The molecular genetics of the bithorax complex of Drosophila: characterization of the products of the Abdominal-B domain

Susan E. Celniker, Dan J. Keelan, and E.B. Lewis

Division of Biology, California Institute of Technology, Pasadena, California 91125 USA

In Drosophila the Abdominal-B (Abd-B) domain of the bithorax complex specifies the identities of several posterior abdominal segments, comprises homeo-protein-coding regions and cis-regulatory regions, and extends from infra-abdominal-5 (iab-5) to iab-8, inclusive. Mutations that eliminate the Abd-B domain act as late embryonic lethals and result in transformations of posterior abdominal segments toward more anterior ones. The Abd-B domain gives rise to a minimum of five homeo-box-containing transcripts, 7.8, 4.7, 4.3, 3.7, and 3.3 kb in length. We examined the structure of the Abd-B domain by sequencing two Abd-B cDNA clones derived from the 4.3- and the 4.7-kb transcripts and the corresponding genomic DNA. The domain spans ~100 kb and contains at least eight exons. The 4.7- and 4.3-kb transcripts contain an open reading frame capable of encoding a 54-kD protein. A portion of the deduced protein-coding sequence common to all of the Abd-B transcripts was cloned into an expression vector. The resultant fusion protein then was used to derive a monoclonal antibody specific to Abd-B. By use of that antibody, we identified two embryonic Abd-B proteins, 54 and 36kD and determined the sum of their segmental distribution by immunohistochemical analysis of whole-mounted embryos and immunofluorescent analysis of dissected embryonic nervous systems. The proteins are distributed in the fourth to the ninth abdominal segments [parasegments (PS) 10–15] inclusive. Embryos homozygous for Polycomb (Pc) show labeling over almost the entire embryo, whereas embryos deficient for the Abd-B domain show no detectable labeling.

[Key Words: Embryonic development; bithorax complex; immunohistochemistry; DNA sequence; homeo box; Polycomb]

Received April 14, 1989; revised version accepted June 16, 1989.

A number of determinative events that control the early embryonic development of Drosophila are well characterized genetically. Two types of zygotically active genes are required: segmentation genes, that control the number and polarity of the body segments (Nüsslein-Volhard and Wieschaus 1980) and homeotic genes that control segmental identity. The latter are represented by genes of the Antennapedia complex [ANT-C] (Kaufman et al. 1980; Garber et al. 1983; Scott et al. 1983) and the bithorax complex [BX-C] (for review, see Duncan 1987). Certain segmentation genes appear to be required for the initiation of homeotic gene expression (Ingham and Martinez-Arias 1986; Ingham et al. 1986; White and Lehmann 1986; for review, see Akam 1987), whereas genes such as Polycomb [Pc] (Lewis 1978; Duncan and Lewis 1982) and extra sex combs [esc] (Struhl 1981) seem to be involved in maintaining the spatial pattern of homeotic gene expression once it is established (Struhl and Akam 1985; Wedeen et al. 1986).

The BX-C is essential for specifying the identity of a body region that extends from the posterior portion of the second thoracic segment to the anterior portion of the ninth abdominal segment. The BX-C is organized into three domains designated Ultrabithorax (Ubx), abdominal-A (abd-A), and Abdominal-B (Abd-B) (Sanchez-Herrero et al. 1985a, b; Tiong et al. 1985) spanning over 300 kb (Bender et al. 1983; Karch et al. 1985). In the BX-C there are three homeo box sequences, one in each domain (Regulski et al. 1985). The homeo box encodes a homeo domain that is a conserved sequence of 60 amino acids with DNA binding abilities (Laughon and Scott 1984; Shepherd et al. 1984, Desplan et al. 1985). The sequential expression of the BX-C genes and their cis-regulatory regions along the anterior/posterior body axis is...
colinear with their order in the chromosome (Lewis 1978). Each of the three homeo-box-containing transcription units is transcribed in a distal to proximal orientation.

The Abd-B domain occupies the distal one-third of the complex (from approximately +100 to +200 on the BX-C walk) and is required for specifying the identities of posterior abdominal segments. Mutations that inactivate all the functions of the Abd-B domain are designated Abd-B alleles (nomenclature of Duncan 1987). Mutations that inactivate specific subfunctions are designated infra-abdominal-5 [iab-5], iab-6, iab-7, or iab-8 (Karch et al. 1985). The iab-7 and iab-8 subfunctions also have been referred to as Abd-B'' and Abd-B', respectively (Casanova et al. 1986). The effects of these mutations in embryos are defined most readily in terms of a parasegmental unit (PS), which consists of the posterior portion of one body segment and the anterior portion of the next most posterior segment (Lawrence and Martinez-Arias 1985). Abd-B' function commences in PS 10 (i.e., the posterior portion of the fourth and anterior portion of the fifth abdominal segment) and extends to PS 15 (i.e., the posterior portion of the eighth and anterior portion of the ninth abdominal segment).

In situ hybridization to embryonic and imaginal tissues reveals that Abd-B transcripts are localized in regions where the Abd-B domain is required in development. Thus, Abd-B transcripts have been detected in the embryonic nervous system in PS 10–PS 14 (Harding et al. 1985; Levine et al. 1985; Regulski et al. 1985; Wedeen et al. 1986). We have shown that a probe derived from the 3' portion of an Abd-B cDNA detects transcripts in genital discs but not other imaginal discs (Celniker and Lewis 1987). Genital discs are the only imaginal discs found in the abdomen.

