



Structure and expression of a family of Ultrabithorax mRNAs generated by alternative splicing and polyadenylation in *Drosophila*.

K Kornfeld, R B Saint, P A Beachy, et al.

Genes Dev. 1989 3: 243-258

Access the most recent version at doi:[10.1101/gad.3.2.243](https://doi.org/10.1101/gad.3.2.243)

References

This article cites 66 articles, 15 of which can be accessed free at:
<http://genesdev.cshlp.org/content/3/2/243.refs.html>

Article cited in:

<http://genesdev.cshlp.org/content/3/2/243#related-urls>

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](#)

To subscribe to *Genes & Development* go to:
<http://genesdev.cshlp.org/subscriptions>

Structure and expression of a family of *Ultrabithorax* mRNAs generated by alternative splicing and polyadenylation in *Drosophila*

Kerry Kornfeld, Robert B. Saint,¹ Philip A. Beachy,² Peter J. Harte,³ Debra A. Peattie,⁴ and David S. Hogness

Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305 USA

The 77-kb primary transcript of the homeotic *Ultrabithorax* (*Ubx*) gene is alternatively spliced to yield at least five different coding regions. Each is restricted to either a 3.2- or a 4.3-kb size class generated by alternative polyadenylation. The pathways for splicing and polyadenylation are therefore coordinately regulated, and because the relative abundance of the respective mRNAs varies throughout development, these pathways also appear to be developmentally regulated. Translation of these mRNAs yields a family of *Ubx* proteins characterized by constant amino- and carboxy-proximal regions of 247 and 99 amino acid residues, respectively. Members of this family are distinguished by a short variable region that links the constant regions and consists of different combinations of three optional elements of 9, 17, and 17 residues. Only four amino acid residues separate this variable region from the 60-residue homeo domain of the carboxy-terminal constant region. This proximity suggests that functional differences among the *Ubx* proteins derive from the differential effects of their variable regions on the DNA-binding capacity of the homeo domain. An argument is made that these functional differences are tissue specific.

[Key Words: *Drosophila* development; homeotic gene; Bithorax Complex; *Ultrabithorax*; alternative splicing; alternative polyadenylation]

Received September 19, 1988; revised version accepted December 13, 1988.

Epidermal development in *Drosophila melanogaster* is characterized by the formation of three kinds of metameric units: compartments, segments, and parasegments (Garcia-Bellido et al. 1973; Lawrence 1981; Struhl 1984; Martinez-Arias and Lawrence 1985). The epidermis can be viewed as a linear array of alternating anterior (a) and posterior (p) compartments. Each is generated from a discrete set of embryonic founder cells so as to create lineage-restricting p/a and a/p boundaries that correspond to the segment (a + p) and parasegment (p + a) boundaries, respectively. The developmental fates of these founder cells and, hence, the identities of the respective metameres are determined by the regulatory functions of homeotic genes—referred to as identity functions. We are concerned here with one of these

homeotic genes, *Ultrabithorax* (*Ubx*) and, more specifically, with the nature of the *Ubx* gene products that effect its identity functions.

Ubx is one of three homeotic genes that comprise the Bithorax Complex (BX-C), the other two being *abdominal-A* (*abd-A*) and *Abdominal-B* (*Abd-B*) (for review, see Duncan 1987). Mutational analyses of these three genes indicate that together they provide identity functions in epidermal metameres that extend posteriorly from the a/p boundary in the second thoracic segment (T2) to include parasegment 5 (ps5) through parasegment 14 (ps14). Of the three, only *Ubx* provides identity functions in ps5 and ps6 or, equivalently, compartments T2p + T3a and T3p + A1a, where T3 and A1 represent the third thoracic and first abdominal segments, respectively (Lewis 1978, 1981, 1982; Morata and Kerridge 1981; Minana and Garcia-Bellido 1982; Hayes et al. 1984; Struhl 1984). *Ubx* also provides identity functions in ps7–ps13 of the larval epidermis (Lewis 1978, 1981; Struhl 1984; Bender et al. 1985), but here these functions act in combination with those provided by *abd-A*, or by both *abd-A* and *Abd-B*, to elicit the appropriate metameric identities (Sanchez-Herrero et al. 1985a,b; Tiong et al. 1985). Although less studied, the identity transformations produced in the central nervous system (CNS)

¹Present addresses: Department of Biochemistry, University of Adelaide, Adelaide, South Australia 5001; ²Howard Hughes Medical Institute and Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205 USA; ³Department of Genetics, Case Western Reserve University Medical School, Cleveland, Ohio 44106 USA; ⁴Department of Tropical Public Health, Harvard School of Public Health, Boston, Massachusetts 02115 USA.

and the somatic musculature by mutations in the *Ubx* gene indicate that it also provides identity functions for these tissues (Teugels and Ghysen 1985; Hooper 1986).

Previously, we showed that the *Ubx* gene is divisible into two functionally and mutationally distinct parts by molecular mapping of its mutations and transcription units (Bender et al. 1983a,b; Beachy et al. 1985; Hogness et al. 1985; Lipshitz et al. 1987). One part consists of a long transcription unit (*Ubx* unit) that includes and is virtually coextensive with a 73-kb region delineated by the sites of the *Ubx* mutations—a class of recessive mutations that inactivate all of the known identity functions of the gene. The other part consists of an approximately 45-kb *cis*-regulatory region lying immediately upstream of the *Ubx* unit. It includes and is virtually coextensive with the region delineated by the sites of the *bithoraxoid* (*bx*d) and *postbithorax* (*pb*x) mutations, whose effects are restricted to the *cis*-inactivation of the ps6 identity functions and the ps7–ps13 larval functions. Another group of recessive mutations (*anterobithorax* or *ab*x, and *bithorax* or *bx*) also *cis*-inactivates a subset of the *Ubx* functions—in this case, those required for the appropriate identity of ps5. The mutational sites of this group cluster near the center of the *Ubx* unit within its 3'-proximal intron and appear to define a downstream *cis*-regulatory region (Figs. 1 and 3).

Our initial analysis of the *Ubx* unit indicated that its primary transcript is spliced in different ways to yield at least two mRNA forms, labeled Ia and Ib, that contain different AUG-initiated open reading frames (ORFs), or coding regions (Beachy et al. 1985; Hogness et al. 1985). Each coding region is divisible into four parts: a 5' exon, two microexons, and a 3' exon. Those parts contributed by the two microexons and the 3' exon, which contains the *Ubx* homeo box (Regulski et al. 1985), are the same for forms Ia and Ib. They differ according to which of two 5' splice sites is used to join the 5' exon to the first microexon. These two sites are 27 bp apart and define an optional element present in Ib but not in Ia, which we therefore call the b element.

These results, in combination with the observation that *Ubx* mutations generally interrupt transcription or alter the sequence of the coding region contributed by the 3' exon and/or the common part of the 5' exon, led to the following two propositions (Beachy et al. 1985; Hogness et al. 1985). All identity functions of the *Ubx* gene require one or more proteins encoded by the *Ubx* unit. This proposition and its corollary, that the *bx*d/*pb*x and *ab*x/*bx* regions provide *cis*-regulatory elements controlling expression of the *Ubx* unit, are consistent with the spatial distributions of *Ubx* expression observed in wild-type and mutant *Drosophila* (Akam 1983; Akam and Martinez-Arias 1985; White and Wilcox 1984, 1985a, 1985b; Beachy et al. 1985; Cabrera et al. 1985; Hogness et al. 1985). The second proposition is that different combinations of *Ubx* proteins provide the identity functions for different metameres and/or tissues—a proposition that includes the suggestion that the family of *Ubx* proteins includes more than the two members defined above.

In this paper we extend our initial analysis of the structure and expression of the *Ubx* unit and its mRNAs with particular reference to this second proposition. We show that alternative splicing creates a set of *Ubx* mRNAs that include at least five coding sequences whose differences define a short variable region that links two constant regions derived from the 5' and 3' exons. The variable region consists of different combinations of three optional coding elements: the b element and the two microexons noted above. Nucleotide sequence analyses of genomic and cDNA clones combined with 5'- and 3'-end analyses have defined the sequences and boundaries of all exons in this set of mRNAs. These boundaries consist of the 5' termini common to all members of the set, the splice sites for each exon, and alternative polyadenylation sites that define two size classes of 3' exons, which account for the 3.2- and 4.3-kb mRNA size classes of the set. The lengths of the *Ubx* transcription unit required for the 3.2- and 4.3-kb mRNAs were thereby defined as 76 and 77 kb, respectively. Analysis of fractionated 3.2- and 4.3-kb mRNAs demonstrated that each of the five coding sequences is restricted to one of these two size classes, indicating that the splicing and polyadenylation pathways are coordinately regulated. Furthermore, the relative abundance of the respective mRNAs varies throughout development, indicating that their expression is also developmentally regulated so as to provide different combinations of *Ubx* proteins pertinent to the identity functions of different metameres and/or tissues. A variable region for the *Ubx* mRNAs like that defined here has recently been reported by O'Connor et al. (1988).

Results

A large Ubx primary transcript is processed into discrete RNA size classes

To analyze transcriptional activity in and adjacent to the 73-kb *Ubx* region defined by the sites of *Ubx* mutations (Bender et al. 1983b), strand-specific probes spanning 100 kb were hybridized to filters containing RNA from early, mid-, and late embryogenesis. Figure 1 shows that two RNA size classes of 3.2 and 4.3 kb contained sequences from both ends of the region and were transcribed in the rightward direction (blots 3 and 16). In contrast, a third size class of 4.7 kb appeared to contain sequences only from the 5' end of the *Ubx* region (Fig. 1, blot 3; Fig. 5, 3- to 6-hr panel). The 4.7-kb class also differed from the 3.2- and 4.3-kb classes in two other characteristics: It was not polyadenylated, whereas the other two were (data not shown); and it was transiently expressed only during the first few hours of embryogenesis, whereas the other two continued to be expressed in the larval, pupal, and adult stages (Fig. 5). In this paper, we focus on the 3.2- and 4.3-kb poly(A)⁺ RNAs, which comprise the *Ubx* mRNAs. Probes between the ends of the *Ubx* region hybridized only to a heterogeneous 'smear' of high-molecular-weight RNA, which we interpret to be a combination of nascent transcripts and processed or partially degraded intron RNAs (blots 6–13). Probes just outside the

