The EMBO Journal vol.7 no.2 pp.435-445, 1988

Alternative RNA products from the *Ultrabithorax* domain of the bithorax complex

Michael B.O'Connor, Richard Binari, Lizabeth A. Perkins[1] and Welcome Bender

Department of Biological Chemistry and Molecular Pharmacology and [1]Department of Genetics, Harvard Medical School and Howard Hughes Medical Institute, Boston, MA 02115, USA

Communicated by W.J.Gehring

The homeotic gene, *Ultrabithorax* (*Ubx*) is involved in specifying the identities of several segments in the fly *Drosophila melanogaster*. The structures of over 60 independent *Ubx* cDNAs have been examined. There are two major species of transcripts, 3.2 and 4.3 kb in length, which are produced by alternate sites of polyadenylation. Differential splicing gives rise to at least five variant *Ubx* proteins. The variant forms share common 5′ and 3′ exons but differ in their small internal 'micro' exons. Additional variation is generated by two separate splice donor sites at the end of the common 5′ exon, situated 27 bp apart. Northern hybridization and S1 nuclease protection studies of RNA from various developmental stages and tissue types reveal that the alternate splicing and the choice of polyadenylation site are each differentially regulated in both a temporal and a tissue specific manner. Additional transcripts were found just downstream of the *Ubx* transcription unit, which may be products of the *lethal left of bithorax* gene (*llb*).

Key words: cDNA/*Drosophila*/homeotic/segmentation/*Ultrabithorax*

Introduction

The fruit fly *Drosophila melanogaster* has a body composed of a number of segments. The process of segmentation is controlled by maternal gene products, which specify the spatial coordinates of the embryo (Nusslein-Volhard, 1979), as well as zygotic gene products, which determine the number and polarity of the segments (Nusslein-Volhard and Wieschaus, 1980). The individual identities of the segments are then controlled through the action of two loci, the Antennapedia and bithorax complexes (reviewed by Duncan, 1987; Kaufmann *et al.*, 1980). Mutations within these two loci transform parts of one segment into those of another segment.

Genetic studies (Lewis, 1978; 1981; 1982) have shown that there are at least five distinct classes of mutations associated with lesions within the thoracic part of the bithorax complex. The most severe, termed *Ubx*, are embryonic or larval lethals; they cause transformations from the posterior compartment of the second thoracic segment (pT2) through to the anterior portion of the first abdominal segment (aA1); i.e. the primary realm of *Ubx* action is in parasegments 5 and 6 (Hayes *et al.*, 1984; Martinez-Arias and Lawrence, 1985). More limited effects are observed on abdominal segments (parasegments 7-13). There are also less severe recessive mutations which produce a subset of the *Ubx* transformations. Typically, these mutations cause strong transformations within a single parasegment. The effects of *abx* and *bx* mutations are primarily confined to pT2 and aT3 (parasegment 5) while *pbx* and *bxd* alleles transform pT3 and aA1 (parasegment 6) (Casanova *et al.*, 1985; Peifer and Bender, 1986). Thus, the *Ubx* phenotype can be considered as the sum of the four partial phenotypes exhibited by the *abx, bx, pbx* and *bxd* alleles.

The transformations due to mutations within the *Ubx* domain are not just limited to structures derived from the epidermis. Effects on both the central (CNS) and peripheral (PNS) nervous systems, as well as somatic musculature, have been documented (Ghysen *et al.*, 1985; Dambly-Chaudiere and Ghysen, 1987; Hooper, 1986). The specific lesions of over 30 of these mutations have been identified and located on the molecular map (Bender *et al.*, 1983a,b; 1985; Peifer and Bender, 1986), and some are shown in Figure 1.

It remains unclear how the various phenotypes associated with mutations within the *Ubx* domain are related to products coded for by the locus. Molecular studies have identified three discrete transcriptional units within the *Ubx* domain. One, with a 75 kb transcript from between −110 and −30 kb on the molecular map (Figure 1), is called the *Ubx* unit, since it spans the positions of all breakpoints with a *Ubx* phenotype (Bender *et al.*, 1983a,b; Hogness *et al.*, 1985). A second, called the early *bxd* transcript, covers a region of about 20 kb between −20 and 0 kb (Lipshitz *et al.*, 1987). This RNA does not appear to code for a protein, and its function, if any, remains unknown. A third short transcript from the *bxd* region at about −15 kb is found in pupae (Lipshitz *et al.*, 1987). It could encode a small protein, but its function is also unknown.

Three major size classes of RNA transcripts have been identified from the region of *Ubx* mutant lesions: a 4.7 kb poly(A)$^-$ species, which appears within the first 6 h of development and then recedes; and two poly(A)$^+$ species, 3.2 and 4.3 kb in size, which first appear at ~3 and 6 h of embryogenesis respectively and then persist throughout all subsequent stages of development (Hogness *et al.*, 1985). A transient poly(A)$^+$ species of 1.6 kb has also been observed (Struhl and Akam, 1985). *In situ* hybridization studies have shown that these transcripts are expressed in a spatially complex manner within all three germ layers during embryogenesis (Akam and Martinez-Arias, 1985).

Two *Ubx* cDNAs, believed to represent the 3.2 kb species, have been partially characterized (Beachy *et al.*, 1985). Each contains four exons: two larger 3′ and 5′ exons and two internal 51 bp micro exons. These two cDNAs differ by 27 bp due to alternative use of two splice donor sites in the 5′ exon. Antibodies raised against a portion of the 5′ exon's open reading frame have been used to detect several protein species resolved on SDS-PAGE (White and Wilcox, 1984; Beachy *et al.*, 1985; White and Wilcox, 1985a,b).

