Regulation of a *decapentaplegic* midgut enhancer by homeotic proteins

J. Robert Manak, Laura D. Mathies and Matthew P. Scott*  
Departments of Developmental Biology and Genetics, Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305-5427, USA  
*Author for correspondence

SUMMARY

The clustered homeotic genes encode transcription factors that regulate pattern formation in all animals, conferring cell fates by coordinating the activities of downstream ‘target’ genes. In the *Drosophila* midgut, the *Ultrabithorax (Ubx)* protein activates and the *abdominalA (abd-A)* protein represses transcription of the *decapentaplegic (dpp)* gene, which encodes a secreted signalling protein of the TGFβ class. We have identified an 813 bp *dpp* enhancer which is capable of driving expression of a *lacZ* gene in a correct pattern in the embryonic midgut. The enhancer is activated ectopically in the visceral mesoderm by ubiquitous expression of *Ubx* or *Antennapedia* but not by *Sex combs reduced* protein. Ectopic expression of *abd-A* represses the enhancer. Deletion analysis reveals regions required for repression and activation. A 419 bp subfragment of the 813 bp fragment also drives reporter gene expression in an appropriate pattern, albeit more weakly. Evolutionary sequence conservation suggests other factors work with homeotic proteins to regulate *dpp*. A candidate cofactor, the *extradenticle* protein, binds to the *dpp* enhancer in close proximity to homeotic protein binding sites. Mutation of either this site or another conserved motif compromises enhancer function. A 45 bp fragment of DNA from *dpp* in the enhancer correctly responds to both UBX and ABD-A in a largely tissue-specific manner, thus representing the smallest in vivo homeotic response element (HOMRE) identified to date.  

Key words: homeotic, homeodomain, *decapentaplegic*, TGFβ, midgut development, *Drosophila*

INTRODUCTION

Dramatic changes in embryonic development occur when the function of one of the clustered homeotic genes is lacking or when one of the genes is active in the wrong place (reviewed by McGinnis and Krumlauf, 1992). In flies, worms, and mice, each of the genes is normally transcribed in a specific region of the animal along the anterior-posterior body axis. Eight *Drosophila* homeotic genes are located in two clusters called the Antennapedia and bithorax complexes and are believed to have originally been part of a single linked group. In mice and humans, a single cluster has been duplicated twice and some genes have apparently been lost, leading to four clusters called the HOX complexes with a total of 38 genes (reviewed by Mavilio, 1993). Nematodes appear to have a single corresponding cluster (Wang et al., 1993). The parallels in gene organization and function in vertebrates, insects, and worms strongly supports the view of the clusters as homologous (Duboule and Dolle, 1989; Graham et al., 1989).  

Homeotic proteins determine cell fates and the organization of appropriate structures by acting as transcription factors. Their powerful influence on developmental events is therefore thought to be due to regulation of arrays of downstream ‘target’ genes, although only a few such genes have been identified (reviewed by Andrew and Scott, 1992; Botas, 1993; Morata, 1993). Targets for *Antp* include *connectin*, *apterous*, *spalt major*, and *teashirt*. Targets for *Ubx* include *apterous*, *connectin*, *decapentaplegic*, *Distal-less*, *scabrous*, *teashirt*, and *β3 tubulin*. *abd-A* regulates many of the same genes regulated by *Ubx*, and also activates *wingless*. Some homeotic target genes encode transcription factors; others encode structural proteins, cell-surface molecules, and growth factors.  

The homeotic proteins have in common a 61 amino acid DNA-binding domain called the homeodomain (reviewed by Scott et al., 1989; Gehring et al., 1990). Each homeotic protein confers a particular pathway of development on the embryo, even when misexpressed in parts of the animal where the gene is normally repressed (reviewed by Hayashi and Scott, 1990). Using chimeric proteins, several investigators have shown that the homeodomain is crucial for determining the activity of homeotic proteins (Kuziora and McGinnis, 1989; Gibson et al., 1990; Lin and McGinnis, 1992; Chan and Mann, 1993; Zeng et al., 1993). This is particularly striking because the homeodomains of many of these proteins are highly similar in sequence. To understand better how homeodomain proteins affect transcription and, in turn, morphogenesis, it is necessary to define DNA sequences that are bound and regulated by these proteins in vivo.  

DNA binding of homeotic proteins in vitro has been examined using random oligonucleotide selection and fragments of DNA from genes likely to be genuine targets (Beachy et al., 1988; Müller et al., 1989; Affolter et al., 1990; Ekker et al., 1991; Regulski et al., 1991; Capovilla et al., 1994; Zeng et al., 1994). Only slight differences in DNA binding
specificity have been observed for different homeotic proteins in vitro (e.g. Capovilla et al., 1994), although these differences may be significant in vivo. The best defined homeotic response element known to be active in a proper pattern in vivo is the Deformed autoregulatory element, which has been defined as a 120 bp element containing a Deformed binding site and a site of action of other undefined factors (Zeng et al., 1994). Both sites are required for enhancer function. In other cases as well, DNA binding proteins may interact with the homeotic proteins to facilitate their action, perhaps by providing additional specificity in binding site selection. The best candidate for such a cofactor is the product of the extradenticle (exd) gene, a gene which affects the outcome of homeotic protein function without altering where the homeotic genes are expressed (Peifer and Wieschaus, 1990). exd encodes a homeodomain protein closely related to the mammalian proto-oncogene Pbxi (Rauskolb et al., 1993) and its family members. Recently, two research groups have provided evidence that UBX and EXD bind cooperatively to DNA (Chan et al., 1994; van Dijk and Murre, 1994). One group suggests that these proteins collaborate in activating the dpp midgut enhancer (Chan et al., 1994).

