The EMBO Journal vol.14 no.3 pp.520–535, 1995

Ultrabithorax protein is necessary but not sufficient for full activation of *decapentaplegic* expression in the visceral mesoderm

Benjamin Sun, Deborah A.Hursh[1], Donald Jackson and Philip A.Beachy[2]

Howard Hughes Medical Institute, Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205 and [1]Laboratory of Biochemistry, NCI/NIH Building 37, Room 4A01, Bethesda, MD 20892, USA

[2]Corresponding author

B.Sun and D.A.Hursh contributed equally to this work

Communicated by J.Tooze

To elucidate the mechanisms by which homeotic selector (HOM) genes specify the unique features of *Drosophila* segments, we have analyzed the regulation of *decapentaplegic* (*dpp*), a transforming growth factor (TGF)-β superfamily member, and have found that the *Ultrabithorax* (*Ubx*) HOM protein directly activates *dpp* expression in parasegment 7 (PS7) of the embryonic visceral mesoderm. Other factors are also required, including one that appears to act through homeodomain protein binding sites and may be encoded by *extradenticle* (*exd*). The *exd* protein binds in a highly co-operative manner to regulatory sequences mediating PS7-specific *dpp* expression, consistent with a genetic requirement for *exd* function in normal visceral mesoderm expression of *dpp*. A second mechanism contributing to PS7 expression of *dpp* appears not to require *Ubx* protein directly, and involves a general visceral mesoderm enhancer coupled to a spatially specific repression element. Thus, even in an apparently simple case where visceral mesoderm expression of the *dpp* target gene mirrors that of the *Ubx* HOM protein, full activation by *Ubx* protein requires at least one additional factor. In addition, a distinct regulatory mode not directly involving *Ubx* protein also appears to contribute to PS7-specific expression.

Key words: *decapentaplegic*/development/*Drosophila*/*extradenticle*/*Ultrabithorax*

Introduction

Drosophila segments owe their distinctive morphologies to the action of the homeotic selector (HOM) genes (for a review see McGinnis and Krumlauf, 1992). HOM proteins each contain a distinctive form of an autonomous, sequence-specific DNA binding domain, the homeodomain, whose broad sequence conservation has spurred the characterization of HOM clusters in many other metazoa. Although the unique segmental identities specified by HOM proteins are thought to result from the regulation of distinct sets of downstream target genes, the molecular mechanisms underlying target specificity remain unclear. Whereas most genetic studies of chimeric protein function implicate the homeodomain itself as the major repository of segment identity information within *Drosophila* HOM proteins (Kuziora and McGinnis, 1989, 1991; Gibson *et al.*, 1990; Furukubo-Tokunaga *et al.*, 1993; Zeng *et al.*, 1993), at least one study suggests a role for sequences outside the homeodomain (Chan and Mann, 1993).

At the mechanistic level, the DNA sequence preferences of some HOM proteins are distinctive while others display overlapping or identical specificities (Ekker *et al.*, 1994). Even small differences could yield significant discrimination when summed by co-operative binding to multiple sites, and co-operative binding has been reported for proteins encoded by *Ultrabithorax* (*Ubx*; Beachy *et al.*, 1993) and several murine HOM genes (Galang and Hauser, 1992). The extent to which differential sequence recognition by HOM proteins can account for target specificity, however, remains to be demonstrated.

Another potential mechanism underlying target specificity would involve specific interactions of HOM proteins with other protein factors, and several lines of argument support this idea. First, the biochemically indistinguishable sequence specificities of certain HOM proteins, such as those encoded by *Ubx* and *Antennapedia* (*Antp*), support the existence of other mechanisms, such as differential interactions with other protein factors, as possibilities for enhancing target specificity. Second, chimeric constructs with the *Ubx* and *Antp* HOM genes suggest a role for sequences C-terminal to the *Ubx* homeodomain (Chan and Mann, 1993). These sequences do not contribute to the specificity of DNA sequence recognition (Ekker *et al.*, 1992), but they may contribute to protein–protein interactions and are predicted to form a coiled-coil (Lupas *et al.*, 1991; Beachy *et al.*, 1993). Third, reporter constructs containing multiple copies of sequences recognized by HOM proteins (and by other homeodomain-containing proteins) fail to be expressed in patterns like those of the HOM genes, implying that binding by HOM proteins alone is insufficient to activate expression (Vincent *et al.*, 1990; Nelson and Laughon, 1993; Zeng *et al.*, 1994). Finally, specification of three distinct cell types by yeast mating type homeodomain proteins involves crucial interactions with each other or with specificity-enhancing partners, thus providing a well-studied paradigm for the interactions of homeodomain proteins with other protein factors (Goutte and Johnson, 1993; Vershon and Johnson, 1993; Ho *et al.*, 1994; Stark and Johnson, 1994).

As a step towards understanding the interaction of HOM proteins with their DNA target sites and with other *trans*-acting factors, we have initiated a study of the midgut expression of *decapentaplegic* (*dpp*), a transforming growth factor (TGF)-β family member, and its regulation by the HOM protein encoded by *Ubx*. During

© Oxford University Press

midgut formation, *Ubx* and *dpp* are expressed in the mesoderm surrounding the endodermally derived midgut, at a location which underlies parasegment 7 (PS7) in the ectoderm and corresponds to the future site of the secondary midgut constriction (St Johnston and Gelbart, 1987; Bienz and Tremml, 1988; reviewed by Bienz, 1994). Both *Ubx* null mutations and most of the *shortvein* class of *cis*-regulatory *dpp* mutations prevent the formation of the secondary constriction (Immerglück *et al.*, 1990; Panganiban *et al.*, 1990; Hursh *et al.*, 1993; Masucci and Hoffmann, 1993); these mutations are associated with reduced *dpp* protein expression in PS7, thus indicating that *dpp* protein is critical for the formation of the secondary constriction and suggesting that *Ubx* is required for normal activation of *dpp*. A role for *Ubx* in the activation of *dpp* is further reinforced by the observation that the ectopic expression of *Ubx* protein under hsp70 promoter control causes coincident ectopic expression of *dpp* in anterior portions of the visceral mesoderm (Reuter *et al.*, 1990). This apparently simple regulatory circuit is complicated by the fact that *Ubx* expression also requires *dpp* and *wingless* (*wg*) in an indirect autoregulatory mechanism (Hursh *et al.*, 1993; Thüringer and Bienz, 1993; Thüringer *et al.*, 1993), and the posterior boundaries of both *Ubx* and *dpp* appear to be independently controlled by the homeotic gene *abdominal-A* (Bienz and Tremml, 1988; Tremml and Bienz, 1989; Reuter *et al.*, 1990; Hursh *et al.*, 1993).

We chose to study *Ubx* regulation of *dpp* for the simplicity of *dpp* and *Ubx* expression patterns in the midgut, for the prior genetic definition of regulatory regions physiologically important for *dpp* expression, and for the extensive biochemical characterization of *Ubx* protein interactions with DNA. We demonstrate here for the first time that *Ubx* protein directly regulates *dpp* expression, as was claimed but not proven by Capovilla *et al.* (1994; see below). We also find, however, that at least one other activity is required in conjunction with *Ubx* protein for PS7-specific activation. Our evidence suggests that this additional activity is a homeodomain protein, possibly encoded by *extradenticle* (*exd*; Rauskolb *et al.*, 1993). Our analysis also indicates the existence of a distinct regulatory mode for visceral mesoderm expression of *dpp* that involves general activation within the visceral mesoderm coupled to a spatially specific repressing activity. Thus, while the simple pattern of *dpp* midgut expression reflects the expression of *Ubx* in the midgut and indeed requires direct activation by *Ubx* protein, other tissue- and positionally-specific activities also contribute.

Results

Regulatory sequences required for PS7 expression of dpp in the visceral mesoderm

Expression of *dpp* in the visceral mesoderm is first visible at the end of germ band retraction as a lateral pair of blocks of expression underlying the PS7 region of the ectoderm (Figure 1B). With dorsal closure, the visceral mesoderm thins as it extends around the endoderm, and *dpp* products continue to accumulate and appear as a circumferential belt surrounding the midgut. The position of this band defines the incipient middle lobe of the tripartite gut, with the secondary constriction forming at its posterior boundary. In late embryogenesis the gut takes on a convoluted appearance, with *dpp* expression persisting in the middle section of a continuous tube. Expression of *dpp* is also seen at the site of the evaginating gastric caecae in PS3, although expression in this domain is not *Ubx*-dependent (Reuter *et al.*, 1990).

Previous studies suggested that the portion of the *dpp cis*-regulatory region which controls *Ubx*-responsive expression in PS7 of the visceral mesoderm is located within an 812 bp region to the 5′ side of the coding sequence (Hursh *et al.*, 1993). This region is defined by the overlap between two *dpp* reporter gene constructs, RD1 and RD2 (Figure 1A), both of which direct visceral mesoderm expression in a pattern indistinguishable from that of *dpp*. The significance of this region is supported by the loss of PS7 *dpp* expression in individuals homozygous for the mutation dpp^{s6} (Hursh *et al.*, 1993), which deletes 4845 bp including the 812 bp region of overlap between RD1 and RD2 (Figure 1A; see also Materials and methods). To confirm that this region is sufficient to activate transcription in PS7 of the visceral mesoderm, we introduced into the *Drosophila* germ line a *lacZ* reporter construct containing this overlap region [the *Bam*HI–*Eco*RI (BE) region; Figure 1A]. This construct reproduces the wild-type pattern of *dpp* transcription in PS3 and PS7 of the visceral mesoderm (Figure 1C), although PS3 expression is lower than in wild-type.

As a basis for further analysis of regulation, we determined the nucleotide sequence of the 812 bp BE segment. As shown in Figure 2, this sequence contains a total of 12 TAAT tetranucleotides representing core sequences for potential homeodomain binding sites. Based upon previous systematic analyses of *Ubx* protein binding preferences (Ekker *et al.*, 1991, 1992) and actual binding studies by DNase I footprinting (data not shown), sites 2, 3, 5, 7, 9 and 10 have reasonably high relative affinities while the others are somewhat lower in affinity. In addition, this fragment contains four conserved nonamers that do not contain TAAT cores but otherwise fit the *Ubx* base preferences well (see Figure 2); these sites are bound at intermediate to low affinities (data not shown).

Multiple regulatory inputs for localized visceral mesoderm expression

To further define the requirements for wild-type expression, five reporter constructs derived from the BE region were analyzed. Figure 3B–D demonstrates that deletion of a sequence from either the 3′ or both the 5′ and 3′ ends of the BE element [*Bam*HI–*Xho*I (BX), PX and SX, respectively (see Figure 3 for descriptions of constructs)] incrementally reduces but does not eliminate PS7 expression. We conclude that SX contains minimal sequence information sufficient for localized expression. Since the spatial extent and intensity of SX expression is reduced relative to other larger constructs (compare, for example, PX and SX in Figure 3C and D), we conclude that sequences within BP, XE and PS contribute to normal levels of expression. PX was the smallest construct (413 bp) that yielded readily detectable expression; Figure 3E shows that expression of the PX construct, like that of endogenous *dpp* (Reuter *et al.*, 1990) or *dpp* reporter constructs (Hursh *et al.*, 1993), is attenuated but spatially

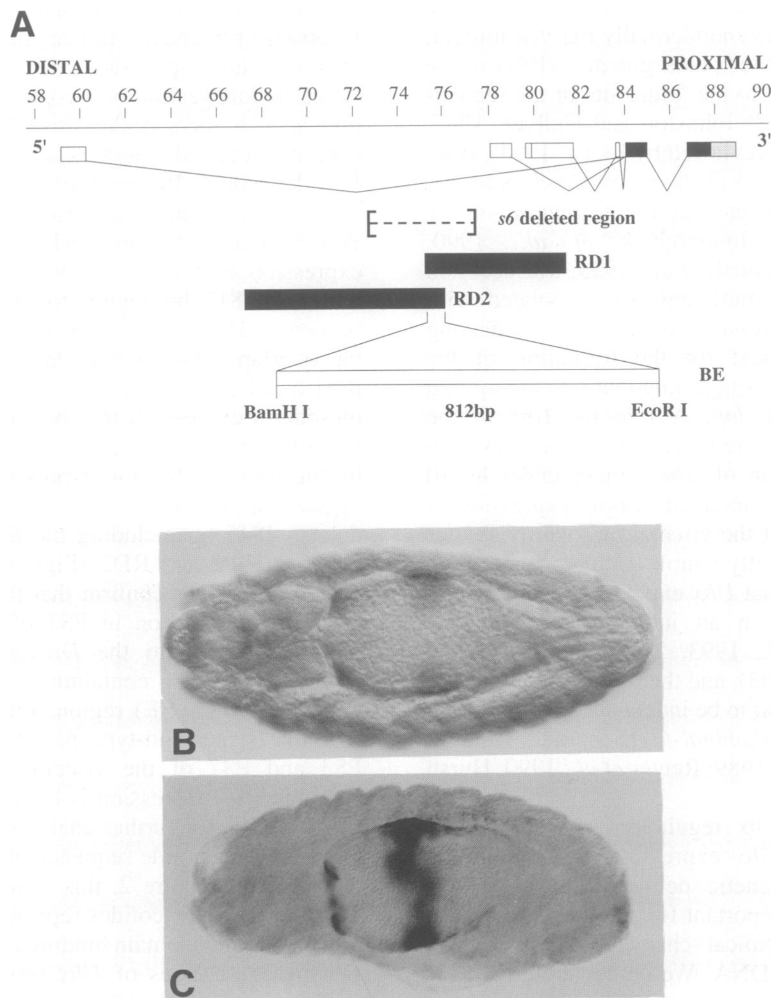


Fig. 1. An 812 bp regulatory region of the *dpp* gene directs accurate expression in PS7 of the visceral mesoderm. (**A**) Structure of the *dpp* gene. Molecular coordinates (St Johnston *et al.*, 1990) are indicated along the top (in kb). Exons are represented by boxes immediately below the line, with filled boxes representing protein coding regions, open boxes representing five alternative 5′ exons and the stippled box representing the 3′ untranslated region. The region deleted in dpp^{s6} is indicated by the dashed line in brackets. Reporter constructs RD1 and RD2 are shown as black bars, with the overlap, an 812 bp *Bam*HI−*Eco*RI fragment (BE), represented as an expanded white bar at the bottom. (**B**) Expression of *dpp* in a wild-type embryo as visualized by whole-mount *in situ* hybridization: horizontal view, stage 14. (**C**) Antibody localization of β-galactosidase expression directed by the BE reporter gene: lateral view, stage 15. In these and all subsequent embryos anterior is to the left.

accurate in a *Ubx* mutant background. Note that the embryo in Figure 3E is overstained to show this residual expression (see arrows); a comparably stained wild-type embryo at the same stage would display high level expression of the PX reporter within a fully formed second lobe.

