Published in final edited form as:

Mech Dev. 1993 May ; 41(2-3): 155–161. doi:10.1016/0925-4773(93)90045-y.

Dpbx, a new homeobox gene closely related to the human proto-oncogene *pbx1* Molecular structure and developmental expression

Willy A. Flegel, **Andrew W. Singson**, **Jonathan S. Margolis**, **Anne G. Bang**, **James W. Posakony**, **Cornelis Murre**

Department of Biology and Center for Molecular Genetics University of California, San Diego La Jolla, CA 92093, USA

Abstract

Recently, a new class of homeodomain containing proteins, *pbx1, pbx2,* and *pbx3* has been described, *pbx* proteins are most closely related to two yeast regulatory proteins, *a1* and *α2*. Here, we identify and characterize the *pbx* homolog in *Drosophila,* designated *Dpbx. Dpbx* is 95% identical to the *pbx* proteins within the homeodomain and, more remarkably, is 85% to 88% identical within a 201 amino acid region adjacent to the homeodomain. Cytologically, the *Dpbx* gene is located on the X chromosome at 14A. mRNA expression is both maternal and zygotic and occurs throughout the life cycle. Prior to full germband retraction, *Dpbx* is rather ubiquitously present and variations are minor. The most notable feature of *Dpbx* expression is that after germband retraction, high levels of *Dpbx* are observed in the anterior portion of the ventral nerve cord.

Keywords

Dpbx; *pbx* gene family; Proto-oncogene; Homeobox gene; Homeodomain; Central nervous system; *Drosophila melanogaster* development

Introduction

A new family of homeodomain proteins, designated *pbx,* was recently identified. The *pbx* family contains three members, *pbx1, pbx2* and *pbx3* (Monica et al., 1991). The *pbx1* gene is rearranged in pre-B acute lymphoblastic leukemias (ALL), involving a t(1;19) (q23;p13) translocation. The t(1;19) translocation results in a fusion between the N-terminal region of *E2A* and the C-terminal domain of *pbx1,* replacing the *E2A* DNA binding domain with a homeodomain derived from *pbx1* (Nourse et al., 1990, Kamps et al., 1990). The N-terminal domain of *E2A* contains a transactivation motif (Henthorn et al., 1990; Quong et al., 1993). Thus the t(1;19) chromosomal translocation fuses the N-terminal transactivation domain of *E2A* to a heterologous DNA binding domain derived from *pbx1.*

Correspondence to: W.A. Flegel, Department of Biology 0116, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0116, USA.

Present address: Carnegie Institution of Washington, Department of Embryology, 115 West University Parkway, Baltimore, MD 21210, USA.

Pbx1, pbx2 and *pbx3* are almost identical in their amino acid sequence (92% to 94% identity in a protein region of 273 amino acids). In contrast to *pbx1,* however, *pbx2* and *pbx3* are not known to function as oncogenes. The *pbx1* protein is expressed in all tissues except in cells of the B and T cell lineage (Kamps et al., 1991). *pbx2* and *pbx3* are ubiquitously expressed (Monica et al., 1991).

The homeodomain of the *pbx* proteins is extremely well conserved and is closely related to the homeodomain of yeast *a1* and *α2* (Kamps et al., 1990; Nourse et al., 1990). However, *a1/α2* and the *pbx* proteins exhibit no sequence similarity outside of their homeodomains.

The biological function of *pbx* proteins is unknown. As a first approach to this question we have identified and characterized the *Drosophila melanogaster* homolog of the *pbx* family. The protein sequence of *Dpbx* is extremely well conserved and, most surprisingly, the high level of amino acid similarity extends well beyond the homeodomain and suggests important, conserved functions for other portions of the molecule. Interestingly, *Dpbx* mRNA is most prominently present in the anterior portion of the ventral nerve cord of the developing *Drosophila* embryo.

Results

Isolation and characterization of Dpbx

A lambda cDNA library derived from adult flies was screened using a fragment containing the *pbx1* homeobox region as a probe. One cDNA, designated *Dpbx,* was isolated and characterized further. We sequenced the *Dpbx* cDNA and compared the translation of its open reading frame to that of the mammalian *pbx* proteins (Figs. 1 and 2). There is a high degree of similarity between the homeodomains of *Dpbx* and the human *pbx* proteins (Fig. 2). Only three substitutions occurred, none are at positions known to be involved in DNA binding. Previously, it was shown that the homeodomain of the mammalian proteins is most closely related to that of yeast *a1* (Kamps et al., 1990; Nourse et al., 1990). Similarly, the homeodomain of *Dpbx* is also homologous to the yeast *a1* homeodomain (36% identity, Fig. 3).

