Evolutionary Conservation of the Structure and Expression of Alternatively Spliced Ultrabithorax Isoforms From Drosophila

Howard M. Bomze and A. Javier López

Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Manuscript received July 30, 1993
Accepted for publication November 8, 1993

ABSTRACT

In Drosophila melanogaster, alternatively spliced mRNAs from the homeotic gene Ultrabithorax (Ubx) encode a family of structurally distinct homeoprotein isoforms. The developmentally regulated expression patterns of these isoforms suggest that they have specialized stage- and tissue-specific functions. To evaluate the functional importance of UBX isoform diversity and gain clues to the mechanism that regulates processing of Ubx RNAs, we have investigated whether the Ubx RNAs of other insects undergo similar alternative splicing. We have isolated and characterized Ubx cDNA fragments from D. melanogaster, Drosophila pseudoobscura, Drosophila hydei and Drosophila virilis, species separated by as much as 60 million years of evolution, and have found that three aspects of Ubx RNA processing have been conserved.

1. These four species exhibit identical patterns of optional exon use in a region adjacent to the homeodomain.
2. These four species produce the same family of UBX protein isoforms with identical amino acid sequences in the optional exons, even though the common amino-proximal region has undergone substantial divergence. The nucleotide sequences of the optional exons, including third positions of rare codons, have also been conserved strongly, suggesting functional constraints that are not limited to coding potential.
3. The tissue- and stage-specific patterns of expression of different UBX isoforms are identical among these Drosophila species, indicating that the developmental regulation of the alternative splicing events has also been conserved. These findings argue for an important role of alternative splicing in Ubx function. We discuss the implications of these results for models of UBX protein function and the mechanism of alternative splicing.

THE body plans of many animals are subdivided along the anteroposterior axis into metameric units such as the segments of insects and other arthropods. In the fruit fly Drosophila melanogaster, the unique characteristics that distinguish one segment from another are controlled by homeotic genes (Lewis 1978). Mutations that reduce the expression of a homeotic gene or modify its expression pattern do not disrupt segmentation itself, rather they alter the development of particular segments so that their morphology reiterates that of other segments. Although many homeotic loci are scattered throughout the Drosophila genome, a set of eight structurally related genes residing in the Antennapedia and bithorax complexes occupy a central position in this system of genetic controls [reviewed by Duncan (1987) and Kaufman et al. (1990)]. Referred to collectively as the Homeotic complex, or HOM-C [reviewed by Aram (1989)], these eight genes encode proteins that share similar versions of the homeodomain, a 60 amino acid DNA-binding motif [reviewed by Scott et al. (1989)]. A variety of studies indicate that HOM-C proteins function as transcriptional regulators (Thali et al. 1988; Krasnow et al. 1989; Winslow et al. 1989; Johnson and Krasnow 1990). HOM-C proteins control the expression of downstream genes directly and indirectly (Gould et al. 1990; Immegrüll et al. 1990; Reuter et al. 1990) and also participate in autoregulatory circuits (Kuziora and McGinnis 1988; Regulski et al. 1991; Bienz and Tremml 1988) and in cross-regulatory interactions (Hafen et al. 1984; Struhl and White 1985) that refine and maintain their own patterns of expression. Molecular studies of several vertebrates and invertebrates have revealed that the system of HOM-C genes predates the divergence of protostome and deuterostome lineages and that a primitive role of these genes in specifying positional identities along the anteroposterior body axis may have been conserved among animals as diverse as nematodes, fruitflies, and mammals [reviewed by McGinnis and Krumlauf (1992)].

The Ultrabithorax (Ubx) gene, one of three HOM-C genes in the bithorax complex of Drosophila, plays a major role in specifying segmental identity in the epidermis and nervous system of the posterior mesothorax (pT2), the metathorax (T3), and the anterior half of the first abdominal segment (aA1) (Lewis 1978; Hayes et al. 1984; Teugels and Gysen 1985; Hartenstein 1987; Heuer and Kaufman 1992). The Ubx gene also plays a minor role in more posterior metameres, where its function overlaps with that of other genes in the bithorax complex (Lewis 1978; Sánchez-Herrero et al. 1985). Ubx is also required in parasegment 7 of the visceral mesoderm (Bienz and Tremml 1988) and in the somatic
mesoderm of abdominal segments (Hooper 1986). Accordingly, Ubx RNAs and proteins are expressed from posterior T2 through A7 in the epidermis and nervous system, with peak abundance in pT3 plus aA1 (parasegment 6), from pT3 through A6 in the somatic mesoderm, and in parasegment 7 in the visceral mesoderm (White and Wilcox 1984, 1985; Alam and Martinez-Arias 1985; Beachy et al. 1985; Brower 1987; Canal and Ferrus 1987).