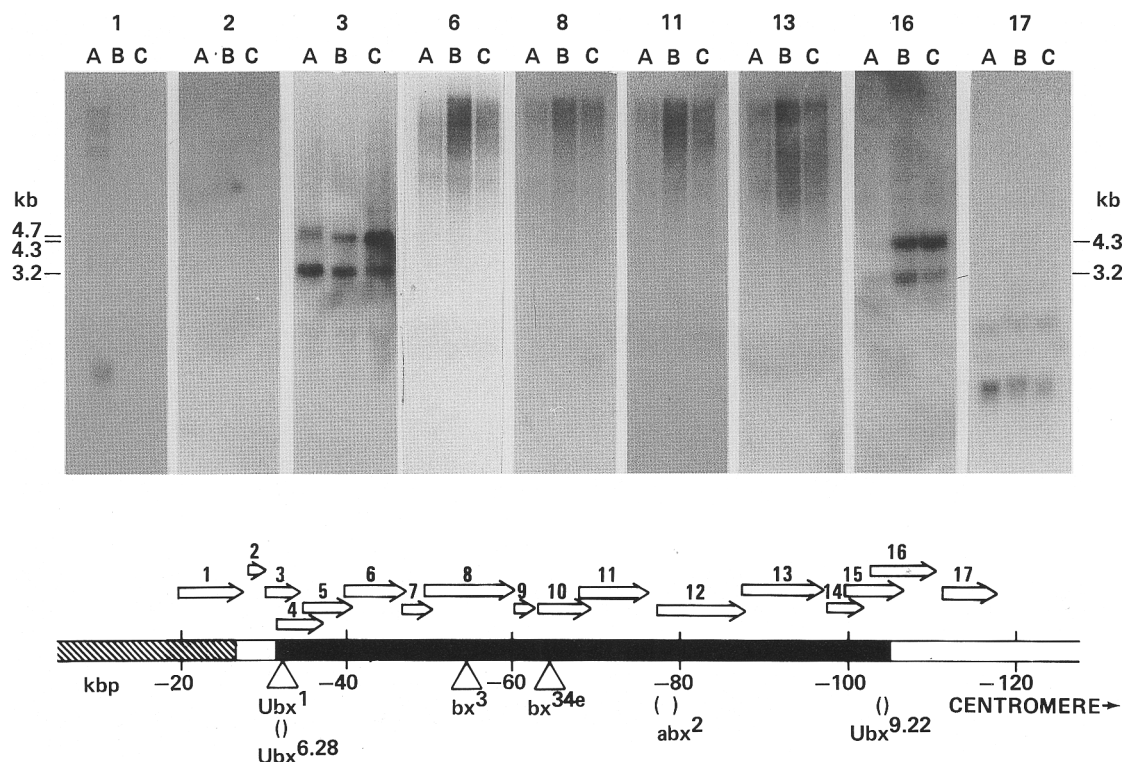


Figure 1. Transcriptional activity in the *Ubx* domain. Total RNA from embryos 3–6, 9–12, and 18–21 hr after egg deposition (A, B, C, respectively) was electrophoresed and transferred to paper. Numbered arrows show the extent and orientation of single-stranded DNA probes and their location in the *Ubx* gene (coordinates in kilobase pairs from Bender et al. 1983b). Numbers above each blot correspond to numbered hybridization probes. Blots 3 and 16 were underexposed relative to the others to show the discrete RNA species whose sizes are indicated at the margins. Probes 6, 8, 11, and 13 lie between the ends of the *Ubx* unit and detected a heterogenous smear of high-molecular-weight RNAs. Other probes from this region (4, 5, 7, 9, 10, 12, 14, and 15) also detected this smear (data not shown). Probes 1–17 can detect only rightward transcription; 17 single-stranded probes with opposite orientation (not shown) detected no leftward transcription (data not shown). Triangles and parenthesis indicate the position of transposable element insertions and small deletions, respectively, found in the indicated BX-C alleles (Bender et al. 1983b; Weinzierl et al. 1987). The solid black and hatched portions denote the regions defined by the location of *Ubx* and *bxd* rearrangement breakpoints, respectively (Bender et al. 1983b; Bender et al. 1985). Probe 1 detected *bxd* transcripts whose sizes and developmental patterns distinguish them from the *Ubx* RNAs (Lipshitz et al. 1987). Probe 2 did not hybridize to the smear, indicating that transcription does not extend from the *bxd* into the *Ubx* region. Probe 17 detected RNAs that we believe are unrelated to *Ubx* function, based on their novel size, presence in maternal RNA, and lack of differential expression during development (data not shown).

Ubx region did not hybridize to the smear (blots 2 and 17), and probes of opposite orientation did not reveal any leftward transcription within the 100-kb region examined (legend to Fig. 1).

Alternative splicing generates a family of *Ubx* mRNAs

cDNA analysis To characterize the structure of the poly(A)⁺ *Ubx* RNAs, cloned cDNAs that contain both 5'- and 3'-proximal *Ubx* sequences were isolated. cDNA libraries made with 1.5- to 5-hr and 3- to 12-hr embryonic RNA yielded 2 and 35 clones, respectively. A preliminary analysis of the two clones from the 1.5- to 5-hr library has been reported (Beachy et al. 1985; Hogness et al. 1985). A nearly full-length clone (3.1 kb) from the 1.5- to 5-hr library was sequenced along with the genomic DNA that overlaps each exon (Fig. 2). The cDNA contained long 5' and 3' exons, predicted by the transcript analysis shown in Figure 1, linked by two 51-bp mi-

croexons. Although these microexons were not detected in the experiments shown in Figure 1, presumably because the homologous region was short and the stringency of the wash high, they were mapped to -40.7 ± 1.1 kb and -52.9 ± 1.7 kb, as described in Materials and methods and illustrated in Figure 3.

The other 36 cDNAs were analyzed by restriction enzyme mapping to uncover heterogeneity, particularly in the microexon region. Comparisons between cDNAs digested with *Ava*I, which cleaves at sites located just outside the microexon region, revealed four different exon patterns. Each novel pattern was confirmed by digests with *Bgl*II and *Dde*I, which cleave at sites located in microexon 1 and 2, respectively, and by sequencing the variable region of a representative cDNA from each class. Figure 3 shows that each cDNA consisted of 5' and 3' constant regions linked by some combination of three optional sequence elements. The optional element farthest upstream is a 27-bp region that lies between two 5'

splice sites in the 5' exon called the a and b sites (Fig. 2). It is called the b element, as it is included in the mRNA only if the b site is used. The other two optional elements are the 51-bp microexons 1 and 2. Each of the eight potential combinations of optional elements is designated by a Roman numeral indicating the microexon pattern followed by a or b to indicate which 5' splice site is used. Thus, forms Ia and Ib contain both microexons, and, respectively, lack or contain the b element; forms IIa and IIb lack microexon 1; forms IIIa and

IIIb lack microexon 2; forms IVa and IVb lack both microexons. Figure 3 shows that among 37 cDNAs isolated from two embryonic libraries, forms Ia (59%) and Ib (27%) predominated over IIa (11%) and IVa (3%).

The consensus splice sites that flank each optional element and the demonstration that the entire *Ubx* unit is transcribed into high-molecular-weight RNAs (Fig. 1) strongly suggest that a 77-kb *Ubx* primary transcript is alternatively spliced when one of two possible 5' splice sites (a or b) is joined with one of three possible 3' splice

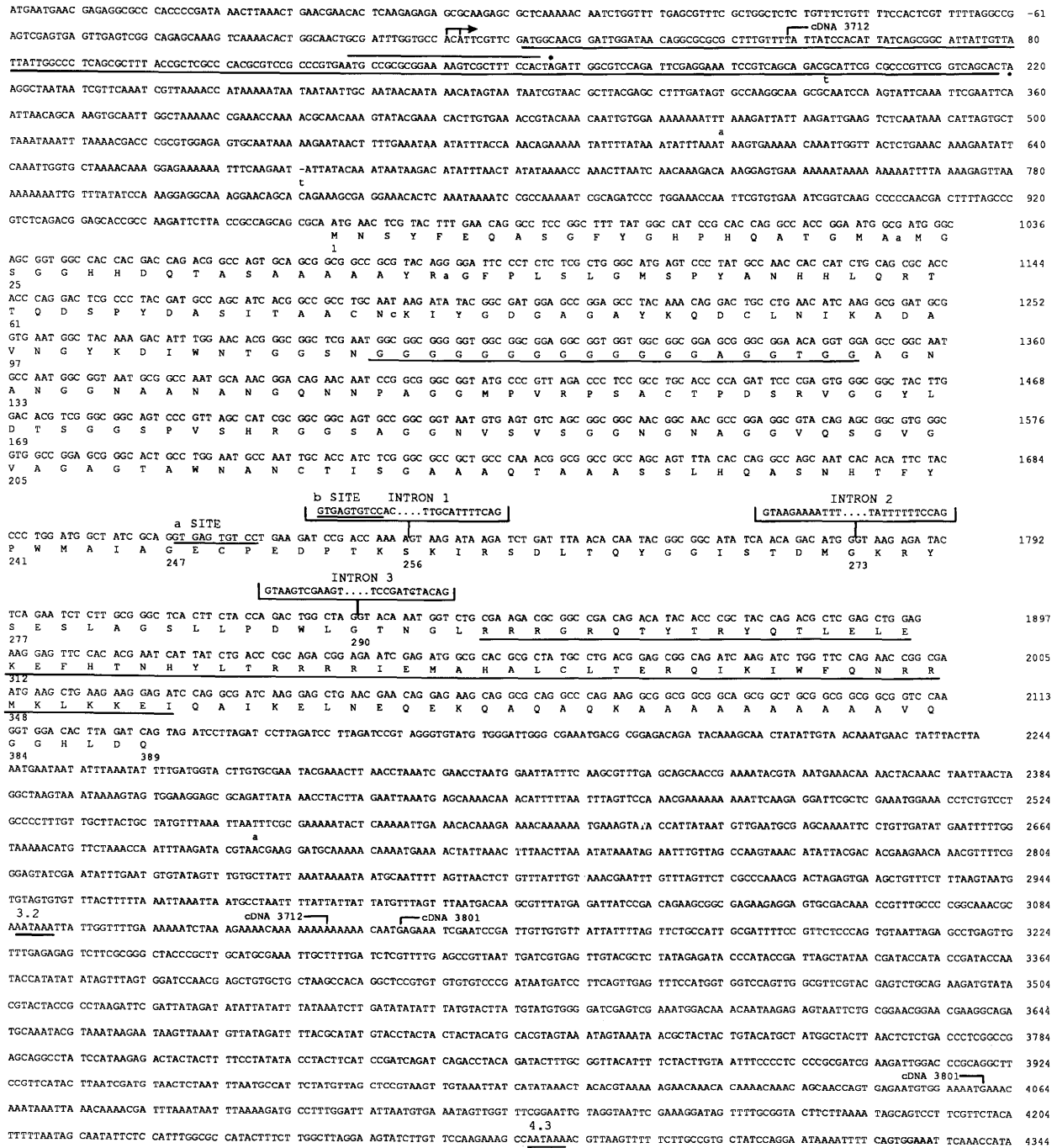


Figure 2. (See facing page for legend.)

sites (adjacent to microexon 1, microexon 2, or the 3' exon). The unlikely alternative explanation that this family of *Ubx* mRNAs results from genomic DNA rearrangements rather than RNA splicing is rendered even less likely by a Southern analysis of restriction fragments of DNA prepared from isolated haltere disks, where nearly every cell contains *Ubx* protein (White and Wilcox 1984; Beachy et al. 1985). No rearrangements were detected in the region of genomic DNA extending 3.3 kb downstream from the *EcoRI* site located at +354 bp (Fig. 2) within the 5' constant region (R.B. Saint and M. Goldschmidt-Clermont, unpubl.). Because this region includes the b element and 1.9 kb of adjacent intronic DNA, as well as the remaining 1349 bp of the 5' constant region, this observation indicates that DNA rearrangement alone cannot account for the *Ubx* mRNA structures.

A long ORF initiates in the 5' exon and terminates in the 3' exon. Because the length of each optional element is a multiple of 3 bases, any combination will maintain an ORF. We therefore predict that the family of *Ubx* mRNAs generated by alternative splicing would encode a corresponding family of *Ubx* proteins containing a maximum of 389 amino acids (Ib) and a minimum of 346 amino acids (IVa). Figure 3 shows the amino acid sequence of the variable region for each of the forms identified by the cDNA analysis.