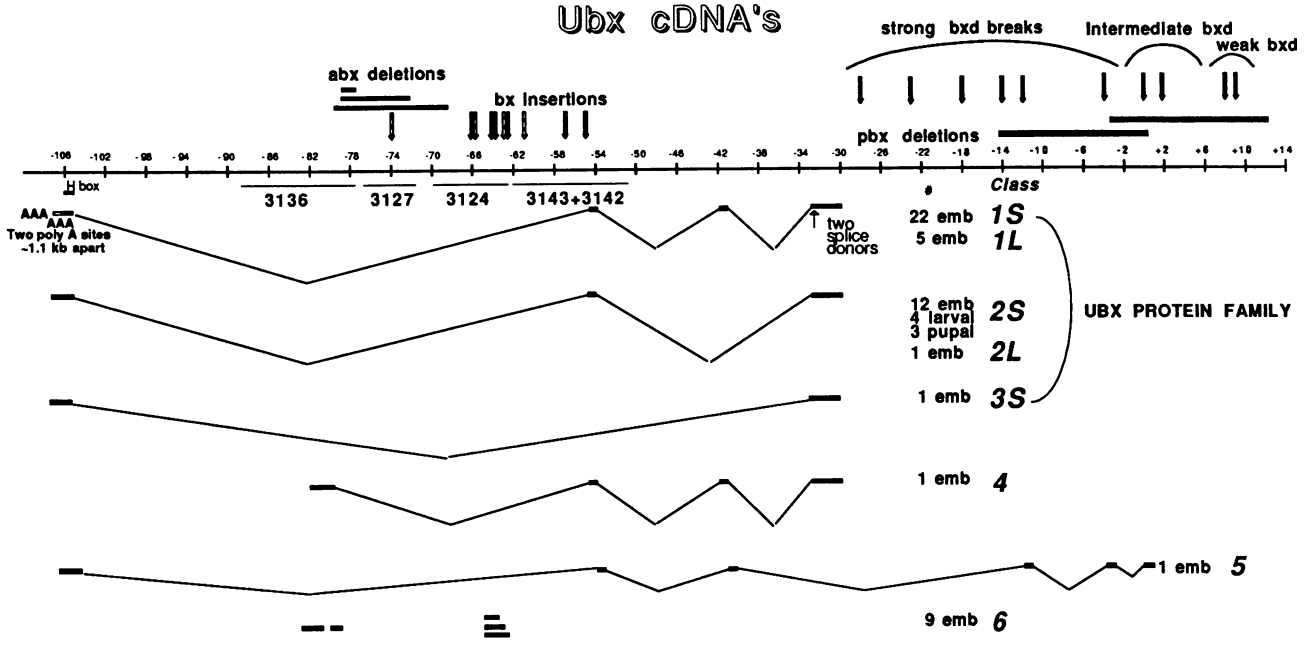

Fig. 1. *Ubx* cDNAs. The top line represents a 90 kb segment of genomic DNA of the bithorax complex, with the locations of some *abx, bx, pbx* and *bxd* mutant lesions shown above the line. The different structures of *Ubx* cDNAs are indicated below the line, with black bars to show the approximate locations of exons. The designation of each cDNA class is shown to the right of each structure diagram, along with the numbers of isolates found in each class and the time period of the library from which they were isolated. The S (short) and L (long) class designations indicate usage of the first and second donor sites respectively. The locations of the genomic subclones 3136, 3127, 3124, 3143 and 3142, which were used to obtain the class 6 cDNAs, are shown immediately below the genomic DNA line.

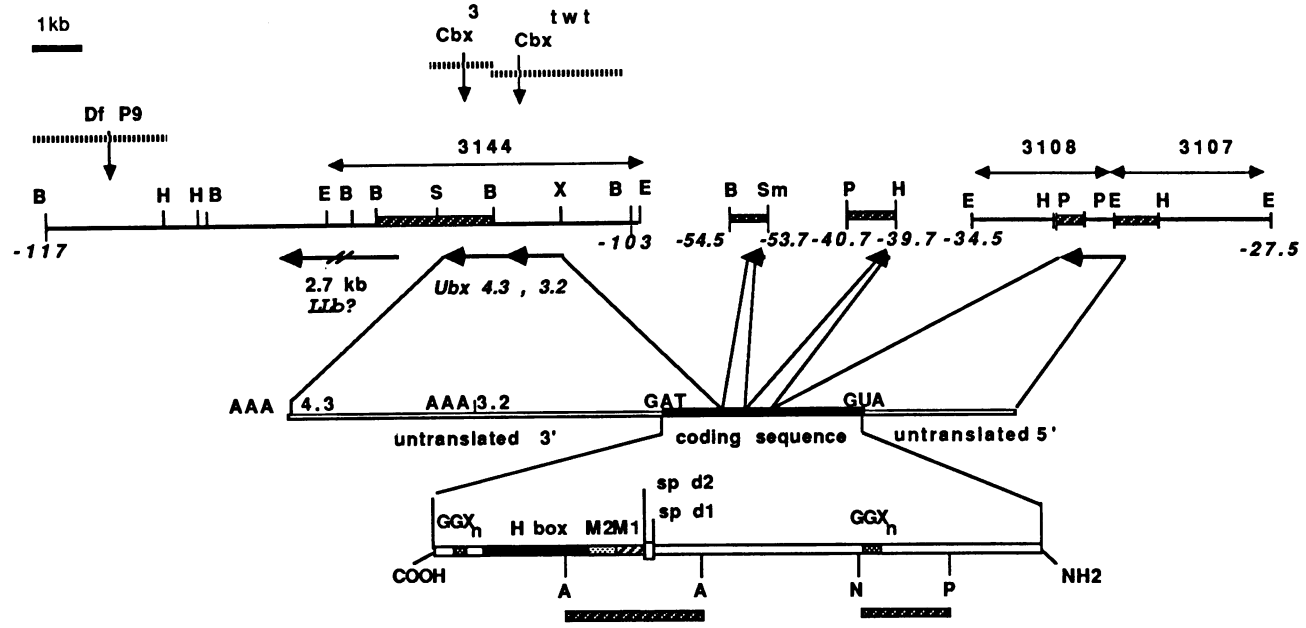

Fig. 2. Detailed structure of the class 1L cDNA. The top set of lines represents fragments of *Ubx* genomic DNA. The coordinates are shown in kilobases, as in Figure 1. Fragments from which exon-specific probes were made are indicated with thick diagonal striped bars. The breakpoints of the Cbx^3 and Cbx^{twt} mutations and the left endpoint of the Df(3R)P9 deletion are shown with vertical arrows above the genomic map, with horizontal hatched lines indicating the uncertainties of these positions. The extents of subclones 3144, 3108 and 3107 are represented by the light double headed horizontal arrows. The regions of DNA from which mature transcripts are derived are shown with heavier horizontal arrows, pointing in the direction of transcription. The bar in the middle of the figure represents a mature *Ubx* transcript. The filled-in portion shows the coding region, while the open portions represent untranslated sequence. The two alternative 3' poly(A) sites, which give rise to the 3.2 and 4.3 kb messages, are indicated by the letters AAA. An expanded diagram of the coding sequences is at the bottom, showing the relative positions of the two splice donor sites (sp); the −40 micro exon (M1); the −50 micro exon (M2); the homeobox (H box); and the GGX_n repeats.

Heterogeneity of the antigens was assumed to reflect either post-translational modifications or alternatively spliced products.