More recently, by use of a number of genomic DNA restriction fragments from the Abd-B region as probes, different sets of transcripts have been detected with differential domains of expression (Kuziora and McGinnis 1988; Sanchez-Herrero and Crosby 1988). The transcript sizes and corresponding spatial distribution are as follows: 4.3- and 4.7-kb transcripts localized to PS 10 to PS 15; 3.3-, 3.7-, and 7.8-kb transcripts localized to PS 14 and PS 15. A 4.1-kb cDNA, corresponding to a portion of the 7.8-kb transcript, has been sequenced by DeLorenzi et al. (1988) who show that it can encode a 30-kD or 36-kD hemo protein. This predicted protein would be responsible for proper development of PS 14 of the embryo and the genital disc and its adult derivatives.

Our goal has been to obtain antibodies to proteins from the Abd-B domain and to determine the spatial and temporal distribution of such proteins in the developing organism. We report here the results of the use of a monoclonal antibody to follow the pattern of protein expression in: (1) wild-type embryos; (2) embryos that lack entirely the Abd-B domain; and (3) embryos that overexpress the BX-C genes owing to the virtual lack of Polycomb* [Pc*], a negative trans-regulator of the BX-C.

### Results

#### Sequence analysis of Abd-B cDNA clones and the corresponding genomic DNA

To determine the structure of the Abd-B domain and transcripts, we isolated and sequenced embryonic and pupal cDNA clones. We described previously the isolation of cDNA clones from 3- to 12-hr embryonic and 5.5- to 7.5-day pupal libraries (Celniker and Lewis 1987). A fragment of genomic DNA, containing the Abd-B homeo box (Regulski et al. 1985) was used to isolate these cDNA clones. The clones hybridize to three genomic EcoRI restriction fragments (5013, 5014, and 5015) that span ~9 kb from about +150 to +159 on the BX-C map (Fig. 1A).

Results from our restriction-fragment analysis and sequencing of genomic and partial cDNA clones (P5 and P3) identify a transcription unit that contains at least five exons. The Abd-B sequence shown in Figure 2, includes the entire P5 cDNA sequence, 5' sequence from cDNA P3, and genomic sequence from the region that immediately flanks the 5' end of the cDNAs. The 5' exon (exon 4) is at least 888 bp, the three internal exons (exons 5, 6, and 7) are 212, 202, and 215 bp, and the last exon (exon 8) is 1913 bp (Fig. 1B). The donor and acceptor sequences are all consensus with the exception of the donor sequence following exon 5, which begins with a nonconsensus GGT (Table 1). Two of the introns (between exons 5 + 6 and 7 + 8) are micro introns that contain only 84 and 56 bp, respectively. The homeo box is encoded by portions of exons 7 and 8. Thus, the first 132 bp of the homeo box are at the 3' end of exon 7. The intron position found in the Abd-B homeo box is identical to that found in the homeo boxes of the labial gene (Mlodzik et al. 1988) and the distal-less gene (Cohen et al. 1989).

The sequence of the 3' end of cDNA clone P5 contains a stretch of 17 A residues. Thirty-six base pairs upstream of the poly(A) tail is a consensus polyadenylation signal sequence, AATAAA (Proudfoot and Brownlee 1976), which presumably corresponds to one of the 3' ends of Abd-B.

#### Developmental expression of Abd-B RNAs and their relation to the Abd-B cDNAs

We examined the temporal expression of Abd-B RNAs by use of a fragment from the 3' end of cDNA P5. This fragment was chosen because it does not contain the homeo box or M repeat and thus should only detect mRNAs transcribed from the Abd-B domain. We detect five RNAs 7.8, 4.7, 4.3, 3.7, and 3.3 kb in length. The 7.8-kb RNA is only weakly detectable between 4 to 10 hr of embryonic development (Fig. 3B). The 4.3- and 3.3-kb RNAs are expressed predominantly in 4- to 7-hr embryos but can be detected in later embryonic stages as well as in larvae, pupae, and adults (Fig. 3A and B).
4.7- and 3.7-kb RNAs are expressed late in embryonic development at 10–13 hr relative to the 4.3- and the 3.3-kb RNAs and do not appear to persist beyond embryogenesis.

These RNAs have been detected previously by use of genomic restriction fragments from the Abd-B domain [Kuziora and McGinnis 1988; Sanchez-Herrero and Crosby 1988]. The estimated sizes of the five transcripts vary within 200 bp but the temporal patterns of expression are in good agreement.

The cDNA clones that we characterized correspond to the 4.3-kb Abd-B transcripts [Fig. 3A] found by Northern analysis to hybridize specifically to the genomic region from +155.5 to +157 kb [Kuziora and McGinnis 1988; Sanchez-Herrero and Crosby 1988]. The last four exons [labeled 5–8, Fig. 1A] are common to all five Abd-B RNAs [Kuziora and McGinnis 1988; Sanchez-Herrero and Crosby 1988, S. Celniker, D. Keelan, and E. Lewis, unpubl.] and are known to encode the 3' portion of the 7.8-kb transcript as sequenced by DeLorenzi et al. [1988] (shown in Fig. 1C). The ~400-bp difference at the 3' ends of the 4.7- and 4.3-kb (4-kb set) and the 3.7- and 3.3-kb (3-kb set) transcripts is the result of alternate choices in the site of polyadenylation [data not shown]. The 3' end of the 7.8-kb transcripts has not been determined. We attribute the large differences in sizes of the transcripts (7.8, 4, and 3 kb) to differences at their 5' ends, although we have not carried out primer extensions. The 4-kb transcripts are not the product of differential splicing because no acceptor splice sites exist in front of exon 4 (S. Celniker and D. Keelan, unpubl.). As a result, the 7.8-kb transcript [Fig. 1C] splices exon 3 directly to exon 5 as shown by Delorenzi et al. [1988]. Kuziora and McGinnis [1988] used restriction mapping of cDNAs to group the Abd-B transcripts into four classes. They predicted that the 3-kb transcripts also splice exon 3 to exon 5, and we confirmed this by sequencing (S. Celniker and D. Keelan, unpubl.). The cDNAs we analyzed are similar in structure to their α class with the exception that, by sequencing, we mapped the presence of two additional exons (exons 6 and 7).