Alternative processing of the primary Ubx transcript generates six different mRNAs (O’Connor et al. 1988; Korfeld et al. 1989). Each of these mRNAs encodes a structurally distinct protein isoform (Figure 1). The six isoforms share common amino proximal and carboxy terminal domains that are separated by a differential region consisting of three small elements (B, mI and mII) used in different combinations. The B element comprises 9 amino acids encoded between alternative donor splice sites “a” (most 5′) and “b” (most 3′) at the end of the common 5′ exon, while mI and mII, each 17 amino acids long, are encoded by separate internal exons (Figure 1). The homeodomain is located in the common carboxyl terminal region of the UBX proteins and is separated from the differential elements by only 4 amino acids. A similar arrangement of the homeodomain relative to alternatively spliced internal elements is present in the predicted protein products encoded by the HOM-C genes Antennapedia (Stroehler et al. 1986; Bermingham and Scott 1988), labial (Mlodzik et al. 1988), and proboscipedia (Cribbs et al. 1992), suggesting that alternative splicing might modulate the complex functions of several homeotic genes in different tissues and stages. A differential element has been identified at a similar location in cDNAs from the distantly related homeobox gene bicoid, which functions as a morphogen to specify anterior positional values in the blastoderm embryo (Berleth et al. 1988).

The developmental patterns of UBX isoform expression suggest that alternative splicing may modulate stage- and tissue-specific functions of the Ultrabithorax gene (López and Hogness 1991; Artero et al. 1992). Isoforms containing both mI and mII account for most of the Ubx RNA and protein expressed in the epidermis, mesoderm, and peripheral nervous system (PNS), but are barely detectable in the central nervous system (CNS) (López and Hogness 1991; Artero et al. 1992). In contrast, isoforms lacking mI are expressed primarily (if they contain mII) or almost exclusively (if they also lack mII) in the CNS (López and Hogness 1991; Artero et al. 1992). Proteins containing the B element in combination with mI and mII are detected in ectoderm and mesoderm but only during early stages of embryogenesis, while other isoforms containing the B element are too rare to detect at any stage by Western blotting (Gavis and Hogness 1991), immunohistochemical staining with isoform-specific antibodies (A. J. López, unpublished results), or in situ hybridization with isoform-specific probes (Artero et al. 1992). These observations are in general agreement with the relative proportions and temporal profiles of alternative Ubx mRNA expression as deduced from nuclease protection analyses (O’Connor et al. 1988; Korfeld et al. 1989).

Although the expression patterns suggest that different UBX isoforms have different tissue- and stage-specific functions, alternative explanations are also possible. The Ubx introns are unusually large, raising the possibility that the alternative splicing events are fortuitous by-products of developmental variations in the splicing machinery that are tolerated because they do not affect UBX protein function. The partial transformation of haltere into wing exhibited by flies homozygous for the mutant allele Ubx^m17, which lacks mII, has been attributed to a reduction of UBX expression in the capitelum (Busturia et al. 1990), but evidence of functional distinctions between isoforms has been reported (Krassnow et al. 1989; Mann and Hogness 1990; Subramaniam et al. 1994). For technical reasons, however, the detailed analyses have focused on the functional limitations of UBX-Iva with respect to development of the peripheral nervous system, and the documented differences have been quantitative (Subramaniam et al. 1994). Thus, the full extent and biological importance of functional differences among UBX isoforms are not known.

To gain a better understanding of the role of alternative splicing in Ubx function, we have investigated
whether the Ubx RNAs of other species undergo similar regulated processing. Aspects of this processing that are important for Ubx function should be conserved among species with similar developmental programs, but functionally irrelevant features should be lost or altered substantially over long evolutionary periods. We report that the developmentally regulated processing of Ubx RNAs has been conserved strongly among Drosophilids separated by at least 60 million years of evolution. The species D. melanogaster, Drosophila pseudoobscura, Drosophila hydei and Drosophila virilis exhibit identical patterns of optional exon use between the homeodomain and highly conserved amino-proximal sequences of Ubx. These species produce the same family of UBX protein isoforms with identical mI and mII amino acid sequences, even though the common amino-terminal exon has undergone substantial divergence. Furthermore, the tissue- and stage-specific patterns of expression of mI and mII are identical among these same Drosophila species, indicating that the developmental regulation of the alternative splicing events has also been conserved. Given that the genus Drosophila is morphologically conservative but exhibits high rates of nucleotide sequence change (MORIYAMA 1987), these findings argue for an important role of alternative splicing and of the differential elements in Ubx function. The high degree of conservation of the alternative splicing pattern in different Drosophila species should facilitate further study of the mechanisms by which it is regulated.

MATERIALS AND METHODS

Insect stocks: D. melanogaster (wild-type strain Ore-R) was from our regular stock collection. D. pseudoobscura, D. virilis and D. hydei were obtained from the National Drosophila Species Resource Center at Bowling Green. These four species were raised at 25 °C on standard molasses-cornmeal-yeast medium.