The *Dpbx* and the human *pbx* proteins are virtually identical in a large portion of the protein flanking the homeodomain. There was about 85% identity between *Dpbx, pbx1, pbx2,* and *pbx3* in a region comprising 201 residues N-terminal of the homeodomain and between 77% and 92% identity in the amino acids C-terminal of the homeodomain (Table I). The overall similarity, including conservative changes, was 94% compared to *pbx1* and *pbx3,* and 92% compared to *pbx2.* The extreme termini of the mammalian *pbx* proteins are highly divergent (Monica et al., 1991). Likewise, the 37 N-terminal amino acids and the C-terminal amino acids beyond position 313 of the *Dpbx* sequence have no similarity to the mammalian *pbx. Pbx1, pbx2* and *pbx3* are almost as closely related to *Dpbx,* as they are to each other (Fig. 2). We attempted to identify other genes related to *Dpbx* by searching protein sequence data bases. However, no related proteins were found.

Expression of Dpbx mRNA during Drosophila development

To investigate the temporal pattern of *Dpbx* expression during embryonic development, we performed a Northern blot analysis. Poly(A) + RNA was extracted at various stages of development, blotted and probed with a *Dpbx* fragment. *Dpbx* transcripts could be detected between 0–2 h after oviposition (Fig. 4). Thus, it is likely that maternal *Dpbx* transcripts are present in the embryo. Two mRNAs of approximately 2.4 kb and 2.8 kb are present throughout the life cycle. During embryonic development the 2.4 kb transcript is predominant, whereas in adult flies both transcripts are present in equal amounts (Fig. 4).

Next, we examined the spatial distribution of *Dpbx* transcripts by in situ hybridization. Whole embryos were hybridized in situ using a digoxigenin-labeled anti-sense RNA probe made from *Dpbx* cDNA. Maternally contributed *Dpbx* mRNA is uniformly distributed in newly laid eggs (Fig. 5A). In the cellular blastoderm the message is present at a low level, uniformly distributed but enriched adjacent to the nuclei (Fig. 5B). After germband elongation, there is a high level of *Dpbx* message in all cells of the germband (Fig. 5C). At stage 11 the pattern of mRNA expression differs within the germband (Fig. 5D). Most of the message seems to be in the mesoderm and developing nervous system. After germ band retraction, mRNA is present in the ventral nerve cord (VNC) in a segmentally repeated manner, in the midgut and hindgut and in the stomodeum (Fig. 5E). Most prominent is the expression in the developing central nervous system. Within the segments of the VNC, *Dpbx* mRNA is expressed at higher levels in the anterior segments. At stage 15, *Dpbx* becomes clearly detectable in the brain (Fig. 5F). Transcripts are still present in the midgut and stomodeum. Interestingly, the VNC showed the same anterior-posterior asymmetry as in stage 13 (Fig. 5E), whilst the border of high level message was moved somewhat more to the anterior (Fig. 5F). During subsequent stages of embryonic development, the *Dpbx* message is clearly confined to the brain and the anterior part of the ventral nerve cord (Fig. 5G). The ventral view of stage 15 shows the asymmetry in message through the VNC (Fig. 5H). There are also several small clusters of segmentally repeated cells visible which express *Dpbx* at high levels. In the final stages of embryonic development, the dorsal view with the focal plane in the brain shows that the *Dpbx* message is predominantly located in the brain (Fig. 5I). In situ hybridization of imaginal discs of third instar larvae showed a uniform pattern of *Dpbx* message in all imaginal discs and in tissues giving rise to the central nervous system (data not shown).

Chromosomal localization of Dpbx

To determine the cytological location of the *Dpbx* gene, a *Dpbx* probe was hybridized to spreads of larval salivary gland polytene chromosomes. The probe hybridized exclusively to a single locus on the X chromosome (chromosome 1) at the proximal end of 14A (data not shown).

Discussion

A Drosophila gene, *Dpbx,* was identified and characterized that is closely related to the human *pbx* genes, *pbx* proteins are a family of proteins containing a homeodomain which is most similar to yeast *a1* and *α2.* One of the *pbx* proteins, *pbx1,* is involved in pre-B acute

lymphoblastic leukemias. However, the biological role of *pbx* proteins is unknown. The purpose of these experiments was to obtain a preliminary characterization of *pbx* in a genetically accessible system like *Drosophila melanogaster*.

The embryonic expression pattern of *Dpbx* is interesting. There is a uniform distribution of mRNA in the earliest stage when the transcript is maternally derived, and in the cellular blastoderm. Throughout the embryonic development *Dpbx* was transcribed in the germband and later at high levels in the ventral nerve cord. Within these tissues the distribution was initially uniform but became more pronounced in the anterior portions and was later confined to the anterior parts of the ventral nerve cord. The gradual loss of *Dpbx* transcript from the posterior portion of the ventral nerve cord is suggestive of the action of a repressor in *Dpbx* transcription. Possible candidates for such a repressor include the abdominal-A and/or abdominal-B proteins encoded by the Bithorax Complex.

Dpbx represents the first description of a homeobox gene in *Drosophila* whose expression is limited to the anterior section of the central nervous system at some stages of the embryonic development. The embryonic distribution pattern of *Dpbx* transcripts is not typical for homeobox genes. Homeobox genes are often segmentally expressed along the anterior-posterior axis of the early embryo (McGinnis and Krumlauf, 1992).