Analysis of the derived amino acid sequence

We identified a long continuous open reading frame [ORF] from nucleotide 513 to 2177 [Fig. 2]. There are three in-frame methionines in this ORF. Translation initiated at these methionines would result in proteins of 54,902, 43,877, and 30,198 kd. The codon usage [Staden 1984] of the ORF matches well with the preferred codon bias for Drosophila [K. Burtis, pers. comm.].

The protein sequences contain several hydrophilic regions as revealed by hydrophobicity plots [Hopp and Woods 1981], and there are no predicted membrane spanning domains. The isoelectric points for the predicted proteins are basic (9.5, 9.7, and 9.9), which results because of the high proportion of lysine, arginine, and histidine residues (14%). The predicted 54-kD protein also contains a large number of repetitive amino acid sequences. The amino half of the protein is glutamine-
Abd-B domain products

Figure 2. DNA sequence and predicted amino acid sequence of Abd-B cDNAs 3 and 5 and genomic sequences upstream of cDNA 3. The limits of all exons and cDNAs are marked by arrowheads. The homeo domain is shown in boldface type. The prospective genomic sequences correspond to the previously described M or opa (amino acids 386-445) of the Abd-B rich, containing two regions (amino acids 2-30 and 108-131, inclusive) predominantly (20 of 27 and 18 of 24) glutamine, in addition to interspersed small runs (4-8 amino acids) of glutamines (amino acids: 57-64, 76-81, 153-156, and 183-189, inclusive). These sequences correspond to the previously described M or opa repeats that are found in several hundred locations in the Drosophila genome (Wharton et al. 1985).

As reported previously (Regulski et al. 1985) Abd-B proteins contain a homeo domain. Our sequence data place this homeo domain near the carboxy-terminal end (amino acids 386-445) of the Abd-B proteins.
Table 1. Abd-B donor and acceptor splice-site sequences

<table>
<thead>
<tr>
<th>Exon splice</th>
<th>cDNA position</th>
<th>Donor sequence</th>
<th>Acceptor sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>4–5</td>
<td>1359</td>
<td>AG/GT^AGT</td>
<td>TTCCAG/A</td>
</tr>
<tr>
<td>5–6</td>
<td>1571</td>
<td>GGT/GTAAGT</td>
<td>CCCGAC/C</td>
</tr>
<tr>
<td>6–7</td>
<td>1773</td>
<td>CAG/GTGAGT</td>
<td>TTTCAG/G</td>
</tr>
<tr>
<td>7–8</td>
<td>1988</td>
<td>CAG/GTGAGT</td>
<td>TTTCAG/G</td>
</tr>
</tbody>
</table>

* Consensus-derived donor and acceptor sequences (Mount 1980).

**Abd-B** homeo domain is no more than 55% identical with any found in Drosophila. As a consequence, it has been placed in a separate class by Scott et al. (1988). The homeo domain is predicted to contain three helixes (Laughon and Scott 1984). The first two helixes of **Abd-B** show poor homology with those of other BX-C genes. The third helix of **Abd-B** (amino acids 425–436, inclusive, Fig. 2), the putative DNA-binding domain (Laughon and Scott 1984), however, differs from that of **Ubx** and **abd-A** only by substitution at position 429 (Fig. 2) of isoleucine for valine.

**Expression of Abd-B sequences in Escherichia coli**

To identify and localize the **Abd-B** protein products we generated **Abd-B**-specific monoclonal antibodies. We cloned a 312-bp Ncol restriction fragment from cDNA 5 (underlined sequence shown in Fig. 2) into an E. coli expression vector containing the trpE gene (Spindler et al. 1984). The Ncol fragment encodes an hydrophilic region of the putative **Abd-B** protein. Furthermore this fragment contains neither the homeo box nor any of the highly repetitive amino acid runs. The fusion protein was partially gel-purified and injected into mice to develop monoclonal antibodies. Three monoclonal antibodies were found to be specific for **Abd-B**: one, 1A2E9, has been used for most of our studies. Its specificity for the **Abd-B** portion of the fusion protein is shown in Figure 4.

**Abd-B nuclear embryonic proteins**

Because developmental Northern analysis (Fig. 3, Kuziora and McGinnis 1988; Sanchez-Herrero and Crosby 1988) shows that the **Abd-B** transcripts are abundant during embryogenesis and because the presence of a homeo box suggests a nuclear localization, we used embryonic nuclear extracts prepared by the method of Heberlein and Tjian (1988) to identify **Abd-B** proteins. The extracts were subjected to one-dimensional, sodium dodecyl sulfate (SDS) gel electrophoresis and transferred electrophoretically to nitrocellulose. Specific proteins were identified by immunochemistry. The 1A2E9 antibody consistently detects two proteins of M, 54 and 36 kD (Fig. 5E). The sizes of the observed protein products are consistent with the coding capacity of the **Abd-B** transcripts. The extracts also were subjected to a two-dimensional analysis. In the first dimension we employed a nonequilibrium pH-gradient gel because of the predicted basic nature of the proteins. In the second dimension we employed an SDS–polyacrylamide gel. At each relative mobility, we resolved families of proteins with subtly different molecular weights and isoelectric points (Fig. 6). The observed isoelectric points of the **Abd-B**