Preparation of RNA: Total RNA was isolated from 0–18 hr-old embryos. To avoid any possibility of cross-contamination, embryos from different species were collected and processed on separate days, using fresh working solutions and factory-sterile plasticware. Glassware was washed with alkali and acid and baked for 16 hr at 250 °C to destroy RNases and eliminate possible contamination by exogenous nucleic acids. The washed embryos were suspended in 5 volumes of GIT Buffer (4 M guanidinium thiocyanate; 0.1 M Tris-HCl, pH 7.5; 0.14 M 2-mercaptoethanol) and disrupted in Potter-Elvehjem homogenizers. After addition of Sarkosyl to a final concentration of 0.5%, the lysate was clarified by centrifugation for 18 min at room temperature in an Eppendorf microfuge, and the supernatant was transferred to a fresh tube and extracted five times with phenol:chloroform (1:1). The reaction was incubated for 10 min at 25 °C, followed by 60 min at 42 °C. To amplify the cDNA products, the reaction mixture was heated at 95 °C for 10 min and chilled on ice, then diluted with 80 μl of PCR reaction mix containing 1 × Taq polymerase buffer (Boehringer Mannheim), 10% DMSO, 50 μl of 1 M TrxS1 primer, and 1 unit of Taq polymerase (Boehringer Mannheim). The sample was subjected to thirty cycles of amplification (30 sec denaturation at 95 °C, 60 sec annealing at 60 °C, 30 sec annealing at 60 °C, 30 sec annealing at 75 °C, and 30 sec extension at 75 °C). Ten-microliter aliquots of the amplified products were analyzed by electrophoresis on a 5% NuSieve GTG agarose gel or subjected to three further cycles of amplification under identical conditions except that the 1 M TrxS1 was replaced with 1 M TrxS2 that had been labeled at the 5′ end with [γ-32P]ATP and polynucleotide kinase. Three microliters of the resulting amplified products were analyzed by electrophoresis through a denaturing gel (6% acrylamide, 6 μM urea) followed by autoradiography.

Cloning and nucleotide sequence analysis of amplified cDNAs: The cDNA fragments obtained from second round amplification using non-labeled primer UBX-3A2 were cloned into pcDNA3 (+) vector (Stratagene). The amplified samples were extracted with phenol:chloroform and precipitated with ethanol. The pellets were resuspended in 10 mM Tris (pH 8.0), 5 mM EDTA, 0.5% SDS and then treated with 50 μg/ml proteinase K at 37 °C for 30 min (COWET et al. 1991). After the proteinase K was inactivated at 70 °C for 10 min, the sample was extracted again with phenol:chloroform and precipitated with ethanol. Upon resuspension, the sample was treated with 4 units of T4 DNA polymerase in the presence of 0.1 mM dNTPs and 5 mM MgCl2 at 37 °C for 20 min in order to make the fragments blunt ended. After the T4 DNA polymerase was heat inactivated the fragments were cut at the Xhol site located within the UBX3A2 primer. These fragments were ligated into the pKS (+) vector digested with Xhol and Smal. Positive clones were identified and classified according to the size of the fragment released by PvuII, which cleaves at vector sites flanking the insert. The sequences of the Ubx cDNA clones were determined on both strands by the dideoxy chain termination method using the Sequenase Kit (U.S. Biochemical Co.) and the forward and reverse sequencing primers complementary to pKS (+).

Immunohistochemical staining: Embryos were dechorionated in 50% Clorox bleach for 2 min, then fixed by shaking for 25 min in a 1:1 mixture of heptane and 4% paraformaldehyde in PEM (0.1 m Pipes, pH 6.9, 1 mM EGTA, 1 mM MgCl2). The
embryos were devitellinnized by replacing the fixative with 80% ethanol and mixing gently at room temperature. Devitellinnized embryos were washed three times with 80% ethanol, rehydrated in BST (10 mM Tricine, pH 7.0, 40 mM NaCl, 55 mM KCl, 10 mM MgSO₄, 1 mM CaCl₂, 20 mM glucose, 50 mM sucrose, 0.2% BSA, 0.02% sodium azide, 0.1% Triton X-100) for 2 hr, then blocked in BSN (BST without Triton X-100, but containing 0.2% saponin and 5% normal goat serum) for 1 hr. After blocking, the embryos were incubated in BSN containing antibodies against UBX protein (see RESULTS). The primary antibodies were detected using the Vectastain Elite Kit (Vector Labs) with an avidin conjugated goat anti-mouse antibody followed by biotin-conjugated horseradish peroxidase (HRP). Diaminobenzidine was used as the HRP substrate. After color development, the embryos were washed and dehydrated in ethanol before clearing and mounting in a 1:1 mixture of methylsalicylate and Canada balsam. Embryos were viewed and photographed under Nomarski optics.

RESULTS

Analysis of alternative splicing patterns of Ubx RNAs by reverse transcription and amplification of cDNAs:

To determine whether the Ubx transcripts of other Drosophila species undergo alternative splicing similar to that in D. melanogaster, we compared the structure of the Ubx mRNAs in the region between the conserved YPWM box, located in the common 5’ exon, and the homeodomain, located in the common 3’ exon. To accomplish this, total RNA was isolated from each species and used as template for reverse transcription using a synthetic oligonucleotide primer that anneals to a region of the Ubx homeobox from D. melanogaster (primer UBX-3A1; Figure 1 and MATERIALS AND METHODS). We expected the amino acid sequence of the Ubx homeodomain to be highly conserved among different insect species, and primer UBX-3A1 was designed to minimize possible mismatches due to codon degeneracy while maximizing discrimination between the homeobox of Ubx and those of other genes in D. melanogaster. The resulting Ubx cDNAs were then amplified by means of the polymerase chain reaction, using primer UBX-3A1 plus an opposing primer that was complementary to nucleotides -133 to -113 relative to donor splice site “b” in the common 5’ exon (primer UBX-5S1; Figure 1 and MATERIALS AND METHODS). The sequence to which UBX-5S1 anneals is known to be perfectly conserved between D. melanogaster and D. pseudoobscura and differs by only 2 internal nucleotides in Drosophila funebris (Wilde and Akam 1987). Successful amplification was only expected for cDNAs templated on fully processed RNAs because the primer target sequences are separated by the three Ubx introns, which are very large in D. melanogaster (measuring 8, 12 and 50 kb, respectively). Following reamplification using a second primer that anneals within the homeodomain (UBX3A2; see Figure 1 and MATERIALS AND METHODS), alternative splicing patterns similar to those in D. melanogaster were predicted to produce amplified cDNA fragments of specific sizes as indicated in Figure 1. These predictions were confirmed by performing similar amplification reactions using each of the five cloned Ubx cDNAs from D. melanogaster as templates, both individually and in mixtures (shown for a mixture of the five in Figure 2, lane 1).