The regulation of Drosophila midgut development has been useful for identifying targets of homeotic genes. Four of the clustered homeotic genes, Sex comb-reduced (Scr), Antennapedia (Antp), Ubx, and abd-A, plus exd, are expressed in the midgut mesoderm. In the mesoderm, the homeotic genes are expressed in non-overlapping anterior-posterior domains (Fig. 1). The homeotic genes are required for the formation of draw-string-like constrictions in three parts of the midgut and for the outgrowth of pockets called caeca at the anterior of the midgut, thus dividing the midgut into compartments (Tremml and Bienz, 1989; Reuter and Scott, 1990). These morphological changes may be visible manifestations of additional differentiation events that functionally distinguish midgut cells along the anterior-posterior axis.

The center constriction forms at the junction between the Ubx and abd-A expression domains. Ubx activates transcription of decapentaplegic (dpp), which encodes a TGFβ homolog, while in the adjacent more posterior cells abd-A activates wingless (wg), which is a member of the Wnt class of signalling molecules (Fig. 1; Immerglück et al., 1990; Reuter et al., 1990; Hursh et al., 1993). UBX activation of dpp appears to be direct, based on compensatory mutations constructed in a dpp enhancer and in the regulating protein (Capovilla et al., 1994); it is unclear whether ABD-A directly represses dpp by binding to its regulatory regions. Both dpp and wg proteins move into the underlying endoderm where they are required for the proper expression and localization of the homeotic gene labial (Immerglück et al., 1990; Reuter et al., 1990; Tremml and Bienz, 1992). The center constriction fails to form in the absence of either dpp or wg. Therefore, in addition to being regulated by homeotic proteins, these targets are clearly implicated in carrying out morphogenetic functions.

We have identified a small part of the dpp gene capable of directing expression of lacZ in the midgut cells where dpp is expressed. We report our analysis of this enhancer element.

**MATERIALS AND METHODS**

**Construction of dpp-lacZ and bacterial expression gene fusions**

A 3.1 kb KpnI/EcoRI fragment of plasmid pC7276 (from J. Masucci and F.M. Hoffmann) was subcloned into pTZ19U (USBiochemical), which produced pTZCS K/R. From this plasmid, a BamHI/EcoRI fragment (813 bp; this fragment was sequenced by the Sanger dideoxy method), a KpnI/EcoRI fragment (3.1 kb), a Sph/BamHI fragment (2.0 kb), a SacII/KpnI fragment (1.0 kb), and a BamHI/SacII fragment (1.2 kb), were individually subcloned into the P-element transformation vector pC4PLZ. pC4PLZ (K. Wharton and S. Crews, unpublished data) contains a basal P-element promoter fused to the lacZ gene and is marked with the white gene. To make the Xhol/PstI fragment fused to lacZ, pTZ CS K/R was digested with Xhol and blunted with DNA polymerase I, Klenow fragment. Phosphorylated PstI linkers (GCTGCAGC) were ligated onto the blunt ends. The 0.4 kb PstI fragment was subcloned in both orientations into pTZ18U cut with PstI. One of the resulting plasmids, pTZ18 P/X-, was cut with HindIII (in the pTZ polylinker) and blunted with Klenow. BamHI linkers (CGGATCCG) were ligated onto the blunt end and the plasmid was digested with BamHI. This produced the 413 bp fragment which was
subcloned in both orientations into the vector pC4PLZ. The 261 bp fragment was produced by digesting pTZ18 P/X- with BssH2 and XhoI, blunting and religating. The resulting construct was digested with BamHI and the 261 bp fragment was subcloned in the forward orientation into pC4PLZ. pC4PLZZX45mer was produced by ligation of annealed and phosphorylated oligos (GATCCAAATTGCA-GCGCGCATCAAATTATTATTTAACTTGAAGTTTTTGAATTGATGCTGCAATTG) into the BamHI site of pC4PLZ. A double insert was then selected and sequenced. The resulting plasmid contains two direct repeats of the 45mer in reverse orientation with relation to the transcription start site. The delta DR construct was made as follows: the 0.4 kb PstI fragment from pTZ18 P/X- was subcloned into the PstI site of pBS II S/K+ so that the the 5’ end of the enhancer fragment was situated toward the KpnI end of the polylinker. The polylinker EcoRI site was destroyed by digestion with EcoRI, blunting with Klenow and religation. The plasmid was digested with Sphi, which cuts in the direct repeat, and the Sphi site was destroyed by chew-back blunting with T4 DNA polymerase. Three EcoRI linkers were subcloned onto the blunt ends. The sequence thus reads GCTG(GGAATTCC)3 CTGC from nucleotide 185 in Fig. 9. A 0.4 kb KpnI/BamHI fragment from this plasmid was subcloned into pC4PLZ in the forward orientation.

Plasmid ptd48-3, which contains exd cDNA (Rauskolb et al., 1993), was digested with Sphi and the 1 kb exd fragment, encoding amino acids 176 to 376, was ligated into pUC19 digested with Sphi. The plasmid was digested with PstI, which cuts in the direct repeat, and the Sphi site was destroyed by chew-back blunting with T4 DNA polymerase. Three EcoRI linkers were subcloned onto the blunt ends. The sequence thus reads GCTG(GGAATTCC)3 CTGC from nucleotide 185 in Fig. 9. A 0.4 kb KpnI/BamHI fragment from this plasmid was subcloned into pC4PLZ in the forward orientation.

Library screening

A Drosophila virilis genomic phage library constructed in lambda EMBL3 by John Tamkun was screened at moderate stringency (hybridization at 65°C in 4× SSPE, 1% sodium dodecyl sulfate, 0.5% non-fat dry milk; washes at 65°C in 2× SSPE, 0.2% sodium dodecyl sulfate, 0.05% sodium phosphate) using an 813 bp BamHI to EcoRI fragment isolated from plasmid pTZCS K/R (see below). PstI/HindIII and PstI/EcoRI fragments of the phage clones that hybridized to the probe were subcloned into pUC19 to produce pUCvirP/H and pUCvirP/R. The inserts in both plasmids were fully sequenced on both strands.