The sequence relationships between *Dpbx* and yeast *a1/α2* indicate related functions. The relationship between *Dpbx* and the mammalian *pbx* proteins is one of greater than 80% identity. For *Dpbx,* but not for *a1* and *α2,* the similarity of sequence extends far beyond the homeodomain (Fig. 2 and 3) Thus clearly, *Dpbx* qualifies to be the *Drosophila* homolog of the mammalian *pbx* proteins. The conserved regions outside the homeodomain have no similarity to any other known protein. But its exceptional conservation suggests an important function. Examination of the sequences of *Dpbx* and the mammalian *pbx* proteins did not show any region that contain motifs previously identified in transactivation domains, such as proline rich, glutamine rich, or acidic blob like motifs. However, *Dpbx* and the mammalian *pbx* proteins contain a stretch of eight alanines interupted by a serine (Fig. 2, amino acids 127–135). Poly-alanine stretches have been noted previously in transcriptional repressors (Licht et al., 1990; Monuki et al., 1990). We would like to consider the possibility that *Dpbx* and mammalian *pbx* proteins normally function as repressors. In pre-B ALL the repressor domain of *pbx* is replaced with an activator domain derived from *E2A*. The conversion of a repressor to an activator may be the crucial step converting a regulatory protein into an oncogene.

The extreme conservation of the *pbx* amino acid sequence is such that they must have been under exceptionally stringent evolutionary conservation and an obvious suggestion is that they have a common function. The *pbx* proteins are most closely related to yeast proteins *a1* and *α2. α2* interacts with *a1* in diploid cells to repress haploid-specific genes (Porter and Smith, 1986; Goutte and Johnson, 1988). In *α* cells, *α2* interacts with *MCM1* to turn off a-specific genes (Smith and Johnson, 1992). We suggest that *Dpbx* is involved in cell lineage specific gene expression, similar to that of its yeast homologs *a1* and *α2,* which are involved in the control of cell type determination. They act in different combinations establishing distinct patterns of gene expression. In mammals at least three *pbx* proteins are present that

are most closely related to *a1* and *a2*. It is conceivable that *Dpbx* similarly associates with related proteins to form different combinations of protein complexes in different tissues. Identification and characterization of *Drosophila* mutants in the *Dpbx* gene should allow us to gain more insight into the biological function of the *pbx* proteins.

Experimental Procedures

Molecular cloning of the Drosophila homolog of the human pbx genes

A lambda ZAPII cDNA library (polydT- and random-primed) derived from adult flies (0–4 days old; Stratagene, La Jolla, CA) was screened under low stringency conditions using a 616 nucleotide DNA probe containing the homeobox of the human *pbx1* gene (*Pvu*II-restriction enzyme fragment, nucleotides 1527–2153 in Fig. 3b, Kamps et al. 1990). Nitrocellulose filters (BA85; Schleicher & Schuell, Keene, NH) with denatured phage DNA (0.5 M NaOH/1.5 M NaC1 for 5 min; vacuum oven 80°C for 2 h) were prehybridized overnight at 41°C in 30% formamide, 5 × Denhardt's solution, 0.1% sodium dodecyl sulfate (SDS), and 5 × SSPE. The *pbx1* probe radiolabelled with [^{32}p]dCTP by random priming was hybridized for 72 h at 41°C in the same solution (10^6 cpm/ml). Low stringency washing was done for 20 min at 42°C in 2 × SST/0.5% SDS (1 × SST is 0.15 M NaC1, 2.5 mM EDTA, 5 mM Tris, pH 7.5) and for 15 min at 42°C in 1 × SST/0.5% SDS. The filters were autoradiographed for 20 h at −70°C with an enhancer screen. After two rounds of phage purification the *Eco*RI-insert of the phage was cloned into pBluescript SKII and shown to specifically hybridize to the probe by Southern blotting. The sequence of the cDNA insert was determined using the dideoxy chain termination method on double-stranded plasmid DNA (Sequenase II sequencing kit; United States Biochemical Co., Cleveland, OH).

A set of internally deleted plasmids was generated (double-stranded nested deletion kit; Pharmacia, Piscataway, N J) which allowed sequencing of the cDNA insert with independent overlapping clones in both directions for most of its length. DNA sequence data were processed using MacVector 3.5 software (International Biotechnologies, Inc., New Haven, CT) and programs described by Devereux et al. (1984) and Smith (1989).

Northern blotting, in situ hybridization, and chromosomal location

Polyadenylated (Poly(A) + RNA was isolated from *D. melanogaster* (PolyATract kit; Promega, Madison, WI) at various stages of development, electrophoretically fractionated on an agarose-formaldehyde gel and transferred to nitrocellulose. Hybridization with a 667 nucleotide fragment of *Dpbx* (*Cla*I/*Eco*RI-restriction enzyme fragment, nucleotides 453–1119 in Fig. 1), ^{32}p-labeled by random priming was done using the conditions described above. The filters were washed under high stringency conditions twice for 10 min at room temperature in 2 × SST/0.1% SDS and twice for 10 min at 55°C in 0.2 × SST/0.5% SDS. Autoradiography took 108 h.