S1 nuclease protection analysis Four of the possible eight combinations of optional elements were not detected by the above analysis of cDNA clones. The RNA populations examined by that analysis were, however, limited to the first 12 hr of embryogenesis, and the alternative splicing patterns may vary with developmental stage. Variation in the structure of the variable region

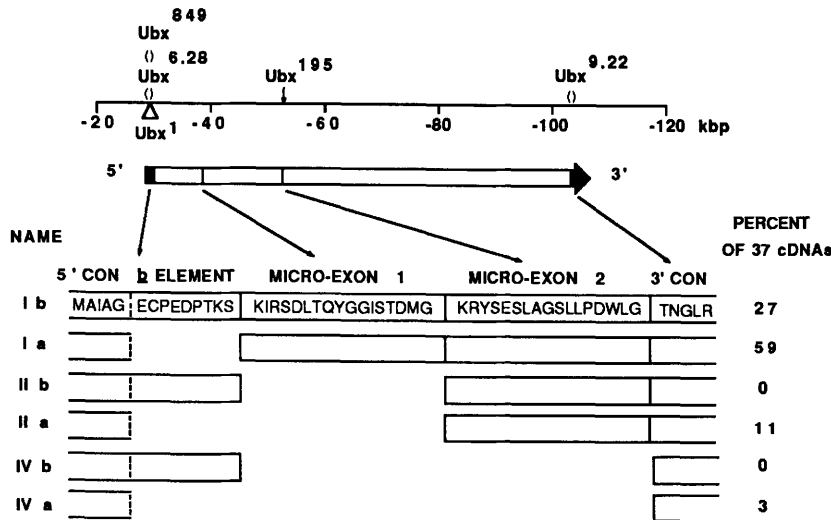
was therefore examined by the rapid and sensitive nuclease S1 protection method in RNA populations from all stages of embryogenesis, from all three larval stages or instars (L1–L3), and from the pupal and adult stages.

Figure 4A shows the structure of the uniformly labeled, single-stranded DNA probe used for these analyses, as well as the expected structure and length of the probe fragments protected from nuclease S1 digestion by *Ubx* mRNAs with different variable regions. Because the probe derives from a Ib cDNA subclone containing all three optional elements, the *Ubx* Ib mRNA will protect all *Ubx* sequences within the probe from nuclease S1 digestion to yield a single protected fragment that will be slightly shorter than the probe because its 48-base polylinker will be digested. In contrast, *Ubx* mRNAs that lack one or more of the optional elements will yield two fragments protected by RNA sequences lying upstream (5' fragment) or downstream (3' fragment) of the missing element(s), as shown in Figure 4A.

Figure 4B shows the results obtained when the protected fragments generated by each RNA population were separated by gel electrophoresis and their lengths determined from their mobilities. Comparison of the determined lengths for the six significant fragment bands observed in Figure 4B with those expected from protection by the different *Ubx* mRNA forms given in Figure 4A shows that the observed bands correspond to the full-length fragment generated by Ib (970 versus 980 bases); the 3' fragments generated by Ia (370 versus 380 bases), by IIa and/or IIb (325 versus 329 bases), and by IVa and/or IVb (275 versus 278 bases); and the 5' fragments generated by all a forms (560 versus 573 bases) and by IIb and/or IVb (590 versus 600 bases). The first five of these bands were expected from the Ia, Ib, IIa, and

Figure 2. Sequence of *Ubx* cDNAs and genomic DNA and conceptual translation of the ORF. cDNA 3712, a nearly full-length Ib clone derived from a 3.2-kb RNA, was entirely sequenced and extends from +48 to +3129 (limits marked). cDNA 3801, a partial clone derived from a 4.3-kb RNA, was entirely sequenced and extends from +3139 to +4060 (limits marked). All other sequences (5' and 3' extensions, introns, and the 9-bp region between the cDNAs) derive from genomic DNA clones (Bender et al. 1983b). Transcription start sites are marked by arrows and are numbered +1 and +3. They apply to both the 3.2- and 4.3-kb mRNAs (see text). Saari and Bienz (1987) have also observed two start sites but assign them to positions -3 and +1 in this sequence. Nucleotide numbers (right) apply to the sequence of cDNA Ib, which contains all three optional elements. The two upstream ATG initiated ORFs, which we do not believe are translated (Beachy et al. 1985), are underlined and overlined; dots indicate stop codons. The conceptual translation of the long ORF is shown in one-letter code. Amino acid residue numbers (left) apply to protein form Ib. A glycine-rich region, which may constitute a flexible protein hinge, is underlined. Residues 295–354 comprise the *Ubx* homeo domain and are also underlined. Introns, numbered in reference to form Ib, are shown as insertions at the splice junctions. Each intron encodes consensus splice signals at both ends. The a and b 5' splice sites are marked, and 10 bp of identical sequence following each is underlined. Consensus polyadenylation signals used in 3.2- and 4.3-kb RNAs are underlined and labeled. The site of polyadenylation for 3.2-kb RNAs is within a 13 A residue stretch derived from genomic DNA (+3122 to +3134), as determined from the 3'-terminal sequence of cDNA clone EC14, which contains a poly(A) tail of >50 residues. The C-S genomic sequence for each exon was determined and found to differ from the O-R cDNA sequence at three positions in the long ORF (all conservative changes) and four in the untranslated region, perhaps due to strain variation. The genomic sequence at these positions is noted below in lowercase letters. In addition, cDNA 3712 exhibited a 22-bp deletion not found in another cDNA clone (3602) nor in the genomic DNA at +769 to +791 (not marked), which may be due to replication slippage during cloning, as may also have caused an AAT insertion at +276 (not marked) in cDNA 3712. A cloning artifact at the 5' end of cDNA 3712 caused by self-primed second-strand synthesis is not shown. Artifacts of this kind have been described by Laughon and Scott (1984). The sequence of C-S genomic DNA from -200 to +988 (Saari and Bienz 1987) and +965 to +3390 (Weinzierl et al. 1987) was reported recently. Differences between our data and that of Weinzierl et al. are T:A (2458), C:— (2645), A:— (3134), C:— (3241), G:— (3242), where our finding precedes theirs and the number in parenthesis is our residue number. Single-letter amino acid code: (A) Ala, (C) Cys, (D) Asp, (E) Glu, (F) Phe, (G) Gly, (H) His, (I) Ile, (K) Lys, (L) Leu, (M) Met, (N) Asn, (P) Pro, (Q) Gln, (R) Arg, (S) Ser, (T) Thr, (W) Trp, (Y) Tyr, (V) Val. Restriction enzyme recognition sites noted in text: *AvaI* (1618, 1886), *AvaII* (2107), *BamHI* (3385), *BglII* (1739), *DdeI* (1792), *EcoRI* (354), *HpaI* (2866), *PstI* (1133), *TaqI* (1289), *XmnI* (2315).

Kornfeld et al.



the 37 *Ubx* cDNAs isolated from the 1.5- to 5-hr and 3- to 12-hr embryonic libraries. The dotted line separating the 5' constant element from the b element indicates that both derive from the same exon, whereas the other elements are discrete exons. The amino acid codons assigned to each element are indicated by the single-letter amino acid code. The first amino acid of the homeo domain is the fifth (R) encoded by the 3' constant region.

IVa forms detected by the cDNA analyses (Fig. 3). The sixth band demonstrates the existence of at least one and possibly two additional forms: IIb and/or IVb. This conclusion is consistent with the finding of a IIb cDNA clone by O'Connor et al. (1988).

A band corresponding to the 651-base 5' protected fragment expected from form IIIb was not detected in any of the RNA populations examined in Figure 4B, indicating that little if any of this form is produced at any stage of development. The absence of form IIIb indicates that the splice between microexon 1 and the 3' exon occurs rarely, if at all. If this rule applies to the a forms as well as the b forms, then IIIa should also be absent or in low abundance throughout development, an expectation consistent with the absence of IIIa and IIIb cDNA clones in searches carried out by us and by O'Connor et al. (1988). The 3'-protected fragment expected from IIIa and IIIb is the same 278-base fragment expected from IVa and IVb (Fig. 4A), and a band corresponding to this fragment was generated only by RNAs from late embryos (12 ± 4 hr and 20 ± 4 hr) and first-instar larvae (Fig. 4B,C). Although our data exclude only IIIb from significantly contributing to this band, the above arguments suggest that the contribution of IIIa is also small.

Temporal developmental profiles for the *Ubx* mRNAs

We have seen that members of the *Ubx* mRNA family are divisible into two length classes and at least five protein-coding forms formed by alternative splicing. We now compare the temporal developmental profiles for the coding forms with those for the length classes to test for possible correlations.

Figure 4C shows the relative molar abundance of the different coding forms, or pairs of those forms, as a function of developmental stage. Ignoring differences of form

Figure 3. Summary of *Ubx* cDNA analysis. An arrow representing the *Ubx* primary transcript, with four exons colored black, is drawn beneath a map of the *Ubx* gene (coordinates as in Fig. 1). Parentheses, a triangle, and an arrow indicate the position of deletions, a transposable element insertion, and a nonsense base change, respectively, in the indicated *Ubx* alleles (Beachy et al. 1985; Weinzierl et al. 1987). Each *Ubx* mRNA encodes a 247-amino-acid residue 5' constant region (5' CON) and a 99-amino-acid residue 3' constant region (3' CON) linked by a variable region that contains some combination of three optional elements: the 9-amino-acid residue b element and 17-amino-acid residue micro exons 1 and 2. Among the eight possible combinations of the three optional elements, the six that are shown have been identified as cDNAs or are consistent with the nuclease S1 protection analyses. (Right) The percentage of each spliced form among

due solely to the presence or absence of the b element for the moment, one sees that form I predominates during early embryogenesis. Forms II and IV then increase while form I decreases so that they predominate during late embryogenesis (20 ± 4 hr sample). During the larval, pupal, and adult stages, form I again predominates, with form IV decreasing dramatically. These data suggest that the choice of which 3' splice site is joined to the 5' exon is developmentally regulated.

We have also determined the relative abundance in the same set of RNAs of forms that contain or do not contain the b element, using a single-stranded genomic DNA probe that spans the two splice sites of the 5' exon (see Materials and methods). The fraction of *Ubx* mRNAs that contains the b element progressively declines from a maximum of 30% in the 2 ± 2 -hr embryonic sample to 10% in adults (data not shown). This result is consistent with the decrease in form Ib seen in Figure 4C and the virtual absence of forms IIb + IVb in the larval, pupal, and adult stages (Fig. 4B). These data suggest that the choice of the 5' splice site in the 5' exon is developmentally regulated. The difference in the developmental profiles for the choice of the 5' and 3' splice sites suggests that they are independently regulated.

Figure 5 shows the results of Northern analyses used to obtain the developmental profiles for the 3.2- and 4.3-kb-length classes. A comparison of the signal intensities within a given lane indicates predominance of the 3.2-kb class during early embryogenesis, followed by predominance of the 4.3-kb class during the last half of embryogenesis followed, in turn, by a return to 3.2-kb class predominance during the remainder of development. A comparison of these profiles with those shown in Figure 4C indicate a correspondence between the 3.2-kb and form I profiles and between the profiles for the 4.3-kb mRNAs and the sum of forms II and IV.