Fly stocks and P-element injections

All P-element vector constructs were co-injected with pA2-3 transposase DNA as helper (D. Rio, unpublished data) into Df[1]w1118 embryos as previously described (Rubin and Spradling, 1982). At least three independent transformant lines of each construct were generated.

Immunohistochemistry and antibody sources

Immunostaining was as described in Mathies et al. (1994). Antisera to β-gal protein were generated in rats and rabbits. The rabbit antibody was affinity purified. The antibody to dpp protein was the kind gift of F. M. Hoffmann.

Heat shock induction of homeotic proteins

Virgin females from a stock containing the 813 bp enhancer constructs homozygous on the third chromosome were crossed to males containing heat shock inducible homeotic protein constructs (Mann and Hogness, 1990; Zeng et al., 1993). Embryos were collected for 10 hours from this transient cross, then induced by heat shock. The embryos were transferred to 37°C three times for 30 minutes with a recovery at RT for 30 minutes between heat pulses. The embryos were then aged for 2.5 hours, fixed, and stained for β-gal expression as described in Mathies et al. (1994).

Expression and purification of proteins

DNA-affinity purified UBX protein was provided by B. Johnson and M. Krasnow (Johnson and Krasnow, 1990). ABD-A protein was prepared as described in Appel and Sakonju, 1993 using plasmid pNB40 abd-A 65 UAC except that a Pharmacia Sephadex column was used in the FPLC purification. ABD-A protein was further purified by DNA-affinity column chromatography as described (Kadonaga, 1991). Briefly, two oligonucleotides of the sequence GATCCA(TTA)11 TTG and GATCCA(ATA)11 ATG were annealed and ligated to form concatamers which were then attached to CNBr-activated sepharose CL-2B beads (Pharmacia). ABD-A protein extract was applied to the DNA affinity column and the ABD-A protein was subsequently eluted with a series of concentrations of KCl ranging from 0.2 M to 1 M. Based on SDS-polyacrylamide gel electrophoresis followed by Coomassie Brilliant Blue staining, we estimate the final protein preparation is greater than 90% pure.

Expression and purification of the 69×10^3 M_r MBP/EXD fusion protein from pMal c2 exd Sph was performed exactly as described in the Protein Fusion and Purification System protocol from New England Biolabs, but using the bacterial strain DH5a. Bacterial cultures harboring the expression plasmid were induced with IPTG. The fusion protein was purified on an amylose column and eluted with maltose. Similar DNAase I footprint results were obtained with the MBP/EXD fusion protein as well as with protein cleaved with Factor Xa, which separates the MBP domain from the EXD polypeptide.

RESULTS

Identification of a dpp midgut enhancer

dpp is transcribed in three regions of the midgut visceral mesoderm, overlying where the gastric caeca (GC) form, just anterior to where the central constriction forms (parasegment 7, PS7) and, nearly undetectably, close to the posterior constriction site (St. Johnston and Gelbart, 1987; Fig. 3A). Previous work (Hursch et al., 1993; Masuccio and Hoffmann, 1993) led to the identification of a midgut expression control region within map coordinates 71.3-75.9, several kilobases upstream of the dpp promoters (Fig. 2). Fragments from this region were joined to a short relatively inactive promoter fused to lacZ and introduced into flies using a P element vector. At
least three independent lines were established for each fragment, and the pattern of expression was observed with antibodies against β-galactosidase (β-gal) protein.

Three fragments (3.1 kb, 813 bp, 419 bp) containing the 419 bp sequence drive appropriate midgut visceral mesoderm expression (Fig. 2), although expression from the 419 bp construct is somewhat reduced compared with the larger two. Three fragments upstream of the 3.1 kb did not produce detectable β-gal protein. We shall hereby refer to the 813 or 419 bp constructs as dpp813 and dpp419.

Expression of dpp813 and dpp419 is shown in Fig. 3B.C. The precise correspondence of the endogenous dpp pattern and the transgene expression is demonstrated by a double label experiment for dpp813 (Fig. 3D). Both enhancer-lacZ transgenes are expressed more strongly in the PS7 domain than in the gastric caeca domain, while the endogenous gene is detected at roughly equal levels in both domains (compare Fig. 3A with 3B,C). Both dpp813 and dpp419 were used to examine regulation by homeotic proteins.

Regulation of the enhancer element by homeotic proteins

Like dpp itself (Reuter et al., 1990), dpp813 is ectopically activated in the visceral mesoderm by ubiquitous expression of Ubx (Fig. 4A,B). Also, like the dpp gene, the ectopic activation is only detected in visceral mesoderm cells that do not contain ABD-A. Ubiquitous expression of abd-A causes a decrease in the expression of the dpp-lacZ construct in the visceral mesoderm (Fig. 4E,F). Ubiquitously expressed Scr, which does not normally affect dpp expression in the midgut, produced only a very slight ectopic induction just anterior to PS7 in some embryos (data not shown).

In contrast ubiquitous Antp ectopically induces the dpp enhancer through most of the anterior midgut in a patchy pattern, albeit much less strongly than Ubx (compare Fig. 4B with 4D). Expression of dpp419 in Ubx and abd-A mutant embryos is consistent with the results obtained with dpp813 in the experiments above (data not shown).

Isolation of the corresponding DNA fragment from Drosophila virilis

Comparisons of enhancer sequences from D. melanogaster and D. virilis have been useful for identifying important control elements (Kassis et al., 1986; Fortini and Rubin, 1990; Maier et al., 1990). The two fly species are thought to have had a common ancestor about 60 million years ago (Beverley and Wilson, 1984). A single-copy band was detected upon probing restriction enzyme-D. virilis DNA (data not shown), and a corresponding fragment was obtained in phage clones from a genomic D. virilis library. The D. melanogaster and D. virilis fragments were sequenced. The two sequences were aligned by computer and eye to optimize the number of matches.