In view of the variety of phenotypes shown by the *abx, bx, bxd* and *pbx* alleles, it was appealing to imagine a variety of protein products made by alternative splicing. The

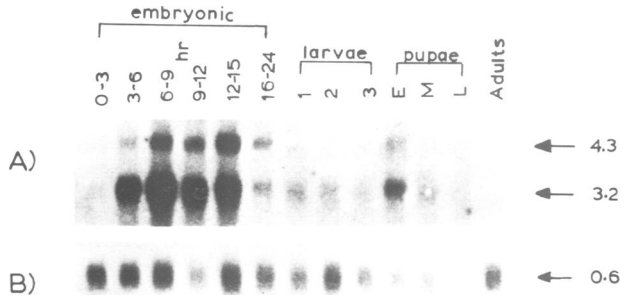

Fig. 3. Developmental profile of the *Ubx* poly(A)$^+$ transcripts. (**A**). Each lane contains 5 μg of polyadenylated RNA which was isolated from whole organisms of the indicated stages. The RNAs were transferred to nitrocellulose and hybridized with the *Ava*I fragment (Figure 2) from a class 1L cDNA. The sizes of the two prominent transcripts were determined by comparison with *E.coli* and calf liver rRNAs, which served as mol. wt standards (not shown). (**B**) The same blot is shown rehybridized to a probe for the RP49 ribosomal protein gene.

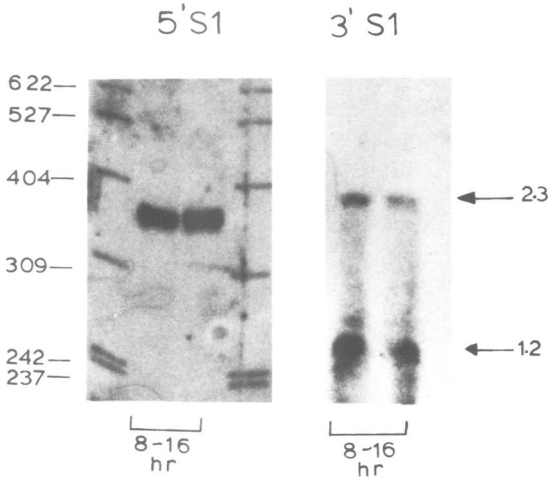

Fig. 4. Mapping the 5′ and 3′ ends of the *Ubx* transcripts by S1 nuclease protection. (**A**) The 1 kb *Eco*RI−*Hin*dIII fragment from subclone 3107 (Figure 2) was cloned into mp19. Single strands of this subclone were hybridized to 5 μg of poly(A)$^+$ RNA from 8−16 h old embryos and treated with S1. The protected fragment was electrophoresed on 6% sequencing gels, blotted to a nylon membrane and hybridized to ^{32}P-labeled subclone 3107. The markers are ^{32}P-end-labeled *Msp*I fragments of plasmid pBR322. (**B**) The 3.2 kb *Xho*I−*Sal*I fragment from subclone 3144 (Figure 2) was cloned into phage mp18. Single stranded copies were hybridized with 5 μg of poly(A)$^+$ RNA from 8−16 h embryos and treated with S1. The protected fragments were electrophoresed on an alkaline agarose gel, blotted to a nylon membrane and hybridized to ^{32}P-labeled subclone 3144. End-labeled digests of pBR322 served as mol. wt markers (not shown). The sizes of the two protected fragments are indicated.

failure of complementation of these mutations with *Ubx* alleles could be explained if the weaker alleles each affected a different variable exon, while the *Ubx* alleles affected exons constant to all the protein products. We therefore sought additional RNA products from the *Ubx* domain. We present evidence here for four additional *Ubx* RNAs with the potential to encode protein products, and for several molecules that probably do not. We also show that the spectrum of RNA products changes with time and tissue, although we now doubt that this variation in products is sufficient to explain the range of mutant phenotypes.

Results

Isolation and characterization of cDNAs

We chose to look for alternative splicing in the *Ubx* domain by examining the structures of a large number of cDNAs. To date we have screened several embryonic libraries from L.Kauvar (Poole *et al.*, 1984) and from N.Brown (N.Brown and Kafatos, personal communication). We have performed more limited screenings of several of the pupal and larval libraries from L.Kauvar. The fragments from the *Ubx* domain used as probes are shown in Figures 1 and 2. A 680 bp *Pst*I genomic fragment, containing much of the 5′ exon's coding region, was used to identify the first 18 clones. Subsequently 26 additional clones were obtained by using a representative cDNA as a probe. Each isolated cDNA clone was mapped by restriction endonuclease digestion. Since most of the clones obtained from the Kauvar libraries were less than full length, we could assign clones as being independent if they were different in size. The libraries made by N.Brown had lower complexities, but they did contain several clones of the full length 3.2 kb RNA. Three of the full length isolates of the 2S class (Figure 1) are not necessarily independent cDNAs, since we cannot tell them apart by size.

Independent clones were hybridized to restriction fragments covering the entire *Ubx* 'walk' (−106 to +25, Bender *et al.*, 1983a,b). The hybridization profiles allowed us to identify roughly the numbers and positions of exon sequences. Subsequently, 20 cDNAs were subcloned into pEMBL vectors (Dente *et al.*, 1983) and partially sequenced. These sequence results established a general trend which allowed us to classify the remaining cDNAs on the basis of hybridization profiles and restriction endonuclease mapping data.

Figure 1 outlines the type and number of cDNA products which we have found. The majority of products, termed the *Ubx* family in Figure 1, represent alternative splicing variants which share common 5′ and 3′ exons but vary in their small internal exons. The most frequent class, 1S (short), shows four points of hybridization at approximately −106, −54, −40 and −30 kb on the map. Comparison of the sequence of five isolates of this type with the published genomic sequences for these regions (Weinzierl *et al.*, 1987) showed that each contains the two 51 bp 'micro' exons situated between two larger 5′ and 3′ exons, as did the cDNAs previously described (Beachy *et al.*, 1985). We have also found five examples of the 1L (long) type. As described by Beachy *et al.* (1985), the first exon of this variant has a splice donor site situated 27 bp distal to that of the 1S form, and thus it codes for a protein which is nine amino acids longer than the 1S protein.

The remaining classes represent novel cDNA products which have not been previously described. The most abundant of these are the class 2 type. They have a structure similar to that of the class 1 cDNAs, but they are missing the first micro exon from the −40 kb region. Clones representing both the possible 5′ splice donor sites have been found, and again, it is the upstream (first) donor site that is preferentially used. The third class of cDNAs found contains only one member; it is similar to the first two classes except that it contains neither micro exon. The single clone obtained uses the first donor site; however, we believe that

the second donor site can also be used (see nuclease protection data). We have found no examples of cDNAs which contain only the −40 kb micro exon and not the −50 kb micro exon. Our nuclease protection experiments indicate that such a species is probably not made to any significant extent (see below). The three classes of alternatively processed transcripts code for a family of related proteins which vary between 389 amino acids (the 1L form) and 346 amino acids (the 3S form). These three classes have also been identified recently by K.Kornfeld and D.Hogness (personal communication).