The apparent activation by Ubx^+ in PS7 is not the only regulatory input by which *dpp* expression is localized. Additional inputs are suggested by the loss of spatial restriction in the BA and *Bam*HI−*Stu*I (BS) constructs (Figure 3F and G); these reporters still contain multiple *Ubx* binding sites and expression in PS7 remains strong, but in addition lower level expression extends beyond this domain. The sequences missing in the BA and BS constructs (the AX fragment) are thereby implicated in spatially specific repression. In the BA and BS constructs, the loss of repression outside PS7 reveals the existence of a general visceral mesoderm enhancer element within the BS segment. This enhancer element appears to reside within the *Bam*HI−*Msc*I (BM) fragment, as demonstrated by the general visceral mesoderm expression of the BM reporter construct, albeit with stronger expression in the posterior (Figure 3H).

To investigate further the properties of this generalized visceral mesoderm enhancer, we appended the BM segment to the upstream regulatory portions of the PX and SX reporter constructs. Figure 4D and E shows that the BM fragment enhances expression levels relative to the PX and SX constructs (Figure 4A and B). This enhancement preserves the larger spatial domain of PX expression in relation to SX (see above), suggesting that the PX and SX segments contain sequences capable of responding positively to BM within their spatial domains while also suppressing the general expression outside of their spatial domains. The general visceral mesoderm-enhancing properties of the BM fragment are thus entrained in a spatially specific manner by nearby stimulatory and repressive regulatory elements.

In attempting to identify elements responsible for augmenting PX expression relative to SX, we noticed

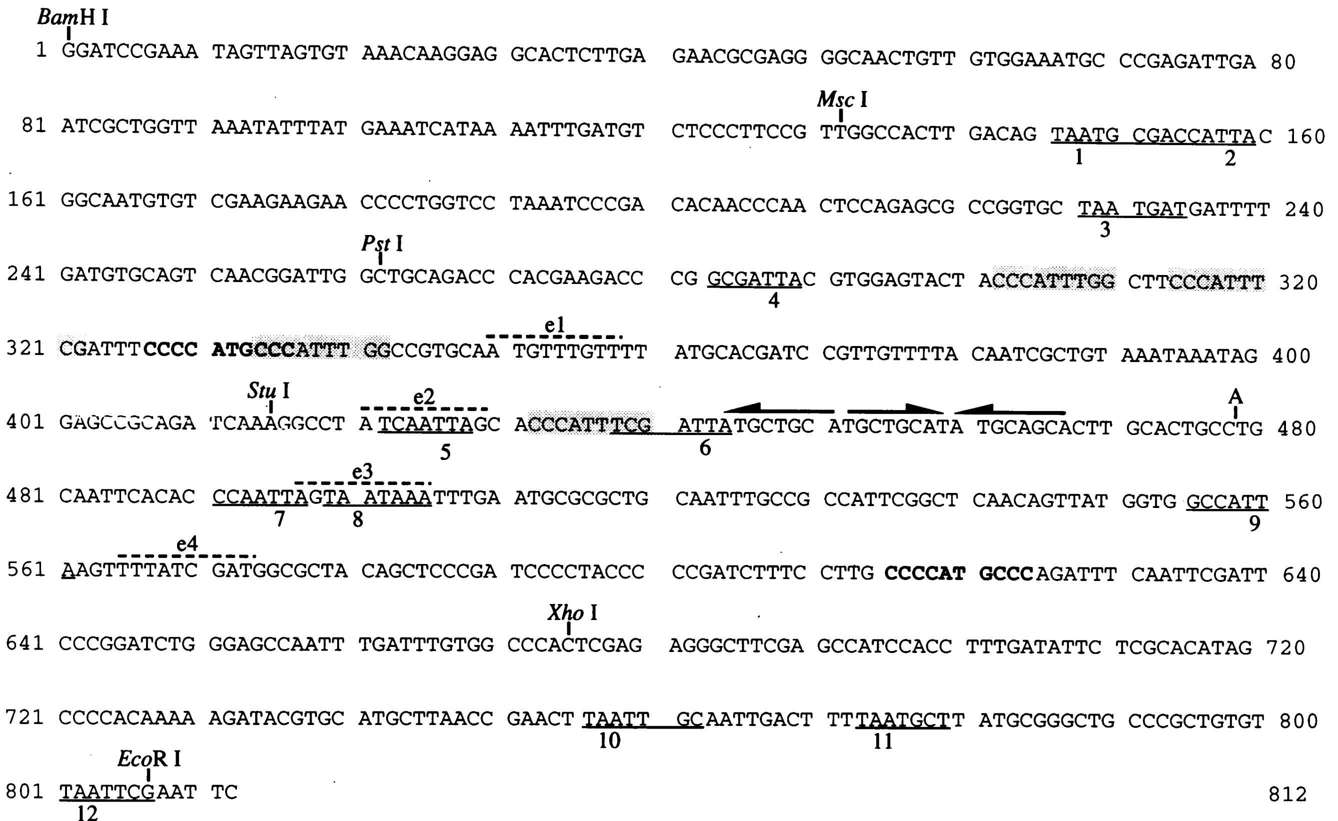


Fig. 2. Sequence analysis of the BE regulatory region. Restriction sites are indicated at the beginning of their recognition sequences. The 'A' between positions 478 and 479 indicates the beginning of a *BglII* site introduced by PCR. TAAT-containing sites are underlined and numbered, and four regions protected by ExdHD in footprinting assays are overlined and designated e1–e4. Four repeats of the nonamer motif CCCATTT(G/C)G motif are shaded. Three inverted repeats of an 8 bp motif are indicated by arrows and two sequences matching the binding site for the transcription factor AP2 (Williams *et al.*, 1988) are shown in bold type (at positions 327 and 615). Disruption of neither the inverted repeats (PX_{PAL}) nor the AP2 sites (PX_{AP2}) alters the pattern or intensity of expression from a PX reporter construct (data not shown).

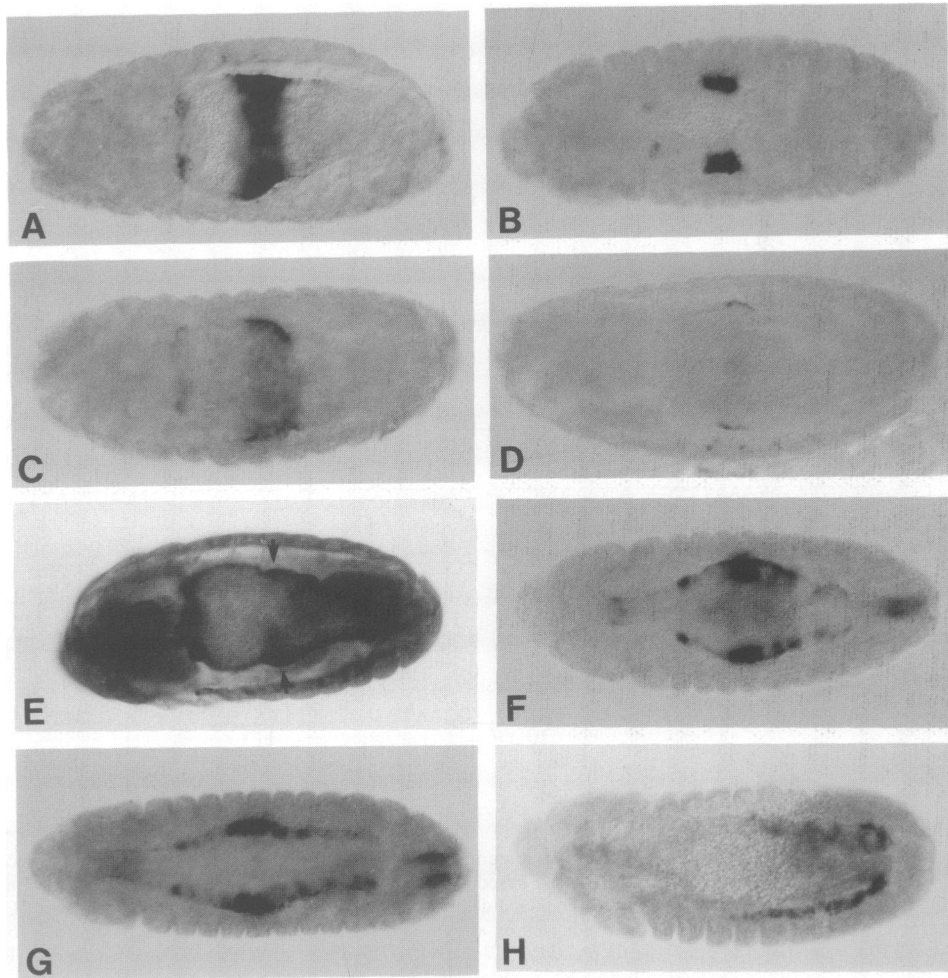
three repeats of the nonamer CCCATTTGG occurring within the PS fragment and a fourth nonamer repeat just downstream of the *StuI* site (Figure 2). Disruption of these four repeats (see Materials and methods) results in strongly reduced early expression (data not shown), with some residual expression appearing in later embryogenesis (Figure 4F). Strong reduction of early expression is also observed for the SX construct, suggesting that the major contribution of the PS fragment to overall expression is accounted for by the presence of these nonamer sequences. The nonamer repeat does not contain a TAAT core sequence but otherwise closely resembles the optimal *Ubx* protein binding site (Ekker *et al.*, 1991, 1992), and these sequences are weakly bound by *Ubx* protein (data not shown; Capovilla *et al.*, 1994). The major contribution of the PS segment could thus be to augment activation by *Ubx* protein; alternatively, other factors may act through the nonamer sites. The PS segment does not on its own produce significant visceral mesoderm expression in a reporter construct (data not shown), and it is not absolutely required for PS7 expression, although its absence in the SX construct produces weaker expression relative to PX; SX expression is further weakened by the mutation of *Ubx* site 9 (Figure 4C), a perfect match to the optimal site for *Ubx* homeodomain binding.

Localized visceral mesoderm expression of *dpp* in PS7 thus appears to depend on several types of regulatory input. The first, as demonstrated by the reduction of *dpp* expression in *Ubx* mutants (Figure 3E), is a stimulatory input from Ubx^{+} that is specific to PS7. A second input, also stimulatory, is mediated by an enhancer element which, although not segmentally specific, is restricted in its activity to the visceral mesoderm. This general stimulatory input is subject to a segmentally localized repression outside of PS7.

Ubx directly activates *dpp* expression

To demonstrate that *Ubx* protein acts directly on *dpp* regulatory sequences, we inactivated *dpp* reporter expression by altering potential *Ubx* protein target sites and then attempted to restore expression using a correspondingly modified *Ubx* protein capable of recognizing the altered targets (Schier and Gehring, 1992, 1993). Previous studies demonstrated that homeodomain residue 50 is critical for the recognition of bases 3′ to the TAAT core (Treisman *et al.*, 1989; Percival-Smith *et al.*, 1990; Hanes and Brent, 1991). To characterize the specificity of a *Ubx* protein with altered specificity, we purified a *Ubx* homeodomain peptide in which lysine was substituted for glutamine at position 50 (the K50 homeodomain). Figure 5B shows that in binding selection experiments this protein binds preferentially to sites carrying the bases CCCC 3′ to the TAAT core; the wild-type protein, in contrast, prefers the bases (G/T)(G/A)CC at these positions (Figure 5A; Ekker *et al.*, 1991, 1992).