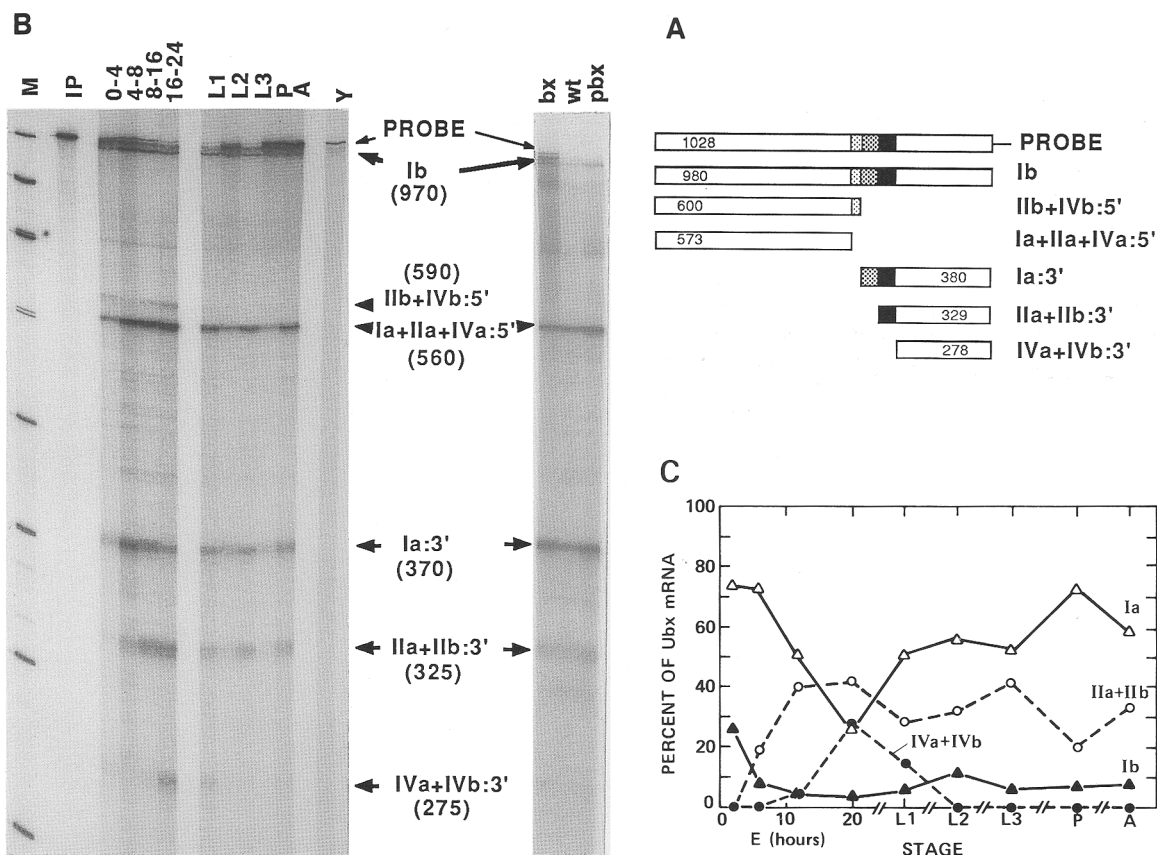


Figure 4. Nuclease S1 protection analysis of the variable region of *Ubx* mRNAs with RNA from wild-type flies at each developmental stage and RNA from *bx³/bx³* and *pbx¹/pbx²* mutants. (A) Schematic of the intact probe is at the top, with schematics for the fragments protected by the indicated mRNAs (or parts thereof) at the bottom. The 1028-base probe is a subclone from an Ib cDNA that extends from the *Pst* site in the 5' exon to the *Ava*II site in the 3' exon (Fig. 2). Sizes and shading of probe elements are 573-base 5' constant (open), 27-base b element (lightly stippled), 51-base microexon 1 (heavily stippled), 51-base microexon 2 (black), 278-base 3' constant (open), and 48-base polylinker (line). The Ib mRNA protects the indicated 980-base fragment. Other *Ubx* mRNAs lack an optional element found in the probe and protect the indicated 5' and 3' fragments. Several probe fragments are protected by two or more different spliced mRNAs. Fragments protected by the form III mRNAs are defined in the text. (B) Lanes: (M) size markers; (IP) intact probe. (Time intervals) Four embryonic periods indicated in hours after egg deposition. (L1, L2, L3) First, second, and third instar larvae, respectively. (P) Pupae; (A) adults; (Y) yeast total RNA, which controls for nonspecific probe protection. (bx) *bx³/bx³* pupae; (wt) wild-type pupae; (pbx) *pbx¹/pbx²* pupae. Poly(A)⁺ RNA was annealed to a single-stranded, uniformly labeled DNA probe and cleaved with nuclease S1. Protected probe fragments were separated in a denaturing acrylamide gel and visualized by autoradiography. The uppermost arrow marks residual, intact probe typically seen in these experiments. Other bands marked by arrows result from probe protection by spliced mRNAs, which are listed. Calculated sizes (shown in parentheses) correspond closely to predicted sizes listed in A. The 5' end of IIIb would protect a 651-base fragment that was not observed. (See the text for a consideration of this and other fragments expected to be protected by form III mRNAs.) A different amount of RNA was used in each lane, so it is meaningful to compare signal intensities within lanes but less so between lanes. Sizes (in bases) of fragments in lane M: 993, 823, 673, 552, 444, 369, 310, 249. (C) The gel was exposed briefly with an intensifying screen and preflashed film at -70°C , and signal intensities were quantitated by scanning densitometry. Signal intensity is proportional to the number of mRNA molecules, which protect the uniformly labeled probe fragment corrected for the number of labeled residues that it contains. Corrected values for Ib, Ia:3', IIa + IIb:3', and IVa + IVb:3' were summed and expressed as a percent of total *Ubx* mRNA such that the four values sum to 100%. IIb + IVb:5' and Ia + IIa + IVa:5' were not used for these calculations. The abscissa is linear for the embryonic (E) stage (24 hr); values from embryonic samples are plotted at the midpoint of the collection period. Breaks indicate a discontinuous time scale at later stages. Note that IVa + IVb and IIa + IIb peak during late embryogenesis. Poly(A)⁺ RNA was used in these experiments to minimize the effects, if any, of nascent and partially processed RNAs.

The 3.2- and 4.3-kb size classes result from different polyadenylation sites

This temporal correspondence between length classes and coding forms suggests that form I is restricted to the

3.2-kb class and that forms II and IV are restricted to the 4.3-kb class—a suggestion that is confirmed in the next section. Given these restrictions, the variable regions in the 3.2-kb mRNAs will be 102 or 129 bases long,

Kornfeld et al.

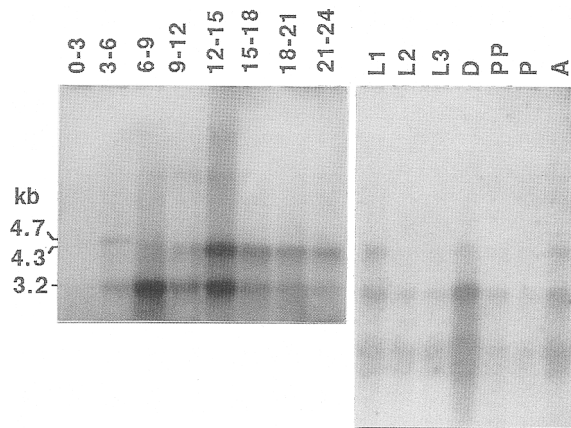


Figure 5. Developmental Northern analysis of *Ubx* mRNAs. Lanes: (Time intervals) Eight embryonic periods indicated in hours after egg deposition. (L1, L2, L3) First, second, and third instar larvae respectively. (D) Imaginal discs; (PP) prepupae; (P) pupae; (A) adults. Total cellular RNA was electrophoresed, transferred to paper, and hybridized with a single-stranded genomic DNA probe that includes the *Ubx* 5' exon and has an identical region of homology with the 3.2- and 4.3-kb size classes (probe 3 of Fig. 1). The 4.7-kb size class is only detected in 3- to 6-hr RNA. The 4.3-kb size class predominates only during late embryogenesis. Low-molecular-weight RNAs seen at L1 and later stages were also detected in embryonic samples (data not shown) and hybridize only to probe fragments containing a (GGX)_n repeat, suggesting that they derive from loci other than *Ubx* (S. Munroe and D.S. Hogness, unpubl.).

whereas those in the 4.3-kb mRNAs will be 0, 27, 51, or 78 bases long. Differences in the variable region therefore contribute little to the length difference between the two size classes. What then accounts for this length difference? We show here that the transcription initiation sites are the same for both classes and that their length difference results from different poly(A) sites.

The 5' terminus for each size class was determined by exploiting the observation that the 3.2- and 4.3-kb mRNAs predominate in early and late embryos, respectively (Fig. 5). Nuclease S1 protection and primer extension 5'-end determinations (Materials and methods) yielded the same result with early and late embryonic poly(A)⁺ RNAs, indicating that the 3.2- and 4.3-kb mRNAs both initiate at either of the two A residues labeled +1 and +3 in Figure 2 (data not shown).

The poly(A) site for the 3.2-kb mRNAs has been mapped to +3128 (± 6) in Figure 2 by determining the 3'-terminal sequence of a cDNA clone with an intact 3' end. (The imprecision of this localization results from the fact that this site is located within a stretch of 13 A residues; see legend to Fig. 2.) A consensus AATAAA polyadenylation signal (Proudfoot and Brownlee 1976) is appropriately located 37 ± 6 nucleotides upstream from this site (Fig. 2).

The poly(A) site for the 4.3-kb mRNAs has not been mapped with similar precision as cDNA clones with intact 3' ends have not been isolated and sequenced for

this class. The sequence of a partial cDNA clone from this class (cDNA 3801) has, however, been determined and is identical to the genomic DNA sequence from position +3139 to +4060 in Figure 2. The 4.3-kb mRNA poly(A) site should therefore be downstream from +4060, a conclusion consistent with the observation of Akam and Martinez-Arias (1985) that a genomic DNA probe extending 1.6 kb downstream from the *Bam*HI site at +3385 (Fig. 2) hybridizes to the 4.3-kb, but not the 3.2-kb, mRNAs. A consensus polyadenylation signal is located at +4277 (Fig. 2), 1191 nucleotides downstream from the polyadenylation signal for the 3.2-kb mRNAs. This difference correlates nicely with the 1.1-kb difference between the lengths of the two size classes. We therefore expect that the poly(A) site for the 4.3-kb mRNA is located within a few dozen nucleotides downstream from this polyadenylation signal (labeled 4.3 in Fig. 2).

The position of the poly(A) sites for the 3.2- and 4.3-kb mRNAs that were determined and predicted, respectively, from the nucleotide sequence data have been confirmed by nuclease S1 protection mapping of the 3' ends of both size classes, again taking advantage of the differences in their relative abundance during embryogenesis. Figure 6 shows that a single-stranded genomic DNA probe extending 1.7-kb downstream from the +2866 position in Figure 2 generated protected fragments of 0.27 and 1.4 kb that attained maximum relative intensities during early and late embryogenesis, respectively (lanes 2 and 4); they are therefore presumed to result, respectively, from the 3.2- and 4.3-kb mRNAs. These data place the poly(A) sites for the 3.2- and 4.3-kb mRNAs at +3.14 and +4.3 kb on the sequence shown in Figure 2, in good agreement with the values obtained by analyzing the sequence data.

Alternative splicing and polyadenylation are coordinately regulated

As noted above, comparison of the temporal developmental profiles for the *Ubx* mRNA size classes and spliced forms suggests that the 3.2- and 4.3-kb mRNAs consist of forms I and II + IV, respectively. Here, we test and confirm this suggestion.

