-negative (red arrow) and β-gal-positive, Dfd-negative (green arrow) nuclei are observed. In H, weak *lacZ* expression can be seen in the labial segment (green arrows).

al., 1994; van Dijk and Murre, 1994; Chang et al., 1995; Phelan et al., 1995; Pöpperl et al., 1995; Chan and Mann, 1996; Chan et al., 1996). In addition, PBC proteins can also cooperatively bind with non-Hox homeodomain proteins such as engrailed (Peltenburg and Murre, 1996). However, the relevance of these in vitro interactions to homeoprotein specificity in vivo is less clear. The experiments presented here strongly suggest that an interaction with PBC proteins is important in vivo for the specificity of at least two different Hox proteins, labial and Deformed. Because *exd* is required for many segment identities in flies (Peifer and Wieschaus, 1990; Rauskolb and Wieschaus, 1994; Gonzalez-Crespo and Morata, 1995; Rauskolb et al., 1995), and because PBC proteins have the capacity to cooperatively bind with many Hox proteins in vitro, we suggest that the findings presented here can be extended to additional homeodomain proteins as well. Thus PBC proteins may be unusual because, depending on small differences between binding sites, they appear able to interact with and contribute to the specificity of multiple homeodomain proteins.

Hox-PBC binding sites are instructive

These results bring up an important question, that is, how can a single cofactor such as extradenticle contribute to the specificity of multiple Hox proteins? Based on the results presented here, it is likely that the DNA binding site plays an important role. Specifically, our results demonstrate that subtly different Hox-PBC binding sites generate different Hox-dependent transcriptional responses in vivo. In the example described here, the relevant differences are in the central two basepairs (NN) in the Hox-PBC consensus binding site, 5′-TGATNNAT[g/t][g/a]. From in vitro studies, it has been proposed that these specificity-conferring basepairs contact the Hox N-terminal arm (Chan and Mann, 1996; Chang et al., 1996; Lu and Kamps, 1996; Mann and Chan, 1996). We suggest that Hox-PBC binding sites are instructive because they promote the formation of heterodimers between PBC proteins and specific subsets of the Hox protein family. Further, we suggest that in Hox-PBC heterodimers, it is the Hox N-terminal arm that is primarily responsible for reading these differences between Hox-PBC binding sites. Consistent with this model, N-terminal arm residues tend to be among the most variable in Hox homeodomains (Burglin, 1994; Mann, 1995), are important for Hox specificity in vivo (Chan and Mann, 1993; Furukubo-Tokunaga et al., 1993; Zeng et al., 1993), but contribute little DNA binding specificity to Hox monomers (Ekker et al., 1991; Laughon, 1991; Ekker et al., 1994; Gehring et al., 1994). In summary, we suggest that extradenticle cooperates with multiple Hox proteins in vivo because, for a particular Hox-PBC binding site, it is able to form productive heterodimers with only a distinct subset of Hox proteins.

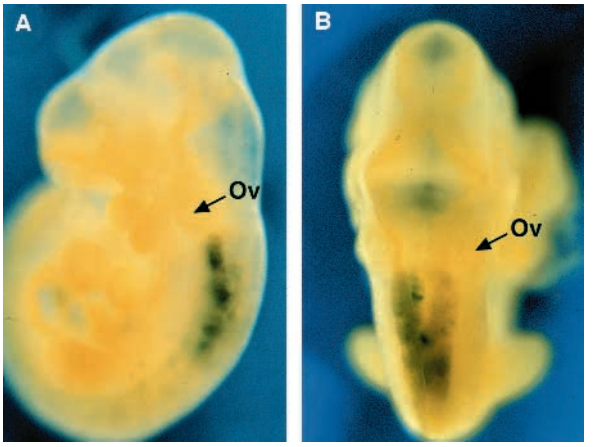

Fig. 5. A Dfd-PBC binding site from the mouse *Hoxb-4* gene functions in mouse embryos. (A,B) Lateral and dorsal views, respectively, of a 10.5 dpc mouse embryo transgenic for a *Dfd-exd[Hoxb-4]-lacZ* construct. Expression is confined to the central nervous system with an anterior boundary at the r6/7 junction, immediately posterior to the otic vesicle (Ov, arrow).