The specificity of the K50 *Ubx* protein was confirmed



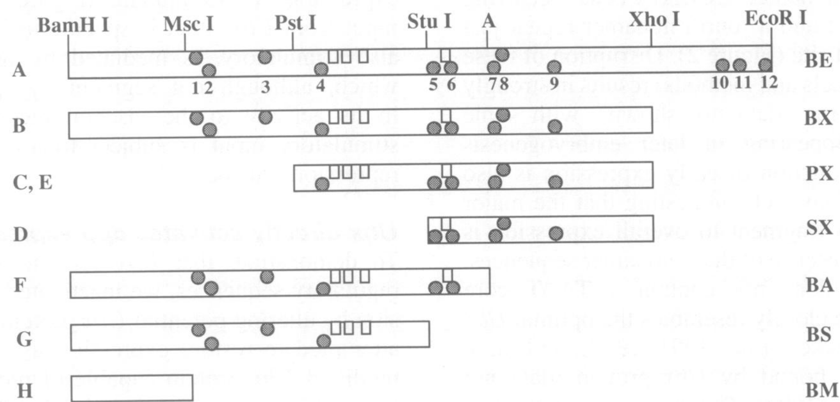
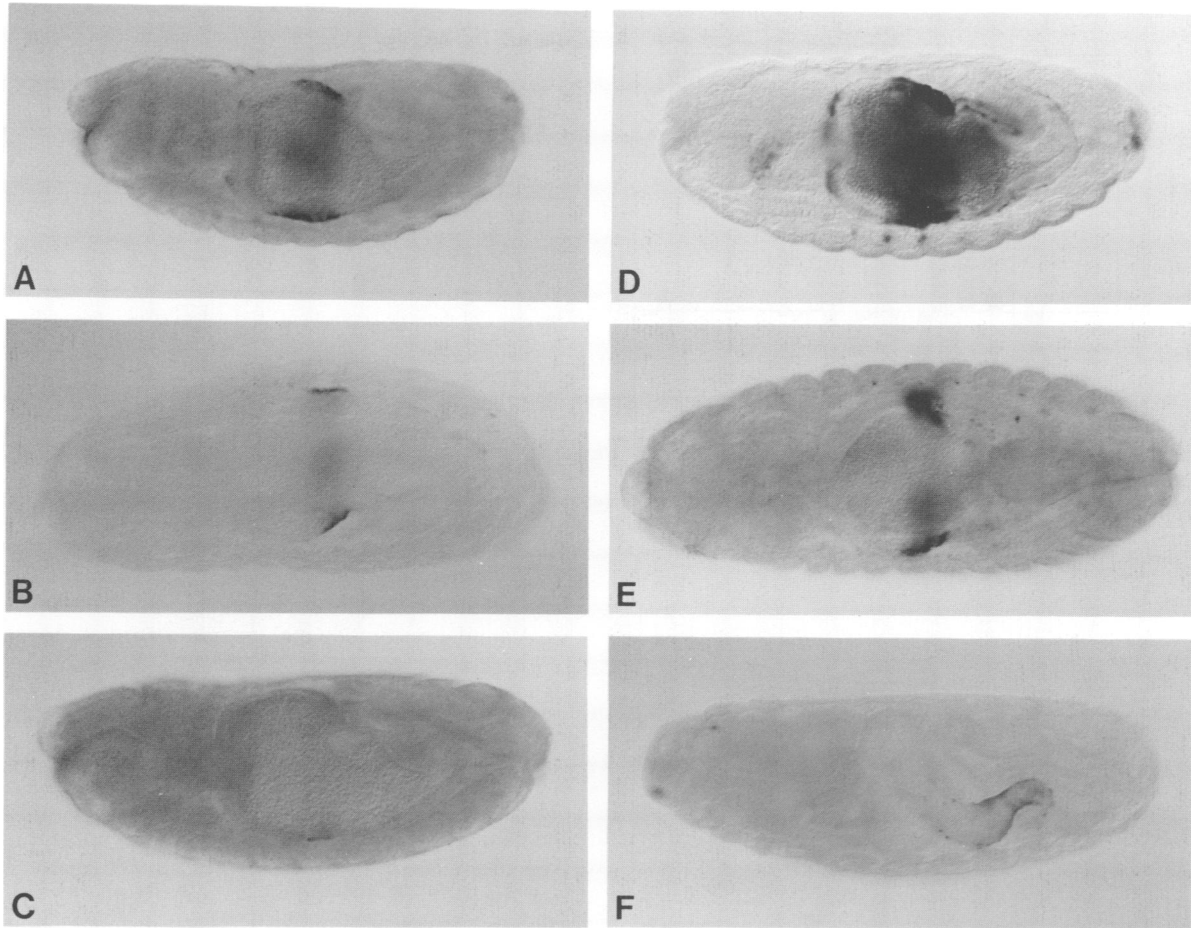


Fig. 3. β-Galactosidase expression resulting from reporter constructs carrying truncations of the BE fragment. Regulatory sequences present in reporter constructs derived from BE (see Figures 1 and 2) are depicted schematically. (**A**) BE, stage 14. (**B**) BX, stage 13. The visceral mesoderm has not yet extended to surround the mesoderm as seen in (A). (**C**) PX, stage 15. (**D**) SX, stage 14. (**E**) PX in a $Ubx^{9.22}$ homozygote, stage 15. The embryo is overstained to demonstrate residual staining in PS7 (arrows; see text). (**F**) BA, stage 13. (**G**) BS, stage 13. (**H**) BM, stage 12. In this and all subsequent figures, shaded circles denote TAAT-containing sites and rectangles represent nonamer sites.

using a DNA fragment in which *Ubx* binding sites 4–9 (Figure 2) were substituted by the sequence 5′-TAATCCC-3′. As shown in Table I, the levels of β-galactosidase expressed from a yeast reporter construct carrying this substituted fragment were 12-fold higher with the K50 *Ubx* protein than with wild-type *Ubx*; the unaltered fragment, in contrast, responded to the wild-type *Ubx* protein with 7.5-fold higher levels of β-galactosidase relative to the K50 construct. From the magnitudes of the β-galactosidase activity levels, the K50 protein appears to activate efficiently via the substituted sites and the two proteins display an ~90-fold overall difference in relative preference for the two reporters. The differential specificities of these proteins were further confirmed by footprinting



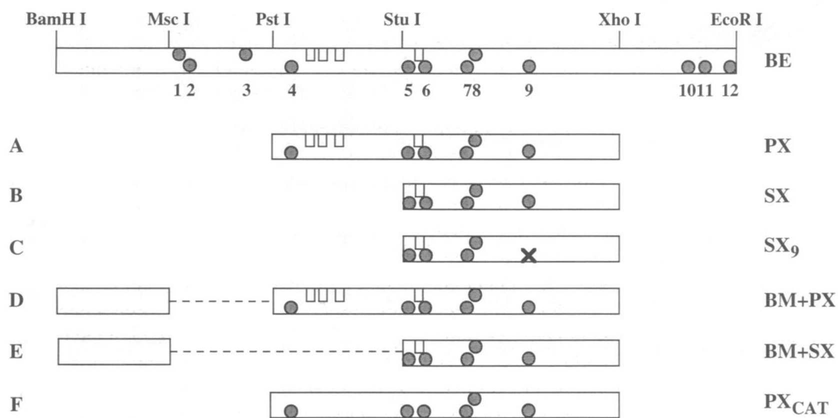


Fig. 4. Quantitative effects upon reporter construct expression. β-Galactosidase expression is detected by antibody staining in embryos expressing the reporter gene constructs depicted schematically. Expression levels of PX (**A**) and SX (**B**) reporter constructs are augmented by the presence of the BM fragment in BM + PX (**D**) and BM + SX (**E**). The mutation of site 9 reduces further the expression level normally seen for SX (SX_9 in **C**). Mutation of the four nonamer sites (**F**) reduces expression (as compared with PX) to a residual level observable only late (stage 17) in the midgut. Stage 15 embryos in (A) and (D) were stained in parallel, as were stage 14 embryos in (B) and (E).

experiments with wild-type and K50 mutant *Ubx* homeodomains (data not shown).

The PX reporter construct suffices to generate attenuated but spatially accurate expression in PS7 (Figure 6C). This construct also responds to ectopically expressed UbxIa protein in the same manner as the endogenous *dpp* gene (Reuter *et al.*, 1990), i.e. by activation in regions anterior but not posterior to PS7 (Figure 6D). The PX fragment therefore was chosen as the context for the identification of sites that mediate *Ubx* protein activation of *dpp*. This fragment includes *Ubx* sites 4–9, all or subsets of which were substituted by the sequence 5′-TAATCCC-3′ to test for *Ubx* expression. PS7 expression was strongly reduced or lost in PX reporter constructs carrying substitutions at sites 4–6 and 9 (Figure 6A), sites 5, 7 and 9 (data not shown), or sites 4–9 (data not shown). This loss of

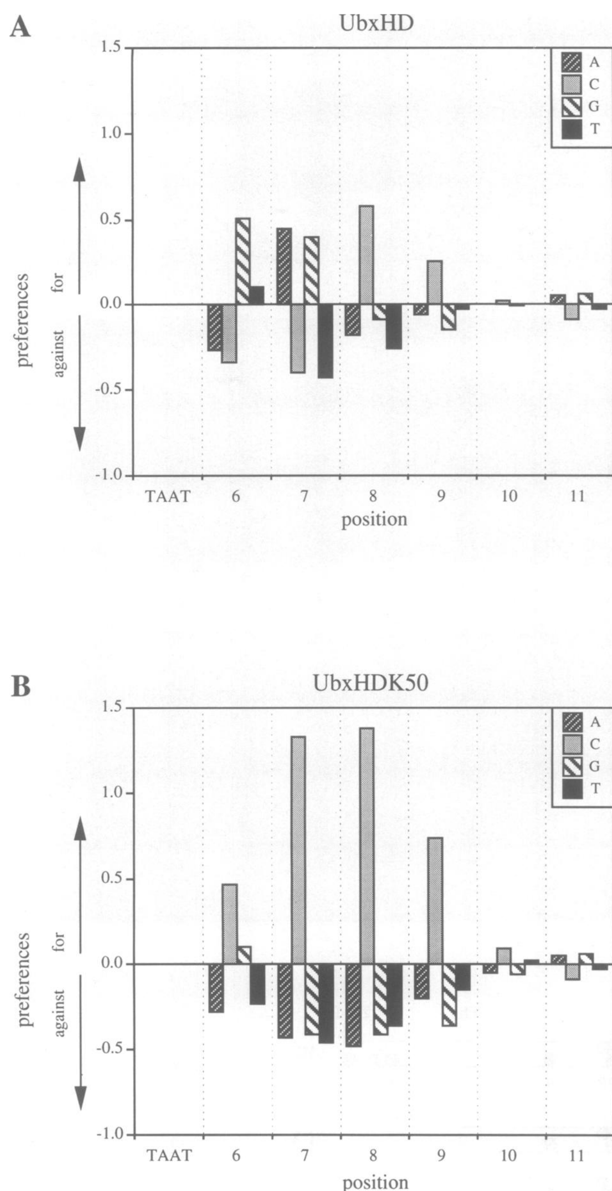


Fig. 5. DNA sequence preferences for a *Ubx* homeodomain with a glutamine (**A**) or a lysine (**B**) residue at position 50. Relative preferences for A, C, G and T are shown at each position by bars extending above or below zero (indicating preferences for or against a base, respectively). Positions are shown in relation to the fixed 5′-TAAT-3′ core. No significant differences were observed in base preferences at positions 5′ to the core (results not shown).

expression suggests that each of the three constructs is missing sites essential for reporter expression in PS7.

To determine whether insufficient binding of *Ubx* protein was responsible for the loss of expression, each of the three reporters was combined genetically with a construct for heat-inducible expression of a *Ubx* protein of modified specificity (hsUbxIaK50; this protein will also be referred to as K50). Indeed, for the $PX_{4-6,9}$ and $PX_{5,7,9}$ reporters, expression in PS7 can be restored by the provision of a *Ubx* protein capable of recognizing the substituted sites (Figure 6B, E and F; see below for consideration of the PX_{4-9} reporter); the wild-type hsUbxIa construct, in contrast, was unable to restore expression of these two reporters (data not shown). The previously demonstrated genetic requirement of *Ubx* function for PX

Table I. Distinct specificities of wild-type and mutated *Ubx* proteins in yeast

Activator	β-Galactosidase activity[a]		
	Wild-type reporter	Substituted reporter	Ratio
UbxIa	1.8 ± 0.2	0.24 ± 0.01	7:1
UbxIaK50	1.3 ± 0.1	16.0 ± 3.0	1:12
pRS314GU (vector)	0.04 ± 0.04	0.02 ± 0.01	–

[a]Average of three determinations ± standard error.

reporter activation (Figure 3E) did not distinguish between a direct versus an indirect mode of action for *Ubx* protein on *dpp* regulatory sequences. Restoration of mutant reporter expression by compensatory changes in *Ubx* protein specificity, however, eliminates the possibility that the *Ubx* effect is due to regulation by *Ubx* of an intermediary factor, and strongly suggests that *dpp* activation by *Ubx* protein in PS7 is due to a direct interaction with *dpp* regulatory sequences.