Fourteen regions of dramatic sequence conservation stand out (Fig. 5A,B). About 0.5 kb of the D. virilis sequence is not represented in the D. melanogaster sequence. All the patches of sequence similarity are colinear and identically oriented, and the spacing between them reasonably well preserved, excepting the 0.5 kb insert. Because conserved sequences from regions 5′ and 3′ of the 419 bp fragment are present, it is likely that the region of D. virilis DNA studied contains all the sequence relevant to the dpp419 expression pattern.

Fig. 2. The decapentaplegic midgut control region. The general region required for expression of dpp in the embryonic midgut was previously defined (Hush et al., 1993; Masucci and Hoffmann, 1993). There are several different dpp promoters; the one closest to midgut regulatory elements is located at position 78 (Masucci and Hoffmann, 1993), about 3 kb from the position of the 419 bp enhancer shown in the present study to control midgut expression. Several different fragments (A) from the region were tested for their ability to direct lacZ expression in the midgut. The fragments shown in B were tested in the vector shown in C. The orientations tested are indicated; ‘forward’ indicates that the 5′ to 3′ orientation of the fragment relative to the transgene promoter was the same as found in the dpp gene, while ‘reverse’ indicates the fragment was flipped end for end. PS7 stands for visceral mesoderm expression near the central constriction; GC stands for the expression of dpp near the gastric caeca. The pC4PLZ vector contains a P element promoter driving expression of a β-gal protein containing a nuclear targeting signal (Materials and Methods).
Locations of binding sites for UBX and ABD-A proteins
The 419 bp fragment is sufficient to direct proper midgut expression, so we tested whether homeotic proteins can bind the fragment. DNase I footprint experiments on this fragment were done using bacterially produced full-length UBX and ABD-A proteins purified to near homogeneity (Figs. 6A-D). Both proteins bind particularly well to a cluster of sites (sites F-I, Figs. 6 and 9) located within about 100 bp of DNA. Another region (site E) binds both proteins, but with considerably lower affinity. UBX protein binds better to the assorted enhancer sites than ABD-A: roughly four times as much ABD-A is required to achieve the extent of protection provided by UBX. However, in contrast to the conclusions reached by Capovilla et al. (1994), we do not observe obvious differences in affinity of UBX versus ABD-A with respect to different binding sites. Moreover, we observe ABD-A binding to site B (site 4a as described in Capovilla et al., 1994). Binding sites E, F, and I, which correspond to sites 3, 2, and 1, respectively in Capovilla et al. (1994), are located in highly conserved blocks of sequence, and each contains at least one TAAT motif, a sequence known to be a common motif in homeodomain binding sites (Laughon, 1991). Binding sites A and B, which correspond to sites 4b and 4a (Capovilla et al., 1994) are located in a region of poor homology. Neither site A nor B has a TAAT motif.

Partial deletion of the enhancer causes ectopic expression
The 3’ region of the 419 bp fragment was removed to produce the 261 bp construct (dpp261; Figures 2 and 9), deleting homology blocks 11 and 12, including one of the strongest binding sites for homeotic proteins. dpp261 embryos show striking derepression throughout the midgut mesoderm, suggesting that abd-A and other factors can no longer repress expression in the posterior and anterior midgut (Fig. 7). Bringing the 261 bp enhancer closer to the promoter region of the lacZ vector (see Materials and Methods) could cause an altered expression pattern, but the dramatic changes in expression suggest the deletion, rather than position, is responsible. Expression remains highest in the central and anterior parts of the mesoderm, with strong ectopic expression immediately posterior to the normal gastric caeca domain. Strong expression is also observed in cells where abd-A is expressed, in a gradient tapering to a low level at the posterior. While the tissue specificity is preserved in the midgut, in the hindgut...
derepression is observed in the endoderm (Fig. 7D). The strong binding site deleted in constructing dpp261, Site I, may therefore be a target for repression by ABD-A.

A direct repeat sequence is required for enhancer-driven high-level expression

The conservation of a direct repeat sequence located around nucleotide 192 suggested that this sequence might be important for enhancer function. We compared five lines harboring a mutation in this sequence in the context of dpp419 (dppΔDR; see Materials and Methods) with five lines harboring dpp419. The dppΔDR lines showed reduced expression in the visceral mesoderm (in both gastric caeca and PS7) compared with the dpp419 lines (compare Fig. 7A with C). However, we cannot rule out the possibility that the decreased expression is due to the insertion site of the constructs.

A 45 bp fragment of the dpp midgut enhancer responds to homeotic genes

Site I and/or its surroundings are required for repression in the anterior and posterior visceral mesoderm. Site F is the other strong homeodomain protein binding site. Based on the cross-species homology data, an oligonucleotide was synthesized containing 45 bp centered around site F (Fig. 8). Two copies of the sequence were built into the lacZ vector, producing the construct dpp45. Expression, especially early, is mainly located within PS7 of the visceral mesoderm (Fig. 8A). Some patchy ectopic expression is observed anteriorly, as well as around the third constriction at later stages of development. Expression of the dpp 45mer in PS7 is narrower than that observed for either dpp419 or dpp813 (compare expression indicated by closed arrow in Fig. 8A with PS7 expression in Fig. 3C). Although tissue specificity is generally maintained, some expression is observed in glial cells as well as transverse muscles of abdominal segments 1 to 3 (open arrows in Fig. 8A). In embryos that contain ubiquitous UBX driven by a heat shock promoter, expression from dpp45 is observed within as well as anterior to PS7 (Fig. 8B). No expression is observed posterior to PS7. This is exactly the behavior of dpp813 and the dpp gene itself, and suggests that the 45mer responds to both UBX and ABD-A.