The in vitro specificity of Hox-PBC binding sites only partially accounts for their in vivo specificity

The difference between the expression patterns generated by repeat 3 and repeat 3[TA] is very striking: while nearly all of the transcription generated by repeat 3 is *labial*-dependent, nearly all of the transcription generated by repeat 3[TA] is *Deformed*-dependent. However, in vitro, the difference between these two Hox-PBC binding sites is less absolute. Repeat 3, which has GG as its central two basepairs, is highly specific for labial-exd heterodimers. In contrast, although repeat 3[TA] has a preference for Deformed-exd heterodimers, this oligonucleotide still binds labial-exd heterodimers as well as heterodimers between exd and other Hox proteins (data not shown). Thus, by changing the central two basepairs of the Hox-PBC binding site in repeat 3 we have dramatically changed specificity in vivo but only partially changed specificity in vitro.

We suggest that other, as yet unidentified, factors contribute to the highly specific expression patterns driven by these oligonucleotides in vivo. Additional evidence for such factors comes from the limited transcriptional response of *3Xrpt3[TA]-lacZ* to ubiquitous Dfd expression. In this case, even when Dfd is present throughout the embryo the repeat 3[TA] reporter gene is activated in only a limited number of cells. Thus, there must be additional factors that limit the activation of this reporter gene in vivo. One such limiting factor is probably exd, because its subcellular localization, in either the cytoplasm or nucleus, is regulated during embryogenesis (Mann and Abu-Shaar, 1996). However, when Dfd is ubiquitously expressed *3Xrpt3[TA]-lacZ* is still only activated in a subset of the cells that contain nuclear-localized exd. Thus, there are probably other factors, in addition to Hox and exd, that bind to these oligonucleotides in vivo and contribute to the specific expression patterns they generate. Another possibility is that Hox or PBC proteins are subject to post-translational modifications that influence their ability to cooperatively bind with each other in vivo. In future experiments, one important goal will be to account for the highly specific behavior of these oligonucleotides in vivo by assembling similarly specific protein-DNA complexes in vitro.

Hox-PBC binding sites in target genes versus autoregulatory enhancers

Interestingly, although the *Dfd* EAE and NAE are largely tissue specific (Bergson and McGinnis, 1990; Lou et al., 1995), the Dfd-exd[EAE] and Dfd-exd[NAE] binding sites activate transcription in both neural and epidermal tissues (Fig. 4). Similarly, repeat 3 activates transcription in neural, epidermal and endodermal cell types (Chan et al., 1996). Thus, Hox-PBC binding sites by themselves are not tissue-specific. However, in the context of the entire EAE or NAE the activity of these binding sites is apparently restricted by tissue-specific factors. Although the Hox-PBC binding sites described here are derived from autoregulatory enhancers, similar binding sites to these might also be used in the regulation of the downstream target genes of *labial* and *Deformed*. As is the case for the *Dfd* EAE, in the context of a target gene enhancer the activity of these Hox-PBC binding sites may be modulated or limited by additional factors that provide tissue or cell type specificity.

Conclusions

In summary, these experiments strongly suggest that PBC proteins are important for the specificity of multiple Hox proteins in vivo, in both vertebrates and insects. Moreover, we demonstrate that subtly different Hox-PBC binding sites generate distinct HOX-dependent expression patterns in vivo. Given these findings, we suggest that Hox-PBC binding sites will be an important component of many Hox-regulated enhancers. Thus, the identification of Hox-PBC binding sites may be a valuable approach for identifying Hox-regulated target genes. Although these results demonstrate a correlation with Hox binding in vitro we note that, in vivo, additional proteins might contribute to the highly specific behavior of these binding sites. Nevertheless, these results demonstrate the critical role of the central two basepairs in the Hox-PBC consensus site in defining Hox target specificity in vivo. In the future, the characterization of additional Hox-PBC binding sites will determine if these findings can be extended to other members of the Hox complex.

We are very grateful to W. McGinnis for providing fly stocks and antibodies, and to T. Kaufman for antibodies. We thank S. Goff, T. Jessell and G. Struhl for their comments on this work and/or this manuscript. This work was supported by grants from the NIH and Searle Foundation to R. S. M.

REFERENCES

Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P. W. J., Krumlauf, R. and Brenner, S. (1995). Detecting conserved regulatory elements with the model genome of the Japanese puffer fish (Fugu rubripes). *Proc. Natl. Acad. Sci. USA* **92,** 1684-1688.

Bergson, C. and McGinnis, W. (1990). An autoregulatory enhancer element of the Drosophila homeotic gene Deformed. *EMBO J.* **9,** 4287-97.