A similar restoration making use of an altered specificity *Ubx* protein was reported recently by another group (Capovilla *et al.*, 1994). We note, however, that expression of the altered specificity reporter construct (similar to our $PX_{5,7,9}$) occurred not in the visceral mesoderm but in the endoderm (see Figure 7B in Capovilla *et al.*, 1994; M.Bienz, personal communication), a tissue in which neither wild-type *Ubx*, wild-type *dpp*, nor the altered specificity *Ubx* protein was expressed. We cannot account for the lack of visceral mesoderm expression by these authors' reporter construct. The restored expression reported here, however, was unambiguously mesodermal (see Figure 6) and was reproducible in multiple experiments with multiple lines for each reporter and for the K50 activator (see Materials and methods).

Ubx activation of dpp requires at least one additional factor

Our results extend beyond the conclusion that *Ubx* protein is directly involved since in our studies the K50 protein was supplied ubiquitously through the use of the hsp70 promoter. Despite this ubiquitous expression of K50, the restoration of expression for the $PX_{4-6,9}$ and $PX_{5,7,9}$ reporters was restricted to PS7, suggesting a requirement for another factor whose activity is localized to PS7. In contrast, the restoration reported by Capovilla *et al.* (1994) utilized a PS7-specific promoter to supply K50 protein, and the requirement for another factor outside of PS7 thus was not detected. Further support for the suggestion that another factor is required comes from the inability of PX_{4-9} construct expression to be restored anywhere in the visceral mesoderm in response to the K50 protein (Figure 6G). The PX_{4-9} reporter carries a more extensive set of substitutions than the reporter used by Capovilla *et al.* (1994), and more than $PX_{4-6,9}$ and $PX_{5,7,9}$; the inability to restore expression of PX_{4-9} thus suggests that a binding site(s) for a protein(s) other than *Ubx* has been affected (see Discussion for a fuller consideration of these arguments).

Specific and co-operative binding of the exd protein to dpp regulatory sequences

Because our results suggested a requirement for another factor and because the sites altered in the PX_{4-9} construct

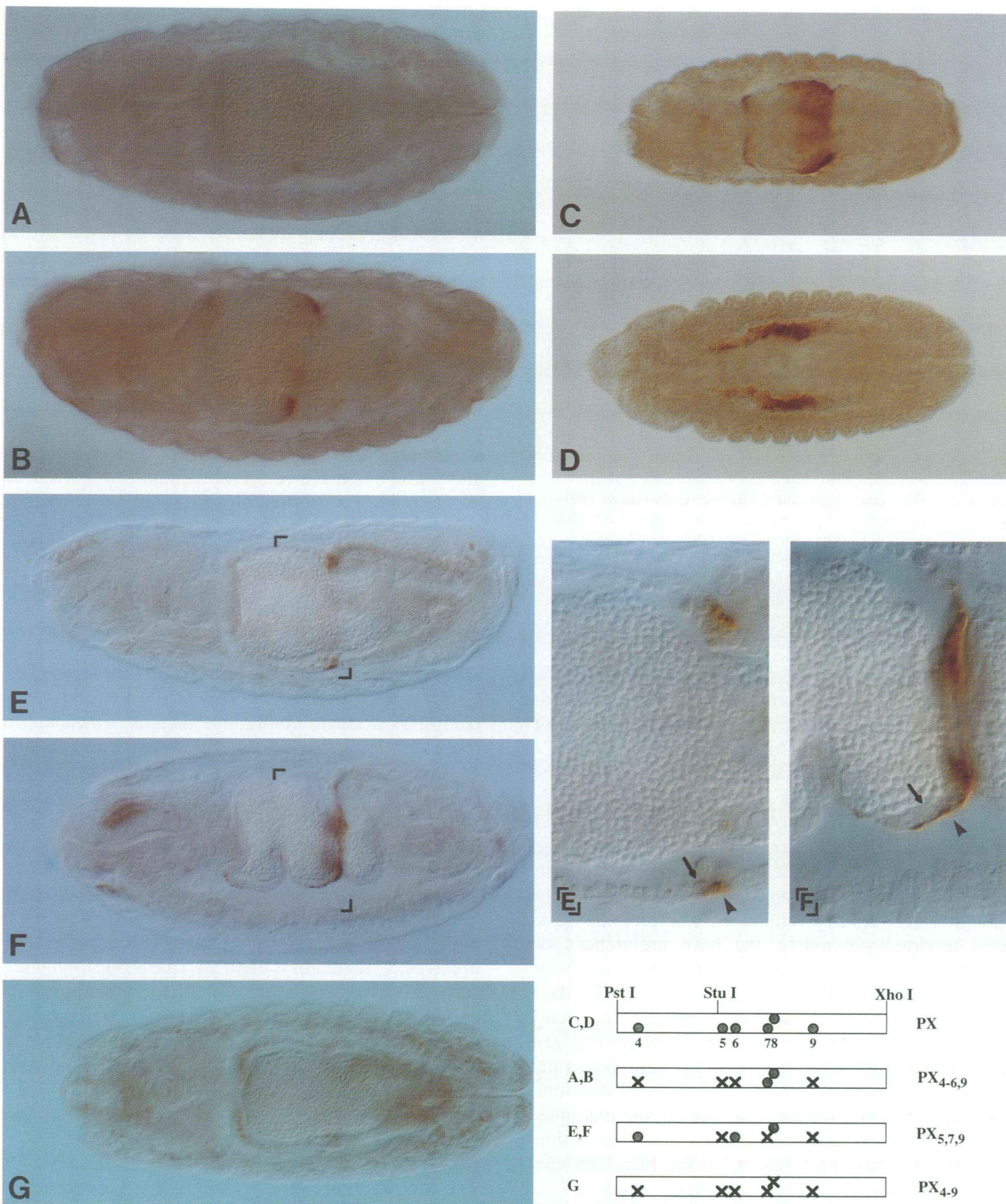


Fig. 6. Effects of binding site substitution and altered specificity *Ubx* proteins upon reporter construct expression. β-Galactosidase expression patterns from wild-type (**C** and **D**) and substituted sequence reporter constructs (**A**, **B** and **E−G**) are shown in a wild-type background (A and C) or in the presence of UbxIa (D) or UbxIaK50 (B and E−G) proteins ubiquitously expressed under the control of the hsp70 promoter (see text). Reporter constructs are represented schematically, with the symbol X denoting substitution of TAATCCC in place of the wild-type sequence TAATNNN. The normal pattern of PX expression (stage 15 embryo in C) expands anteriorly in the presence of ectopic UbxIa (stage 14 embryo in D). Expression in PS7 is lost for substituted site reporters $PX_{4-6,9}$ (stage 14 embryo in A). Expression is restored by UbxIaK50 for $PX_{4-6,9}$ (stage 15 embryos in B) and $PX_{5,7,9}$ (stage 15 embryo in E and stage 16 in F), but only in PS7. The bracketed areas in (E) and (F) are enlarged and shown on the right of (E) and (F). Arrowheads point to the visceral mesodermal cells that are stained, and arrows indicate the endodermal cells that are not stained. The expression of PX_{4-9} cannot be restored by UbxIaK50 (stage 14 embryo in G). β-Galactosidase was detected by antibody staining.

are all homeodomain binding sites, we tested the homeodomain protein encoded by the *exd* gene for binding to *dpp* regulatory sequences. *exd* mutations interact genetically with *Ubx* and other HOM gene mutations (Rauskolb *et al.*, 1993) and exd^+ is specifically required for the activation of *dpp* in PS7 of the visceral mesoderm (Rauskolb and Wieschaus, 1994). Using an *Escherichia coli* expression system as a source of protein (Studier

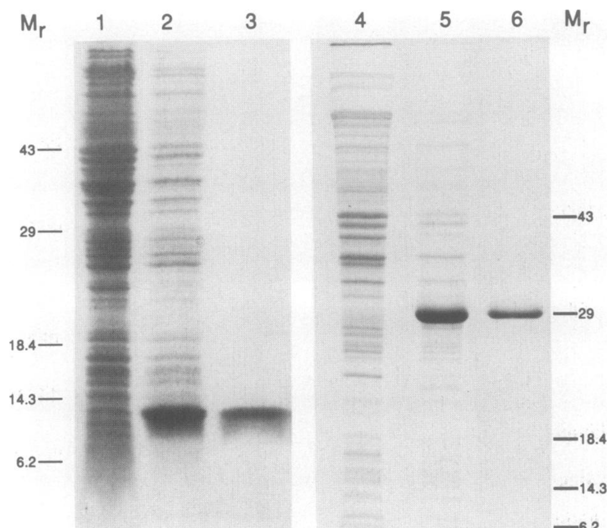


Fig. 7. Purification of bacterially expressed *exd* proteins. ExdHD (amino acids 237–313; lanes 1–3) and ExdHDcc (amino acids 149–377; lanes 4–6) were expressed and purified as described in Materials and methods. Shown are: Coomassie-stained SDS–polyacrylamide gels loaded with total protein from cells carrying the expression vector without induction (lanes 1 and 4); proteins from IPTG-induced cells (lanes 2 and 5); and peak fractions from phosphocellulose chromatography (lanes 3 and 6). Relative molecular masses (M_r) and mobilities of markers are indicated.

et al., 1990), two truncated forms of the *exd* protein were purified (Figure 7). One of these contained the *exd* homeodomain plus an initiator methionine and 13 C-terminal flanking residues (ExdHD), and the other contained the homeodomain plus 89 N-terminal and 75 C-terminal residues (ExdHDcc). As shown in Figure 8A and C, the binding of ExdHD protects four regions of the PX fragment from DNase I cleavage (e1–e4; see also Figure 2).

Although DNA binding studies of *exd* proteins have not been reported previously, the base preferences of Pbx-1, a tightly conserved human homolog, have been systematically studied (Van Dijk *et al.*, 1993). The data for the Pbx-1 protein are tabulated in Figure 8E, along with sequences from sites e1–e4. The e2 and e3 sites are especially noteworthy since they overlap with *Ubx* sites 5 and 8 and are therefore directly affected by substitutions at these sites. Substitution at *Ubx* site 5, for example, reduces from seven to four the number of matches within e2 to the primary base preferences for the Pbx-1 protein (Figure 8E). The ExdHD protein normally binds to the e2 site with a K_D of 50–100 nM under conditions of approximately physiological ionic strength; the sequence alteration completely abolishes this interaction (Figure 8A), thus indicating that binding to e2 is specific and sequence-dependent. Sequence substitutions at site 8 also cause a reduction in the number of matches to the Pbx-1 consensus within e3 (Figure 8E), and this is accompanied by a corresponding reduction in binding affinity for ExdHD (Figure 8C). The behavior of e4 is somewhat more puzzling since binding appears to be affected (Figure 8C) by substitutions within PX, including *Ubx* site 9 which is immediately adjacent to but does not overlap with the sequences protected at site e4 (see Figure 2). This effect upon binding could be mediated by structural distortions of DNA induced by the adjacent sequence changes at site 9 or, alternatively, by residual binding co-operativity in the ExdHD protein (see below) that would serve to reduce the apparent affinity for site e4 when other *exd* protein sites are altered.

The binding behavior of ExdHD, although displaying a distinct specificity and lower affinity, is much like that of other previously studied homeodomains (Figure 8A and C). The larger ExdHDcc protein, in contrast, shows a surprisingly sharp transition from no apparent binding to extensive protection of a large DNA region (Figure 8B and D). The transition occurs at a protein concentration at least 10-fold lower than the concentration required for binding of ExdHD. The extensive footprint and higher affinity are suggestive of a highly co-operative mode of binding that involves multiple proteins interacting with each other through protein–protein interaction surfaces present in ExdHDcc (see Discussion). DNA from PX_{4-9} is also bound by this protein, albeit with reduced affinity and a less extensive footprint. Binding to this substituted site template is presumably mediated by primary interactions with sites e1, e4 and other remaining low affinity sites which are not altered directly by *Ubx* site substitutions.

Discussion

Dual modes of dpp regulation in PS7

Expression of *dpp* within PS7 of the visceral mesoderm comprises a single contiguous domain which appears to reflect directly the visceral mesoderm domain of *Ubx* expression. Indeed, our work demonstrates the direct action of *Ubx* protein upon *dpp* regulatory sequences. Yet even in this apparently simple system, the activity of at least one other protein is required for spatially specific activation by *Ubx* (discussed below). In addition, our analysis uncovered an unexpected second mode of regulation which contributes to normal *dpp* expression in the midgut. This 'general activation/specific repression' mode involves a regulatory element (the BM fragment) that activates expression throughout the visceral mesoderm coupled to elements that restrict general expression to the appropriate spatial domain. This unexpected mode of *dpp* regulation will be considered first, to be followed by a consideration of the *Ubx*-mediated mode.