The remaining cDNA classes lack at least one of the large 5′ or 3′ common exons characteristic of the *Ubx* family. The single class 4 isolate contains the same 5′ exon and two micro exons as the class 1 *Ubx* cDNAs, but it has a new 3′ exon. Hybridization mapping indicates that the position of the 5′ end of this exon is at about −81 kb on the genomic map; this is within 1 kb of the end points of the three known *abx* deletions (Peifer and Bender, 1986). Sequence determination at the junction of the second micro exon with the 3′ exon indicates that the *Ubx* open reading frame continues into the 3′ exon for only 62 bp before a translation stop signal is encountered. If translated, this RNA would result in a protein which is significantly shorter than the normal *Ubx* products and which would not contain a homeobox. The codon usage within this region does not conform with that tabulated for known *D.melanogaster* structural genes, although 21 amino acids is a relatively small sample. The 3′ exon was used to probe Northern blots containing poly(A)$^+$ RNAs from various times in development, but no discrete transcripts were found. The most likely source of this product is an aberrant splice to a fortuitous splice acceptor site located in the large *Ubx* intron.

The single class 5 isolate has the micro exons and large 3′ exon like those of the class 1 *Ubx* cDNAs, but it lacks the *Ubx* 5′ common exon. Instead, it hybridizes to three blocks of sequences from within the *bxd/pbx* region. These three exons are identical, with one modification, to exons 1,2 and 3 of the *bxd* RNA, as described by Lipshitz *et al.* (1987). These three exons lack a significant open reading frame, and they do not have a methionine codon from which translation could be initiated in-frame with the *Ubx* exons. Lipshitz *et al.* (1987) never found all three of these exons together in one *bxd* cDNA, and exon 2 in our isolate is 75 bp longer on the 5′ end than they described. We presume that their cDNA beginning with exon 2 (No. 3604) is incomplete at the 5′ end. The class 5 cDNA is novel in that it shows linkage of *bxd* and *Ubx* transcription units. As with the class 4 product, however, we were never able to detect on Northern blots a band corresponding in size to this cDNA. This failure, together with the lack of a continuous open reading frame, suggests that it is not an important product involved in either *bxd* or *Ubx* function. This clone was probably generated by rare readthrough of the *bxd* transcription termination signal, followed by splicing to the first *Ubx* micro exon.

The final group of cDNAs, class 6, were obtained by hybridizing libraries with a mixture of probes which cover the *abx/bx* region (−87 to −50 kb). They are small and show no evidence of splicing. Sequence analysis of several of the class 6 examples showed no significant open reading frames. When these clones were hybridized to Northern blots, no discrete signals were observed. They were probably generated by internal priming at fortuitous poly(A) sequences in fragments of *Ubx* nuclear RNA, which would be included in the poly(A)$^+$ RNA preparations used to make the libraries.

Ubx RNAs comprise two major bands on a Northern blot, at 4.3 and 3.2 kb (Akam and Martinez-Arias, 1985; Hogness *et al.*, 1985). This size difference was shown to reflect a difference in the 3′ endpoint of *Ubx* mRNAs, and so we expected to find this variation in the 3′ ends of the *Ubx* cDNAs of classes 1−3. We found no cDNAs from the embryonic libraries of Kauvar, or the 12−24 h embryonic library of Brown, which represent the 4.3 kb *Ubx* RNA. The distinguishing feature of the 4.3 kb species is that it hybridizes to the 2.8 kb *Bam*HI fragment of subclone 3144, whereas the 3.2 kb species terminates in the 3.2 kb *Bam*HI fragment (Figure 2). We sought to isolate a cDNA representative of the 4.3 kb message by using the 2.8 kb *Bam*HI fragment as a probe. We found two clones from Brown's 8−12 h embryonic library that hybridize to both the 2.8 kb *Bam*HI fragment and the 3.2 kb fragment, suggesting that they are candidates for the 4.3 kb *Ubx* mRNA species. The longer of these two clones (2.5 kb) contains homeobox sequences, but it ends 38 bp short of the splice acceptor site of the 3′ *Ubx* exon. Restriction analysis suggests that it contains no additional exons at the 3′ end, and thus the 4.3 species of *Ubx* mRNAs appears to be a simple 1.1 kb extension of the 3.2 kb species produced by alternative polyadenylation. This assertion is supported by nuclease protection studies described below. The absence of many positive 4.3 kb cDNA clones is somewhat curious; it may be that the large strings of A present in the sequence of this region (Weinzierl *et al.*, 1987) inhibited proper priming and/or extension during synthesis of the cDNA libraries.

The screen for cDNAs with the 2.8 kb *Bam*HI fragment also turned up eight previously unidentified clones from the 0−8 h library of Brown. All eight clones were ~2.7 kb in length and all had similar restriction maps, which were unlike the canonical *Ubx* clones. A representative example was hybridized to restriction fragments of the *Ubx* 'walk'. As illustrated in Figure 2, this cDNA is totally confined within the region of −112 to −108, downstream of the *Ubx* 3′ exon. The direction of transcription is from right to left as determined by restriction enzyme mapping of the cDNA and genomic fragments. We have not yet determined its splicing pattern. Hybridization to Northern blots revealed that a 2.7 kb transcript is present throughout development and that a 4.2 kb species is also expressed during 6−15 h of embryogenesis (data not shown). These transcripts could be the products of the *llb* (*lethal left of bithorax*) gene. Sanchez-Herrero *et al.* (1985) and Tiong *et al.* (1985) have shown that the deficiency P9 (DfP9) eliminates *llb* function. Mark Peifer (personal communication) has mapped the left-hand endpoint of this deletion within the 3 kb *Bam*HI−*Hin*dIII fragment (Figure 2), and so DfP9 completely eliminates the 2.7 kb transcription unit. We are currently examining *llb* mutations for evidence of changes in this transcript.

Mapping the 5′ and 3′ ends of Ubx mRNA species

Struhl and Akam (1985) found several minor *Ubx* RNAs in addition to the canonical 3.2 and 4.3 kb species. The most prominent of these is a 1.6 kb species, which appears in late embryos, that hybridizes to a 5′ *Ubx* probe but not a 3′ probe. Since we did not find cDNA representatives of these other transcripts, we investigated how many *Ubx* RNA