Figure 3. Developmental expression of Abd-B RNAs. Ten micrograms of poly(A)* RNA isolated from different developmental stages were electrophoresed on a 1% formaldehyde–agarose gel and transferred electrophoretically to a nylon membrane filter. (A) The filter-bound RNA was hybridized with a uniformly labeled probe synthesized from a 1.2-kb, 3' fragment of cDNA 5 (see text). The four transcripts detected at this exposure, 4.7, 4.3, 3.7, and 3.3 kb in length are indicated by the arrows. (B) The same filter shown in A exposed 5× longer. The 7.8-kb transcript weakly detectable in 4- to 7-hr and 7- to 10-hr embryos is indicated by the arrow. (C) The same filter shown in A, washed and rehybridized with a nick-translated, ribosomal protein clone (RP49).
proteins are more acidic than the predicted ones [data not shown]. Both the fushi tarazu (ftz) and engrailed (en) homeo proteins have been reported to be phosphorylated at threonine and serine residues (Gay et al. 1988; Krause et al. 1988). Presumably, the Abd-B proteins are phosphorylated also so that the observed charge differences among the Abd-B proteins may reflect different degrees of phosphorylation.

As a control to determine the quality of our extracts, Ubx proteins were visualized using a Ubx antibody [FP3.38]. Our extracts contained only the intact Ubx proteins all of which are known to cluster from 41 to 45 kD (Gavis and Hogness, pers. comm.) and no lower molecular weight Ubx degradation products [Fig. 5D].

**Embryonic localization of Abd-B proteins**

We used horseradish-peroxidase labeling to detect the segmental distribution of Abd-B proteins in whole mounts of embryos [Fig. 7] and in dissected embryonic nervous systems [Fig. 8]. Although Abd-B transcripts are detected by the stage of cellular blastoderm, we do not detect any Abd-B protein expression above background prior to germ-band extension; that is, mid to late stage 8 embryos [as defined by Campos-Ortega and Hartenstein 1985]. Abd-B proteins are visible first in a band of ectodermal cells at the posterior tip of the extended germ band as well as in cells of the proctodeal primordium [Fig. 7A]. The band of ectodermal cells derive from PS 13 and PS 14 as determined by double-labeling with an en antibody [data not shown]. By the time germ-band extension is completed, in a lateral view, we see labeling of the posterior ectoderm and mesoderm in PS 13 to PS 15, inclusive, and in the mesoderm, including the cells surrounding the tracheal pits, but not in the ectoderm of PS 11 and PS 12 [Fig. 7B and H]. In a dorsal view of an embryo at germ-band extension, the posterior border of Abd-B labeling does not include the proctodeal opening and the labeling in PS 12 is in a row of cells that borders the neurogenic region [Fig. 7C]. At germ-band extension, the Malpighian tubules start to grow out of two diverticula of the hindgut. These cells are labeled and can be seen in a dorsal view though at a different plane than that shown in Figure 7C. During germ-band retraction when the posterior spiracles form, we see labeling of the nuclei in a subset of cells in the sixth, seventh, and
eighth abdominal segments including the cells surrounding the posterior spiracles [Fig. 7D]. At germ-band retraction, in a ventral view, we detect strong labeling at the posterior end of the developing central nervous system (CNS). Abd-B expression is most abundant in the last neuromere, PS 14; diminishes in PS 12 and PS 13 and further tapers off more anteriorly [Fig. 7E, F, and I].

In a dorsal view, a subset of cells of the visceral mesoderm surrounding the hindgut are labeled [Fig. 7F and G]. Also, we detect label in a subset of cells of the Malpighian tubules [data not shown]. Double-labeling with DAPI and rhodamine-conjugated secondary antibodies reveals that the staining is nuclear [data not shown].

As a positive control for these experiments, wild-type embryos were stained with Ubx antibodies [Fig. 7J]. Ubx expression, as reported previously (White and Wilcox 1984; Beachy et al. 1985), is first seen in PS 5, peaks in PS 6, and then tapers off but is still detectable in PS 7–PS 12.

Also, we investigated the distribution of Abd-B proteins in homozygous Pc embryos. These animals die in late embryogenesis, and the morphology suggests that most of the body segments are transformed towards the eighth abdominal segment (PS 13). At germ-band extension in homozygous Pc embryos we detect Abd-B expression commencing weakly in PS 5, extending at moderate levels from PS 6 to PS 12 and at nearly normal levels in PS 13 and PS 14. A dorsal view of the expression in PS 9 to PS 14 is shown in Figure 7K. By germ-band extension, Abd-B expression can be detected in all the neuromeres of the CNS and in the ectoderm of all the segments [Fig. 7L].

Although labeling with 1A2E9 is seen in the CNS, the mesoderm and the epidermis, the most prominently labeled structure in embryos is the CNS. For this reason we dissected the CNS of wild-type and mutant embryos to determine more accurately the limits of Abd-B expression. As expected, 1A2E9 is localized to neuromeres 10 to 14, inclusive, [posterior fourth abdominal to anterior ninth abdominal segments, inclusive] of the CNS [Fig. 8A]. The labeling pattern is complex. There is a gradient of expression, weakly detectable in the tenth neuromere which becomes more intense in each more posterior neuromere with the highest expression in the most posterior neuromere, neuromere 14. The expression, in each neuromere, has at least two components: (1) it is bilaterally symmetric as was observed for Ubx (White and Wilcox 1984; Beachy et al. 1985), and (2) the nuclei in the anterior portion of each neuromere are labeled more intensely than those in the posterior portion resulting in a striped appearance.