The results of this analysis using RNA from different Drosophila species are shown in Figure 2. Total RNA from D. melanogaster embryos aged from 0–18 hours produced 5 bands corresponding to the predicted sizes for isoforms Ib, Ia, IIb, Ila and IVa (Figure 2, lane 2). This pattern was highly reproducible among different batches of RNA and between different strains of D. melanogaster (Oregon-R and Canton-S; data not shown). A faint band corresponding in size to that predicted for the rare isoform IVb was also observed upon longer exposures, but not consistently. Given that the age distribution of the embryos was strongly biased towards late stages, the relative intensities of the bands were generally consistent with previous estimates of the relative abundances of the corresponding isoforms, except that isoform Ia appeared less abundant and isoform IVa more abundant than expected on the basis of nuclease
A. Amino Acid Sequence

<table>
<thead>
<tr>
<th>5' exon</th>
<th>B</th>
<th>mI</th>
<th>mII</th>
<th>3' exon</th>
</tr>
</thead>
<tbody>
<tr>
<td>219</td>
<td>247</td>
<td>256</td>
<td>273</td>
<td>290</td>
</tr>
</tbody>
</table>

D.mel SGAAGTAAASSLHRQSNHTYFWMAIGECPEDTPKSKIRSDLQYGGISTDMKRYSESLAGSLPDLWIYNGLRERRGRQTYTR
D.pse ----A---------------STA--S---------------
D.vir ----A---------------STA--I---------------
D.hyd ----A---------------STA--T---------------

B. Nucleotide Sequence

<table>
<thead>
<tr>
<th>5' exon</th>
<th>microexon I</th>
<th>microexon II</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.mel</td>
<td>TCGGGCCGCCGCCT</td>
<td>GGGCAAAACGGGCAGCAGTTACACCAGGCAGGGCCAAATCAGATTCTACCTCCCTGGATTGCCGATTCGAG</td>
</tr>
<tr>
<td>D.pse</td>
<td>----G--A--A--C--</td>
<td>--C--A--C--B--</td>
</tr>
<tr>
<td>D.vir</td>
<td>----T--GGCG----A--</td>
<td>----C--A--C--B--</td>
</tr>
<tr>
<td>D.hyd</td>
<td>----T--GGCG----A--</td>
<td>----C--A--C--B--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>h element</th>
<th>microexon I</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.mel</td>
<td>GTGAGTCTCAGATACTGAGAACAAGCAGAGGCTCATTACATACATGCAGATTCCCTGGATTGCCGATTCGAG</td>
</tr>
<tr>
<td>D.pse</td>
<td>----C--A--C--</td>
</tr>
<tr>
<td>D.vir</td>
<td>----CCA--C--</td>
</tr>
<tr>
<td>D.hyd</td>
<td>----CCA--C--</td>
</tr>
</tbody>
</table>

Figure 3.—Conservation of differentially spliced elements in the Ubx genes of different Drosophila species. (A) Predicted amino acid sequences of the differential elements and amplified regions of the common exons, shown as spliced in isoform Ib. Sequences corresponding to the primers at either end are not included. The Dayhoff single letter amino acid symbols are used. Dashes indicate identity with D. melanogaster. Numbers refer to amino acid positions in UBX isoform Ib from D. melanogaster. Dots indicate the positions of splice junctions; B, mI, and mII indicate the start of each differential element. (B) Nucleotide sequences, shown separately for each exon and the B element. Sequences corresponding to the primers at either end are not included. Dashes indicate identity with D. melanogaster. Negative numbers indicate nucleotide coordinates relative to splice donor site "b" at the end of the common 5' exon; positive numbers indicate nucleotide coordinates relative to the splice acceptor site at the beginning of the common 3' exon.

RNA from each of the other Drosophila species produced a pattern of amplified cDNA fragments that was virtually identical to that of D. melanogaster in the number, size, and relative intensity of the bands, except that each fragment was slightly larger than its counterpart in D. melanogaster (Figure 2, lanes 3–5). As described below, this size difference reflects the absence of one alanine codon in the Ubx 5' exon of D. melanogaster compared to the other three species. These patterns were highly reproducible among batches of RNA for each species, suggesting that the Ubx transcripts of all four species are processed to produce a similar set of alternative mRNAs. Differences in the relative intensities of the bands probably reflect differences in the developmental stage distribution of the embryos, since D. melanogaster exhibited the most rapid development of the four species compared.