The expression pattern of Dpbx mRNA in embryos and imaginal discs was determined by whole mount in situ hybridization using a digoxigenin-UTP-labeled anti-sense RNA probe of *Dpbx* (RNA labeling kit, Boehringer Mannheim Co., Indianapolis, IN) as described in Tautz and Pfeifle (1989) with modifications described in Jiang et al. (1991). After

phosphatase staining, the tissues were dehydrated in ethanol and mounted in Epon. Staging was done in accordance with the criteria of Campos-Ortega and Hartenstein (1985).

The chromosomal location was determined as described in Ashburner (1989). The *Dpbx* DNA was transcribed using biotinylated-deoxyuridine and hybridized to salivary gland polytene chromosomes prepared from third instar Canton S larvae. The probe was detected by alkaline phosphatase-labeled biotin with nitro blue tetrazolium and BCIP (X-Phos) as the substrate.

Acknowledgments

We thank James N. Kadonaga for providing the *Drosophila* cDNA library. W.A.F. was supported by research grant F1173 from the Deutsche Forschungsge-meinschaft. This work was supported by NIH grants CA54198 and GM41100, the Searle Family Trust, and the Council for Tobacco Research.

References

Ashburner M (1989) *Drosophila*: A Laboratory Handbook. Cold Spring Harbor Laboratory, Cold Spring Harbor, Chapter 27, 44 PP.

Campos-Ortega JA and Hartenstein V (1985) The Embryonic Development of *Drosophila melanogaster*. Springer, Berlin.

Devereux J, Haeberli P and Smithies O (1984) Nucleic Acids Res. 12, 387–395. [PubMed: 6546423]

Goutte C and Johnson AD (1988) Cell 52, 875–882. [PubMed: 3127056]

Henthorn P, Kiledjian M and Kadesch T (1990) Science 247, 467–470. [PubMed: 2105528]

Jiang J, Kosman D, Ip YT and Levine M (1991) Genes Dev. 5, 1881–1891. [PubMed: 1655572]

Kamps MP, Look AT and Baltimore D (1991) Genes Dev. 5, 358–368. [PubMed: 1672117]

Kamps MP, Murre C, Sun X. h. and Baltimore D (1990) Cell 60, 547–555. [PubMed: 1967983]

Licht JD, Grossel MJ, Figge J and Hansen UM (1990) Nature 346, 76–79. [PubMed: 2114551]

McGinnis W and Krumlauf R (1992) Cell 68, 283–302. [PubMed: 1346368]

Monica K, Galili N, Nourse J, Saltman D and Cleary ML (1991) Mol. Cell. Biol 11, 6149–6157. [PubMed: 1682799]

Monuki ES, Kuhn R, Weinmaster G, Trapp BD and Lemke G (1990) Science 249, 1300–1303. [PubMed: 1975954]

Nourse J, Mellentin JD, Galili N, Wilkinson J, Stanbridge E, Smith SD and Cleary ML (1990) Cell 60, 535–545. [PubMed: 1967982]

Porter SD and Smith M (1986) Nature 320, 766–768. [PubMed: 3517656]

Quong MW, Massari ME, Zwart R and Murre C (1993) Mol. Cell. Biol, 13, 792–800. [PubMed: 8423802]

Smith DL and Johnson AD (1992) Cell 68, 133–142. [PubMed: 1732062]

Smith DW (1988) Computer Applications in the BioSciences 4, 212. [PubMed: 3382998]

Scott MP, Tamkun JW and Hartzell GW III (1989) Biochem. Biophys. Acta 989, 12–48.

Tautz D and Pfeifle C (1989) Chromosoma 98, 81–85. [PubMed: 2476281]