Locations of binding sites for the EXD protein

Previous studies suggest that the exd gene product might work in conjunction with a subset of the homeotic proteins, including Ubx and abd-A, to regulate homeotic target genes (Peifer et al., 1991; Rauskolb et al., 1993). dpp expression in PS7 is greatly reduced in exd mutant embryos (Rauskolb and Wieschaus, 1994), suggesting that Ubx and exd might collaborate to regulate dpp in PS7. An EXD fusion protein binds to
Fig. 5. Comparison of the sequences of the D. melanogaster and D. virilis dpp enhancers. The sequences of the 813 bp D. melanogaster midgut enhancer and the 1413 bp D. virilis fragment were determined and compared. Dot matrix computer comparisons and further refinements by eye were used to identify similar sequences. Conserved regions are interspersed with blocks of very similar sequence. A summary of the positions of the similar blocks (A) is correlated by numbers with the specific sequences shown in B. The ratios in the parentheses above homology block numbers indicate exact identity matches over total number of nucleotides within the homology blocks of D. melanogaster. The assignments of numbers to conserved blocks are made only to facilitate discussion of the sequences. In most cases the divisions between the numbered blocks are placed where differences in spacing between conserved sequences occur in the two species. The large brackets indicate the extent of the 261 and 419 bp enhancers. These two smaller enhancer fragments have the same 5' end and different 3' ends. The box at the bottom contains the block of D. virilis sequence which belongs at the insertion site shown. This is the only major difference in spacing or orientation between the two species. The orientation shown is 5' to 3' with respect to the dpp gene.
the 419 bp fragment (Figs 6D and 8, EXD) overlapping the E binding site recognized by UBX and ABD-A. The central sequence in this site (ATCAATTAG) is almost identical to the consensus binding site of the \textit{exd} human homolog PBX1 (ATCAATCA[A]) determined by in vitro analyses (van Dijk et al., 1993; LeBrun and Cleary, 1994).

Enhancer function is affected in flies that harbor a mutation of this sequence in the context of the 419 bp (419exd) or 261 bp (261exd) enhancers (see Fig. 7B and F, respectively). Gastric caeca expression is abolished in the 419exd flies and is greatly reduced in 261exd flies (compare loss of gastric caeca expression as indicated by an arrow in Fig. 7B and F with Fig. 7A and E, respectively). To test whether expression in PS7 is affected, expression levels in five lines of 419exd flies were compared with those in five lines of dpp419 flies. Overall levels of expression in PS7 of flies bearing the mutant enhancer
were reduced (data not shown), although, again, it is impossible to completely rule out lowered expression due to position effects. The level of dpp enhancer expression in both the gastric caeca and PS7 regions therefore appears to be dependent upon the integrity of the EXD binding site, in keeping with loss of dpp expression and dpp enhancer expression in those locations in exd mutants (Rauskolb and Wieschaus, 1994).

**DISCUSSION**

The dpp gene performs diverse roles in development and is expressed in many tissues and at several developmental stages. The structure of the dpp gene reflects its complex regulation. At least four promoters contribute to the overall pattern of expression of a single protein-coding sequence, and controlling transcriptional enhancers are distributed over 50 kb. Two groups (Hursh et al., 1993; Masucci and Hoffmann, 1993) identified a region of the promoter needed to activate dpp transcription in the gastric caeca and central constriction of the midgut visceral mesoderm. Within that region we have identified a 419 bp sequence sufficient to direct appropriate lacZ expression in the midgut mesoderm.

**Characteristics of homeotic protein binding sites in the dpp promoter**

Few direct target sequences of homeotic proteins have been characterized. Examples of such targets include Distal-less (Dll) as well as the clustered homeotic genes Deformed (Dfd) and Antp. The Dll and Antp genes are both repressed in the epidermis of the abdominal segments by Ubx and abd-A. Dfd protein is required for expression of a Dfd autoregulatory element. Mutation of the homeotic binding sites identified within the control regions of these genes compromises the enhancer function. (Vachon et al., 1992; Appel and Sakonju, 1993; Regulski et al., 1991; Zeng et al., 1994). Recently, Capovilla et al (1994) have identified a dpp midgut enhancer
that overlaps with the enhancer described in this report. The authors expressed a mutant form of UBX in flies that harbor an altered form of the dpp enhancer fused to a lacZ reporter gene. The mutant UBX is capable of recognizing the altered enhancer in vitro. Based on the observed expression pattern of lacZ in the midgut, they argued for a direct interaction between UBX and the dpp enhancer. Although this strategy has been used in a few cases (e.g. Schier and Gehring, 1992), and can provide strong evidence for direct interaction, the altered UBX protein could recognize the altered site even if there is normally an intermediary between UBX and the dpp enhancer. A stronger case for direct interaction would come from altering one of the components, e.g. the protein sequence, and screening for compensatory mutations in the enhancer. In this way an intermediary could be identified.

Most homeotic protein binding sites studied to date, including those identified in the Dll and Antp enhancers, contain TAAT sequences (reviewed by Laughon, 1991). TAAT serves as a core part of the recognition sequence in the crystal structures of the engrailed and MATa2 homeodomains bound to DNA (Kissing et al., 1990; Wolber et al., 1991). TAAT alone does not define a homeotic protein binding site, as strong preferences for two nucleotides just 3′ to the TAAT have been found. The favored nucleotides 3′ to the TAAT for the UBX homeodomain include GG, TA, AG, and TG in order of decreasing preference (Ekker et al., 1991), and CC for the quite different bicoid homeodomain (Driever and Nüsslein-Volhard, 1989). The binding sites for UBX and ABD-A in the dpp enhancer (Table 1) are largely consistent with these in vitro binding studies. The two nucleotides 3′ to the TAAT in each dpp binding site are GG or TG, while the block 5 site that is not bound in vitro contains the sequence TAATCG. One high affinity site (I) contains the TAATGG motif whereas another (F) contains the motif TAATTG. The medium affinity site E also contains the sequence TAATTG, so this motif alone is clearly not sufficient for high affinity DNA binding in vitro.