To confirm that the observed bands were amplified fragments of Ubx cDNAs, the non-labeled amplification products were electrophoresed on agarose gels and analyzed by Southern blot hybridization using a D. melanogaster probe specific for the amplified region of the Ubx cDNAs but excluding primer sequences. To ensure that all Ubx isoforms would be detected, the probe consisted of pooled Ntl-XhoI restriction fragments spanning the differential region from each of the five cloned Ubx cDNAs (Figure 1). The five amplified fragments from each species hybridized to the probe under stringent conditions (washed at 68°C in 0.1 × SSC, 0.1% SDS), and similar signal intensities were obtained with equal amounts of amplified DNA from each source, demonstrating that the amplified fragments share a high degree of sequence similarity with Ubx cDNAs from D. melanogaster (data not shown). These conclusions were confirmed by cloning and determining the nucleotide sequences of the amplified fragments from each species, as described below.

Sequence analysis of amplified Ubx cDNAs: The amplified Ubx cDNA fragments from Drosophila pseudobscura, D. virilis and D. hydei were gel purified and cloned separately into pBluescript KS(+) for nucleotide sequence analysis (see MATERIALS AND METHODS). At least
two clones derived from different amplifications and from different batches of RNA were analyzed for each isoform, and the resulting sequences were compared to identify possible discrepancies due to nucleotide misincorporation by reverse transcriptase or by Taq DNA polymerase.

The nucleotide sequences of the amplified cDNA fragments confirmed that each Drosophila species encodes isoforms Ib, Ia, Ib, Ia and IVa. As anticipated from the extremely low abundance of presumptive isoform IVb, no clones of this type were isolated from any species. The nucleotide sequences and deduced amino acid sequences of the amplified common and differential elements of each species are shown in Figure 3. The deduced amino acid sequences of both mI and mII are identical in all four Drosophila species (Figure 3A). In contrast, the B element had the same length in all four species, but three sequence variants were observed: one shared by D. melanogaster and D. pseudoobscura, one in D. virilis, and one in D. hydei. The B element from D. melanogaster differed from those in D. virilis or D. hydei at 4 of 9 positions, whereas the B elements from D. virilis and D. hydei differed from each other at only one position. Notable aspects of the changes relative to D. melanogaster are the elimination of one negative charge and one alanine residue in the common 5' exon has diverged at 7 of 86 positions between D. melanogaster and D. pseudoobscura, at 7 of 86 positions between D. melanogaster and D. hydei. In contrast, the sequence of the amplified region of the common 5' exon has diverged at 7 of 86 positions between D. melanogaster and D. pseudoobscura, at 7 of 86 positions between D. melanogaster and D. virilis, and 9 of 86 positions between D. melanogaster and D. hydei. In addition, as noted above, the amplified region from D. melanogaster contains a deletion of 3 nucleotides relative to the other three species. The B element has diverged at 27 of 27 positions between D. melanogaster and D. pseudoobscura, and at 7 of 27 positions between D. melanogaster and D. virilis or D. hydei. The “a” donor site in the 5' exon, which is a perfect match to the consensus CAG/GUGAGU, is conserved in all four species.

Developmental expression patterns of differential elements in the UBX proteins: The finding that different Drosophila species produce the same set of alternative Ubx RNAs in roughly similar proportions suggested that the developmental regulation of Ubx splicing has been conserved. To determine whether these alternative RNAs actually encode developmentally regulated UBX protein isoforms in all four species, we used immunohistochemical staining of whole mount embryos to analyze the stage- and tissue-specific pattern of expression of each predicted exon. The perfect conservation of amino acid sequence in regions of the 5' exon, the 3' exon, and all of mI and mII facilitated this analysis, since many antibodies raised against different exons of the UBX proteins from D. melanogaster (López and Hogness 1991) were expected to react with the appropriate UBX isoforms of other species.

As predicted, monoclonal antibodies against epitopes in conserved regions of the common 5' exon or 3' exon reacted with the UBX proteins in all four Drosophila species, revealing identical dynamic patterns of expression (illustrated with a 3' exon epitope in Figure 4, A, B, G and H, and Figure 5, A, B, G and H). These conserved patterns included expression in epidermis, somatic and visceral mesoderm, peripheral nervous system, and the central nervous system, and were the same as described previously for D. melanogaster using other monoclonal and polyclonal antibodies that recognize the entire UBX protein family (White and Wilcox 1985; Beachy et al. 1985; López and Hogness 1991). Monoclonal antibodies directed against epitopes in mI and mII from D. melanogaster also produced staining patterns showing identical stage and tissue specificity in all four species (Figure 4, C-F and I-L; Figure 5, C-F and I-L). Antibodies against mI (which reveal the pattern of expression of isoforms Ia and Ib) stained nuclei of the epidermis, mesoderm and peripheral nervous system strongly, but stained the central nervous system only weakly and only at the time of maximum UBX expression in that tissue. In contrast, antibodies against mII stained nuclei of the central nervous system strongly (isoforms Iib and Iia) in addition to the epidermis, mesoderm, and peripheral nervous system (isoforms Ib, Ia, Iib and Iia). In addition, the dynamics of tissue-specific expression were identical (compare Figures 4 and 5). In all species the expression of mI and mII in epidermis and mesoderm declined during development while the expression of mII in the CNS increased. Furthermore, mII expression appeared to account for most of the pattern and quantity of UBX expression in the CNS during mid-embryogenesis, but not during late stages, reflecting the appearance of isoform IVa (and presumably IVb) in a partially overlapping but distinct pattern in all species, as has been determined directly for D. melanogaster (López and Hogness 1991; A. J. López, unpublished results).
Evolution of Ultrabithorax Isoforms

These results demonstrate that the developmental patterns of expression of the alternatively spliced exons mI and mII have been conserved over at least 60 million years of evolution, the time since the lineages leading to D. melanogaster/D. pseudoobscura and to D. virilis/D. hydei diverged (THROCKMORTON 1977; HENNIG 1991; BEVERLEY and WILSON 1984). Presumably, this conservation of expression patterns reflects similar conservation.
of the regulatory mechanisms that control alternative splicing of the \textit{Ubx} RNAs.