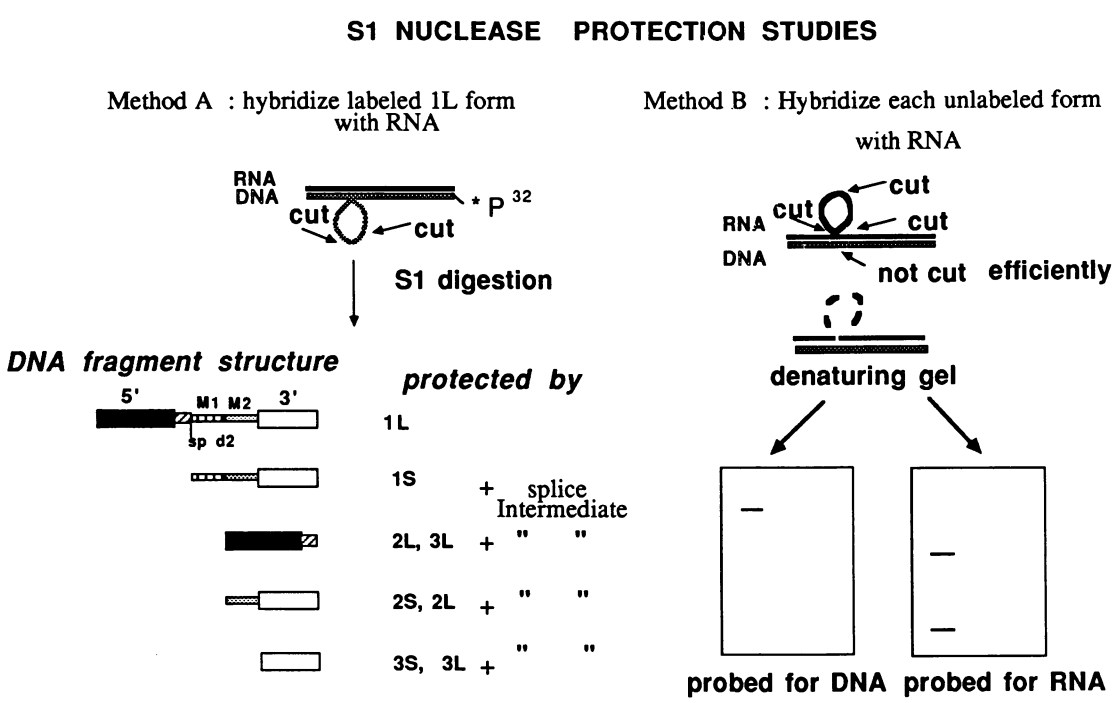

Fig. 5. Strategies for S1 nuclease protection studies. The left-hand portion of the figure shows some of the fragments which might be observed when a uniformly labeled fragment equivalent to the 1L species is used in a traditional S1 protection experiment. RNA of the 1L class will give the full length protected product, but other fragments arise from RNA of classes 1S, 2L, 2S, 3L, 3S, 4 and 5, as well as from splicing intermediates. The source of these fragments cannot be unambiguously identified. They could also arise from other RNA species which do not have the four canonical exons. The right-hand portion illustrates the method used in these experiments. An unlabeled DNA homolog of each cDNA class is used separately. The digestion products are electrophoresed, blotted and then hybridized with strand specific probes. If a probe for the DNA strand is used, anomalous results are obtained because S1 does not cleave efficiently across from the looped out RNA. This can lead to full length protection even if there is no RNA species exactly equivalent to the DNA homolog. We therefore probe for the RNA strand of the hybrid.

species exist and how many different transcription start sites and polyadenylation signals are used.

Figure 3 shows the results of a Northern hybridization experiment using poly(A)$^+$ RNA. In the example shown, a 268 bp *Ava*I (Figure 2) fragment from a class 1L cDNA was used as a probe. This fragment spans the splice junction sites and contains both micro exons, as well as 112 bp from the 5′ exon and 54 bp from the 3′ exon. Only two bands of hybridization are observed: a 3.2 kb transcript which initially appears at about 3 h of development, and a 4.3 kb species which is first detected at about 6 h of embryogenesis. The 4.3 kb transcript peaks between 12−15 h of development about 3−6 h after the 3.2 kb species peaks. Both transcripts are relatively abundant until late in embryogenesis (>16 h), at which time the levels of each drop significantly. We see the same profile when a 5′ exon probe (the 1.4 kb *Eco*RI−*Hin*dIII fragment of subclone 3107, Figure 2), or a 3′ exon probe (the 3.2 kb *Bam*HI fragment from subclone 3144) is used. Only when we use an internal *Nae*I−*Pst*I fragment from the 5′ exon do we see hybridization to additional RNA species, including a broad band of about 1.6 kb (data not shown). We note that this *Nae*I−*Pst*I fragment contains 16 repeats of the sequence GGX (termed the hinge region by Beachy *et al.*, 1985). We believe that the 1.6 kb species, as well as the other minor bands, are due to cross-hybridization with GGX sequences in mRNAs derived from other genomic loci. This repeat is present at numerous sites within the *Drosophila* chromosome (Haynes *et al.*, 1987) and we have detected at least six of these GGX-rich regions within the BX-C walk (data not shown).

In Figure 4, panels A and B show, respectively, S1 nuclease protection studies of the 5′ and 3′ ends of the *Ubx* transcription unit. At the 5′ end, we see protection of a single fragment, indicative of one major transcription start site located ~360 bp upstream of the *Eco*RI site at −32 kb on the molecular map. Two of our cDNAs end at an A located 354 bp from the *Eco*RI site. Recently Saari and Bienz (1987) reported the location of the *Ubx* transcription start site on the basis of S1 nuclease protection studies and primer extension experiments. They find major and minor initiation sites located at 357 and 354 bp upstream of the *Eco*RI site. The two cDNAs mentioned above both appear to represent initiation from the minor start site.

At the 3′ end (panel B) we see two prominent protected fragments of 2.3 and 1.2 kb in size. The former is consistent with protection of the input DNA strand by the 3.2 kb RNA species, while the latter arises by protection with the 4.3 kb species. We interpret these results to indicate that the 4.3 kb RNA species has no additional 3′ exons; its 3′ end is simply a 1.1 kb extension of the 3.2 kb species to an alternative polyadenylation signal. We present additional analysis of the internal splicing patterns of the 4.3 kb species below.

Temporal regulation of alternative Ubx RNAs

We wished to measure the levels of the different *Ubx* splicing products throughout development by a nuclease protection assay. Traditional S1 analysis using the longest 1L, species as a homolog as shown in Figure 5A, can lead to misleading estimates of the various forms if there are signifi-