Mid-embryonic RNA was fractionated by electrophoresis in an agarose-urea gel, extracted from a series of gel fractions, and analyzed for the amounts of the different spliced forms by the nuclease S1 protection assay described in Figure 4. Because complete separation of the two size classes was not achieved, the RNA from each fraction was also assayed for the amount of each size class by nuclease S1 protection (see Materials and methods). Table 1 shows the relative amounts of the different spliced forms in each size class, as calculated from the data obtained for different pairs of fractions by, in effect, 'subtracting out' the cross contamination; see Materials and methods for the data and method of calculation. The striking result of this experiment is that within experimental error, the 3.2-kb size class consists of the mRNAs that contain both microexons (Ia and Ib),

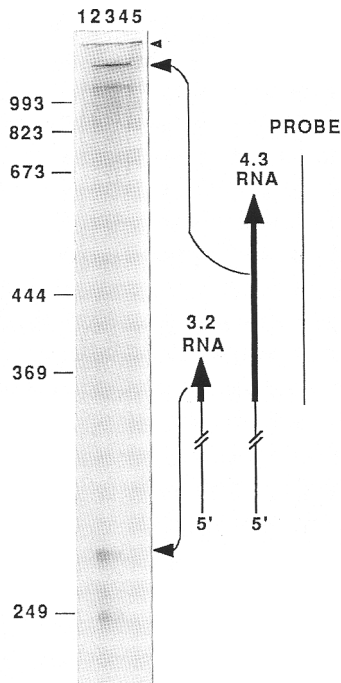


Figure 6. Nuclease S1 protection analysis of the 3' end of *Ubx* mRNAs. Lanes: (1) 0–4 hr; (2) 4–8 hr; (3) 8–16 hr; (4) 16–24 hr poly(A)⁺ embryonic RNA; (5) yeast total RNA. A uniformly labeled, single-stranded, genomic DNA probe that extends from a *Hpa*I site (Fig. 2, position +2866) in the 3' exon to a *Bgl*II site 1.7 kb downstream was annealed to RNA, treated with nuclease S1, and electrophoresed in a denaturing acrylamide gel. (Left) Migration of size markers in nucleotides. Probe (thin vertical line) is protected by the 3' end of *Ubx* mRNAs (arrows, with the protecting region highlighted). Doublets of 275 + 250 bases and 1.4 + 1.1 kb indicate protection by 3.2- and 4.3-kb size classes, respectively. The uppermost band of each doublet is marked by an arrow. Comparing signal intensities within lanes and correcting for probe length shows that the 3.2- and 4.3-kb size class predominated early and late in embryogenesis, respectively. A different amount of RNA was used in each lane so comparisons of intensities between lanes may be less accurate than comparisons within lanes. In each doublet, the ratio of lower band to upper band was increased by incubation with additional nuclease S1, suggesting that the upper bands map the actual 3' ends of mRNAs and the lower bands are due to S1-sensitive sites within the protected probe fragment. The probe fragment protected by the 4.3-kb RNAs, which cannot be accurately sized from these data, was determined to be 1.4 kb by glyoxylation and electrophoresis in a 1.3% agarose gel with appropriate size markers. The uppermost arrowhead marks residual, intact probe that is present in the yeast control lane and is typically seen in these experiments.

whereas the 4.3-kb size class consists of mRNAs that lack one or both microexons (IIa + IIb and IVa + IVb). Figure 7 diagrams this result.

This result is consistent with limited data from the 37 *Ubx* cDNA clones that were analyzed. Two of these clones derived from 1.5- to 5-hr embryonic RNA containing only the 3.2-kb size class of *Ubx* mRNAs, as determined by Northern analysis (M. Goldschmidt-Clermont and D.S. Hogness, unpubl.). One of these cDNAs

exhibits the Ia form, and the other, the Ib form. A third clone, obtained from the 3- to 12-hr library, exhibits the Ia form and is known to derive from the 3.2-kb size class because its 3' end is intact, as demonstrated by the inclusion of sequences from the poly(A) tail of the mRNA. The remaining 34 clones were isolated from the 3- to 12-hr library and could derive from either size class.

Attempts to isolate *Ubx* cDNA clones containing the sequences from the central variable region and sequences from the 3'-proximal region specific to the 4.3-kb mRNAs failed. When libraries made from late embryonic RNA preparations, in which the 4.3-kb size class predominates, were screened with 4.3-kb-specific sequences, only short clones (<1 kb) were isolated, and their sequences were restricted to the 3'-untranslated region unique to the 4.3-kb class. We suspect that during cDNA synthesis, the oligo(dT) primer anneals to an internal stretch of 13 A residues (+3122 to +3134, Fig. 2) in the 4.3-kb RNAs that could (1) block first-strand synthesis initiated at the poly(A) tail and (2) prime first-strand synthesis resulting in 4.3-kb-derived cDNAs with truncated 3' ends. Consistent with this hypothesis, the longest isolated 4.3-kb-derived cDNA terminates 4 bases downstream of the 13 A residues (cDNA 3801, Fig. 2).

The relative abundance of the spliced forms is the same in bx³ and pbx mutants as in wild type

The *bx³* and *pbx* mutations inactivate the T3a and T3p identity function of the adult epidermis, respectively (Lewis 1963) and, correspondingly, prevent the appearance of *Ubx* protein in the respective compartments of the T3 imaginal discs (White and Wilcox 1985a; Cabrera et al. 1985). These negative effects could result either because the mutations inactivate *cis*-regulatory elements required for the transcription of the *Ubx* unit in these compartments or because these compartments contain a subset, perhaps only one, of the different spliced forms and the mutations inactivate the splicing pathway(s) specific to the subset.

If such a specific inactivation of splicing should occur in all tissues generating *Ubx* transcripts, then one would expect that the relative abundance of the different spliced forms would be different in the *bx³* and *pbx* mutants than in wild type. RNA was therefore isolated from *bx³/bx³* and *pbx¹/pbx²* larvae, pupae, and adults and analyzed for the different spliced forms by nuclease S1 protection. Figure 4B shows that the different spliced forms were present in the same relative amounts in the mutant pupae as in the wild-type pupae. The same result was obtained for the larval and adult RNAs (data not shown). It therefore appears that the *bx³* and *pbx* mutations do not generally alter the frequencies of the different splicing pathways, although the possibility remains that they alter such frequencies in a tissue-specific or metamere-specific manner.

Discussion

Here, we consider three questions raised by our results.

Table 1. Calculated composition of each *Ubx* mRNA size class

<i>Ubx</i> mRNA size class (kb)	Percentage of mRNAs containing spliced forms			
	Ib	Ia	IIa + IIb	IVa + IVb
3.2	22 ± 7	68 ± 4	(9 ± 10)	(1 ± 5)
4.3	(4 ± 6)	(-3 ± 4)	66 ± 10	33 ± 2

Embryonic RNA (10- to 18-hr plus 15- to 18-hr) was fractionated in an agarose-urea gel. The spliced forms that comprise the 3.2- and 4.3-kb size classes were calculated by comparing two fractions enriched for 3.2-kb mRNAs with two fractions enriched for 4.3-kb mRNAs. The values listed are the average of four calculations ± 1 s.d. Values in parenthesis are within 1 s.d. of zero (see Materials and methods for the contents of each fraction and the computation that yields these values).

What functional differences might derive from the structural differences among the proteins encoded by the *Ubx* mRNAs? What accounts for the observed correlation between alternative splicing and polyadenylation of the *Ubx* transcripts? Is the great length of the *Ubx* transcription unit, per se, an important factor in the control of *Drosophila* development?

Structure and function of the members of the *Ubx* protein family

It has long been assumed that the products of homeotic genes such as *Ubx* specify the identities of epidermal metameres by regulating the expression of other genes that are required to effect these identities (Lewis 1964). More recently, it has become apparent that the postulated set of 'effector' genes regulated by *Ubx* proteins must be expanded to include not only those required for epidermal identities but also those required for the quite different identities of metameres in the CNS and musculature (Teugels and Ghysen 1985; Hooper 1986).

At present, no member of this set of effector genes has been identified. Another set of genes has, however, been identified as likely regulatory targets of the proteins en-

coded by homeotic genes—namely, the homeotic genes themselves. This identification derives from observations of the effects of mutations in one homeotic gene upon the spatial distribution of the products of other homeotic genes. Thus, the distribution of the RNA and protein products of the *Antennapedia* (*Antp*) gene (a homeotic gene contributing identity function for metameres in the anterior thorax) is perturbed by *Ubx* null mutations in a manner indicating that *Ubx* proteins repress *Antp* expression in the wild type (Hafen et al. 1984; Carroll et al. 1986). Similar experiments indicate that the proteins encoded by the *abd-A* and *Abd-B* homeotic genes of the BX-C repress *Ubx* expression (Struhl and White 1985) and that *abd-A* proteins, like *Ubx* proteins, also repress *Antp* expression (Carroll et al. 1986). The effect of this cross regulation is to refine the patterns of homeotic gene expression after their initial determination by the earlier and transiently expressed segmentation genes (for review, see Akam 1987; Scott and Carroll 1987).

More direct lines of evidence leave little doubt that the *Ubx* proteins are indeed genetic regulatory proteins and that they act by binding to specific DNA sequences associated with the genes that they regulate. The *Ubx* Ib

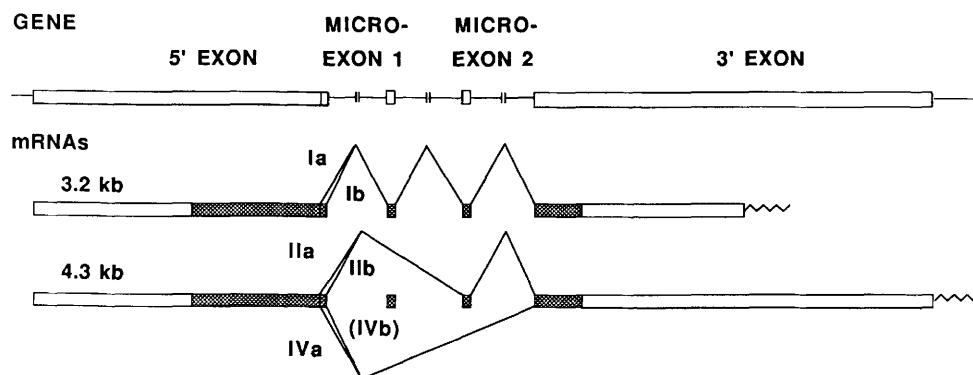


Figure 7. Summary of *Ubx* mRNA structures. *Ubx* genomic DNA (GENE) is represented above: exons (named, open boxes) are drawn to scale, whereas introns (lines) are not drawn to scale. *Ubx* mRNAs are illustrated below. Exons are shown as above; the coding region is stippled and a wavy line represents the poly(A) tail. Lines mark splice sites joined in the indicated spliced forms. The 3.2-kb size class contains predominantly spliced forms Ia and Ib, whereas the 4.3-kb size class contains predominantly IIa + IIb and IVa + IVb. Form IVb is shown in parenthesis because a corresponding cDNA has not been identified although O'Conner et al. (1988) detected this form by nuclease S1 protection analysis in cultured neuroblasts.

protein (Fig. 2) purified from an overproducing strain of *Escherichia coli* has been shown to bind to specific DNA sequences near the *Ubx* and *Antp* P1 promoters (Beachy et al. 1988). The regulatory role of this protein has been further indicated by the observation that its expression from constructs introduced into cultured *Drosophila* cells results in the induction and repression, respectively, of the *Ubx* and *Antp* P1 promoters introduced into these cells in the form of hybrid genes consisting of a reporter gene fused to the respective promoter regions (M. Krasnow and D.S. Hogness, in prep.).

One of us (K.K.) has used this cell culture system to show that expression of the other members of the *Ubx* protein family (i.e., Ia, IIa, IIb, IVa; Figs. 2 and 3) has the same qualitative effect on the *Ubx* and *Antp* P1 promoters as the Ib protein. Presumably, these regulatory functions derive from DNA-binding domains in each of the *Ubx* proteins. This presumption is consistent with the fact that the carboxy-terminal constant region in each protein contains a homeo domain (Figs. 2 and 3) and with studies indicating that the homeo domains in the *engrailed* and *Antp* proteins provide at least part of their DNA-binding domains (Desplan et al. 1985; Gehring 1987, and pers. comm.).

It is tempting to speculate that different members of the *Ubx* protein family may exhibit different DNA-binding and/or transcriptional regulatory capacities as a result of the modulating effects of the variable region on the functions of the homeo domain. Linking the amino-terminal and carboxy-terminal constant regions of these proteins, the variable region consists of different combinations of three optional elements of 9, 17, and 17 amino acids and is separated by only 4 amino acids from the homeo domain (Figs. 2 and 3). This postulated modulation might affect the DNA sequence recognition properties of the individual *Ubx* proteins or their ability to interact cooperatively with other proteins, including members of the *Ubx* family, in effecting their different regulatory functions.