**DISCUSSION**

Our results demonstrate that the alternatively spliced elements of the \textit{Ubx} gene have been conserved among Drosophilids separated by 60 million years of evolution, and that these elements are combined to produce the same set of UBX protein isoforms in the four species examined. The developmentally regulated expression of the alternatively spliced exons mI and mII has also been conserved, suggesting that these species share the same mechanism for alternative splicing of \textit{Ubx} transcripts. Although our analysis does not demonstrate directly that the different \textit{Ubx} RNA structures are generated by alternative splicing in all four species, this is the
most plausible interpretation of the data. In the first place, the nucleotide sequences of all cDNAs from a given species were identical except for the presence or absence of those elements that are known to be alternatively spliced in *D. melanogaster*. Furthermore, previous studies of *D. pseudoobscura*, *D. funebris*, *D. virilis* and *Musca domestica* indicate that each of these species, like *D. melanogaster*, contains a single *Ubx* gene (Wilde and Akam 1987).

**Patterns of amino acid and nucleotide sequence conservation**: Two types of B element sequences were found in this investigation: one shared by *D. melanogaster* and *D. pseudoobscura* and one shared by *D. virilis* and *D. hydei*. The nucleotide sequence of the B element from *D. hydei* was identical to that reported by Wilde and Akam (1987) for the last 27 nucleotides of the *Ubx* 5' exon from *D. funebris*, and the sequence of the B element from *D. virilis* differs at only one position. These sequence relationships concur with the phylogenetic tree of the genus *Drosophila* (Throckmorton 1975; Hennig 1981; Beverley and Wilson 1984). The divergence of the B element may reflect drift due to absence of functional constraints, or selection for different functional properties in the two species groups. *M. domestica* lacks the B element in the 5' exon of Ubx (Wilde and Akam 1987), suggesting that, at least in this species, the B element may not be required for Ubx function. It remains possible, however, that the equivalent of a B element is encoded as a separate exon in the *Ubx* gene from *Musca*.

In contrast, the differential elements mI and mII have been conserved strongly among the four *Drosophila* species examined. The amino acid sequences are identical, as is the nucleotide sequence of mI, while the nucleotide sequence of mII has diverged at only one (*D. pseudoobscura* and *D. hydei*) or two (*D. virilis*) of 51 positions relative to *D. melanogaster*. These results suggest that mI and mII are under strong selective pressures. In the absence of such pressures, the rate of nucleotide substitution at neutral positions, estimated to be about 1.7% per million years for the genus *Drosophila* (Moriyama 1987; Caccione et al. 1988), would have led to 45% divergence of the microexon nucleotide sequences in the 40 million years separating *D. melanogaster* and *D. pseudoobscura*, and to 56% divergence in the 60 million years separating *D. melanogaster* and *D. hydei* or *D. virilis* (correcting the nucleotide difference, K, for multiple substitutions as: $K_e = (-3/4) \ln (1 - (4/3) K)$ (Jukes and Cantor 1969).

The significance of amino acid sequence conservation in mI and mII is highlighted by the pattern of sequence alterations in the rest of the *UBX* protein [illustrated in Figure 6, which also incorporates data for the 5' exon from a previous study by Wilde and Akam (1987)]. Overall, the amino acid sequence of the common 5' exon has diverged by 15% over 60 million years, but highly divergent regions are interspersed with highly conserved regions. Together with the homeodomain, mI and mII form the largest uninterrupted block of conserved amino acids in the *UBX* proteins. Two other unaltered blocks contain the MNSYF and YPWM peptides, which are also conserved in other homeoproteins and across greater evolutionary distances and are suspected to have important functions. Thus, different regions of the *UBX* proteins appear to be under different degrees of functional constraint, and regions in which precise amino acid sequences are not essential for function have had sufficient opportunity to diverge extensively during the period covered by our study.

The nucleotide sequence of mI is conserved perfectly among the four *Drosophila* species analyzed, and that of mII is almost completely conserved. These results suggest that the nucleotide sequence itself may be under selective pressure, since independent silent substitutions would have been possible at 17 nucleotide posi-