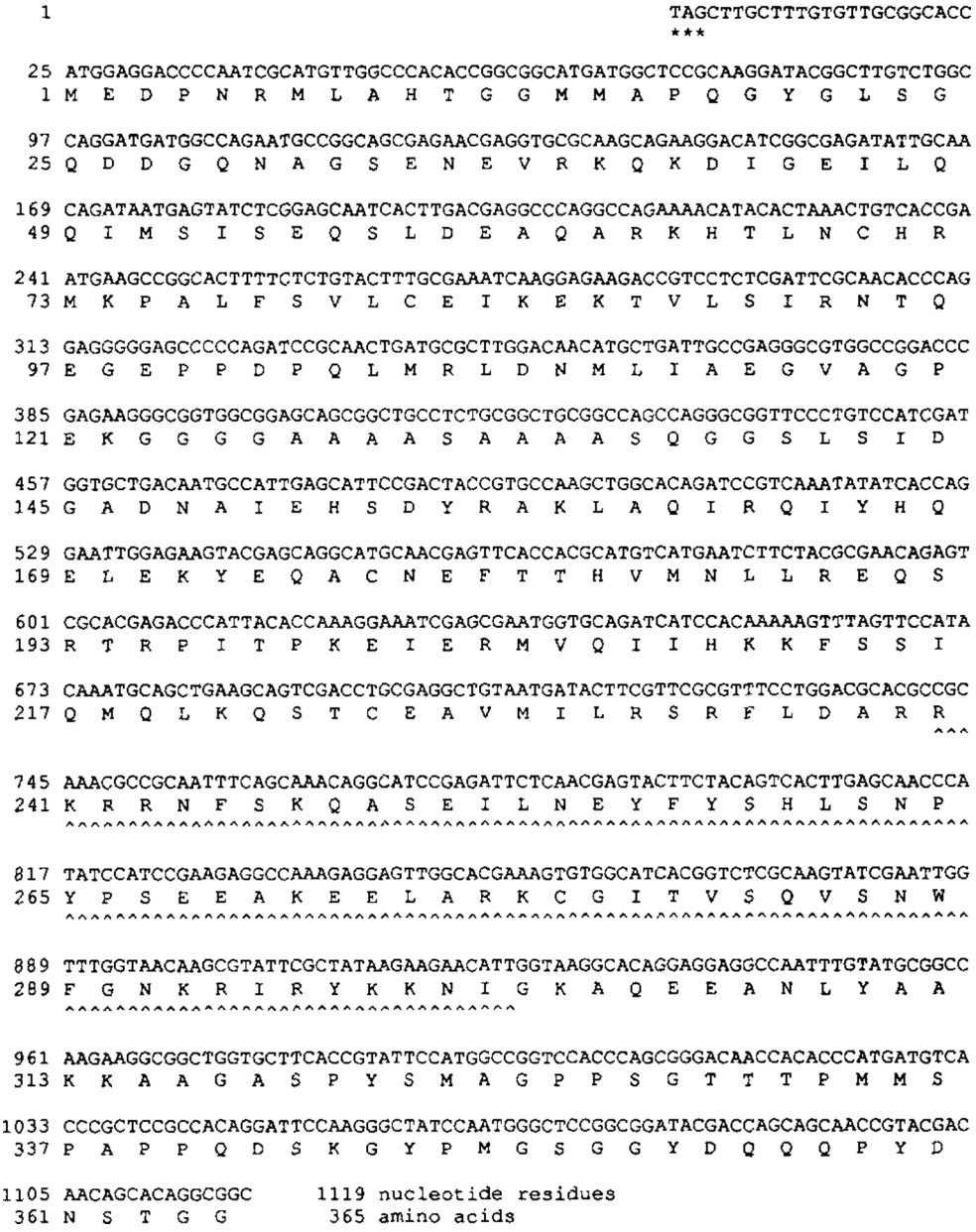

Fig. 1.

Partial nucleotide sequence of *Dpbx* cDNA and its predicted amino acid sequence. A stop codon upstream of the open reading frame and in-frame to the start codon is indicated by three asterisks. The predicted 61 amino acids that comprise the homeodomain are underlined by carets (numbers 240–300). The nucleotide and amino acid sequences are numbered at the left. The sequence data are deposited in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number Z18864.