The sequence conservation between the D. melanogaster and D. virilis dpp midgut enhancers suggests that other factors necessary for enhancer function may bind adjacent to the homeotic proteins. The regions of sequence conservation can extend over 30 bp in length, the conserved blocks generally exceeding the extent of the DNaseI footprints of the homeodomain proteins. Each homeodomain binding site is flanked by conserved sequence which does not contain a TAAT and is not bound by the homeotic proteins. We have shown that one such element, the direct repeat motif located between sites E

---

**Fig. 8.** Expression pattern of dpp45. The dpp45 transgene is expressed in PS7 in a largely tissue-specific manner. A shows dpp45 expression in PS7 (closed arrow) as well as staining in transverse muscles of abdominal segments 1 to 3 (open arrows). B shows expanded dpp45 expression in the presence of ubiquitously expressed UBX. The embryo in B was slightly enlarged photographically to facilitate comparison of panels A and B. The model presented in C summarizes our current knowledge of regulation through dpp419. In the model ABD-A represses dpp419 through at least two sites located within blocks 9 and 11, based on the loss of responsiveness of dpp261 to ABD-A and the sensitivity of dpp45 to it. Other anterior repressors act through blocks 10 and 11. UBX activates dpp419 through at least block 9 and perhaps through other blocks. A general visceral mesoderm activator binds the direct repeat. Either a general or gastric caeca-specific activator, which may or may not be EXD, acts through block 8. DR, direct repeat. Numbers below boxes indicate homology blocks. Numbers at ends indicate enhancer coordinates. Labels to the left indicate various constructs that were tested in this study. Brackets indicate boundaries of constructs.
Table 1. Comparison of homeodomain binding site sequences with TAAT motifs within the 813 bp fragment

<table>
<thead>
<tr>
<th>Site</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site I (site 1)</td>
<td>AAAACCTTAATGGCCACCATTAAC</td>
</tr>
<tr>
<td>Site F (site 2)</td>
<td>TATTTCTTAATTGGTGTGGATT</td>
</tr>
<tr>
<td>Site E (site 3)</td>
<td>GGGGCTTAATTGAGTGGCTTT</td>
</tr>
<tr>
<td>Site E-rev</td>
<td>AATTGTAATTAAATTGTGAATTC</td>
</tr>
<tr>
<td>Site E-3'</td>
<td>GCGAGCAATTGCGAAGGTGTC</td>
</tr>
<tr>
<td>Block 3</td>
<td>TGGACGTAATTGCAACATTACG</td>
</tr>
<tr>
<td>Block 4</td>
<td>CGGGCTTAATTGATGATTGTTAG</td>
</tr>
<tr>
<td>Block 14-5'</td>
<td>CGAACCTTAATTGCAATGGCT</td>
</tr>
<tr>
<td>Block 14-3'</td>
<td>CGTGTTTAATTCGAATTT</td>
</tr>
</tbody>
</table>

Sites E, F, and I are bound by both UBX and ABD-A. Blocks 3, 4, 14-5' and 14-3' (which correlate with blocks shown in Fig. 5B) are TAAT motifs in conserved blocks outside of the 419 bp fragment which were not tested for homoeotic protein binding. Numbers in parentheses correspond to TAAT regions of in vitro homeotic binding sites as described in Capovilla et al. (1994).

and F (indicated in Fig. 9), does not bind either UBX or ABD-A, and yet is required for high-level expression driven by the enhancer. Identification of the proteins that bind this and other elements will lead to a better understanding of how homeotic-responsive enhancers function. Not every conserved element need serve a function in determining the specificity of action of homeotic proteins. In its place within the array of enhancers in the dpp gene, the 419 bp element may serve multiple roles in addition to its role in visceral mesoderm expression. Some conserved elements may serve no regulatory function.

Specificity of homeotic protein action

ANTP and UBX both activate the dpp enhancer in the anterior midgut mesoderm, while SCR has no effect. Even though ANTP does not normally activate the dpp enhancer, perhaps the high level in the heat shock experiment causes it to have this abnormal activity. It may, for example, bind to UBX/ABD-A binding sites from which it is normally excluded. ANTP and UBX differ by 7 amino acids in the homeodomain, while ANTP and SCR differ by only four amino acids near the N-terminal part of the homeodomain. These differences are, however, crucial to the proteins’ specific actions (Kuziora and McGinnis, 1989; Gibson et al., 1990; Chan and Mann, 1993; Zeng et al., 1993). The distinct response to ANTP and UBX versus SCR suggests that binding of SCR in vivo may be ineffective. The binding affinity and/or specificity of ANTP and UBX could be increased by interaction with other proteins. One candidate is the product of the exd gene, which is required for dpp expression in PS7 and interacts genetically with Antp and Ubx, but not Scr. Alternatively SCR may bind to the dpp enhancer but have no effect once bound. In the cuticle SCR is blocked from regulating denticle pattern by UBX and ABD-A, but in the induction of salivary glands the homeotic gene teashirt is the relevant limiting influence (Andrew et al., 1994). The influences limiting SCR function in the mesoderm have not been identified.

Repression versus activation

Removal of homology blocks 11 and 12 causes a dramatic derepression of lacZ expression, especially posteriorly. However, Capovilla et al. (1994) report that site-directed mutation of the TAAT region of block 11 (their site 1) has only a moderate effect on expression. It is unlikely that the removal of block 12 has any effect on expression, because the 303 bp fragment they have characterized does not contain block 12 yet drives normal expression. Therefore it is likely that removal of block 11 in its entirety causes the dramatic derepression. The site-directed mutation of homology block 11 may have been insufficient to fully compromise its function.