What regulatory functions of the *Ubx* proteins might be differentially affected by their different structures? Given their quite similar, if not identical, effects upon the *Ubx* and *Antp* P1 promoters in the cell culture system noted above, we suspect that the ability to regulate homeotic gene expression is a function common to all members of the family. We therefore focus on the possibility that different *Ubx* proteins are required for the regulation of different effector genes and consider two extreme models for such differential regulatory functions. In the first, 'metamere-specific' model, different combinations of *Ubx* proteins are required in different metameres of a given tissue to effect their appropriate identities. In the second, 'tissue-specific' model, different combinations of *Ubx* proteins are required to mediate the *Ubx* identity functions in different tissues.

The existing data argue against the metamere-specific model and support the tissue-specific model. Several observations indicate that the identity of a given epidermal metamere can be transformed by changing the amount rather than the kind of *Ubx* proteins in that metamere.

For example, *bx*d mutations within the *cis*-regulatory region upstream of the *Ubx* transcription unit transform the identity of ps6 to that of ps5. These mutations would not be expected to change the splicing pathways of the *Ubx* transcript and, hence, the kinds of *Ubx* proteins present in ps6. They have, however, been observed to decrease the level of total *Ubx* protein (as detected by antibodies that react with all *Ubx* proteins) toward the lower levels normally found in ps5 (Beachy et al. 1985; Hogness et al. 1985; S. Helfand and D.S. Hogness, unpubl.). Apparently, the differences in the epidermal identities of ps5 and ps6 do not depend on different combinations of *Ubx* proteins but on differences in the amount of a given combination. Similarly, the identity transformation of ps5 toward ps6 observed in flies carrying eight copies of the *Ubx* gene (including its upstream regulatory region) instead of two (S. Smolik-Utlaut, pers. comm.) would be expected to result from an increase in the amount of *Ubx* proteins rather than from a change in kind.

Support for the tissue-specific model derives from the finding that mass isolated embryonic muscle precursor cells are enriched for mRNAs that encode form I proteins and that similarly isolated embryonic neuroblasts are enriched for mRNAs that encode form II and IV proteins (O'Connor et al. 1988). This observation is consistent with an earlier finding that the *Ubx*¹⁹⁵ mutation (Fig. 3), which contains a termination codon in microexon 2 that prevents synthesis of form I and II proteins but not form IV, retains *Ubx* identity functions for ps5 and ps6 of the adult CNS but acts like a *Ubx* null mutation in other tissues (Weinzierl et al. 1987). Apparently, the form IV protein is a sufficient provider of *Ubx* identity function for the adult CNS but not for other tissues. Preliminary immunocytochemical localization studies of embryos with monoclonal antibodies specific for forms I, IV, and I + II show that form I is located primarily in the epidermis and mesoderm, whereas forms II and IV are located primarily in the CNS (J. Lopez and D.S. Hogness, unpubl.).

The existing data thus suggest that although the different *Ubx* proteins may share certain regulatory functions, as may be the case for their regulation of homeotic gene expression, structural differences in their variable regions may be required for appropriate regulation of effector genes in different tissues. This requirement may result from a regulatory mechanism in which the *Ubx* proteins must interact with tissue-specific factors for appropriate control of effector gene expression. Within a given tissue, the different metameric identity functions appear to result from the amount and spatial distribution of a given combination of *Ubx* proteins.

The correlation between alternative splicing and polyadenylation

One of the more interesting results of the experiments reported here is our finding that electrophoretic fractionation of mid-embryonic RNAs for the two size classes of *Ubx* mRNAs resulted in coincident fractionation of the

different spliced forms, demonstrating that the 3.2-kb mRNA consists primarily or exclusively of forms Ia and Ib, whereas the 4.3-kb mRNA consists primarily or exclusively of forms II and IV. The observation that the ratio of the 3.2- to 4.3-kb mRNAs parallels the ratio of (Ia + Ib) to (II + IV) at each developmental stage suggests that this correlation between polyadenylation and splicing pathways persists throughout development.

This distribution of spliced forms between the two size classes of mRNAs nicely parallels the distribution of *Ubx* protein forms among the tissues (J. Lopez and D.S. Hogness, unpubl.), leading to the expectation that *Ubx* mRNAs in the epidermis and mesoderm will be primarily of the 3.2-kb size class, whereas those in the CNS will be primarily of the 4.3-kb size class. This expectation is met in the case of the mass isolated embryonic muscle precursor cells and embryonic neuroblasts examined by O'Connor et al. (1988) and is consistent with the finding of 4.3-kb mRNAs in the embryonic CNS by in situ hybridization (Akam and Martinez-Arias 1985).

To our knowledge, this is the only example of correlations between splicing and polyadenylation pathways where each splicing pathway is theoretically possible irrespective of which polyadenylation site is used. In all other examples where a correlation exists, two processing events are mutually exclusive because cleavage at the upstream poly(A) site, which is in an intron, eliminates one or more downstream exons (for review, see Leff et al. 1986). The distinction is important because the *Ubx* example raises the interesting possibility that the first event (either splicing or polyadenylation) may bias the site choice of the second event.

If the two events are indeed causally related in this manner, it would allow an economy in the mechanism for the tissue-specific regulation of *Ubx* RNA processing. An interesting causal model of this sort proposes that polyadenylation occurs first and that the 4.3-kb poly(A) site is chosen over the 3.2-kb site only in the presence of tissue-specific factors. When the 4.3-kb poly(A) site is chosen, 3'-proximal sequences unique to this class prevent splicing into microexon 1—perhaps, by specific base pairing that blocks the microexon 1 splice site—so that only forms II and IV are produced. In the absence of such factors, the 3.2-kb poly(A) site is chosen and all introns are spliced out to yield form I.

The function of the long transcription units of homeotic genes

The *Ubx* transcription unit, measured from the common initiation sites to the polyadenylation site for the 4.3-kb mRNA, is 77 kb long. Remarkably, this is only the second longest transcription unit among the eight homeotic genes of the Antennapedia and Bithorax complexes which, as a class, exhibit transcription units that are atypically long for *D. melanogaster* genes. They range from 12 kb for *Deformed (Dfd)* (Regulski et al. 1987) to the 103-kb *Antp P1* unit (Laughon et al. 1986) and exhibit a mean length of at least 40 kb. [*Antp P2*

unit, 36 kb; *Abd-B*, >34 kb (Sanchez-Herrero and Crosby 1988; S. Sakonju, pers. comm.); *proboscipedia*, 34 kb (Pultz et al. 1988; D. Cribbs and T. Kaufman, pers. comm.); *Sex combs reduced*, >25 kb (T. Kaufman, pers. comm.); *abd-A*, >23 kb (S. Sakonju, pers. comm.); and *labial*, 17 kb (R. Diederich and T. Kaufman, pers. comm.)].

We propose that the lengths of the homeotic transcription units function as delay timers that stage the times of appearance of the proteins they encode subsequent to the activation of their promoters in a manner critical to segmental development. First, we consider whether the delay times caused by transcription are significant relative to the interval between the time of appearance of transcripts and proteins—a period that we refer to as the 'interval time.'

An elongation rate of 1.1 kb/min at 25°C was determined in vivo for the 60-kb ecdysone-inducible E74 gene of *D. melanogaster* to an accuracy of $\pm 30\%$ (C. Thummel, K. Burtis, and D.S. Hogness, in prep.). Assuming this rate is generalizable to other *Drosophila* genes, transcription of the 77-kb *Ubx* unit would be expected to contribute 70 min to the delay interval between the time its promoter was activated and the time *Ubx* protein first appeared. Akam and Martinez-Arias (1985) detect strong accumulation of *Ubx* transcripts in ps6 of 'late cellular blastoderm embryos' before the onset of gastrulation, some time between 2.5 and 3.0 hr postfertilization, whereas a similar accumulation of *Ubx* proteins in ps6 is detected just before extended germ band at 4.5 hr (White and Lehmann 1986; J. Lopez and D.S. Hogness, unpubl.). This interval time is no more than 2 hr and probably closer to 1.5 hr. Although imprecisely determined, its accuracy is sufficient to indicate that the 70-min delay due to transcription is an appreciable fraction of the interval time.

The difference between these interval and transcription delay times for *Ubx* (35 ± 15 min) is remarkably close to the 30-min interval time for the *Krüppel (Kr)* gene, a gap segmentation gene whose transcription unit of 2.5 kb (Rosenberg et al. 1986) is so short that most of this interval time must result from RNA processing, translation, and differences in sensitivity of the in situ assays for transcript and protein. [*Kr* transcripts were detected shortly after the eleventh syncytial nuclear division (i.e., shortly after 110 min postfertilization; Knipple et al. 1985), and *Kr* proteins were detected at the thirteenth nuclear division (144 min postfertilization; Gaul et al. 1987).] Apparently, these factors contribute approximately equal times for the long and short genes.

We should expect the interval times to decrease linearly with decreasing length of the homeotic transcription unit, so that for the smallest 12-kb *Dfd* unit the interval time should closely approach the half-hour interval observed for the 2.5-kb *Kr* unit. This appears to be the case as the *Dfd* transcripts were first detected after the thirteenth nuclear division (144 min; Chadwick and McGinnis 1987), whereas the *Dfd* protein was first detected in 3-hr cellular blastoderm embryos (Jack et al. 1988).

What functions might be served by the delays imposed by the long transcription units of the homeotic genes on the appearance of the proteins they encode? One function might be to separate the temporal phase when the homeotic genes cross regulate one another and begin to specify segmental identities from that when the earlier and transiently expressed segmentation genes of the gap and pair-rule classes similarly interact to establish highly resolved patterns of spatial expression (for review, see Akam 1987; Scott and Carroll 1987).

As would be expected for such an early and rapidly interactive set, the mapped transcription units among these segmentation genes are strikingly shorter than those for the homeotic genes. [1.5 kb—*even-skipped*, *eve* (MacDonald et al. 1986); 2.1 kb—*fushi tarazu*, *ftz* (Kuroiwa et al. 1984); 2.5 kb—*Kr* (Rosenberg et al. 1986); 3.0 kb—*paired*, *prd* (Frigerio et al. 1986); 3.4 kb—*hunchback*, *hb* (Tautz et al. 1987)]. It can be safely predicted that the other members of this set will exhibit similarly short transcription units because the heterogeneity of response times resulting from the introduction of long transcription units to the set would not be tolerated. As noted earlier, some of the proteins that are rapidly produced by these segmentation genes appear to regulate the homeotic genes to determine their initial expression pattern. Furthermore, some of these proteins carry homeo domains closely related to those encoded by the homeotic genes (e.g., the *ftz* protein) by which that regulation may be presumed to be effected. Hence, the separation of the temporal phases of protein expression for these segmentation and homeotic genes may serve to minimize the number of regulatory proteins that must be accommodated at a particular time.

At a higher level of temporal resolution, the differential delay times effected by the different lengths of the homeotic transcription units may be an important factor in determining the refinement of the patterns of their expression that results from their cross regulation. Finally, we note that little is known about how the post-embryonic changes in homeotic gene expression are effected, and the delay times caused by their long transcription units may again be employed in these transitions.

Materials and methods

Strains

Wild-type strains of *D. melanogaster* were Canton-S (C-S) or Oregon-R (O-R). Mutants *bx³/Xasta*, *pbx¹/Xasta*, and *pbx²/TM1* were provided by E. Lewis (Caltech). *bx³/bx³* and *pbx¹/pbx²* flies were selected by adult phenotype and maintained as stocks.