Embryos deficient for the distal end of the complex [Df P9/Df C4] including all of the Abd-B domain provide a negative control. They are identifiable because they have an extra segment and were found to have only background levels of antibody staining [Fig. 8B]. Thus, the lack of antigen in Abd-B~ embryos shows that the 1A2E9 antibody is specific for the Abd-B proteins and confirms that the ORF we identified corresponds to that of the Abd-B proteins.

**Figure 6.** Distribution of embryonic nuclear Abd-B protein products as a function of molecular weight and relative isoelectric point. Total nuclear proteins from 4- to 24-hr embryos were electrophoresed in the first dimension by NEPHGE with a gradient of pH 6–10, and in the second dimension by a gradient of acrylamide from 5 to 15%. The proteins were transferred electrophoretically to nitrocellulose and incubated with Abd-B monoclonal antibody. The large arrows indicate the direction of electrophoresis, whereas the small arrows identify two sets of protein products. Size standards are indicated on the left.

**Discussion**

**Genomic organization**

Among the homeotic genes that have been characterized, the genomic organization of the Abd-B domain is one of the most complex. It spans some 100 kb, contains at least eight exons, and produces five major transcripts that differ in structure as well as in their molecular lengths (7.8, 4.7, 4.3, 3.7, and 3.3 kb) [Kuziora and McGinnis 1988; Sanchez-Herrero and Crosby 1988, this paper]. We presume that the 7.8-kb transcript arises from a distal promoter (P1), the 3-kb transcripts from an internal promoter (P2), and the 4-kb transcripts from the most proximal promoter (P3) [Fig. 9]. The Abd-B transcription units are contained within a 45-kb region. On the basis of genetic data, the remaining 50 kb that lie 3' to the transcription units and correspond to iab-5, iab-6, and iab-7, seem to be required for proper spatial expression of the P3 transcripts [Duncan 1987].

Transcripts from the P1 and P2 promoters contain a coding region that produces a 30-kD homeo protein [or a 36-kD, if the protein is initiated at a GUG as proposed by Delorenzi et al. 1988], whereas transcripts from the P3 promoter contain a coding region that produces a 54-kD homeo protein. The predicted small homeo protein, whether 30 kD or 36 kD, lacks 221 amino acids. This stretch of amino acids forms the amino-terminal end of the 54-kD homeo protein and includes multiple copies of the M repeat.

Akam et al. [1988] suggested that the ancient ancestor
Abd-B domain products

Figure 7. (See following page for legend.)
of the myriapod-insect lineage contained a labial-like and an Abd-B-like homeotic gene. We find that Abd-B has a splice site in the recognition helix that is in exactly the same position as the splice site in the labial recognition helix. This identity of splice sites suggests that the labial and Abd-B homeo domains derive from a common ancestral homeo domain that may have predated the myriapod-insect lineage.

**Abd-B protein expression**

Embryos that are homozygous deficient for the Abd-B domain (Df-C4/Df P9) have PS 10 to PS 15, inclusive, transformed toward more anterior parasegments. Presumably this phenotype results from the absence of Abd-B expression. Therefore, in wild type we would expect to see Abd-B expression throughout these six parasegments. First, we detect Abd-B expression in PS 13 and PS 14 of early germ-band-extended animals. By the time the intersegmental grooves form, we see ectodermal and mesodermal tissues of PS 13 to PS 15 labeling, as well as mesodermal labeling of the more anterior parasegments PS 11 and PS 12. Only in the CNS of germ-band-retracted embryos do we see labeling of PS 10. Late in embryogenesis we detect labeling of the visceral mesoderm surrounding the hindgut and in two of the four Malpighian tubules. The gene caudal, which lies outside the BX-C in chromosome 2 and is expressed in more posterior structures of the embryo than is Abd-B, also is reported to be expressed in the Malpighian tubules [MacDonald and Struhl 1986]. It will be interesting to see if the expression of these two homeo domain proteins is mutually exclusive with respect to these tubules.

With our Abd-B antibody, IA2E9, we detect two embryonic nuclear proteins, 54 kD and 36 kD in size. Because the 4-kb Abd-B transcripts are primarily restricted spatially to PS 10–13 [with possibly lower expression in PS 14 and PS 15] and the 7.8- and 3-kb Abd-B transcripts are restricted spatially to PS 14 and PS 15 [Kuziora and McGinnis 1988; Sanchez-Herrero and Crosby 1988], we expect that the 54-kD and the 36-kD Abd-B proteins will be correspondingly restricted spatially. Our antibody detects only the sum of the distribution of both proteins since it recognizes an epitope common to both proteins. To address the issue of spatial distribution of these proteins directly would require the development of antibodies specific to each. However, we can determine whether the two proteins are differentially expressed by studying various mutants that affect development of the posterior abdominal region. Thus, we expect strong iab-8 mutants, such as 1065 (Fig. 9) that disrupt the P1 and P2 promoters, to be unable to make the 36-kD protein but still retain the ability to make the 54-kD protein. We would expect certain iab-7 mutants, such as D14 (Fig. 9), to be unable to make the 54-kD protein but retain the ability to make the 36-kD protein. Currently, we are testing these predictions using the IA2E9 antibody.