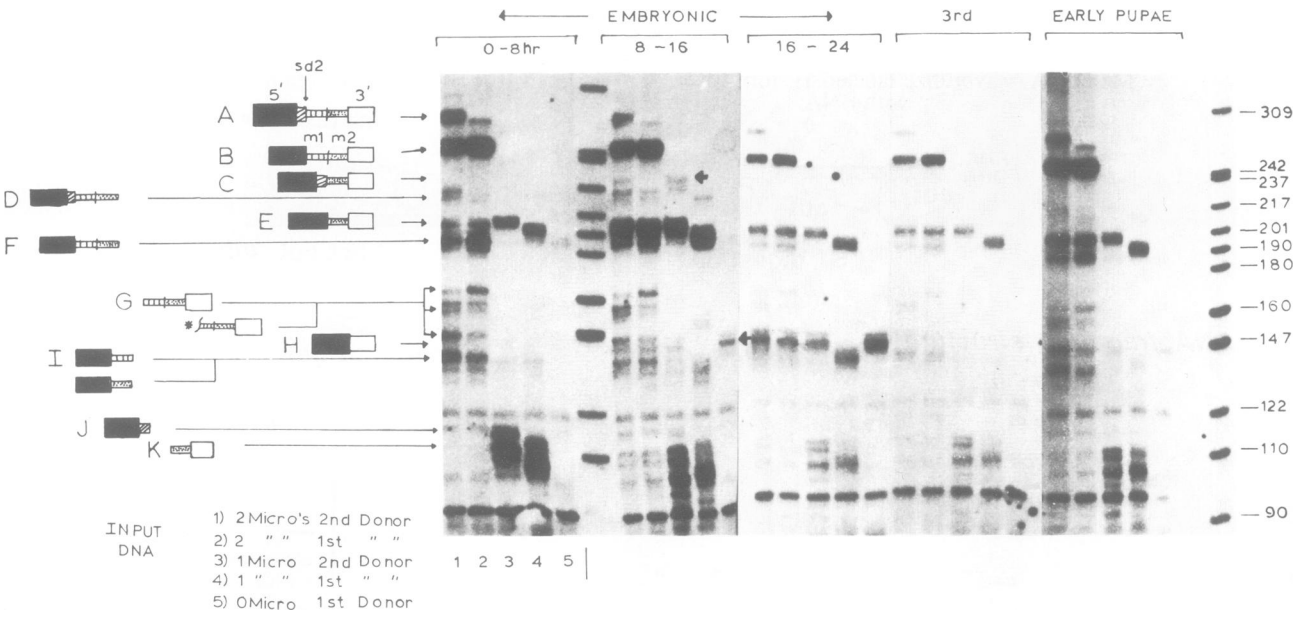

Fig. 6. S1 nuclease protection studies of the five alternate *Ubx* cDNAs. The internal *AvaI* fragment (Figure 2) from a representative of each of the five alternate types of *Ubx* cDNAs (1L, 1S, 2L, 2S and 3S) was cloned into phage mp18. Single strands, prepared from each class, were hybridized with 5 μg of poly(A)$^+$ RNA from each of the indicated stages. The hybrids were treated with S1 and the digestion products fractionated on 6% sequencing gels. The fragments were blotted onto a nylon membrane and hybridized with ^{32}P-labeled single stranded probe specific for the RNA strand of the hybrid. For each stage there are five lanes. Each lane shows the digestion products derived from one of the five alternate DNA templates. The order of the templates, from left to right within each stage, is 1L, 1S, 2L, 2S and 3S. The structure of each fragment is illustrated schematically on the left of the figure. The structures were deduced from size measurements and from rehybridization experiments using probes specific for each exon. The exact structures of the fragments labeled G are uncertain since the fragment shown should yield only one band. The lower bands may be partial degradation products. The arrows in the 8–16 and 16–24 h panels show the positions of the 2L (C) and 3S (H) forms respectively. The slight shift in the 2S (E fragment) form seen in the fourth lane of each panel arises because a small deletion of a few base pairs was generated at the end of the cDNA *AvaI* fragment during the construction of the mp18 derivative. End labeled *MspI* digested pBR322 fragments served as mol. wt markers; their sizes in base pairs are indicated on the right. Between the 0–8 h and 8–6 h lanes there is an additional marker lane.

cant levels of splicing intermediates present. Also, one cannot unambiguously distinguish the relative levels of the 2L, 3L and 3S forms since no unique representative fragments exist for these species. We therefore used a representative cDNA from each class and examined the fully protected species. This too can lead to a misleading estimate of the various forms, as shown in Figure 5B. If an RNA–DNA heteroduplex with an insertion loop is treated with S1, the nuclease does not efficiently cleave the shorter strand, since there are no mismatched bases opposite the insertion loopout. The alternative *Ubx* RNAs differ in the presence or absence of internal exons. If a labeled DNA strand of one of the shorter species is hybridized with RNA of a longer species, then the RNA–DNA hybrids with precise insertion loops will be generated, and the labeled DNA strand will be protected.

We attempted to optimize the S1 conditions using *Ubx* mRNA molecules generated *in vitro*. Increasing the S1 concentration over a 5-fold range does not enhance the cleavage opposite a loop, but increasing the temperature of the digestion does have dramatic effects. At 37°C, <5% of the hybrids show DNA cleavage across from the RNA loop, while at 55°C, ~80% are digested. Unfortunately, at this high temperature the products are partly degraded, presumably due to digestion at partially melted regions (data not shown).

To surmount these difficulties, we have followed the RNA strand after S1 digestion. Unlabeled mp18 templates, containing the diagnostic *AvaI* fragments from the five distinct *Ubx* family members, were hybridized individually with poly(A)$^+$ RNA from different developmental stages. The hybrids were treated with S1, electrophoresed on denaturing polyacrylamide gels, blotted onto nylon membranes and hybridized with strand specific probes complementary to the *Ubx* RNA.

Figure 6 shows representative blots containing RNA from five stages of development which illustrate the changes in splicing combinations. For each time point there are five lanes. Each lane shows the RNA products of S1 digestion of RNA–DNA hybrids with DNA representing one of the splicing possibilities. The lanes are ordered according to the size of the DNA probe, with the longest 1L form (two micros 2nd donor site) on the left and the smallest 3S (no micros 1st donor) on the right. In each lane the largest fragment represents full length protection of the RNA species equivalent to the input DNA strand. The intensity of this band accurately reflects the level of that particular species at that time in development.

Each input DNA species will quantitatively protect its full length RNA homolog, but it will also largely protect all smaller related RNA species. The identities of the smaller fragments have been determined by measuring their sizes, and by rehybridizing the blots with probes specific for each exon. The deduced structures for these smaller fragments are illustrated on the left hand portion of the figure. As an example, consider lane 1 of the 0–8 h embryonic time point. Band (A) corresponds to the full length RNA of the 1L class. Bands B and E represent protected RNA of the 1S and 2S forms, respectively. Band F represents RNA with only the