RNA preparation

Adult flies were fed 3 hr prior to egg collection to minimize egg withholding; staged embryos were maintained at 25°C. Imaginal discs, prepared by a batch isolation procedure (Eugene and Fristrom 1978), were a gift of J. Fristrom. Total cellular RNA was prepared by homogenization of tissues in 10 volumes of 6 M guanidine hydrochloride, 0.1 M NaOAc (pH 5.2) and centrifuga-

tion through 4.8 M CsCl in 10 mM EDTA (pH 8.5). Pellets were resuspended, ethanol precipitated, and dissolved in H₂O. Poly(A)⁺ RNA was selected using columns of oligo(dT)-cellulose (Collaborative Research).

Preparation and hybridization of Northern blots

Thirty micrograms of total cellular RNA was denatured by incubation in 1 M glyoxal at 50°C for 1 hr and electrophoresed in a 1% agarose gel in 10 mM NaH₂PO₄ (pH 7.0) at 1 V/cm overnight. Deglyoxylation and transfer to activated cellulose paper were carried out according to Alwine et al. (1977), except that DPT-cellulose paper (Seed 1982) replaced DBM-cellulose paper.

For synthesis of single-stranded hybridization probes, pBR322 (Bolivar et al. 1977) was modified by replacing sequences between the *Hind*III and *Bam*HI sites with the *Hae*III fragment of ϕ X174 (from 4199 to 4477 bp from the *Pst*I cleavage site) using mixed *Bam*HI and *Hind*III linker ligation (R. Mulligan, unpubl.). This plasmid, named p ϕ X, contains the ϕ X174 origin of viral strand synthesis in an anticlockwise direction on the standard orientation of pBR322. Fragments of Charon 4 genomic clones (Bender et al. 1983b) subcloned into p ϕ X were labeled in vitro by the addition of single-stranded binding protein, gene A protein, *rep* protein, DNA polymerase III holoenzyme, [³²P]dNTPs, and buffered salts (Arai et al. 1980). Purified proteins were generous gifts of members of A. Kornberg's research group. The resulting labeled, covalently closed, single-stranded probes were digested in 0.1% SDS and 0.1 mg/ml proteinase K at 50°C for 1 hr and separated from dNTPs by Bio-Gel P-60 column chromatography.

Filters were incubated with 50% formamide, 5× SSPE, and 0.5 mg/ml sonicated denatured salmon sperm DNA for at least 4 hr. Hybridizations with labeled probe were carried out in 5× SSPE, 50 μ g/ml sonicated denatured salmon sperm DNA, 0.2% SDS, and 10% (wt/vol) dextran sulfate at 42°C for 4–12 hr. Filters were washed briefly in 2× SSPE, 0.2% SDS at 25°C and three times in 0.2× SSPE, 0.2% SDS at 68°C for a total of 3–6 hr, blotted dry, and exposed to Kodak X-Omat R film with screen intensification. Single-stranded probes were >10 times more sensitive than nick-translated, double-stranded probes of similar specific activity used under the same conditions (data not shown). Preparation of phage and plasmid DNA and construction of plasmid subclones followed standard protocols (Davis et al. 1980; Maniatis et al. 1982).

cDNA isolation and analysis

Lambda gt10 cDNA libraries made with 1.5- to 5-hr and 3- to 12-hr O-R embryonic RNA (gifts of M. Goldschmidt-Clermont and L. Kauvar, respectively; Poole et al. 1985) were screened with nick-translated, genomic DNA inserts of plasmids p ϕ X 3108 and p ϕ X 3144, which overlap the 5' and 3' *Ubx* exons, respectively (Bender et al. 1983b). Clone 3712 (1.5- to 5-hr) was completely sequenced, 100% on both strands (form Ib). Clone 3602 (1.5- to 5-hr) was sequenced at its 5' and 3' termini and in the variable region on one strand (form Ia). Thirty-five cDNAs (3- to 12-hr) were restriction mapped, and parts of three were sequenced on one strand: EC13 variable region (form IIa), EC1 variable region (form IVa), and EC14 3' terminus [poly(A) stretch, form Ia]. We made a λ gt10 cDNA library with 12- to 18-hr, poly(A)⁺ embryonic RNA (Huynh et al. 1984). This and a 12- to 24-hr O-R library (gift of L. Kauvar) were screened with a nick-translated, genomic DNA fragment that hybridizes only to 4.3-kb RNAs (Akam and Martinez-Arias 1985). cDNA 3801 (12- to 18-hr) was sequenced entirely, 100% on both strands.

Kornfeld et al.

Determination of exon positions within the Ubx unit

Labeled cDNAs were hybridized to DNA clones that span the *Ubx* unit (Bender et al. 1983a; data not shown). Microexons 1 and 2 mapped to positions -39.6 to -41.8 , and -51.2 to -54.7 , respectively. Small libraries in M13 mp8 (Messing 1983) were generated with sonicated genomic DNA from these regions, plaques were selected using a labeled cDNA fragment as a hybridization probe, and the inserts were sequenced on one strand.

DNA sequence analysis

DNA fragments subcloned by random or directed methods were sequenced using the dideoxy chain termination method (Sanger et al. 1977; Biggin et al. 1983). Genomic DNA overlapping the 3' exon was sequenced 63% on both strands, whereas that overlapping the 5' exon was sequenced 78% on both strands. Whenever a genomic sequence derived from only one strand, several independent clones were sequenced.

Nuclease S1 protection and primer extension

Nuclease S1 mapping and primer extension were done with uniformly labeled, single-stranded M13 extension products, as described previously (Lipshitz et al. 1987). Single-stranded probes were made with [32 P]dCTP and unlabeled dGTP, dATP, and dTTP. The S1 protection analysis probe that spans the 5' exon splice sites was made from a 655-bp, genomic *Taq* fragment, starting at position +1289 (Fig. 2); RNAs containing the b element protect a 444-base probe fragment, whereas RNAs lacking this element protect a 417-base probe fragment. To map the transcription initiation site, a 270-base *HaeIII*-*EcoRI* primer starting at position +354 (Fig. 2) was used. Extension products were compared with a dideoxy sequencing ladder to determine the initiation sites at +1 and +3. A nuclease S1 protection experiment using a 1454-base probe extending from the *HindIII* site (-1100) to the *EcoRI* site (+354, Fig. 2) protected fragments that confirmed this assignment.

Size separation, extraction, and analysis of embryonic RNAs

An agarose-urea gel system (Locker 1979) was used to separate 4.3-kb RNAs from 3.2-kb RNAs because RNA can be efficiently recovered from these gels. Sixty micrograms of poly(A) $^{+}$, 15–18 hr embryonic RNA was mixed with 45 μ g of poly(A) $^{+}$, 10–18 hr embryonic RNA and electrophoresed in a 1% agarose-6 M urea preparative slab gel for 12 hr at 5 V/cm (Sehgal 1981). The gel was sliced at 0.5-cm intervals beginning 4 cm from the well and extending to 8 cm. tRNA (50 μ g) was added to each gel slice as carrier, and the RNA was extracted as a quarternary ammonium salt (3.67 g QN+/100 ml butanol), according to Langridge et al. (1980). The relative amount of each *Ubx* spliced form was quantitated and expressed as a percentage of total *Ubx* mRNA, as described in Figure 4. The relative amount of each *Ubx* mRNA size class was determined by nuclease S1 protection analysis using a single-stranded, uniformly labeled, genomic DNA probe that extended from the *XmnI* site (+2315) to the *BamHI* site (+3385) in the 3' exon (Fig. 2). RNAs of 4.3 kb protect 1070 bases, whereas RNAs of 3.2-kb protect a doublet of 830 and 800 bases (data not shown). Signal intensities for each size class were quantitated by scanning densitometry, corrected for probe length, which is proportional to the number of labeled residues in the probe, and expressed as a percentage of the total *Ubx* mRNA. Data for four

gel slices, which include the peak occurrence of each size class are shown below:

Sample	mRNA size class (kb)		Percentage of mRNAs containing spliced form			
	3.2	4.3	Ib	Ia	IIa + IIb	IVa + IVb
A	23	77	11	15	48	25
B	28	72	7	15	54	24
C	54	46	16	34	33	17
D	61	39	13	41	34	11

Samples A–D are gel slices at distances of 5.5–6.0, 6.0–6.5, 6.5–7.0, and 7.0–7.5 cm from the well, respectively. The first two values—percentage of each size class—sum to 100%. The next four values—percentage of each spliced form—also sum to 100%. The spliced forms that comprise pure 3.2- and 4.3-kb RNAs were calculated from these data. Below is an example of our computations solving for the percentage of mRNAs in each size class that contain spliced form Ia using data from samples A and D:

Let x = Percentage of 3.2-kb mRNAs containing Ia. Let y = Percentage of 4.3-kb mRNAs containing Ia.

$$0.23(x) + 0.77(y) = 0.15$$

$$0.61(x) + 0.39(y) = 0.41$$

Data from any two gel slices are sufficient to solve for the two unknown values x and y . The equations were solved with data from four pairs of gel slices: A and C, A and D, B and C, B and D. The average values are listed in Table 1.

Acknowledgments

We thank Michel Goldschmidt-Clermont and Laurence Kauvar for gifts of cDNA libraries, James Fristrom for a gift of imaginal discs, and members of Arthur Kornberg's group for purified enzymes. Louise Prestidge and Ken Relloma assisted in DNA sequencing. We are indebted to past and current members of the group for their advice, assistance, and critical comments about this manuscript. This work was supported by fellowships from the Medical Scientist Training Program (K.K.), the Damon Runyon-Walter Winchell Cancer Fund (R.B.S.), the National Science Foundation (P.A.B.), the Jane Coffin Childs Foundation (P.J.H.), the Helen Hay Whitney Foundation (D.A.P.), and a National Institutes of Health grant to D.S.H.

References

- Akam, M.E. 1983. The location of Ultrabithorax transcripts in *Drosophila* tissue sections. *EMBO J.* **2**: 2075–2084.
- . 1987. The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* **101**: 1–22.
- Akam, M.E. and A. Martinez-Arias. 1985. The distribution of Ultrabithorax transcripts in *Drosophila* embryos. *EMBO J.* **4**: 1689–1700.
- Alwine, J.C., D.J. Kemp, and G.R. Stark. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci.* **74**: 5350–5354.
- Arai, K., N. Arai, J. Schlomai, and A. Kornberg. 1980. Replication of duplex DNA of phage ϕ X174 reconstituted with purified enzymes. *Proc. Natl. Acad. Sci.* **77**: 3322–3326.
- Beachy, P.A., S.L. Helfand, and D.S. Hogness. 1985. Segmental distribution of bithorax complex proteins during *Drosophila* development. *Nature* **313**: 545–551.
- Beachy, P.A., M.A. Krasnow, E.R. Gavis, and D.S. Hogness.