**The role of Abd-B in gene regulation**

Proteins with homeo domains of the Abd-B type presumably exert a master regulatory function by binding to specific DNA motifs in other genes thereby activating or repressing the transcription of such genes, thus, the homeo-domain-containing protein Ubx has been shown to mediate its regulatory effects by binding to a DNA motif composed of a fivefold tandemly repeated TAA motif [Beachy et al. 1988]. Although the homeo domains of Ubx and Abd-B have diverged considerably in the first and second helices, they differ as already noted in only

---

**Figure 7. Expression of Abd-B proteins in embryos.** Embryos were dechorionated, the vitelline membrane removed, fixed, and incubated with Abd-B antibody 1A2E9 and proteins were visualized by use of horseradish-peroxidase-conjugated secondary antibodies. The anterior end of the embryos is always to the left. [A, B, F, G, H, and I] Lateral views, [C, D, and K] dorsal views, and [E, J, and L] ventral views of optical sections using Nomarski optics. [A] An embryo that has completed the fast period of germ-band elongation (late stage 8). Abd-B is expressed in a band of 12–17 ectodermal cells starting at the tip of the extended germ band. It is expressed weakly in the proctodeal primordium but not the posterior midgut primordium. [B] A fully extended germ-band embryo (stage 12). Strong ectodermal and mesodermal Abd-B expression in PS 13, PS 14, and PS 15, with only mesodermal expression in PS 11 and PS 12 is indicated by the arrow. [C] Dorsal view of an embryo the same age as that shown in B. Labeling is seen in the ectoderm of PS 13, PS 14, and PS 15. The posterior border of the label corresponds to cells surrounding the proctodeal opening. The heaviest labeling occurs in the epidermal groove between PS 13 and PS 14. [D] Dorsal view of an embryo in the fast period of germ-band shortening (late stage 12). The tip of the retracting germ band is at about 20% egg length (EL). Strong Abd-B expression is detected in cells surrounding the differentiating posterior spiracles, and lesser Abd-B expression is seen more anteriorly. [E] Ventral view of an embryo that has completed germ-band shortening (stage 14). Strong Abd-B expression is detected in neuromere 14 of the developing CNS with a moderate level of expression detected in neuromeres 11 to 13. [F] Lateral view of an embryo the same age as that shown in E. Strong Abd-B expression is seen in the dorsal epidermis, the posterior spiracles, the posterior neuromeres of the developing CNS, and the visceral mesoderm surrounding the hindgut; no expression is detectable in the anal pads. [G] Higher magnification of the visceral mesoderm surrounding the hindgut of the embryo in F. Abd-B is expressed in lines of cells outside the hindgut. [H] Higher magnification of PS 11 to PS 14, inclusive, of the embryo shown in B. The tracheal pits of PS 11 and PS 12 as well as the mesoderm are stained weakly. [I] Ventral view of a wild-type embryo in stage 14 labeled with Ubx antibody (Fig. 3.38). Nuclei of the CNS are labeled clearly in neuromeres 5–12 with the peak in neuromere 6. [K] Dorsal view of a PsP5 homeozygous embryo (stage 12), the same age and view as the wild-type embryo shown in C. The level of expression of Abd-B in PS 13 and PS 14 appears normal, whereas expression in PS 9–PS 12 inclusive, is elevated uniformly. [L] Ventral view of a PsP5 homeozygous embryo (stage 14), the same age and view as the wild-type embryo shown in E. Abd-B proteins are expressed in all neuromeres of the developing CNS. The highest level of expression is in the most posterior neuromere, with the more anterior neuromeres showing an intermediate level of expression.
or whether they would both bind to Ubx from that of not clear whether AbJ-S would bind to a different motif significant sequence homologies using dot-matrix analyses. Abd-B and both contain stretches of involved in regulating two classes of genes in addition to the possibility that they are autoregulatory. One class would be other homeo-box-containing genes such as Ubx and Abd-B both contain stretches of repeated amino acids, glycine and glutamine, respectively, in roughly the same relative position to the homeo domain. We anticipate that the Abd-B proteins will be involved in regulating two classes of genes in addition to the possibility that they are autoregulatory. One class would be other homeo-box-containing genes such as Ubx and Antp whose expression has been shown to be altered in animals lacking the Abd-B domain [Struhl and White 1985; Casanova and White 1987; Duncan 1988]. The second class would be a specific set of downstream genes that are required for the formation of structures unique to the posterior segments.

Trans-regulation of Abd-B by polycomb

Pc+ acts as a negative trans-regulator of the genes of the BX-C. Embryos homozygous for the Pc mutation show a transformation of most of the body segments toward A8 (Lewis 1978). In germ-band-extended embryos that are homozygous for Pc3, Abd-B expression is highest in neuromere 14, and moderate in all of the more anterior neuromeres. Expression also is elevated in the epidermis of every segment. This pattern of Abd-B expression in late Pc embryos is consistent with the Abd-B transcript distribution previously reported for such animals [Kuziora and McGinnis 1988].