Botas, J. (1993). Control of morphogenesis and differentiation by HOM/Hox genes. *Curr. Op. Cell. Biol.* **5,** 1015-1022.

Burglin, T. (1994). A comprehensive classification of homeobox genes. In *Guidebook to the Homeobox Genes* (ed. D. Duboule), pp. 26-71. Oxford, UK: Oxford University Press.

Chan, S.-K., 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.

Chan, S.-K. and Mann, R. S. (1993). The segment identity functions of Ultrabithorax are contained within its homeo domain and carboxy-terminal sequences. *Genes Dev.* **7,** 796-811.

Chan, S.-K. and Mann, R. S. (1996). A structural model for an extradenticle-HOX-DNA complex accounts for the choice of HOX protein in the heterodimer. *Proc. Natl. Acad. Sci. USA* **93,** 5223-5228.

Chan, S.-K., Pöpperl, H., Krumlauf, R. and Mann, R. S. (1996). An extradenticle-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. *EMBO J.* **15,** 2477-2488.

Chang, C.-P., Brocchieri, L., Shen, W.-F., Largman, C. and Cleary, M. (1996). Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the HOX locus. *Mol. Cell Biol.* **16,** 1734-1745.

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

Chouinard, S. and Kaufman, T. C. (1991). Control of expression of the homeotic *labial* (*lab*) locus of *Drosophila melanogaster*: evidence for both positive and negative autogenous regulation. *Development* **113,** 1267-1280.

Desplan, C., Theis, J. and O'Farrell, P. H. (1988). The sequence specificity of homeodomain-DNA interaction. *Cell* **54,** 1081-90.

Ekker, S., von Kessler, D. and Beachy, P. (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.

Ekker, S. C., Young, K. E., von Kessler, D. P. and Beachy, P. A. (1991). Optimal DNA sequence recognition by the Ultrabithorax homeodomain of Drosophila. *EMBO J.* **10,** 1179-86.

Furukubo-Tokunaga, K., Flister, S. and Gehring, W. J. (1993). Functional specificity of the Antennapedia homeodomain. *Proc. Natl. Acad. Sci. USA* **90,** 6360-4.

Garcia-Bellido, A. (1975). Genetic control of wing disc development in *Drosophila*. In *Cell Patterning, Ciba Found Symp*. (ed. S. Brenner), pp. ?? Associated Scientific Publishers, New York.

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G. and Wuthrich, K. (1994). Homeodomain-DNA recognition. *Cell* **78,** 211-23.

Gonzalez-Crespo, S. and Morata, G. (1995). Control of *Drosophila* adult pattern by *extradenticle*. *Development* **121,** 2117-2125.

Gould, A., Morrison, A., Sproat, G., White, R. and Krumlauf, R. (1997). Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns. *Genes Dev.* **11,** 900-913.

Goutte, C. and Johnson, A. D. (1993). Yeast a1 and alpha 2 homeodomain proteins form a DNA-binding activity with properties distinct from those of either protein. *J. Mol. Biol.* **233,** 359-71.

Goutte, C. and Johnson, A. D. (1994). Recognition of a DNA operator by a dimer composed of two different homeodomain proteins. *EMBO J.* **13,** 1434-42.

Gross, C. T. and McGinnis, W. (1996). DEAF-1, a novel protein that binds an essential region in a Deformed response element. *EMBO J.* **15,** 1961-70.

Harding, K. W., Gellon, G., McGinnis, N. and McGinnis, W. (1995). A screen for modifiers of Deformed function in Drosophila. *Genetics* **140,** 1339-52.

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

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

Kuziora, M. A. and McGinnis, W. (1988). Autoregulation of a Drosophila homeotic selector gene. *Cell* **55,** 477-85.

Laughon, A. (1991). DNA binding specificity of homeodomains. *Biochemistry* **30,** 11357-67.

Lou, L., Bergson, C. and McGinnis, W. (1995). Deformed expression in the Drosophila central nervous system is controlled by an autoactivated intronic enhancer. *Nucl. Acids Res.* **23,** 3481-7.

Lu, Q. and Kamps, M. (1996). Structural determinants of Pbx1 mediating cooperative DNA-binding with pentapeptide-containing HOX proteins. *Mol. Cell Biol.* **16,** 1632-1640.