The derepression observed with dpp261 suggests that activation in the visceral mesoderm is the default state, with a variety of repressors restricting expression. This view is further supported by residual dpp expression seen in mutants completely lacking Ubx function (Reuter et al., 1990). Ubx seems necessary to augment dpp expression, not to initiate it. The regulator(s) which activates dpp in PS7 in the absence of Ubx may also facilitate the activation of dpp by UBX.

ABD-A represses dpp while UBX and ANTP activate dpp, but the in vitro DNA binding data do not explain the different activities of the UBX versus ABD-A proteins. The enhancer is bound by UBX and ABD-A proteins at the same sites in vitro, with our preparation of UBX protein binding more strongly than ABD-A. How then does binding of the proteins produce different actions? The pattern of expression of dpp261 suggests that this deletion construct is no longer efficiently repressed by abd-A in the posterior, or by different factors in the anterior midgut and hindgut. The single strong homeodomain binding site which distinguishes dpp419 from dpp261 appears to be a major site of repression acted upon by a variety of proteins, including ABD-A (Fig. 7D,E). dpp261 is expressed more strongly in PS7 than in PS8 and therefore is sufficient to respond to UBX. However, the homeotic binding site removed to create dpp261 might also be used by UBX in vivo. Similarly, the residual repression of dpp261 in the posterior midgut (the derepression of dpp in abd-A mutants exceeds the derepression seen with dpp261) could be due to weak effects of abd-A on remaining binding sites.

The difference between repression and activation could be determined by the interaction of UBX and ABD-A proteins with distinct cofactors. Ubiquitous cofactors could differentially interact with either UBX or ABD-A, or a cofactor protein capable of interacting with only one homeotic protein may be made only in certain cells. A necessary activator protein might be absent from PS8, preventing ABD-A from activating dpp, or the factor might be present in PS8 but unable to act with ABD-A. Ubiquitous expression of ABD-A prevents dpp expression rather than permitting it, so ABD-A appears unable to cooperate with a PS7 factor to activate dpp. It is also possible that UBX or ABD-A could exert their effects by modulating regulation by a third as yet unidentified regulator. Finally, UBX and ABD-A could interact with the same cofactor but have different effects once bound.

45 bp of the dpp midgut enhancer defines a homeotic response element (HOMRE)

We have identified an element 45 bp in length that responds to both UBX and ABD-A homeotic proteins. The response of only 45 bp of dpp enhancer to these proteins, with tissue specificity largely preserved, suggests that most factors necessary for promoting appropriate visceral mesoderm expression can recognize 45 bp, or less, of DNA. It is now possible to dissect base by base this homeotic response element to help understand how homeotic protein specificity is achieved. There are two TAAT motifs in dpp45. Both ABD-A and UBX may bind.
to the same site, or the two proteins may each have a preferred site of action.

Why is dpp261, which contains the 45bp of dpp45 sequence, posteriorly derepressed while the dpp45 construct is not? Perhaps the ratio between a general visceral mesoderm activator and homeotic repressor (ABD-A) determines whether expression does or does not occur posteriorly. In the case of dpp261, one of the few high affinity ABD-A binding sites has been removed and activation predominates. In the case of the dpp45 construct, ABD-A need only overcome the general activation provided by one or few visceral mesoderm activators, assuming that few proteins can bind 45 bp of DNA, so repression predominates. TGCATGCA-like motifs, which occur abundantly within the 419 bases of DNA (Fig. 9) and can bind a visceral mesoderm activator, are plentiful within the 261 bp fragment but occur only once in dpp45.

The results reported here and by Capovilla et al. (1994) support the existence of tissue-specific information for dpp expression tightly linked to homeotic protein responsiveness. The low level residual expression of dpp in PS7 in Ubx null mutants (Reuter et al., 1990) may be due to the activators whose existence we infer. Ubx activity in PS7 raises the level of expression, while unknown repressors in the anterior and Abd-A in the posterior lower dpp expression. In an evolutionary context, this refinement of a tissue-specific expression pattern into a more elaborate spatial pattern may reflect acquisition of HOMRE’s by the dpp gene.

It is interesting to note that dpp45 does not contain an in vitro EXD binding site nor does it contain a sequence that resembles the ATCAATCA motif (see below). Although it is possible that EXD does interact with dpp45, it is also possible that other as yet unidentified factors are interacting with UBX and ABD-A. Whether such factors also impart tissue specificity remains to be determined.

**Binding by extradenticle protein**

exd has long been the best candidate to encode a homeotic protein cofactor. Careful analyses by Peifer and Wieschaus (1990) demonstrated that although homeotic transformations occur in exd mutants, there is no detectable change in the expression patterns of the clustered HOX class homeotic genes. The findings are in clear contrast to the effects of repressors and activators of the Polycomb (paro, 1993) and trithorax groups (Kennison, 1993), where changes in homeotic gene expression account for the homeotic transformations seen in the mutants. Peifer and Wieschaus (1990) further substantiated the role of exd in homeotic protein function by showing how exd alters the effects of ectopically expressed homeotic proteins. They concluded that exd is likely to encode either a gene involved in carrying out the orders given by homeotic proteins, a target gene, or a cofactor that helps to control the specificity of target gene recognition or regulation.