---

**Figure 6**—Pattern of sequence divergence in the *UBX* proteins of different *Drosophila* species. In the center is a schematic diagram of *D. melanogaster* UBX isoform Ib, represented as a rectangle (N, amino terminus; C, carboxyl terminus). The regions whose sequence has been determined in three or more *Drosophila* species are indicated by continuous lines; regions whose sequence is known only for *D. melanogaster* are indicated by broken lines. Positions of amino acid substitutions between *D. melanogaster* and either *D. pseudoobscura* or *D. funebris* are indicated as vertical lines within the rectangle. Deletions relative to *D. melanogaster* are shown as triangles with their base towards the protein, indicating the position and extent of the deletion. Insertions relative to *D. melanogaster* are shown as triangles with their apexes toward the rectangle, indicating the position of the insertion. Numbers above the triangles indicate the number of residues deleted or inserted. Features of the *UBX* proteins located within conserved regions are indicated by labeled lines below the rectangle. M, the amino terminal MNSYF peptide, which is also shared with several other homeoproteins of *Drosophila* and vertebrates; Pi, major phosphorylated region (Gavis and Hogness 1991); Y, the YPWM tetrapeptide, which is also shared with the proteins encoded by other homeotic genes of *Drosophila* and vertebrates; B, I and II, B element, mI and mII, respectively; HD, homeodomain. The sequences of mI, mII and the 5' exon have diverged at only one position. These residues are under strong selective pressures. In the absence of such pressures, the rate of nucleotide substitution at neutral positions, estimated to be about 1.7% per million years for the genus *Drosophila* (Moriyama 1987; Caccione et al. 1988), would have led to 45% divergence of the microexon nucleotide sequences in the 40 million years separating *D. melanogaster* and *D. pseudoobscura*, and to 56% divergence in the 60 million years separating *D. melanogaster* and *D. hydei* or *D. virilis* (correcting the nucleotide difference, K, for multiple substitutions as: $K_e = (-3/4) \ln (1 - (4/3) K)$ (Jukes and Cantor 1969).
tions in mI and at 18 nucleotide positions in mII, not counting sites constrained by donor and acceptor splice site sequence preferences. This conservation is not explained by generalized Drosophila codon preferences. A comparison of codon usage in mI and mII with the tabulated codon preferences for D. melanogaster (Wada et al. 1991) reveals that the less frequent codons are used at 24 positions out of 30 where a choice between frequent and infrequent synonymous codons is possible. In contrast, codon usage within the common 5' and 3' exons is in good agreement with general D. melanogaster codon preferences.

Implications for UBX isoform function: The conservation of amino acid sequence in mI and mII and of the developmentally regulated isoform expression patterns is consistent with other evidence for the functional importance of UBX isoform diversity. Experiments using transgenic flies that ectopically express particular Ubx cDNAs have shown that UBX-Ia fails to transform certain cell fates in the PNS that can be transformed by UBX-IIa (Mann and Hogness 1990; Subramaniam et al. 1994). These results are in good agreement with the tissue specificity of UBX isoform expression in wild-type flies (Lopez and Hogness 1991; this study) and with the phenotype of Ubx^MX17 mutants, which exhibit a defect in alternative splicing that results in the substitution of isoform IVa for isoforms Ia/b, IIa/b and IVb (Subramaniam et al. 1994). Together, these observations indicate that different UBX isoforms have partially overlapping but distinct functional capabilities, although the full extent and the molecular basis of the differences are not known. The strong conservation of alternative isoform structures and expression patterns suggests that the functional distinctions are more extensive than has been documented thus far.

The differential elements of the UBX proteins are positioned like the "homeodomain extension," a short region that mediates cooperative interactions between the yeast α2 homeoprotein and a second DNA-binding protein, MCM10, leading to specific recognition of functionally relevant target sites (Version and Johnson 1993). The differential elements of UBX might function in a similar capacity. Another possibility is that they serve to vary the spatial relationship between the homeodomain and a constant element responsible for protein-protein or protein-DNA interactions. The highly conserved YPWM tetrapeptide (Mavilio et al. 1986; Wilde and Akam 1987; Karch et al. 1990), which is separated from the homeodomain by differentially spliced elements in the proteins encoded by Ubx, Antp, lab and pb (Strobiher et al. 1986; O'Connor et al. 1988; Bermanham and Scott 1988; Mlodzik et al. 1988; Kornfeld et al. 1989; Grubs et al. 1992), might be such an element. Proteins that interact with UBX have not been identified, although the gene extradenticle (exd) may encode such a protein. Mutations in exd alter the homeotic transformations produced by Ubx mutations but do not affect the pattern or level of Ubx expression (Peifer and Wieschaus 1990).

Comparisons of genomic sequences encoding the Antp gene have shown that the predicted amino acid sequences of the two alternatively spliced elements from D. melanogaster are also conserved in D. subobscura and D. virilis (Hooper et al. 1992). The conservation of nucleotide sequence around the intron-exon boundaries suggests that their splicing is regulated similarly in the three species (Hooper et al. 1992). The differential elements of Antp, like those of Ubx, thus appear to be under selective pressure (Hooper et al. 1992). In contrast, an alternatively spliced element present at an analogous position in cDNAs from the distantly related homeobox gene bicoïd is not conserved in D. pseudoobscura (Seeger and Kaufman 1990). Unlike the situation in Ubx, this optional element is very small (5 codons between two acceptor sites at the end of intron 2), it has only been observed in a minority of bicoïd cDNAs (Belteth et al. 1988; Driever et al. 1990), and the corresponding protein isoform cannot be detected in embryos (Driever and Nüsslein-Volhard 1988). The minor bicoïd CDNA variants may result from aberrant processing of the pre-mRNA, perhaps as a consequence of the unusually small size (55 nucleotides) of the relevant intron in D. melanogaster. Furthermore, the BICOID protein species derived from the most abundant cDNA class (which lacks the optional element) can exert all morphogenetic functions of the bicoïd gene in vivo (Driever et al. 1990). The lack of conservation of the alternatively spliced element in bicoïd mRNAs is thus consistent with its lack of specific biological function in D. melanogaster, and this strengthens our conclusion that alternative splicing of Ubx has been conserved by selection for function.