General activation with spatially specific repression

The central observation regarding the 'general activation/specific repression' mode of *dpp* regulation is the ability of the 130 bp BM fragment alone to drive reporter construct expression throughout the visceral mesoderm (Figure 3H; see also Figure 9A). In constructs which also carry sequences from within the AX region, the activating effect of BM becomes restricted to the PS7 region, thus indicating that these additional sequences mediate a repressing effect outside of PS7 (see Figures 3 and 4). The *abdominal-A* (*abd-A*) HOM gene encodes a candidate effector protein for this repression effect since *abd-A* mutations result in the ectopic expression of *dpp* posterior to PS7 (Immerglück *et al.*, 1990; Reuter *et al.*, 1990). *abd-A* mutations also result in posterior expression of the *Ubx* protein (Bienz and Tremml, 1988). Posterior

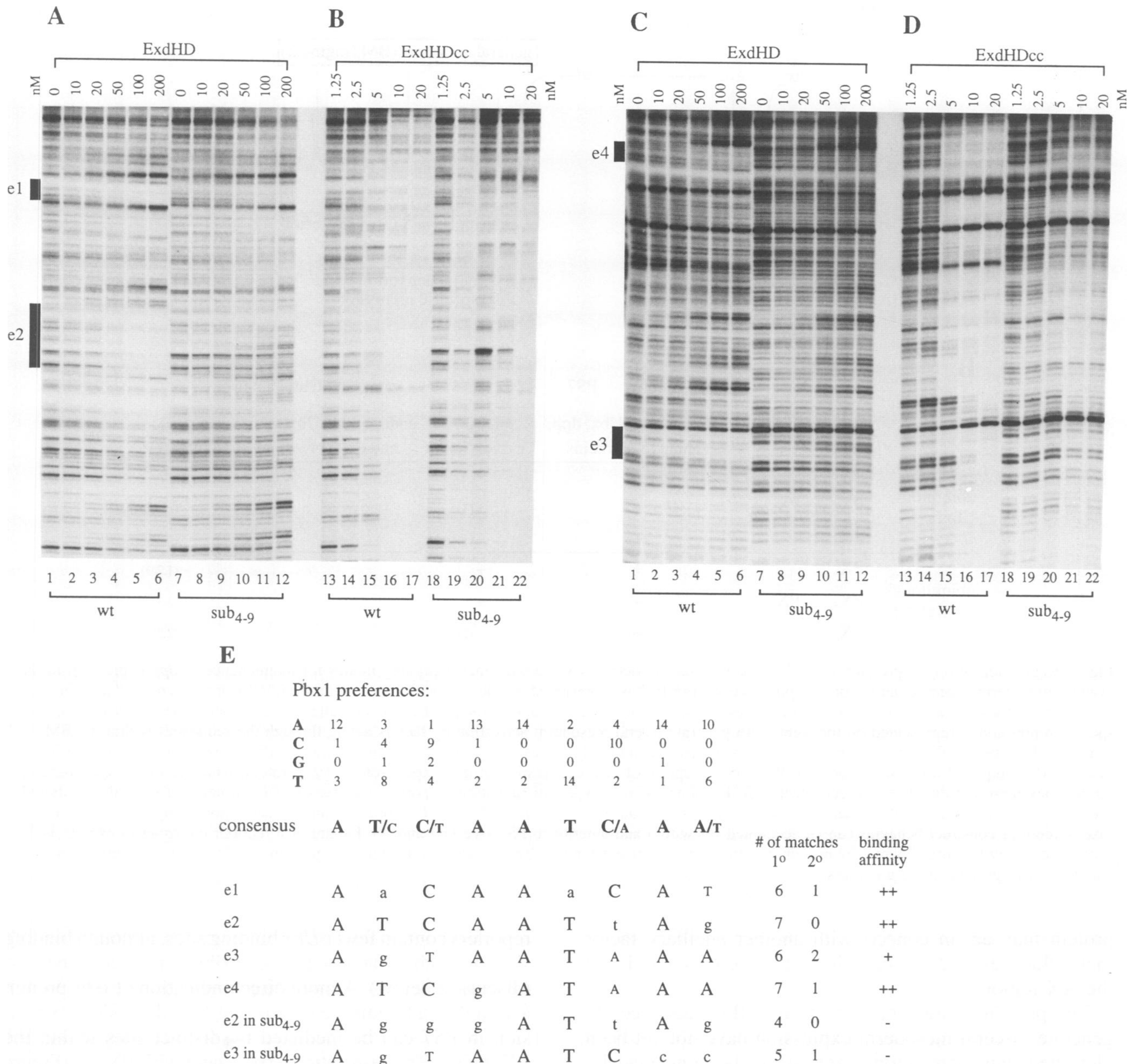
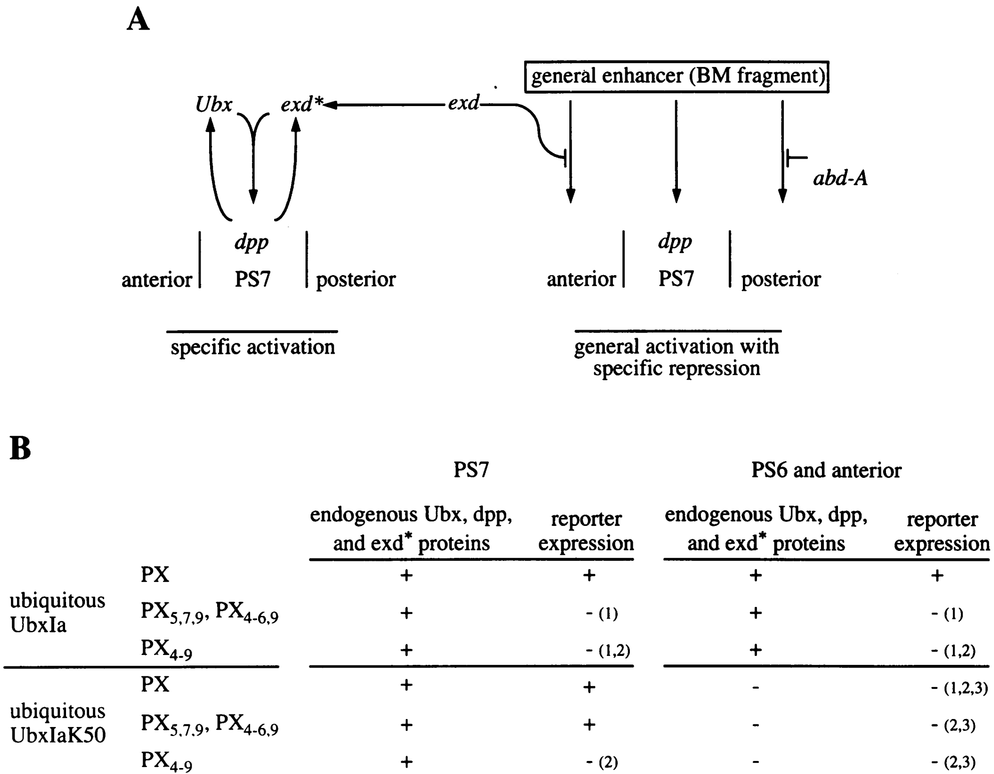


Fig. 8. Binding of *exd* homeodomain proteins to *dpp* regulatory sequences. (**A−D**) DNase I protection assays of ExdHD and ExdHDcc binding to wild-type and substituted sequences within PX. Purified ExdHD and ExdHDcc protein concentrations (nM) are indicated above, and template DNAs are identified below. Wild-type and substituted DNA templates were 3′ end-labeled on the top (A and B) or bottom strands (C and D). Protected regions are marked by solid bars and designated e1−e4 (see also Figure 2). Note the decrease in binding by ExdHD to the e2, e3 and e4 regions in the substituted template, and the more extensive and higher affinity binding observed with the ExdHDcc protein (see text). Under the conditions of these experiments *exd* proteins were in molar excess relative to the labeled fragments, and the protein concentration required to produce half-maximal protection may be considered an estimate of the equilibrium dissociation coefficient. (**E**) Alignment of *exd* binding sites and comparison with DNA sequence preferences of Pbx-1, its human homolog. The top portion shows a tabulation of the data of Van Dijk *et al.* (1993) with the Pbx-1 consensus base preferences shown below (primary and secondary base preferences shown in larger and smaller type, respectively). Sequences within the e1−e4 regions of the wild-type template and of the e2 and e3 regions in the substituted template are aligned, along with the number of matches to primary and secondary bases within the Pbx-1 consensus. Bases in lower case denote mismatches to both primary and secondary bases of the Pbx-1 consensus. Note the rough correspondence between number of matches and binding affinity for ExdHD (right-hand column).

activation of *dpp* is not due solely to *Ubx* derepression, however, since in a *Ubx abd-A* double mutant the low residual level of *dpp* seen in *Ubx* mutants also extends posteriorly, thus revealing a repressing effect of *abd-A* that is independent of *Ubx* (Reuter *et al.*, 1990; Hursh *et al.*, 1993). The sequence specificity of *abd-A* protein is very similar to that of *Ubx* (Appel and Sakonju, 1993; Capovilla *et al.*, 1994; Ekker *et al.*, 1994) and multiple sites for *abd-A* protein binding thus are present in the AX segment but also elsewhere throughout the *dpp* regulatory region. The targeting of the posterior repression activity to the AX fragment therefore suggests that the *abd-A* effect may be indirect, involving an intermediary protein that binds specifically within AX. Alternatively, the *abd-A*



		PS7		PS6 and anterior	
		endogenous Ubx, dpp, and exd* proteins	reporter expression	endogenous Ubx, dpp, and exd* proteins	reporter expression
ubiquitous UbxIa	PX	+	+	+	+
	$PX_{5,7,9}$, $PX_{4\text{-}6,9}$	+	- (1)	+	- (1)
	$PX_{4\text{-}9}$	+	- (1,2)	+	- (1,2)
ubiquitous UbxIaK50	PX	+	+	-	- (1,2,3)
	$PX_{5,7,9}$, $PX_{4\text{-}6,9}$	+	+	-	- (2,3)
	$PX_{4\text{-}9}$	+	- (2)	-	- (2,3)

Fig. 9. Regulation of *dpp* expression in PS7 of the visceral mesoderm. (**A**) Two distinct regulatory modes for maintenance of *dpp* expression in the visceral mesoderm. Specific activation of *dpp* expression in PS7 is represented on the left, with input required by *Ubx* protein and *exd* protein in activation-competent form (exd*; activation requires *dpp* expression; see Discussion for justification and alternative models). General activation with specific repression is represented on the right, with general visceral mesoderm activation by factors acting through the sequences within the BM fragment. Posterior to PS7 this general activation is directly or indirectly repressed by *abd-A* protein; anterior to PS7, expression is repressed by the activation-incompetent form of *exd* protein. With the exception of direct regulation of *dpp* expression by *Ubx* protein, which is solidly established, arrows may represent direct or indirect effects. (**B**) Response of wild-type and substituted reporter constructs to *Ubx* proteins of normal and altered specificities. With the assumption that PX reporter expression requires sufficient binding by *Ubx* proteins and by activation-competent *exd* proteins (exd*), reporter construct behavior can be accounted for within and anterior to PS7 (see text and A). Failure of a reporter construct to express is accounted for by (1) insufficient *Ubx* protein binding, (2) insufficient binding by activation-competent *exd* protein, or (3) absence of activation-competent *exd* protein (no *dpp* expression).

protein may act in concert with another ancillary factor that influences *abd-A* specificity and targets its activity to the AX region.

The protein factors that act through BM sequences to generate visceral mesoderm expression have not yet been identified, but their activity appears to be controlled by repressing factors bound within AX, suggesting possible interactions between proteins bound to the BM and AX regions. In addition, because BM can contribute to specific expression within PS7 (Figure 4), BM binding factors seem likely to interact with specific activating factors bound to other portions of the *dpp* regulatory region.