Figure 8. Localization of Abd-B proteins in the CNS of wild-type and Abd-B- embryos. (A) CNS dissected from a wild-type embryo [stage 14] incubated with Abd-B antibody [1A2E9] followed by FITC-conjugated, goat anti-mouse immunoglobulin, photographed by fluorescence microscopy. Labeling can be detected weakly in the 10th neuromere, increasing in intensity posteriorly with the most intense labeling in the 14th neuromere. (B) Dissected CNS is shown from an Abd-B- [Df C4/Df P9] embryo labeled and photographed as in A. No labeling is detected above background.

one amino acid in the third, or recognition, helix. It is not clear whether Abd-B would bind to a different motif from that of Ubx or whether they would both bind to the same motif but with different affinities. Outside of the homeo domain we have not been able to detect significant sequence homologies using dot-matrix analyses. However, Ubx and Abd-B contain stretches of repeated amino acids, glycine and glutamine, respectively, in roughly the same relative position to the homeo domain. We anticipate that the Abd-B proteins will be involved in regulating two classes of genes in addition to the possibility that they are autoregulatory. One class would be other homeo-box-containing genes such as Ubx and Antp whose expression has been shown to be altered in animals lacking the Abd-B domain [Struhl and White 1985; Casanova and White 1987; Duncan 1988]. The second class would be a specific set of downstream genes that are required for the formation of structures unique to the posterior segments.

Embryos homozygous for Pc3, Abd-B expression is highest in neuromere 14, and moderate in all of the more anterior neuromeres. Expression also is elevated in the epidermis of every segment. This pattern of Abd-B expression in late Pc embryos is consistent with the Abd-B tran

Vertebrate Abd-B homologs

The Abd-B homeo domain is represented in vertebrates by well-conserved homologs. It is greater than 70% identical with murine homeo domains: Hox-1.7 [Rubin et al. 1987], Hox-3.2 [Breier et al. 1988], Hox-5.2 [Duboule and Dolle 1989] and a human homeo domain: HOX 2.5 [Boncinelli et al. 1988]. Indeed, the Abd-B homeo domain is more similar to such vertebrate domains than it is to any homeo domain in Drosophila, even those of the separate clusters, ANT-C and BX-C. Moreover, in the recognition helix of the homeo domain, the amino acid sequence of Abd-B is identical with that of its murine and human homologs. Just as Abd-B is expressed in the more posterior body segments of Drosophila so its homolog in the mouse is expressed in the more posterior somites of the spinal cord [Rubin et al. 1987]. The colinearity of gene order in the chromosome and spatial expression along the antero-posterior (rostral-caudal) axis, first found in Drosophila, occurs in an even more remarkable fashion in mice [Gaunt et al. 1988] and in human beings [Boncinelli et al. 1988] where the ANT-C-like and the BX-C-like genes are united in a single cluster.

Materials and methods

Nucleic acids preparation and DNA sequence analysis

Isolation of genomic DNA from the distal end of the BX-C has been described previously by Karch et al. (1985). For our studies, we subcloned the EcoRI restriction fragments from +150 to +200 into pUC18. We described previously the isolation of cDNA 3 and cDNA 5 (Celniker and Lewis 1987). Additional cDNAs were obtained by screening an embryonic library [kindly provided by Kai Zinn] with purified genomic restriction fragments from +150 to +190 of the BX-C walk.

Genomic and cDNA sequences were generated according to Sanger et al. (1980) using 35S in place of 32P. Ambiguities or polymerase stops were resequenced according to Maxam and Gilbert (1980). Gel-purified genomic DNA fragments were cleaved with Sau3AI and cloned into the BamHI site of M13 double-stranded DNA. Only cDNA-cross-hybridizing genomic clones were sequenced. cDNA 5 was sequenced on both strands using convenient restriction sites and subcloning into either M13mp18 or M13mp19. cDNA 3 was sequenced as described for cDNA 5 but sequenced only on one strand. Dot-matrix analysis was performed with the DNA Inspector IIe [Textco].

Construction of trpE hybrid gene fusion

pATH 1, a plasmid vector used for the construction of trpE hybrid gene fusions was obtained from T.J. Koerner (Hardy and Strauss 1988). By use of standard cloning techniques [Maniatis et al. 1982], restriction fragments from cDNA 5 [Celniker and
Figure 9. Putative transcript organization of the Abd-B domain. The different exons are shown (open boxes). The ORF is indicated (filled boxes). The homeo box (H) spans exons 7 and 8. The M repeat (M) is located in exon 4. Three putative promoters are indicated P1, P2, and P3. The large intron is not drawn to scale, but the size is indicated in kilobases. Mutant breakpoints are designated with arrows and the D14 deletion indicated by the bracket.

Lewis 1987) were filled-in, low-melting-point gel-purified, and ligated into the filled-in EcoRI site of pATH 1. Eight plasmids containing the inserts were identified by colony hybridization (Grunstein and Hogness 1975). After screening by restriction-fragment analysis, three were found to have the proper orientation. The insertion junction of one of these (clone p1Abdb) was sequenced (Maxam and Gilbert 1980) to confirm the reading frame.

Expression and purification of trpE fusion protein

Cultures of MCI061 cells containing the hybrid gene fusion constructs were grown overnight in M9CA (Maniatis et al. 1982) containing 50 μg/ml ampicillin and 10 μg/ml tryptophan at 37°C. The cultures were diluted 1:10 in M9CA containing ampicillin and grown for 1 hr at 30°C. 3 β-indoleacrylic acid (Sigma, 10 mg/ml in 100% EtOH) then was added at a concentration of 5 μg/ml, and the cultures were allowed to grow for an additional 2 hr at 30°C. The cells were pelleted at 5000g for 10 min at 4°C. The supernatant was drained carefully and the pellet was resuspended in cracking buffer [0.01 M sodium phosphate (pH 7.2), 1% β-mercaptoethanol, 1% SDS, 6 M urea]. The mixture was heated at 37°C for 2 hr to solubilize the protein. The yield of fusion protein was estimated to be 0.2–1.0 mg/ml.