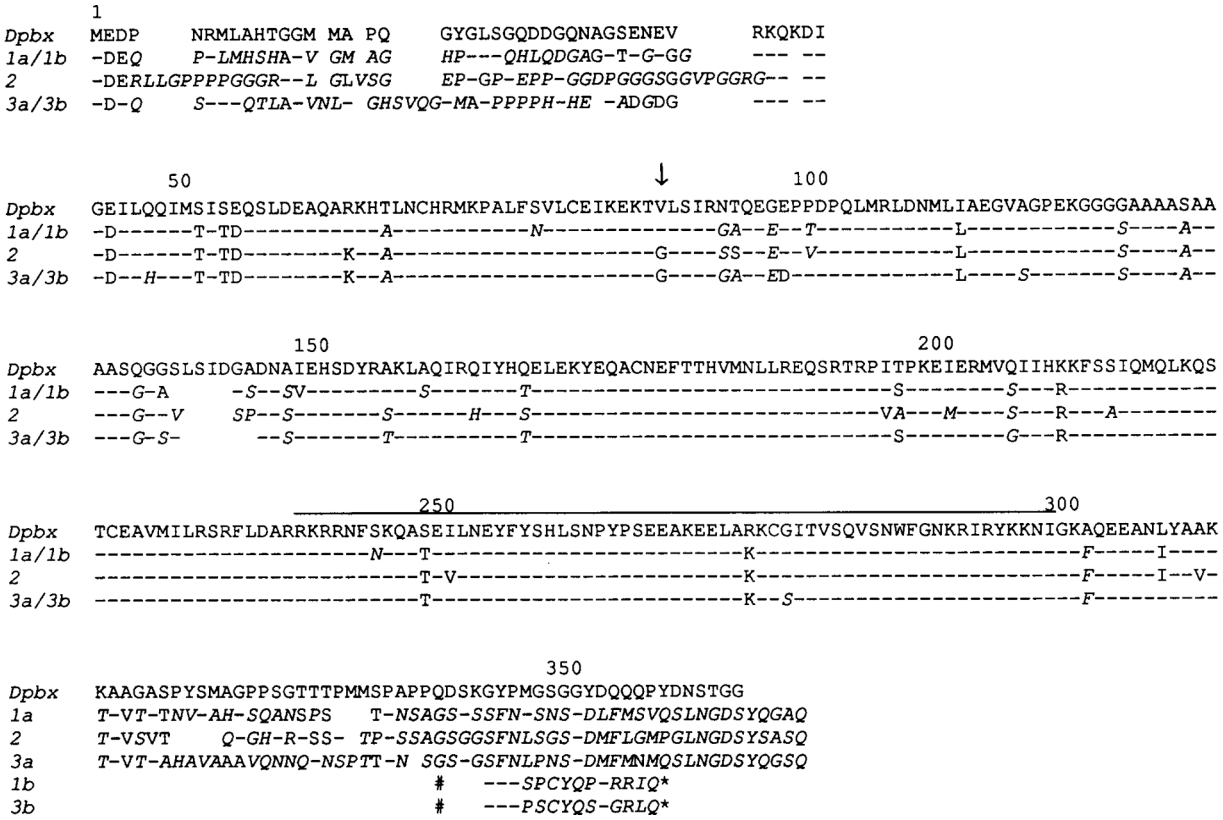

Fig. 2.

Alignment of *Dpbx* with human *pbx1a/1b, pbx2,* and *pbx3a/3b* amino acid sequences. The *Dpbx* sequence (upper line in single-letter code) is compared with that of the human *pbx* proteins (partial amino acid sequences for *pbx1a, pbx2,* and *pbx3a*; complete sequences for *pbx1b* and *pbx3b,* asterisks denote carboxy termini; Monica et al., 1991). *Pbx1a/1b* and *pbx3a/3b* are differentially spliced products of, the *pbx1* and *pbx3* gene and differ only in the carboxy terminal part of the sequences starting at the pound sign. Amino acid numbering corresponds to that shown in Fig. 1. Dashes denote amino acid identity with *Dpbx,* conservative substitutions are designated with plain letters, non-conservative substitutions are given in italics. Spaces allow for insertions or deletions to achieve optimal alignment. The homeodomain is overlined. The arrow indicates the breakpoint at which the carboxyterminal part of *pbx1* is fused to *E2A* forming the oncogene *E2Apbx.*

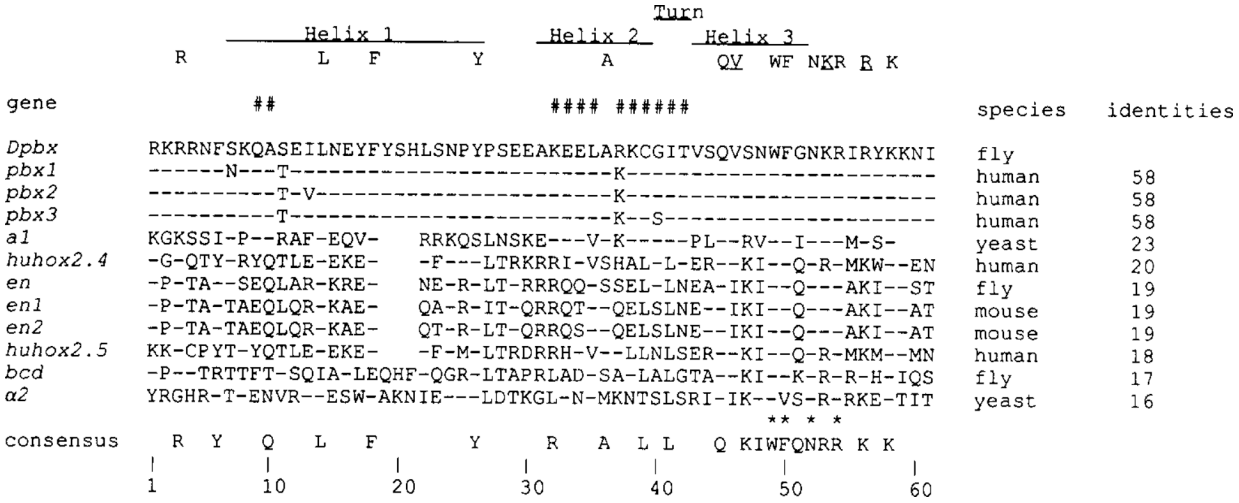

Fig. 3.