1988. An Ultrabithorax protein binds sequences near its own and the Antennapedia P1 promoters. *Cell* **55**: 1069–1081.
- Bender, W., P. Spierer, and D.S. Hogness. 1983a. Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in *Drosophila melanogaster*. *J. Mol. Biol.* **168**: 17–33.
- Bender, W., M. Akam, F. Karch, P.A. Beachy, M. Peifer, P. Spierer, E.B. Lewis, and D.S. Hogness. 1983b. Molecular genetics of the Bithorax Complex of *Drosophila melanogaster*. *Science* **221**: 23–29.
- Bender, W., B. Weiffenbach, F. Karch, and M. Peifer. 1985. Domains of cis-interaction in the bithorax complex. *Cold Spring Harbor Symp. Quant. Biol.* **50**: 173–180.
- Biggin, M.D., T.J. Gibson, and G.F. Hong. 1983. Buffer gradient gels and 35S label as an aid to rapid DNA sequencing. *Proc. Natl. Acad. Sci.* **80**: 3963–3965.
- Bolivar, F., R.L. Rodriguez, P.J. Greene, M.C. Beltach, H.L. Heynecker, H.W. Boyer, J.H. Crosa, and S. Falkow. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**: 95–113.
- Cabrera, C.V., J. Botas, and A. Garcia-Bellido. 1985. Distribution of Ultrabithorax proteins in mutants of *Drosophila bithorax* complex and its transregulatory genes. *Nature* **318**: 569–571.
- Carroll, S.B., R.A. Laymon, M.A. McCutcheon, P.D. Riley, and M.P. Scott. 1986. The localization and regulation of Antennapedia protein expression in *Drosophila* embryos. *Cell* **47**: 113–122.
- Chadwick, R. and W. McGinnis. 1987. Temporal and spatial distribution of transcripts from the Deformed gene of *Drosophila*. *EMBO J.* **6**: 779–789.
- Davis, R.W., D. Botstein, and J.R. Roth. 1980. *Advanced bacterial genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Desplan, C., J. Theis, and P.H. O'Farrell. 1985. The *Drosophila* developmental gene, engrailed, encodes a sequence-specific DNA binding activity. *Nature* **318**: 630–635.
- Duncan, I. 1987. The Bithorax Complex. *Annu. Rev. Genet.* **21**: 285–319.
- Eugene, O.M. and J.W. Fristrom. 1978. The mass isolation of imaginal discs. In *The genetics and biology of Drosophila* (ed. M. Ashburner and T.R.F. Wright), vol. 2a, pp. 121–126, Academic Press, London.
- Frigerio, G., M. Burri, D. Bopp, S. Baumgartner, and M. Noll. 1986. Structure of the segmentation gene paired and the *Drosophila* PRD gene set as part of a gene network. *Cell* **47**: 735–746.
- Garcia-Bellido, A., P. Ripoll, and G. Morata. 1973. Developmental compartmentalization of the wing disc of *Drosophila*. *Nat. New Biol.* **245**: 241–253.
- Gaul, U., E. Seifert, R. Schuh, and H. Jäckle. 1987. Analysis of Krüppel protein distribution during early *Drosophila* development reveals posttranscriptional regulation. *Cell* **50**: 639–647.
- Gehring, W.J. 1987. Homeo boxes in the study of development. *Science* **236**: 1245–1252.
- Hafen, E., M. Levine, and W.J. Gehring. 1984. Regulation of Antennapedia transcript distribution by the bithorax complex in *Drosophila*. *Nature* **307**: 287–289.
- Hayes, P.H., T. Sato, and R.E. Denell. 1984. Homoeosis in *Drosophila*: The Ultrabithorax larval syndrome. *Proc. Natl. Acad. Sci.* **81**: 545–549.
- Hogness, D.S., H.D. Lipshitz, P.A. Beachy, D.A. Peattie, R.B. Saint, M. Goldschmidt-Clermont, P.J. Harte, E.R. Gavis, and S.L. Helfand. 1985. Regulation and products of the *Ubx* domain of the bithorax complex. *Cold Spring Harbor Symp. Quant. Biol.* **50**: 181–194.
- Hooper, J.E. 1986. Homeotic gene function in the muscles of *Drosophila* larvae. *EMBO J.* **5**: 2321–2329.
- Huynh, T.V., R.A. Young, and R.W. Davis. 1984. Constructing and screening cDNA libraries in lambda gt10 and lambda gt11. In *DNA cloning techniques: A practical approach* (ed. D. Glover), pp. 49–78. IRL Press, Oxford.
- Jack, T., M. Regulski, and W. McGinnis. 1988. Pair-rule segmentation genes regulate the expression of the homeotic selector gene, *Deformed*. *Genes Dev.* **2**: 635–651.
- Knipple, D.C., E. Seifert, U.B. Rosenberg, A. Preiss, and H. Jäckle. 1985. Spatial and temporal patterns of Krüppel gene expression in early *Drosophila* embryos. *Nature* **317**: 40–44.
- Kuroiwa, A., E. Hafen, and W.J. Gehring. 1984. Cloning and transcriptional analysis of the segmentation gene fushi tarazu of *Drosophila*. *Cell* **37**: 825–831.
- Langridge, J., P. Langridge, and P.L. Bergquist. 1980. Extraction of nucleic acids from agarose gels. *Anal. Biochem.* **103**: 264–271.
- Laughon, A. and M.P. Scott. 1984. Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. *Nature* **30**: 25–31.
- Laughon, A., A.M. Boulet, J.R. Birmingham, R.A. Laymon, and M.P. Scott. 1986. The structure of transcripts from the homeotic Antennapedia gene of *Drosophila*: Two promoters control the major protein-coding region. *Mol. Cell. Biol.* **6**: 4676–4689.
- Lawrence, P.A. 1981. The cellular basis of segmentation in insects. *Cell* **29**: 3–10.
- Leff, S.E., M.G. Rosenfeld, and R.M. Evans. 1986. Complex transcriptional units; diversity in gene expression by alternative RNA processing. *Annu. Rev. Biochem.* **55**: 1091–1117.
- Lewis, E.B. 1963. Genes and developmental pathways. *Am. Zool.* **3**: 33–56.
- . 1964. Genetic control and regulation of developmental pathways. In *Role of chromosomes in development* (ed. M. Locke), pp. 231–252. Academic Press, New York.
- . 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**: 565–570.
- . 1981. Developmental genetics of the bithorax complex of *Drosophila*. In *Developmental biology using purified genes, ICN-UCLA symposium on molecular and cellular biology* (eds. D.D. Brown and C.C. Fox), vol. 23, pp. 189–208. Academic Press, New York.
- . 1982. Control of body segment differentiation in *Drosophila* by the bithorax gene complex. In *Embryonic development, part A: Genetic aspects* (ed. M. Burger), pp. 269–288. A.R. Liss, New York.
- Lipshitz, H.D., D.A. Peattie, and D.S. Hogness. 1987. Novel transcripts from the Ultrabithorax domain of the bithorax complex. *Genes Dev.* **1**: 307–322.
- Locker, J. 1979. Analytical and preparative electrophoresis of RNA in agarose-urea. *Anal. Biochem.* **98**: 358–367.
- MacDonald, P.M., P. Ingham, and G. Struhl. 1986. Isolation, structure and expression of even-skipped: A second pair-rule gene of *Drosophila* containing a homeobox. *Cell* **47**: 721–734.
- Maniatis, T., E.F. Fritsh, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Martinez-Arias, A. and P.A. Lawrence. 1985. Parasegments and compartments in the *Drosophila* embryo. *Nature* **313**: 639–642.
- Messing, J. 1983. New M13 vectors for cloning. *Methods En-*

Kornfeld et al.

- zymol.* **101**: 20–78.
- Minana, F.J. and A. Garcia-Bellido. 1982. Preblastoderm mosaics of mutants of the bithorax-complex. *Wilhelm Roux's Arch. Dev. Biol.* **191**: 331–334.
- Morata, G. and S. Kerridge. 1981. Sequential functions of the bithorax complex of *Drosophila*. *Nature* **290**: 778–791.
- O'Connor, M.B., R. Binari, L.A. Perkins, and W. Bender. 1988. Alternative RNA products from the Ultrabithorax domain of the bithorax complex. *EMBO J.* **7**: 435–445.
- Poole, S.J., L.M. Kauvar, B. Drees, and T. Kornberg. 1985. The engrailed locus of *Drosophila*: Structural analysis of an embryonic transcript. *Cell* **40**: 37–43.
- Proudfoot, N.J., and G.G. Brownlee. 1976. 3' non-coding region sequences in eucaryotic mRNA. *Nature* **263**: 211–214.
- Pultz, M.A., R.J. Diederich, D.L. Cribbs, and T.C. Kaufman. 1988. The proboscipedia locus of the Antennapedia Complex: A molecular and genetic analysis. *Genes Dev.* **2**: 901–920.
- Regulski, M., N. McGinnis, R. Chadwick, and W. McGinnis. 1987. Developmental and molecular analysis of Deformed, a homeotic gene controlling *Drosophila* head development. *EMBO J.* **6**: 767–777.
- Regulski, M., K. Harding, R. Kostriken, F. Karch, M. Levine, and W. McGinnis. 1985. Homeo box genes of the Antennapedia and bithorax complexes of *Drosophila*. *Cell* **43**: 71–80.
- Rosenberg, U.B., C. Schroder, A. Preiss, A. Kienlin, S. Cote, I. Riede, and H. Jäckle. 1986. Structural homology of the product of the *Drosophila* Kruppel gene with *Xenopus* transcription factor IIIA. *Nature* **319**: 336–339.
- Saari, G. and M. Bienz. 1987. The structure of the Ultrabithorax promoter of *Drosophila melanogaster*. *EMBO J.* **6**: 1775–1779.
- Sanchez-Herrero, E. and M.A. Crosby. 1988. The Abdominal-B gene of *Drosophila melanogaster*: overlapping transcripts exhibit two different spatial distributions. *EMBO J.* **7**: 2163–2173.
- Sanchez-Herrero, E., J. Casanova, S. Kerridge, and G. Morata. 1985a. Anatomy and function of the bithorax complex of *Drosophila*. *Cold Spring Harbor Symp. Quant. Biol.* **50**: 165–172.
- Sanchez-Herrero, E., I. Vernos, R. Marco, and G. Morata. 1985b. Genetic organization of *Drosophila* bithorax complex. *Nature* **313**: 108–113.
- Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**: 5463–5467.
- Scott, M.P. and S.B. Carroll. 1987. The segmentation and homeotic gene network in early *Drosophila* development. *Cell* **51**: 689–698.
- Seed, B. 1982. Attachment of nucleic acids to nitrocellulose and diazonium-substituted supports. In *Genetic engineering principles and methods* (ed. J.K. Setlow and A. Hollander), vol. 4, pp. 91–102. Plenum Press, New York, London.
- Sehgal, P.B. 1981. Procedures to estimate the size of interferon mRNA and use of U.V. irradiation to estimate the size of its primary transcript. *Methods Enzymol.* **79**: 111–124.
- Struhl, G. 1984. Splitting the bithorax complex of *Drosophila*. *Nature* **308**: 454–457.
- Struhl, G. and R.A.H. White. 1985. Regulation of the Ultrabithorax gene of *Drosophila* by other bithorax complex genes. *Cell* **43**: 507–519.
- Tautz, D., R. Lehmann, H. Schnurch, R. Schuh, E. Seifert, K. Kienlin, and H. Jäckle. 1987. Finger protein of novel structure encoded by hunchback, a second member of the gap class of *Drosophila* segmentation genes. *Nature* **327**: 383–389.
- Teugels, E. and A. Ghysen. 1985. Domains of action of bithorax genes in *Drosophila* central nervous system. *Nature* **314**: 558–561.
- Tiong, S., L.M. Bone, and J.R.S. Whittle. 1985. Recessive lethal mutations within the bithorax-complex in *Drosophila*. *Mol. Gen. Genet.* **200**: 335–342.
- Weinzierl, R., J.M. Axton, A. Ghysen, and M. Akam. 1987. *Ultrabithorax* mutations in common and variable regions of the protein coding sequence. *Genes Dev.* **1**: 386–397.
- White, R.A.H. and R. Lehmann. 1986. A gap gene, hunchback, regulates the spatial expression of Ultrabithorax. *Cell* **47**: 311–321.
- White, R.A.H. and M. Wilcox. 1984. Protein products of the bithorax complex of *Drosophila*. *Cell* **39**: 163–171.
- . 1985a. Regulation of the distribution of Ultrabithorax proteins in *Drosophila*. *Nature* **318**: 563–567.
- . 1985b. Distribution of Ultrabithorax proteins in *Drosophila*. *EMBO J.* **4**: 2035–2043.