Specific activation by Ubx protein binding to multiple sites

The loss of PS7 expression in substituted reporter constructs and restoration by *Ubx* protein of a correspondingly altered specificity (Figure 6C–F) clearly rules out an intermediary factor and demonstrates a direct interaction between *Ubx* proteins and *dpp* regulatory sequences. The following arguments suggest that this activation normally occurs through multiple *Ubx* binding sites. First, as illustrated in Figure 3, many reporters carrying truncated portions of the BE fragment retain spatially accurate but reduced levels of expression, thus suggesting a graded and quantitative change in regulation. These truncated reporters contain fewer *Ubx* binding sites, although binding sites for other factors that contribute may also be lost (discussed below). A more direct indication of *Ubx* protein action through multiple sites is that *Ubx*-dependent expression in PS7 can be mediated by distinct sites within the BE region. The restoration experiment with $PX_{5,7,9}$ (Figure 6E and F) thus indicates that *Ubx* protein can act through sites located within the SX fragment (sites 5, 7 and 9). Likewise, BS reporter expression within PS7 is augmented relative to expression levels outside PS7 (Figure 3G), and this augmentation is mediated through *Ubx* protein binding to sites 1–4; this was demonstrated by loss of expression upon substitution of sites 1–4 followed by restoration of expression with the K50 protein (data not shown). Since BS and SX represent non-overlapping portions of the BE fragment and both can mediate *Ubx* activation, we conclude that *Ubx*-responsive *cis*-regulatory information is not confined to one portion of the BE fragment. Furthermore, the loss and restoration of PS7 expression with multiple combinations of substituted sites suggests that activation by *Ubx* protein operates through multiple distinct sites within the PX fragment. In addition, expression levels of particular reporters can be reduced by the mutation of specific *Ubx* binding sites. For example, expression of PX_{CAT} is reduced relative to PX (Figure 4A and F) and expression of SX_9 is reduced relative to SX

(Figure 4B and C). *Ubx*-responsive elements thus appear to be distributed at multiple locations throughout the BE regulatory region.

This topography of *Ubx* response elements is consistent with previous biochemical studies demonstrating that *Ubx* protein is capable of forming a multiprotein complex with numerous binding sites distributed in various arrangements throughout an extensive DNA region (Beachy *et al.*, 1993). The apparent gradation in quantitative levels of expression from various *dpp* reporter constructs is also consistent with a previously observed correlation between the overall quality and quantity of sites present and the stability of the resulting complex (Ekker *et al.*, 1992; Beachy *et al.*, 1993), thus suggesting that *Ubx* protein can act as an integrator of the sequence properties of extensive DNA regions. The apparent involvement of multiple dispersed *Ubx* binding sites, plus the inferred interactions of BM sequences with other sequences mediating repressing and activating effects, thus suggest the existence of a complex nucleoprotein structure with multiple simultaneous or alternative interactions between proteins bound throughout the BE segment.

Specific activation by Ubx in PS7 requires additional factors

Several lines of evidence suggest that an activity other than Ubx^+ is required for the PS7-specific stimulation of *dpp* expression. First, restoration of the $PX_{4-6,9}$ and $PX_{5,7,9}$ constructs is restricted to PS7 (Figure 6C–F), despite ubiquitous expression of UbxIaK50. This behavior might be accounted for if restored expression in PS7 required a collaboration between K50 protein bound to substituted sites and the endogenous wild-type *Ubx* protein bound to unaltered sites; expression outside the PS7 domain would thus be prevented simply by the absence of wild-type protein and the consequently insufficient overall occupancy of *Ubx* target sequences. However, if insufficient binding outside of PS7 were the cause of this inability to activate, one might expect only a reduction of expression; the complete lack of expression observed thus suggests a role for another factor whose activity is restricted to PS7. Such a factor could be expressed solely in PS7, perhaps as an additional downstream target of *Ubx*. Alternatively, a ubiquitously expressed factor could be modified as a consequence of extracellular signaling by activities present in PS7 such as *dpp* itself or the product of the *wg* gene, a Wnt homolog expressed in the anterior part of PS8. *Ubx* expression is dependent on *dpp* and *wg* for both its maintenance and the positioning of its anterior boundary; similarly, *dpp* autoregulates itself and its anterior boundary can be influenced by *wg* signaling (Hursh *et al.*, 1993; Thüringer and Bienz, 1993; Thüringer *et al.*, 1993). We note that the anterior boundary of our *dpp* reporter constructs varies with the extent of sequences included, which may also reflect differences in ability to respond to extracellular signaling factors.

A second and stronger line of argument is based upon the behavior of the PX_{4-9} reporter, whose expression is not activated by the K50 protein in PS7 or at any other location within the visceral mesoderm (Figure 6G). The PX_{4-9} reporter is also unable to respond anywhere within the visceral mesoderm when both the wild-type UbxIa and the K50 proteins are ubiquitously expressed under heat-shock promoter control (data not shown), thus clearly indicating that substitution of sites 4–9 has inactivated binding by some factor other than *Ubx* protein. A comparison of the substitution constructs suggests that this other factor may act through site 8, since this is the only site that is altered in PX_{4-9} and not in either of the other two constructs. We note that site 8 is not a preferred site for *Ubx*, consistent with recognition by a distinct factor, although it does contain a TAAT core characteristic of homeodomain protein binding sites.

Does exd encode the additional homeodomain factor implicated in PS7-specific activation?

Our substitution/restoration experiments clearly establish a role for a factor other than *Ubx* in PS7-specific activation of *dpp*. Since it is alterations of homeodomain binding sites that render PX_{4-9} unable to respond, these experiments suggest that the factor is a homeodomain protein. Two other independent lines of evidence specifically implicate *exd*: the first includes genetic studies that identify an important role for *exd* in the regulation of *dpp* and other genes expressed in the visceral mesoderm (Rauskolb and Wieschaus, 1994). The second consists of the biochemical studies presented here that demonstrate specific binding of *exd* homeodomain to four sites within a *dpp* regulatory region sufficient for expression in PS7 of the visceral mesoderm.

Two potential problems complicate the assignment of *exd* protein as the additional homeodomain factor implicated by substitution/restoration experiments. The first is that site 8 is implicated by the difference between the PX_{4-9} construct and the two constructs $PX_{4-6,9}$ and $PX_{5,7,9}$; the site 8 substitution, however, only affects *exd* protein binding at e3, which is not the only nor even the highest affinity site within PX. A second complication is that while *exd* is transcribed throughout the visceral mesoderm, our substitution/restoration experiments suggest an activity localized to PS7.

With regard to the role of site 8 (or *exd* site e3), it is important to note that the PX_{4-9} substitutions also affect *exd* protein binding to sites e2 and e4 (see Results). Site e3 therefore assumes its critical importance in a context where sites e2 and e4 are already affected, and might have a less dramatic impact if sites e2 and e4 remained intact. Similarly, sites e2 and e4, where binding seems affected by substitution of *Ubx* sites 5 and 9, might appear critical for reporter expression if eliminated in a context where e3 and/or other *exd* sites were already disrupted. The possibility that any of several DNA sites suffice to mediate *exd* protein regulation of *dpp* is supported by the highly co-operative nature of ExdHDcc binding to the PX fragment (see Results and Figure 8B and D). Such co-operativity would lead to an overall integration of the number and quality of sites present within a regulatory region, with the regulatory effect dependent on the overall affinity of the complex rather than the presence or absence of any single site. We note that application of the coiled-coil structure prediction algorithm of Lupas *et al.* (1991) suggests the possible existence of several distinct coiled-coil domains within the *exd* protein (Figure 10). The existence and locations of these predicted domains are highly conserved in evolution, as can be seen from the analysis of homologous human and nematode sequences

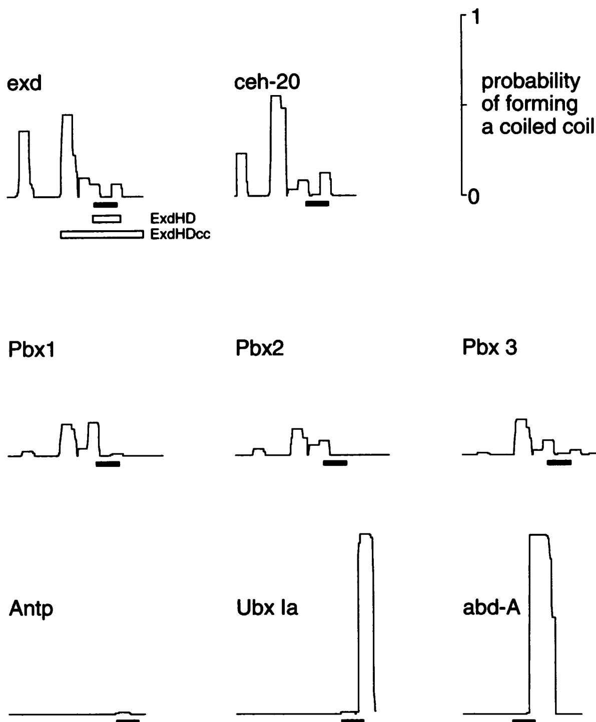


Fig. 10. Predicted coiled-coil forming regions of *exd* and its homologs. The algorithm of Lupas *et al.* (1991) was applied to the amino acid sequences of the protein encoded by *exd*, of the *Caenorhabditis elegans* protein ceh-20, and of the human proteins Pbx-1, Pbx-2 and Pbx-3 (Burglin and Ruvkun, 1992); analyses of HOM proteins Antp, UbxIa and abd-A are also presented (Beachy *et al.*, 1993). The probability of coiled-coil formation is plotted as a function of position within the amino acid sequence (scale shown). The black bars below each graph show the location of the homeodomain (including a three-residue insertion in the homeodomains of exd, ceh-20 and the Pbx proteins). Open boxes below the exd representation show the extent of sequences present in the ExdHD and ExdHDcc proteins (see also Figures 7 and 8); note the presence of multiple domains with some probability of coiled-coil formation in ExdHDcc. All sequences were obtained from release 10.0 of the Entrez sequence database (National Center for Biotechnology Information, Bethesda, MD).

(Figure 10). The ExdHDcc protein contains all but one of these predicted coiled-coil domains, suggesting that the surfaces responsible for the highly co-operative binding behavior of this protein may be coiled-coils. The presence of the predicted coiled-coil structure within the ExdHD protein, in contrast, is minimal.

With regard to the expression of *exd* throughout the visceral mesoderm, the observations of Rauskolb and Wieschaus (1994) suggest that activities of *exd* protein within the visceral mesoderm are not uniform. For example, exd^+ appears to be required for the expression of *wg* specifically in PS8 and PS9, without activating *wg* expression elsewhere. As regards *dpp* expression, exd^+ is required for activation in PS7, but anterior to PS7 it is required for repression. Rauskolb and Wieschaus (1994) proposed that the activities of *exd* protein at distinct locations throughout the visceral mesoderm are determined by the presence of distinct HOM proteins which act together with *exd*. This proposal is consistent with the observations that *wg* expression in PS8 and PS9 also requires $abd\text{-}A^+$ and that expression of *dpp* in PS7 also requires Ubx^+, although the repression of *dpp* at positions anterior to PS7 does not require $Antp^+$ or other HOM gene products (Reuter *et al.*, 1990).

The interaction of *exd* protein as a partner in DNA binding with the *Ubx* and *abd-A* proteins but not the *Antp* protein is an attractive mechanistic explanation for these observations, particularly since the *Ubx* and *abd-A* proteins but not the *Antp* protein contain predicted coiled-coil structures at positions just C-terminal to their respective homeodomains (Figure 10). A highly suggestive analogy can be drawn here to the determination of the yeast diploid cell type, whose specification requires an interaction between homeodomain proteins a1 and α2 (Goutte and Johnson, 1993). The a1 homeodomain protein is the closest yeast relative to the *exd* protein (Rauskolb *et al.*, 1993), and recent work suggests that at least one component of the interaction between a1 and α2 is mediated through surfaces that adopt a coiled-coil structure (Ho *et al.*, 1994).

While differential interactions of the *exd* protein with HOM proteins form an attractive hypothesis to account for a number of observations, this hypothesis alone does not fully explain the paradoxical restriction to PS7 of substituted *dpp* reporter expression in response to ubiquitous K50 protein (Figure 6C–F); endogenous *dpp* and wild-type reporter constructs, in contrast, are expressed ectopically in response to ubiquitous wild-type *Ubx* protein (Figure 6A and B). An attractive model to account for these observations is that the *exd* protein only acquires competence to activate *dpp* (in partnership with *Ubx* protein) as a result of signaling by the *dpp* gene product itself (Figure 9A). This model suggests a mechanism by which the previously reported autoregulation of *dpp* in the visceral mesoderm is accomplished and would restrict the spatial domain in which *exd* protein is active. The competence to act in *dpp* activation could be the result of some form of post-translational modification, such as phosphorylation or dephosphorylation, in response to *dpp* signaling. Such signal-dependent activation of a transcription factor by post-translational modification is well established, e.g. in the case of CREBP (reviewed by Karin and Smeal, 1992). The attraction of this model is that it accounts for the inability of ubiquitous K50 protein to activate substituted site reporters outside of the PS7 domain, since endogenous *dpp*, required for *exd* competence, would not be induced by the K50 protein (Figure 9B). Wild-type reporters, on the other hand, would be ectopically activated by ubiquitous wild-type *Ubx* protein since ectopic expression of endogenous *dpp* would be induced, thereby causing conversion of *exd* protein to the activation-competent form. A second attraction of this model is that the *exd* protein would function simultaneously as a cofactor for *Ubx* activation of *dpp* in PS7 and as the conduit for *dpp* autoregulatory action, thus interlocking these two *dpp* regulatory circuits through a common component, the *exd* protein. A second related model that could account for these observations is that ectopic wild-type but not K50 *Ubx* protein is capable of directly or indirectly inducing expression of an as yet unidentified protein that is required as a cofactor with *Ubx* protein for *dpp* activation in the visceral mesoderm. The resolution of these possibilities must await further studies.