---

**Fig. 9.** Location of binding sites for the UBX, ABD-A, and EXD proteins on the *D. melanogaster* enhancer. The locations of binding sites for homeotic proteins on the 419 bp midgut enhancer are shown in comparison to the pattern of evolutionary conserved sequences. The sequence is shown in its 5’ to 3’ configuration as it is arranged with respect to the dpp promoter in the chromosome. The enhancer is numbered from 1 to 419, with position 1 corresponding to position 262 in Fig. 5B. Gray shading indicates sequences conserved between *D. melanogaster* and *D. virilis*, with the boxed numbers corresponding to the block numbers in Fig. 5. The 261 bp version of the enhancer has the same 5’ end as the 419 bp, but ends at the small bracket within conserved block 10. In addition to the 256 bp indicated, the method of building the fragment left an additional ‘TCGAG’ at the 3’ end for a total of 261 bp. The large brackets demarcate the 45 bp fragment capable of responding to UBX and ABD-A. Notable features of the sequence are identified by boxes (TAAT sequences which form the core of many homeodomain binding sites), underlining (TGCAG repeats which are surprisingly abundant in the sequence), arrows (a direct repeat with dyad symmetrical core elements), or by horizontal brackets (TGCAG-like motifs). The positions of strong, medium, and weak binding sites for UBX and ABD-A proteins are indicated by bars below the sequence. The EXD binding site is indicated. The vertical bars indicate *D. melanogaster* sequence blocks that are conserved but interrupted in *D. virilis*.
The *exd* protein is a homeodomain protein most closely related to the yeast MAT product *a1* (Rauskolb et al., 1993), a cofactor for the MATα2 homeodomain protein in repressing haploid-specific target genes (Johnson, 1992). Although the UBX homeodomain does not look particularly similar to MATα2, the similarity of EXD to MATα1 makes its role as a cofactor plausible. Rauskolb and Wieschaus (1994) tested target genes known to be regulated by homeotic proteins, including *dpp* in the midgut mesoderm, to see whether their expression patterns are dependent on *exd*, and found that they are. Thus the homeotic gene expression patterns are preserved in *exd* mutants but the target gene expression patterns are not. They found a requirement for EXD in activating *dpp* expression in PS7.

EXD is closely related to a mammalian protein, PBX1, implicated in human leukemias by chromosome rearrangements causing a chimeric PBX1/E2A protein to be produced (Kamps et al., 1990; Nourse et al., 1990). PBX1 and EXD are extremely similar within the homeodomain, and quite similar throughout the rest of the proteins (Flegel et al., 1993; Rauskolb et al., 1993). A cDNA encoding the hybrid protein introduced into mice causes lymphomas (Dedera et al., 1993). One possible interpretation of these observations is that the homeodomain or other parts of PBX1 targets the chimeric protein to genes normally regulated by PBX1. Because the hybrid protein causes loss of growth control in blood cells, normal targets of PBX may include growth regulating genes.

Whatever the exact mechanism of leukemogenesis, the high degree of similarity between PBX and EXD suggests PBX may function as a cofactor for the MATα2/E2A protein to genes normally regulated by PBX1. Because the homeodomain or other parts of PBX1 targets the chimeric protein to genes normally regulated by PBX1. Because the hybrid protein causes loss of growth control in blood cells, normal targets of PBX may include growth regulating genes.

The apparent reduction of expression in both the gastric caeca and in PS7 of the visceral mesoderm upon mutation of the EXD binding site is consistent with reduction of *dpp* and *pp* enhancer/lacZ expression from both gastric caeca and PS7 in *exd* mutants (Rauskolb and Wieschaus, 1994). However, loss of expression in the gastric caeca is clearly more dramatic than loss in PS7. This could be due to additional HOMREs in the 419 bp enhancer that can act in the absence of an EXD site in PS7 but not in the caeca. Thus, targeting an EXD site involved in both pathways of regulation might have a greater effect on gastric caeca expression. Two alternative hypotheses are also possible. First, the site where we have mutated could be a control element that may or may not bind EXD in vivo but is exclusively required for gastric caeca expression. Second, the site could be a control element that is not bound by EXD in vivo but is nonetheless necessary for both gastric caeca and PS7 expression.

Our EXD binding results differ from those of Chan et al. (1994). We have not observed EXD binding to the region which they call *dpp*80, even though we are able to observe a clear footprint at the ATCAATTA motif. The *dpp*80 region of the enhancer is relatively poorly conserved between fly species, save for the motif GTTTGTTTTAT (Fig. 9). Moreover, the cooperative binding observed by Chan et al. (1994) involved a several hundred-fold molar excess of EXD over UBX. The results of van Dijk and Murre (1994) show that a region of EXD amino terminal to the homeodomain is absolutely required for cooperative binding with UBX. This domain is absent from the EXD polypeptide used by Chan et al. (1994). Thus, a number of differences in experimental design may explain the different findings reported in the three papers.

We especially want to thank James Masucci, F. Michael Hoffmann, Cordelia Rauskolb, and Eric Wieschaus for communicating results prior to their publication. George Chen and Chen-sen Wu provided valuable help with library screens, injections and other experiments. We thank Drs Brad Johnson, Mark Krasnow, Bruce Appel, and Shige Sakonju for gifts of protein and advice on purification, Dr Volker Hartenstein for helping identify muscle cells, Drs Ernesto Sanchez-Herrero and Gines Morata for the heat shock-inducible *abd-A* flies, the Cell Sciences Imaging Facility for assistance with confocal imaging, and the Indiana University *Drosophila* Stock Center for providing many fly stocks. We thank Bruce Appel, Michelle Lamica, Shige Sakonju, James Masucci, F. Michael Hoffmann, Cordelia Rauskolb, John Tamkun, and Eric Wieschaus for various plasmids, libraries, and antibodies. Finally, we thank the anonymous reviewers for careful reviews and thoughtful suggestions for improvements. This work was supported by N.I.H. grant no. HD18163. M. P. S. is an investigator of the Howard Hughes Medical Institute.

**REFERENCES**


van Dijk, M. A., P. M. Voorhoeve, and C. Murre (1993). Pbxi is converted
into a transcriptional activator upon acquiring the N-terminal region of E2A in pre-B-cell acute lymphoblastoid leukemia. *Proc Natl Acad Sci USA* 90, 6061-6065.


(Accepted 23 September 1994)