Preparation of monoclonal antibodies

Fusion protein was semipurified before injection. From 1 to 5 ml of whole-cell lysate was electrophoresed on a 12.5%, discontinuous, SDS–polyacrylamide gel and visualized by placing the gel in 0.25 M KCl at 4°C for 10 min. The fusion protein band was excised, placed in dialysis tubing, and electroeluted at 30 mA overnight. After extensive dialysis against distilled H2O, containing phenylmethylsulfonyl fluoride (PMSF) at 17.4 μg/ml, the protein was lyophilized. The lyophilized protein was resuspended in 200–400 μl of H2O. Robertsonian mice were immunized with 20–60 μg of fusion protein intraperitoneally in complete Ribi adjuvant. They were boosted two times at three week intervals and again 8 weeks later. A final boost was given after an additional 2 weeks; 4 days later, the spleen was removed for fusion with FOX-NY myeloma cells. A total of 6.24 × 10⁸ spleen cells was fused with 3.12 × 10⁸ myeloma cells. Then 3 × 10⁷ cells per well of a 96-well microtiter dish were plated in RPMI containing 10% fetal calf serum, 10 mM HEPES, 2 mM glutamate, 1 mM sodium pyruvate, and 10 U/ml penicillin and streptomycin. The 420 clones were screened on protein dot blots against both semipurified fusion protein and semipurified trpE protein. The 35 dot-blot-positive clones were rescreened on Western protein blots, three positives of which were cloned subsequently by limited dilution. Culture supernatants were used for immunological assays.

Protein blots

Fusion and 4- to 20-hr embryonic nuclear proteins were isolated in the presence of PMSF [prepared as described by Heberlein and Tjian 1988], were separated in one dimension by SDS–gel electrophoresis (Laemmli 1970) or in two dimensions by non-equilibrium pH gel electrophoresis (NEPHGE) combined with SDS gradient gel electrophoresis (O'Farrell et al. 1977). The NEPHGE gels (pH range: 6–10) were run for ~2000 V-hr followed by a 6-hr run on a 5–15% SDS gradient gel at 30 mA. In both cases, the proteins were transferred electrophoretically to nitrocellulose filters (20 volts overnight in a 0.01% boric acid, 0.015% sodium borate buffer). The filters were blocked for 1–4 hr in 10% fetal calf serum, 0.05% Tween-20, 10 mM Tris (pH 7.5), 129 mM NaCl, 5 mM KCl, 1 mM EGTA, 2% BSA, and 0.03% sodium azide and then incubated with primary antibody.
for 3–15 hr. Then the blots were washed in three changes of gelatin buffer (0.5% gelatin, 12 mM Tris (pH 7.5), 0.75 M NaCl, 5 mM EDTA, and 0.03% sodium azide) and incubated with alkaline phosphatase-conjugated, goat anti-mouse secondary antibodies for 1 hr. The blots were washed again in gelatin buffer and the antigen was localized using a solution containing color development substrates, according to the Promega Protoblott system.

**Immunolocalization methods**

Staged embryos were dechorionated in 50% bleach, washed extensively in 0.02% Triton X-100, and deventrillized using biphasic extraction essentially as described by Mitchison and Sedat (1983). The primary antibody for Abd-B was a mouse monoclonal (1A2E9); Ubx was a mouse monoclonal (White and Wilcox 1985). Horseradish peroxidase was from Cooper Biochemical. Secondary antibodies conjugated to fluorescein isothiocyanate (FITC) and rhodamine were from Cappel. For immunoperoxidase or immunooalkaline phosphatase labeling, we obtained the appropriate ABC kits from Vectastain. Nuclei were stained using DAPI at 1 μg/ml before equilibrating the embryos in the anti-quenching PDA glycerol mounting solution. Immunohistochemically stained embryos were dehydrated in a graded ethanol series, cleared in xylene, mounted in Permount and photographed using Nomarski interference contrast microscopy.

**Acknowledgments**

We are indebted to Susan Ou, of the Caltech Monoclonal Facility, under the direction of Paul Patterson, for producing excellent Abd-B monoclonal antibodies. We thank John Pollock for assistance with the CNS dissections, Seymour Benzer for providing the CNS, and Donn Ballinger, Charles Rozek, Elizabeth Gavis, Welcome Bender, and Howard Lipshitz for helpful discussions, Richard Mann and David Hogness for sending the Ubx antibody (originally from R. White and M. Wilcox), Marianne Bienz for communicating sequence data on the iab-8 cDNA pH189 in advance of publication, Scott Rosenberg for sequencing most of cDNA 5; Tom Kornberg for providing the αt antibody; Joan Roach for technical assistance; and Charles Rozek and Joanne Topol for commenting on the manuscript. This work was supported by research grants to E.B.L. from the National Institutes of Health (HD-06331), and PR-07003 Program Project Grant (NIH GM-40499), the American Cancer Society (000543), the March of Dimes (1-10229) from the National Institutes of Health (HD-06331), and the Lucille P. Markey Charitable Trust.

**References**


Celniker et al.


Note

Sequence data described in this paper have been submitted to the EMBL/GenBank Data Libraries under accession number X16134 Abd-B P3.
The molecular genetics of the bithorax complex of Drosophila: characterization of the products of the Abdominal-B domain.

S E Celniker, D J Keelan and E B Lewis

Genes Dev. 1989, 3:
Access the most recent version at doi:10.1101/gad.3.9.1424

References
This article cites 59 articles, 16 of which can be accessed free at:
http://genesdev.cshlp.org/content/3/9/1424.full.html#ref-list-1

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.