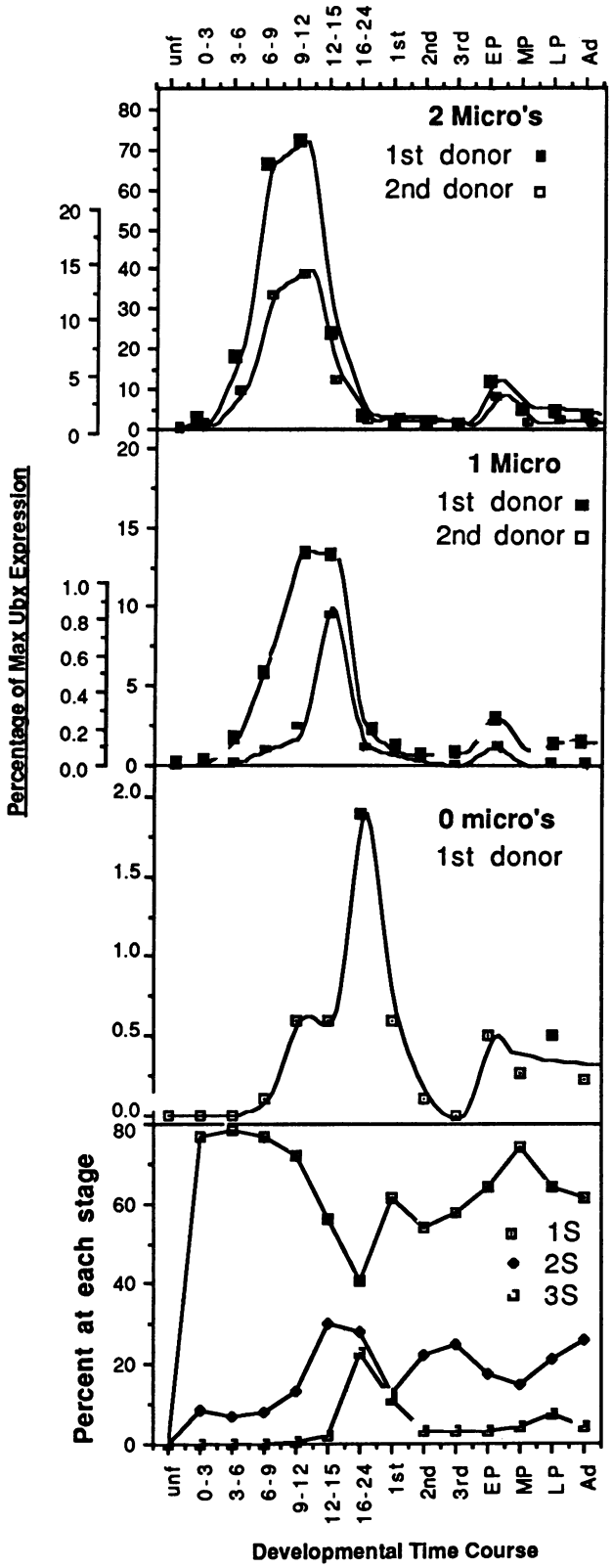

Fig. 7. Developmental profiles of each alternately spliced *Ubx* RNA. Autoradiographs equivalent to those shown in Figure 6 were quantitated by densitometry as described in Materials and methods. For **panels 1–3**, (from the top) amounts are expressed as percentages where 100% is taken to be that level of *Ubx* message (both 4.3 and 3.2 kb species) present in 9–12 h embryos (peak expression). Each RNA class has a different scale; the shorter scales in panels 1 and 2 are for the L (second donor) forms. Note that the classes in each panel peak at a different time in embryonic development. In **Panel 4**, the data are plotted as a percentage of the total RNA at each stage. Only the 1S, 2S and 3S forms are shown.

5′ exon and the two micro exons and not the 3′ exon. This band could represent RNAs like our class 4 example, which have a new exon in place of the normal *Ubx* 3′ exon. We believe it is more likely that this band represents a splicing intermediate in which the last intron has not yet been removed. Fragment I may represent another splicing intermediate in which only the first intron has been removed. The presence of these fragments is consistent with the order of splicing being the same as the order of the exons. The group of fragments labeled G are missing the 5′ exon. Since the intensities of these fragments vary with the temperature of digestion (data not shown), they probably result from limited RNA cleavage across from a loop in the DNA.

Using the type of analysis depicted in Figure 6, we examined RNA from 14 stages of development: unfertilized eggs; 0–3, 3–6, 6–9, 9–12, 12–15 and 16–24 h embryos; 1st, 2nd and 3rd instar larvae; early, mid and late pupae; and adults. The results of densitometric tracings are plotted in Figure 7. We found no *Ubx* RNA in unfertilized eggs, consistent with the absence of any maternal effects of *Ubx* mutations (Kerridge and Dura, 1982). As embryogenesis proceeds, we see that the three major *Ubx* classes peak in successive order, with the class 1 type peaking earliest and the class 3 type last. The 1S form is the most prominent early (see 0–8 h, Figure 6) and remains so until about 12 h of development. At this time, the 2S form (one micro) makes a significant contribution and the 3S form (no micro) is just beginning to appear (see 8–16 h, Figure 6). Between 16–24 h of development the 3S form peaks at a level of about 25% of the total. The levels of the class 1 and 2 RNAs drop rapidly in later embryos, in keeping with the drop in total *Ubx* RNA seen on Northern blots. The relative ratios of all forms remains essentially unchanged throughout larval stages, except that the 3S species appears to drop somewhat. During early pupation all forms rise slightly, and then they regress to the low levels which are observed in adults. Note that the profiles of the short and long forms of each class are comparable. Apparently, the two alternative donor sites do not give rise to species with independent temporal patterns of expression in RNA from whole organisms. Essentially identical temporal profiles have been obtained by K.Kornfeld and D.Hogness (personal communication).

These nuclease digestion experiments allowed us to search for a species which contains only the first micro exon (from −40) and not the second. This species would be protected by input DNA of the 1L or 1S forms, because of the inefficiency of S1 in cutting RNA across from a DNA loop. The variant would have exactly the same size as fragment E (or C if the long form were present) since both micro exons are 51 bp in length. We hybridized the blots with a probe specific for the first micro exon, and found no signal at the E and C positions at any time of development. If such a first-micro-only *Ubx* RNA variant exists, it must be quite rare.

Tissue specificity of alternate Ubx RNAs

Furst and Mahowald (1985) have devised a technique which permits the mass isolation of relatively pure cultures of embryonic neuroblasts and, with slightly less purity, embryonic muscle precursor cells. RNA prepared from such cells was examined to see which of the *Ubx* RNA species are present in these isolated cell types. In Figure 8A, lanes 2 and 3 show the *Ubx* RNAs from 4–6 h old myoblasts and neuroblasts, respectively. The neuroblast RNA shows