Comparison of eukaryotic homeodomains with the *pbx* homeodomain. Identities of amino acids with the *Dpbx* homeodomain are indicated by dashes, the number of identities is given at the right. A putative helix-turn-helix motif and the consensus sequences of eukaryotic homeodomains are shown (Scott et al., 1989). For comparison, the amino acids of the eukaryotic consensus sequence conserved in *pbx* are shown at the top. Underlined characters denote conservative substitutions. Four amino acids marked by asterisks (*) are invariant in all known non-yeast homeodomains and are also found in the *pbx* homeodomain. It may be noted that the amino acid differences between the *pbx* homeodomains do not occur at positions of the eukaryotic consensus. The highest amino acid similarity between *Dpbx* and the eukaryotic consensus is found in the helix 3 and just downstream of it, a protein region implicated in DNA binding. Number signs (#) indicate amino acids which are identical in *Dpbx* and yeast *a1* but not part of the eukaryotic consensus. They are mostly located in the helix 2, which is involved in DNA binding, and in the turn. A three amino acid gap has been introduced in the sequences of *a1, huhox2.4, huhox2.5, en, en1,* and *en2* to optimize sequence alignment.

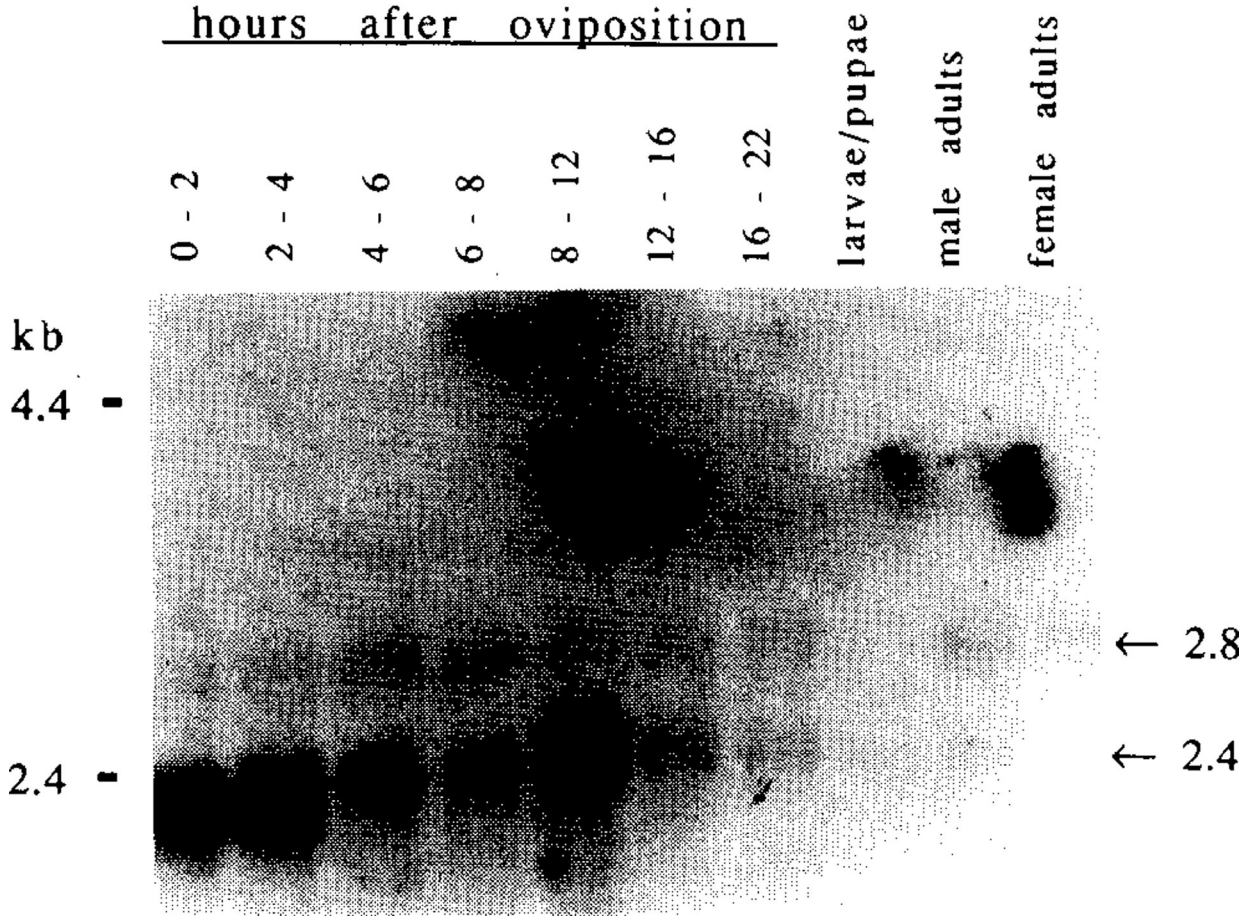

Fig. 4.

Developmental expression of *Dpbx* mRNA. A Northern blot using a random labeled *Dpbx* probe is shown. Embryonic development is divided into h after oviposition (0–2, 2–4, 4–6, 6–8, 8–12, 12–16, and 16–22 h). The other developmental stages are late larvae, early pupae (96–168 h), and male and female adult flies. Approximate RNA sizes are given on the left side. Two transcripts are detected (2.4 kb and 2.8 kb) indicated by arrows on the right side.