Correct spatial expression reinforced through functional overlap

Perhaps the most striking conclusion to emerge from our analysis of *dpp* regulation is the operation of two distinct modes of spatial regulation that both produce expression in PS7 of the visceral mesoderm (Figure 9A). Multiple pathways are also observed in the autoregulation of *Ubx* through *wg* and *dpp* (Thüringer and Bienz, 1993; Thüringer *et al.*, 1993), and a discussion of the potential roles of multiple layers of regulation is provided by these authors. The two modes of *dpp* regulation reported here also illustrate the potentially misleading conclusions that might be drawn by focusing exclusively upon minimal reporter constructs, which might have led to the conclusion that the PX fragment carries binding sites for all of the factors that contribute to *dpp* expression *in vivo*. Even now, we cannot rule out the possibility that additional unidentified regulatory modes involving novel factors may also contribute to PS7 expression of *dpp* in the visceral mesoderm.

In addition to functional overlap between multiple regulatory modes, our detailed analysis of the 'specific activation' mode of *dpp* regulation revealed functional overlap at several other levels. For example, multiple distinct sets of *Ubx* protein binding sites appear to suffice for PS7-specific activation of *dpp*, and it is difficult to identify any single absolutely critical site within the full context of the 812 bp *dpp* regulatory region (see above). In addition to apparently redundant binding sites for individual factors, however, we also observe functional redundancy as represented by spatial restriction of more than one factor. Restriction to PS7 thus appears to be true not only for *Ubx* but also for the additional homeodomain protein activity identified by our studies (possibly the activation-competent form of the *exd* protein; see above and Figure 9). Whatever the ultimate purpose of such functional overlap, the current studies provide a basis for the study of other factors that act with *Ubx* in the regulation of *dpp*.

Materials and methods

Plasmids

An 812 bp *Bam*HI−*Eco*RI fragment (BE) was subcloned from RD1 (Hursh *et al.*, 1993) into Bluescript KS$^+$ (Stratagene) and its sequence determined using Sequenase (USB) to facilitate further cloning steps. This construct, pBE, was used in all subsequent cloning and mutagenesis reactions. The β-galactosidase reporter gene construct pCaSpeR-hs43-LacZ (abbreviated CSR43; Thummel *et al.*, 1988) was linearized with *Xho*I, end-filled with the Klenow fragment of polymerase I, and further digested with *Bam*HI after phenol−chloroform (1:1) extraction and ethanol precipitation. BE and a series of 3′ truncations were prepared for ligation into the linearized CSR43 vector by digestion of pBE with *Eco*RI (the BE construct), *Xho*I (the BX construct), *Stu*I (the BS construct) or *Msc*I (the BM construct), end-filling the 3′ ends with Klenow where necessary to generate blunt ends and digestion with *Bam*HI. An additional truncation, BA, was prepared for ligation into linearized CSR43 by using the T7 and PB547 oligonucleotides (see Table II) as PCR primers with the pBE template; the amplification product was recovered, digested with *Bgl*II, end-filled with Klenow, and digested with *Bam*HI.

For the preparation of two additional truncated constructs, CSR43 was opened at the *Eco*RI site and end-filled with Klenow, followed by digestion with *Xho*I. Truncated inserts were prepared by digestion of pBE with *Pst*I (for the PX construct) or *Stu*I (for the SX construct). Treatment of the *Pst*I digest with T4 polymerase in the presence of deoxynucleoside triphosphates was required to make the *Pst*I end blunt; both fragments were then treated further with *Xho*I prior to ligation. The PX and SX inserts were also combined individually with the BM fragment (described in the previous paragraph) for three-fragment ligations that also included CSR43 digested with *Bam*HI and *Xho*I.

Recombinant PCR (Jones and Winistorfer, 1992) using primers described in Table II was used to alter different site sequences within pBE. Combinations of site alterations were generated by sequential rounds of RPCR. The PX portions of these mutagenized plasmids were then excised and subcloned into CSR43, as described above for the PX construct, resulting in the reporter constructs PX_{4-9}, $PX_{5,7,9}$, $PX_{4-6,9}$, PX_{CAT}, PX_{PAL} and PX_{AP2}.

The expression construct for UbxHDK50 was generated by RPCR using primers PB314 and PB315 (Table II) with wild-type UbxHD in Bluescript as template. The mutated open reading frame was isolated and subcloned into the bacterial expression vector pET3c (Novagen) using *Nde*I and *Bam*HI sites introduced by the PCR. PB314 and PB315 were also used to generate a full-length UbxIaK50 open reading frame carrying a lysine at position 50 within the homeodomain. The UbxIaK50 open reading frame with *Nde*I (blunted) and *Bam*HI ends was inserted into pCaSpeR-hs (Thummel *et al.*, 1988) using vector sites *Eco*RI (blunted) and *Bgl*II. Expression constructs for ExdHD (amino acids 237–312 of the *exd* open reading frame plus an initiator methionine) and ExdHDcc (amino acids 149–376 of the *exd* open reading frame plus an initiator methionine) were amplified by PCR from the *exd* cDNA (Rauskolb *et al.*, 1993) and treated with T4 polynucleotide kinase. An *Nde*I compatible overhang at the 5′ end of the open reading frame was introduced by a two-stage treatment with T4 DNA polymerase, the first in the presence of dCTP and dGTP and the second with added dCTP, dGTP and dATP. This was followed by treatment with *Bam*HI and ligation into the bacterial expression vector pET11c (Novagene) digested previously with *Nde*I and *Bam*HI. The sequences of all PCR-generated portions of all reporter and expression constructs were confirmed using Sequenase (US Biochemicals).

Mapping of the dpp^{s6} mutation

Oligonucleotides PB339 and PB392 (Table II) were used as PCR primers to amplify a fragment from homozygous *dpp*^{s6} (Segal and Gelbart, 1985) genomic DNA. The PCR product was subcloned as a blunt-end fragment into Bluescript and the sequence of the insert showed that the *dpp*^{s6} deletion removes 4845 bp, beginning 2329 bp upstream of the *Bam*HI site and ending 1709 bp downstream of the *Eco*RI site that define BE (all distances given relative to the 5′-most base of the *Bam*HI and *Eco*RI recognition sequences).

Yeast transactivation assays

Yeast reporter plasmids were made by subcloning *Pst*I−*Eco*RI fragments from pBE or pBE with substitutions at sites 4–9 into the yeast centromere vector pSEΔ1′ (Ekker *et al.*, 1992). Both inserts were located upstream of the *lacZ* reporter gene in the same relative orientation as compared with the *dpp* open reading frame. Activator plasmids were made by subcloning UbxIa or UbxIaK50 coding sequences into the yeast centromere vector pRS314GU (provided by P.Hieter). This vector is based on pRS314 (Sikorski and Hieter, 1989) but contains the GAL1 promoter inserted from the *Kpn*I to *Xho*I sites and the URA3 promoter from the *Sac*I to *Bam*HI sites in the polylinker. Both inserts were oriented so that expression of UbxIa was controlled by the GAL1 promoter. The reporter and activator constructs were transformed into the yeast strain YPH500 (Sikorski and Hieter, 1989) using lithium acetate (Gietz *et al.*, 1992) and grown on selective medium to ensure maintenance of the plasmids. Galactose induction of *Ubx* protein expression and β-galactosidase assays were performed as described previously (Ekker *et al.*, 1992).

Drosophila strains

Strains carrying reporter constructs or expression constructs were generated by P-element-mediated germ line transformation according to standard protocols (Spradling, 1986). Constructs (at 400 μg/ml) and pπ25.7wc DNA (at 100 μg/ml; Karess and Rubin, 1984) were co-injected into w^{1118} or *y* w^{1118} embryos. Inserts were mapped to a chromosome and balanced or made homozygous. The $Ubx^{9.22}$ allele was isolated (Kerridge and Morata, 1982) and its genetic and molecular mapping have been described previously (Bender *et al.*, 1983; Akam *et al.*, 1984; Weinzierl *et al.*, 1987). For the analysis of PX reporter expression in a $Ubx^{9.22}$ background, embryos were collected from a stock homozygous for a second chromosome PX reporter and heterozygous for the *Ubx* allele. The hsUbxIa strain has been described previously (González-Reyes *et al.*, 1990).

Table II. Oligonucleotide sequences

Name	5′−3′ sequence[a]	Remarks
T3	TCGAAATTAACCCTCACTAAAG	KS$^+$
T7	TTGTAATACGACTCACTATAG	KS$^+$
PB316	AATGATACCGCGAGACCCACGCTC	RPCR, KS$^+$
PB317	GTATTATCCCGTATTGACGCCGGG	RPCR, KS$^+$
PB437	GAAGACCCGGGGATTACGTGG	site 4
PB438	CCACGTAATCCCCGGGTCTTC	site 4
PB689	AAGGCCTAGGGATTAGCTCCCATTTC	site 5
PB690	GTGCTAATCCCTAGGCCTTTGATCTG	site 5
PB464	AGGGATTAGCACCCATTGGGATTATGCTGCA	sites 5 and 6
PB465	TCCCAATGGGTGCTAATCCCTAGGCCTTTG	sites 5 and 6
PB691	AATTCACACGGGATTAGTAATAAAT	site 7
PB692	ATTACTAATCCCGTGTGAATTGCAG	site 7
PB466	ACGGGATTAGTAATCCCTTTGAATGCG	sites 7 and 8
PB467	AAGGGATTACTAATCCCGTGTGAATTG	sites 7 and 8
PB384	GTTATGGTGGGGATTAAGTTTT	site 9
PB385	AAAACTTAATCCCCACCATAAC	site 9
PB472	TACAAGTGGTATTTCCCCATGACAAGTGGTCCGTGCAATGTTTG	nonamer nos 1–3
PB473	GGGGAAATACCACTTGTAAGACCACTTGTTAGTACTCCACGTA	nonamer nos 1–3
PB484	AGCAACAAGTGCGATTATGCTGCATGCTG	nonamer no. 4
PB485	ATCGCACTTGTTGCTAATTGATAGGCC	nonamer no. 4
PB547	GTGTGAAGATCTGGCAGTGCAAGTGCT	artificial *Bgl*II
PB390	ATCGGGAGCTCGAGCGCCATC	artificial *Xho*I
PB468	CGATGTACACTTTCCCATTTGGCCGTG	AP-2 no. 1
PB469	ATGGGAAAGTGTACATCGAAATGGGAAGC	AP-2 no. 1
PB470	TCCTGTACACTTTCCCAGATTTCAATTC	AP-2 no. 2
PB471	TCTGGGAAAGTGTACAGGAAAGATCGGGGGT	AP-2 no. 2
PB478	ATACTGGATAGGAATAACTTGCACTGCCTGC	palindrome
PB479	GTTATTCCTATCCAGTATGCAGCATAATCGAA	palindrome
PB339	GTCGACAAATGAAAAGTTTCGAC	s6 upstream
PB392	AGCTTATCGGACCACCAGAAAT	s6 downstream
PB66	GTTTTCCCAGTCACGGCGGCC	selection sequencing

[a]Bold bases represent substitutions.

Reporter construct expression assays

Digoxigenin whole-mount *in situ* hybridization was conducted as described previously (Hursh *et al.*, 1993). Antibody detection of β-galactosidase was also as described previously (Hursh *et al.*, 1993) with the following modifications. Either rabbit polyclonal anti-β-galactosidase antibody (Cappel; Figures 1, 3 and 4) or a monoclonal anti-β-galactosidase antibody (Promega; Figure 6) were used. Normal donkey serum (Jackson Laboratory) was used instead of normal goat serum. In cases where the monoclonal antibody was used, Vectastain Elite Kit (Vector) was used instead of secondary antibody directly conjugated to horseradish peroxidase. All control/experimental pairs were stained in parallel and a minimum of five independently derived lines were examined for each construct. Embryos were cleared in methyl salicylate and photographed under Nomarski optics.

Combinations of specific reporter constructs with *Ubx* protein expression constructs were analyzed by collecting embryos from mass matings of virgin females homozygous for the reporter construct with males carrying the *Ubx* protein expression construct (or vice versa). Embryos were collected on a nylon mesh and heat shocked for 1.5–2 h by placing them at the air−water interface of a 37°C water bath. This was followed by a 4 h recovery at 25°C and fixation and staining for β-galactosidase (see above). Five, three and two independent insertions of the PX_{4-9}, $PX_{5,7,9}$ and $PX_{4-6,9}$ constructs, respectively, were combined with each of two different K50 *Ubx* producer lines; results with each of these combinations were similar to those shown in Figure 6.