Implications for the mechanism of alternative splicing: It appears unlikely that codon choice in mI and mII is important for modulating the efficiency of translation of Ubx mRNAs, since expression of individual Ubx CDNAs in transgenic Drosophila embryos (Mann and Hogness 1990) or Drosophila cell lines (Gavis and Hogness 1991) has revealed no obvious differences in translatability of Ubx mRNA variants. It is conceivable that the highly conserved nucleotide sequences of mI and mII could play a role in control of nucleo-cytoplasmic transport, mRNA stability, or transcription, but a more appealing possibility is that they are important for the proper developmental regulation of splicing. Numerous studies in Drosophila and in mammals have shown that internal exon sequences can play an important role in the splicing mechanism (reviewed recently in Dietz et al. 1993 and Watakabe et al. 1993). Insertions, deletions, or nucleotide substitutions within exons can influence splice site selection by altering the optimal distance between 3' and 5' splice junctions, inactivating
Evolution of Ultrabithorax Isoforms

Figure 7.—Model for the pathway of alternative splicing of Ubx pre-mRNAs. Exons are indicated as rectangles and introns as lines. Splice donor sites are underlined and labeled X, Y, and Z if they exist in the unprocessed RNA, or X’, Y’ and Z’ if they are generated by juxtaposition of exons after a splicing event. For clarity, the model is illustrated with X as donor site “a” of the 5’ exon, ignoring donor site “b.” This does not alter the basic concept or working of the model: if X were to represent donor site b, the first intermediate would have three potential donor sites: “a,” X’, and Y, in that order from 5’ to 3’. See text for details.

cis-acting elements, altering secondary structures, or activating cryptic splice sites. Thus, the nucleotide sequences of mI and mII, which are relatively small exons, may be multiply constrained by the need to maintain proper regulation of alternative splicing, as well as by coding requirements. One of the perfectly conserved features of the mI and mII nucleotide sequences is the presence of U1 snRNP recognition site consensus sequences (GURAGU) at the 5’ end of each exon (Figure 3B). Recent studies suggest that binding of U1 snRNP to sequences within some exons may be required to activate the splicing of upstream introns (Watarabe et al., 1993), and it is possible that the putative U1 snRNP recognition sites play this role in mI and mII.

Closer inspection of the cDNA sequences suggests a specific model for the mechanism of alternative splicing of Ubx RNAs. In all four species, splicing of either microexon to donor site “a” of the 5’ exon (the more upstream donor of the two that flank the B element) or to each other regenerates a good match to the donor splice site consensus sequence (C/A)AG/GURAGU (Shapiro and Senapathy 1987; Brunak and Engelbrecht 1991) (see Figure 3B). Splicing of either microexon to donor site “b” also generates a new site that is a better match to the consensus than the original “b” site. If splicing of the large Ubx transcripts occurs cotranscriptionally beginning at the 5’ end, as suggested by in situ hybridization experiments (Axter et al. 1992), the resulting partially spliced intermediates would contain a set of potential donor splice sites spaced 27–51 nucleotides apart and marking the boundaries of the differential elements. The Ubx variants could be generated during successive splicing events by selection among the potential donor sites in these processing intermediates (an “in and out” mechanism of alternative splicing, as illustrated in Figure 7). An attractive feature of this model is that what appears superficially as a complex set of regulatory decisions matching two alternative donor sites and three alternative acceptor sites would be reduced to a single type of choice reiterated during the splicing of each intron: whether or not to use a particular donor site. The regulation of splicing of the B element, mI and mII could thus involve similar molecular mechanisms and even the same trans-acting factors. The temporal and spatial distributions of the normal UBX isoforms, as well as the absence of isoforms containing mI but not mII, could be explained by a hierarchy of intrinsic donor site strengths together with developmentally specific variations in the concentration of a factor required for donor site activation or repression.

A second set of observations is relevant to this model. When Ubx cDNAs encoding isoform Ib or Ia are ex-
pressed in Drosophila tissue culture cell lines (Krasnow et al., 1989; V. Subramaniam and A. J. López, Unpublished observations) or in transgenic embryos (Mann and Hogness, 1990), mI and most of mII are removed from a fraction of the resulting RNAs by splicing reactions that use a cryptic acceptor site at the 5' end of mII and either the "a" donor site in the 5' exon (isofrom IIb) or the new donor site generated at the boundary between the 5' exon and m (isofrom Ia). Therefore, as predicted by the model, the "a" donor site remains active in isofrom b RNAs, and the donor site generated by splicing to the "a" donor site is also functional.

Concluding remarks: The conservation of structure, sequence and regulation of the different elements in the UBX proteins of different Drosophila species strengthens the conclusion that these elements mediate important stage and tissue-specific Ubx functions. Elucidation of this role, and of the alternative splicing mechanism that controls the developmentally specific pattern of expression of UBX isoforms, will require experimental analysis using biochemical and molecular genetic techniques. Interspecies comparisons such as the one presented here can provide useful clues and models to guide these investigations.

We thank Guillermo Donate for technical assistance and John L. Woodruff, Elizabeth W. Jones, and María-Teresa Sáenz-Robles for helpful discussions. This work was supported by a grant to A.J.L. from the National Institutes of Health (ROI HD28664).

LITERATURE CITED


Evolution of Ultrabithorax Isoforms


Communicating editor: R. E. Denell