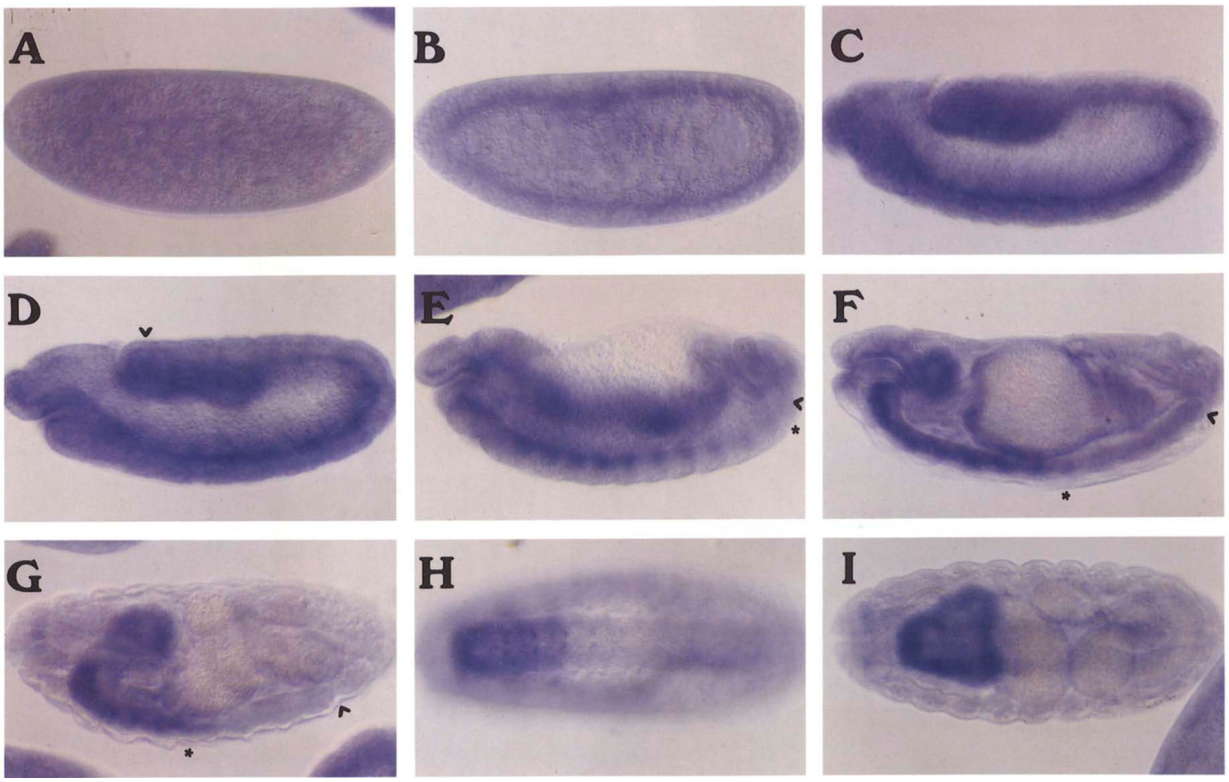

Fig. 5.

Expression pattern of *Dpbx* during embryonic development. In situ hybridization on whole-mount embryos was performed using digoxigenin-labeled anti-sense RNA of *Dpbx* and alkaline phosphatase immunocytochemistry. A purplish reaction product indicates the presence of *Dpbx* transcript. Lateral views of embryos at various stages are shown in panels A-G (approximate time after oviposition in parentheses) with anterior to the left and dorsal up: (A) stage 2 (1 h); (B) stage 4 (2 h); (C) stage 9 (4 h); (D) stage 11 (6 h); (E) stage 13 (10 h); (F) stage 15 (12 h); (G) stage 16 (13 h). (H) Ventral view of stage 15 with the focus in the ventral nerve cord (VNC). (I) Dorsal view of stage 16 with focus in the plane of the brain. The carets 0 indicate the posterior end of the germband or the VNC, the asterisks (*) indicate the approximate point within the VNC where the level of *Dpbx* message drops. This point, initially identical with the posterior end of the VNC, moves to the anterior between stage 11 and stage 16.

TABLE I

Comparison of *Dpbx* with human *pbx1, pbx2,* and *pbx3*

Genes	Similarity with *Dpbx*[a] in various amino acid regions				
	1–37[b]	38–239[b]	240–300	301–313	> 313[b]
pbx1	<30	88 (93)	95 (98)	84 (92)	<20
pbx2	<20	85 (90)	95 (100)	77 (92)	<20
pbx3	<30	88 (93)	95 (98)	92 (92)	<20

[a]Expressed as percentage of identities. Percentage of identical plus similar amino acids are given in parentheses. Amino acid numbers are shown in Fig. 2. Amino acids in position 240–300 comprise the homeodomain. The overall homology (identities and similarities, excluding spaces for optimal alignment) in 276 amino acids (position 38 to 313) was 94% (*pbx1, pbx3*) and 92% (*pbx2*).

[b]Excluding spaces for optimal alignment.