Purification of recombinant proteins

UbxHD and UbxHDK50 were prepared as described previously (Ekker *et al.*, 1992) using the expression construct described above for UbxHDK50. ExdHD and ExdHDcc were prepared using expression constructs described above in the bacterial strain BL21(DE3) pLysS (Studier *et al.*, 1990). Following 3 h of induction with 1 mM IPTG, cleared lysates were prepared and nucleic acids precipitated with polyethyleneimine P as described (Beachy *et al.*, 1988). These extracts were then fractionated by column chromatography with phosphocellulose using a gradient of 0.1–1.0 M NaCl in 40 mM $NaPO_4$ (pH 6.8), 5 mM DTT, 1 mM EDTA and 10% glycerol. The resulting peak fractions contained proteins of the expected electrophoretic mobility at 80–90% homogeneity as determined by SDS−PAGE analysis. The identities of these purified proteins were confirmed by six and ten rounds of automated Edman degradation (Applied Biosystems Model 477A) for the smaller and larger proteins, respectively. Protein concentrations were determined spectrophotometrically by absorbance at 205 and 280 nm (Scopes, 1987).

Optimal binding site selections

A 64 base oligonucleotide with the sequence 5′-GTAAAACGACGCC-AGTGGATCCNNNNNNNATTANNNNNNNGCGGCCGCCGTGGA-CTGGAAAAC-3′ (N indicates use of an equimolar mixture of all four base precursors during synthesis) was used for binding site selections. Two rounds of selection were performed as described previously (Ekker *et al.*, 1992) with either UbxHD or UbxHDK50 protein. Briefly, labeled double-stranded DNA at 1 nM was mixed with wild-type or lysine 50 homeodomain protein at 10 nM in a solution of 20 mM Tris−Cl (pH 7.6), 75 mM KCl, 50 μg/ml BSA, 1 mM DTT, 10% glycerol and allowed to equilibrate for 20 min at room temperature. Double-stranded unlabeled competitor DNA was added to 50 nM and aliquots taken for electrophoresis at various times (0–30 min) afterwards. The competitor used for UbxHD consisted of a double-stranded oligonucleotide of sequence $(TAATGG)_6$. The same competitor was used in the first round of selection with UbxHDK50; for the second round a competitor of sequence $(TAATCC)_6$ was used because the $(TAATGG)_6$ oligonucleotide was ineffective as a competitor. Samples were electrophoresed in 10% polyacrylamide (30:0.8 acrylamide:bisacrylamide) gels containing 0.5× TBE and 3% glycerol. Bands corresponding to protein−DNA complexes were excised, rehydrated and amplified as described previously (Ekker *et al.*, 1992), and subjected to dideoxy chain termination reactions with Sequenase 2.0 and $MnCl_2$ buffer (US Biochemicals) using ^{32}P-labeled primer PB66 (Table II). Reactions were electrophoresed in 10% D600 (J.T.Baker)−7 M urea gels and dried. A digital image of the gel was acquired on a PhosphorImager (Molecular Dynamics). Band intensities were quantified and base preferences calculated as described previously (Ekker *et al.*, 1992).

DNase I footprinting assay

DNA probes were prepared using primers PB437 and PB390 (Table II) for PCR with two different templates, pBE and pBE_{4-9}. The PCR

products were isolated by agarose gel electrophoresis and electroelution followed by ethanol precipitation. In Figure 8A and B the recovered DNA fragments were digested with *Xho*I (introduced at position 578 by primer PB390) followed by agarose gel purification. The resulting DNA fragments (from base pair 274 to 582; Figure 2) were end-labeled at the overhanging *Xho*I site with Klenow using [α-^{32}P]dCTP, chased with cold dCTP, extracted once with phenol–chloroform (1:1) and purified using Nick columns (Pharmacia). In Figure 8C and D the same PCR products were digested with *Sau*3A1 (at position 409) and the resulting DNA fragments (from base pair 409 to 591) were isolated and labeled the same way as described above.

Binding of either ExdHD or ExdHDcc proteins to labeled DNA (estimated at 70 pM concentration) proceeded at 22°C for 25 min in 300 μl binding buffer (10 mM HEPES, pH 7.0, 150 mM KoAc, 2 mM $MgCl_2$, 0.1 mM EDTA, 0.05% Triton X-100, 50 μg/ml BSA and 5 mM dithiothreitol). Each reaction then received 40 μl of a solution containing 21 mM of $CaCl_2$ and 37 mM of $MgCl_2$, immediately followed by 4 ng pancreatic DNase I (in a volume of 10 μl; Worthington Enzymes, DPFF grade). DNase I digestion was stopped after 30 s at 22°C by the addition of 60 μl of stop solution (1.5 M NaCl, 250 mM Tris–HCl, pH 8.0, 100 mM EDTA, 5% sodium lauryl sarcosine and 175 μg/ml yeast tRNA). After extraction twice with phenol–chloroform (1:1), nucleic acids were precipitated by the addition of 1 ml ethanol and resuspended in 6 μl loading buffer (80% deionized formamide, 0.1× TBE, 0.05% bromophenol blue and 0.08% xylene cyanol) for electrophoresis in 0.4 mm, 7.3% polyacrylamide gels. The G/A tracks were prepared as described in Muro *et al.* (1993).

Acknowledgements

We thank R.Padgett, W.Gelbart, C.Rauskolb, C.Murre, J.Manak, and M.Scott for the communication of unpublished data, and C.Rauskolb, C.Murre, P.Hieter, V.Corces and G.Struhl for strains and plasmids. We also thank M.Bienz for communicating unpublished observations and for useful suggestions. J.Porter, D.Chang and C.Chiang were helpful in the initial stages of embryo injection, S.Ekker provided advice on oligonucleotide selections and yeast expression assays, and D.Andrew and M.Mortin assisted with antibody staining. We are grateful to C.Davenport, J.Ptak, C.Riley and W.-S.Liu for oligonucleotide synthesis and protein sequence determination, and to T.Jones for excellent technical assistance. We also thank M.Mortin and S.Ekker for helpful discussions and comments on the manuscript. During the review stages of this manuscript, two articles appeared on the subject of *extradenticle* protein and its potential interaction with *Ultrabithorax* protein (Chan *et al.*, 1994; van Dijk and Murre, 1994).

References

Akam,M., Moore,H. and Cox,A. (1984) *Nature*, **309**, 635–637.
Appel,B. and Sakonju,S. (1993) *EMBO J.*, **12**, 1099–1109.
Beachy,P.A., Krasnow,M.A., Gavis,E.R. and Hogness,D.S. (1988) *Cell*, **55**, 1069–1081.
Beachy,P.A., Varkey,J., Young,K.E., von Kessler,D.P., Sun,B.I. and Ekker,S.C. (1993) *Mol. Cell. Biol.*, **13**, 6941–6956.
Bender,W., Akam,M., Karch,F., Beachy,P.A., Peifer,M., Spierer,P., Lewis,E.B. and Hogness,D.S. (1983) *Science*, **221**, 23–29.
Bienz,M. (1994) *Trends Genet.*, **10**, 22–26.
Bienz,M. and Tremml,G. (1988) *Nature*, **333**, 576–578.
Burglin,T.R. and Ruvkun,G. (1992) *Nature Genet.*, **1**, 319–320.
Capovilla,M., Brandt,M. and Botas,J. (1994) *Cell*, **76**, 461–476.
Chan,S.K. and Mann,R.S. (1993) *Genes Dev.*, **7**, 796–811.
Chan,S.K., Jaffe,L., Capovilla,M., Botas,J. and Mann,R.S. (1994) *Cell*, **78**, 603–615.
Ekker,S.C., Young,K.E., von Kessler,D.P. and Beachy,P.A. (1991) *EMBO J.*, **10**, 1179–1186.
Ekker,S.C., von Kessler,D.P. and Beachy,P.A. (1992) *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) *EMBO J.*, **13**, 3551–3560.
Furukubo-Tokunaga,K., Flister,S. and Gehring,W.J. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6360–6364.
Galang,C.K. and Hauser,C.A. (1992) *New Biol.*, **4**, 558–568.
Gibson,G., Schier,A., LeMotte,P. and Gehring,W.J. (1990) *Cell*, **62**, 1087–1103.
Gietz,D., St Jean,A., Woods,R.A. and Schiestl,R.H. (1992) *Nucleic Acids Res.*, **20**, 1425.
González-Reyes,A., Urquia,N., Gehring,W.J., Struhl,G. and Morata,G. (1990) *Nature*, **344**, 78–80.
Goutte,C. and Johnson,A.D. (1993) *J. Mol. Biol.*, **233**, 359–371.
Hanes,S.D. and Brent,R. (1991) *Science*, **251**, 426–430.
Ho,C., Adamson,J.G., Hodges,R.S. and Smith,M. (1994) *EMBO J.*, **13**, 1403–1413.
Hursh,D.A., Padgett,R.W. and Gelbart,W.M. (1993) *Development*, **117**, 1211–1222.
Immerglück,K., Lawrence,P.A. and Bienz,M. (1990) *Cell*, **62**, 261–268.
Jones,D.H. and Winistorfer,S.C. (1992) *BioTechniques*, **12**, 528–534.
Karess,R.E. and Rubin,G.M. (1984) *Cell*, **38**, 135–146.
Karin,M. and Smeal,T. (1992) *Trends Biochem. Sci.*, **17**, 418–422.
Kerridge,S. and Morata,G. (1982) *J. Embryol. Exp. Morphol.*, **68**, 211–234.
Kuziora,M.A. and McGinnis,W. (1989) *Cell*, **59**, 563–571.
Kuziora,M.A. and McGinnis,W. (1991) *Mech. Dev.*, **33**, 83–94.
Lupas,A., Van Dyke,M. and Stock,J. (1991) *Science*, **252**, 1162–1164.
Masucci,J.D. and Hoffmann,F.M. (1993) *Dev. Biol.*, **159**, 276–287.
McGinnis,W. and Krumlauf,R. (1992) *Cell*, **68**, 283–302.
Muro,A.F., Pesce,C.G. and Kornblihtt,A.R. (1993) *Trends Genet.*, **9**, 337–338.
Nelson,H. and Laughon,A. (1993) *Roux's Arch. Dev. Biol.*, **202**, 341–354.
Panganiban,G.E.F., Reuter,R., Scott,M. and Hoffmann,F.M. (1990) *Development*, **110**, 1041–1050.
Percival-Smith,A., Müller,M., Affolter,M. and Gehring,W.J. (1990) *EMBO J.*, **9**, 3967–3974.
Rauskolb,C. and Wieschaus,E. (1994) *EMBO J.*, **13**, 3561–3569.
Rauskolb,C., Peifer,M. and Wieschaus,E. (1993) *Cell*, **74**, 1101–1112.
Reuter,R., Panganiban,G.E.F., Hoffmann,F.M. and Scott,M.P. (1990) *Development*, **110**, 1041–1050.
Schier,A.F. and Gehring,W.J. (1992) *Nature*, **356**, 804–807.
Schier,A.F. and Gehring,W.J. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 1450–1454.
Scopes,R.K. (1987) *Protein Purification, Principles and Practice*. Springer-Verlag, New York.
Segal,D. and Gelbart,W.M. (1985) *Genetics*, **109**, 119–143.
Sikorski,R.S. and Hieter,P. (1989) *Genetics*, **122**, 19–27.
Spradling,A.C. (1986) In Roberts,D.B. (ed.), *Drosophila: A Practical Approach.* IRL Press, Oxford, UK.
St Johnston,R.D. and Gelbart,W.M. (1987) *EMBO J.*, **6**, 2785–2791.
St Johnston,R.D., Hoffmann,F.M., Blackman,R.K., Segal,D., Grimaila,R., Padgett,R.W., Irick,H.A. and Gelbart,W.M. (1990) *Genes Dev.*, **4**, 1114–1127.
Stark,M.R. and Johnson,A.D. (1994) *Nature*, **371**, 429–432.
Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
Thummel,C.S., Boulet,A.M. and Lipshitz,H.D. (1988) *Gene*, **74**, 445–456.
Thüringer,F. and Bienz,M. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 3899–3903.
Thüringer,F., Cohen,S.M. and Bienz,M. (1993) *EMBO J.*, **12**, 2419–2430.
Treisman,J., Gönczy,P., Vashishtha,M., Harris,E. and Desplan,C. (1989) *Cell*, **59**, 553–582.
Tremml,G. and Bienz,M. (1989) *EMBO J.*, **8**, 2677–2685.
Van Dijk,M.A. and Murre,C. (1994) *Cell*, **78**, 617–624.
Van Dijk,M.A., Voorhoeve,P.M. and Murre,C. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6061–6065.
Vershon,A. and Johnson,A.K. (1993) *Cell*, **72**, 105–112.
Vincent,J.P., Kassis,J.A. and O'Farrell,P.H. (1990) *EMBO J.*, **9**, 2573–2578.
Weinzierl,R., Axton,J.M., Ghysen,A. and Akam,M. (1987) *Genes Dev.*, **1**, 386–397.
Williams,T., Admon,A., Lüscher,B. and Tjian,R. (1988) *Genes Dev.*, **2**, 1557–1569.
Zeng,C., Pinsonneault,J., Gellon,G., McGinnis,N. and McGinnis,W. (1994) *EMBO J.*, **13**, 2362–2377.
Zeng,W., Andrew,D.J., Mathies,L.D., Horner,M.A. and Scott,M.P. (1993) *Development*, **118**, 339–352.

Received on August 2, 1994; revised on November 14, 1994