Lu, Q., Knoepfler, P., Scheele, J., Wright, D. and Kamps, M. (1995). Both Pbx1 and E2A-Pbx1 bind the DNA m,otif ATCAATCAA cooperatively with the products of multiple murine *Hox* genes, some of which are themselves oncogenes. *Mol. Cell Biol.* **15,** 3786-3795.

Mann, R. and Abu-Shaar, M. (1996). Nuclear import of the homeodomain protein Extradenticle in response to Decapentaplegic and Wingless signalling. *Nature* **383,** 630-633.

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

Mann, R. S. and Chan, S.-K. (1996). Extra specificity from *extradenticle*: The partnership between HOX and exd/pbx homeodomain proteins. *Trends Genet.* **12,** 258-262.

McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* **68,** 283-302.

Morrison, A., Chaudhuri, C., Ariza-McNaughton, L., Muchamore, I., Kuroiwa, A. and Krumlauf, R. (1995). Comparative analysis of chicken *Hoxb-4* regulation in transgenic mice. *Mech. Dev.* **53,** 47-59.

Peifer, M. and Wieschaus, E. (1990). Mutations in the *Drosophila* gene *extradenticle* affect the way specific homeo domain proteins regulate segmental identity. *Genes Dev.* **4,** 1209-1223.

Peltenburg, L. T. C. and Murre, C. (1996). Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *EMBO J.* **15,** 3385-3393.

Phelan, M. L., Rambaldi, I. and Featherstone, M. (1995). Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. *Mol. Cell Biol.* **15,** 3989-3997.

Pöpperl, H., Bienz, M., Studer, M., Chan, S.-K., Aparicio, S., Brenner, S., Mann, R. and Krumlauf, R. (1995). Segmental expression of *Hoxb-1* is controlled by a highly conserved autoregulatory loop dependent upon *exd/Pbx*. *Cell* **81,** 1031-1042.

Pöpperl, H. and Featherstone, M. S. (1992). An autoregulatory element of the murine *Hox-4.2* gene. *EMBO J.* **11,** 3673-3680.

Rauskolb, C., Peifer, M. and Wieschaus, E. (1993). *extradenticle*, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene *pbx1*. *Cell* **74,** 1-20.

Rauskolb, C., Smith, K., Peifer, M. and Wieschaus, E. (1995). extradenticle determines segmental identities throughout development. *Development* **121,** 3663-3671.

Rauskolb, C. and Wieschaus, E. (1994). Coordinate regulation of downstream genes by extradenticle and the homeotic selector proteins. *EMBO J.* **13,** 3561-3569.

Regulski, M., Dessain, S., McGinnis, N. and McGinnis, W. (1991). High-affinity binding sites for the Deformed protein are required for the function of an autoregulatory enhancer of the *Deformed* gene. *Genes Dev.* **5,** 278-286.

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

Tremml, G. and Bienz, M. (1992). Induction of *labial* expression in the *Drosophila* endoderm: response elements for dpp signalling and for autoregulation. *Development* **116,** 447-456.

van Dijk, M. and Murre, C. (1994). *extradenticle* raises the DNA binding specificity of homeotic selector gene products. *Cell* **78,** 617-624.

Vincent, J. P., Kassis, J. A. and O'Farrell, P. H. (1990). A synthetic homeodomain binding site acts as a cell type specific, promoter specific enhancer in Drosophila embryos. *EMBO J.* **9,** 2573-2578.

Wilkinson, D. G., Bhatt, S., Cook, M., Boncinelli, E. and Krumlauf, R. (1989). Segmental expression of Hox-2 homoeobox-containing genes in the developing mouse hindbrain. *Nature* **341,** 405-9.

Xue, D., Finney, M., Ruvkun, G. and Chalfie, M. (1992). Regulation of the mec-3 gene by the C. elegans homeoproteins UNC-86 and MEC-3. *EMBO J.* **11,** 4969-79.

Xue, D., Tu, Y. and Chalfie, M. (1993). Cooperative interactions between the Caenorhabditis elegans homeoproteins UNC-86 and MEC-3. *Science* **261,** 1324-8.

Zeng, C., Pinsonneault, J., Gellon, G., McGinnis, N. and McGinnis, W. (1994). Deformed protein binding sites and cofactor binding sites are required for the function of a small segment-specific regulatory element in Drosophila embryos. *EMBO J.* **13,** 2362-77.

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

(Accepted 17 March 1997)