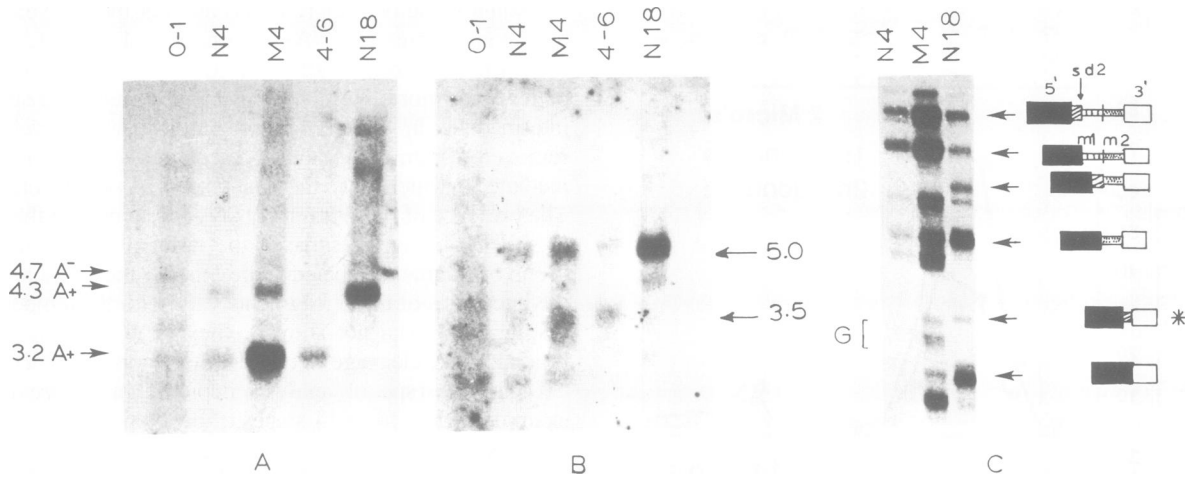

Fig. 8 .Northern and S1 analysis of RNA from isolated cell types. (**A**) Each lane contains 10 μg of total RNA from either 0–1 h embryos, 4–6 h neuroblasts (N4), 4–6 h myoblasts (M4), 4–6 h embryos, or neuroblasts maintained in culture for 18 h (N18). The purity of the N4, M4 and N18 cells was ~95, 85 and 99% respectively as judged by cell morphology. The blot was transferred to nitrocellulose and hybridized with the labeled internal *Ava*I fragment (Figure 2) from a representative *Ubx* cDNA. The 4.3 and 3.2 kb poly(A)$^+$ species are indicated. In the 0–1 h lane, the 4.7 kb species is probably the poly(A)$^+$ RNA product described by Hogness *et al.* (1985). (**B**) The same blot was rehybridized with a labeled *Antp* cDNA. The faint band immediately below the 5.0 kb transcript in the N18 lane may be caused by cross hybridization to the *Ubx* 4.3 kb species, due to the homeobox sequences. (**C**) S1 nuclease digestions were used to measure the relative amounts of the different *Ubx* products in isolated cell types. The experiment was done as described in Figure 6 except only the 1L class of template mp18 was used. The fragments corresponding to full length protection of the differently spliced variants are indicated. The asterisk denotes a band that probably represents the 3L form, for which we did not obtain a cDNA. The bands labeled G appear the same as those described in Figure 6.

a definite enrichment for the 4.3 kb species relative to an equal amount of RNA from 4–6 h embryos (lane 4). Myoblasts contain predominantly the 3.2 kb species. Once plated, the neuroblast cultures can be kept alive for many hours. Over a period of 18 h, the cells divide, extend axons, and look much like differentiated neuronal cells (Furst and Mahowald, 1985). As shown in lane 5, RNA from these differentiated neurons contains the 4.3 kb *Ubx* mRNA almost exclusively. Some preparations of such RNA do contain low levels (<5%) of 3.2 kb RNA, perhaps due to low level contamination by non-neural cells.

Studies by Laughon *et al.* (1986), Streoher *et al.* (1986) and Schneuwly *et al.* (1986) have demonstrated that transcription of the *Antennapedia (Antp)* gene also produces two classes of mRNAs which differ in their choice of polyadenylation sites. As with *Ubx*, the longer messages are simple 3′ extensions of the shorter variety by about 1.2 kb. We rehybridized the blot shown in Figure 8A with a probe specific for *Antp* mRNA. As illustrated in Figure 8B, the longer forms of the *Antp* transcripts are enriched in the aged neuronal cells. Each size class of *Antp* message is actually composed of two different transcripts which initiate from separate promoters. We do not yet know whether both of the *Antp* promoters contribute to the signal seen with neuronal RNA. Streoher *et al.* (1986) noted a similar enrichment of the longer *Antp* transcript in isolated ventral nerve cords. Curiously, the muscle cells (panel B, lane M4) fail to show a marked enrichment for the smaller species of *Antp* transcript, as was true for *Ubx* RNA.

We next examined, by S1 nuclease protection, the *Ubx* splicing patterns utilized in these cell types. As illustrated in Figure 8C, the aged neuronal cells show a marked preference for the 2S (one micro) and 3S (no micro) forms, while the myoblast RNA is highly enriched for the 1L and 1S (two micros) species. Note that in the neuronal lane there is a fragment of about 167 bp labeled with an asterisk. This fragment probably represents the 3L (no micro, second donor site) form for which we have no cDNA example. This fragment hybridizes to both 3′ and 5′ exon probes but not to probes specific for the micro exons. In contrast, the G fragments in the myoblast lanes do not hybridize to the 5′ end probe, and they probably represent products due to limited cleavage across from a DNA loop. Our inability to identify the 3L species in RNA from whole embryos suggests that there could be other rare *Ubx* mRNA variants which may be expressed only in particular cell types.

Discussion

How many Ubx products are there?

The *Ubx* gene is regulated in a complex spatial and temporal manner (Akam and Martinez-Arias, 1985), and many different tissues are transformed in *Ubx* mutant animals. The different responses of different tissues to *Ubx* expression could reflect the use of alternate protein products. It therefore seems essential to catalog the different *Ubx* gene products, and to learn where and when they are expressed.

Among *Ubx* cDNAs, we have found five distinct splicing products, and we believe that a sixth form is quite likely. The cDNA variability is more limited than we expected, given the large size of the transcription unit and the variety of mutant phenotypes. The products form a family of related proteins which share common 5′ and 3′ exons but differ with respect to their numbers of internal 'micro' exons (2, 1 or none). This exon heterogeneity, combined with two alternative splice donor sites located on the 5′ exon, and two polyadenylation sites, can result in the production of as many as twelve distinct mRNA species. We could identify all six forms of the 4.3 kb RNA in neuronal cells, but could only definitively identify two forms of the 3.2 kb RNA (1L and 1S) in myoblasts. Note that since the myoblast cultures are only about 85% pure, we cannot be sure whether the low level of 4.3 kb RNA seen in these samples represents the true ratio in expression of these species in these cells, or

whether it is due to contamination. Similarly, the low level of 2S RNA seen in these myoblast samples cannot unambiguously be assigned as having come from the 3.2 kb species or the low level of 4.3 kb RNA which is present. Whether in other tissues, or at other times, the 3.2 kb species also produces the 2L, 3L and 3S forms is unresolved.

Since the splicing alternatives described can generate only six alternate *Ubx* protein products which do not contain variable exons from the *abx, bx, pbx* and *bxd* regions, we doubt that different products can account for the diversity of phenotypes among these mutations. We have not, however, examined directly whether parasegmental and/or compartment specific splicing of the identified products occurs. We did find two cDNAs with exons in the *abx* and *bxd* regions, but we do not believe that they contribute important functions. Neither cDNA contains a significant open reading frame in the alternate exons, and there were no RNA transcripts corresponding to these cDNAs observed on developmental Northern blots. It seems likely that both of these products result from rare errors in transcription termination and splicing. If these products do represent rare errors, they suggest that we have probably identified all of the major *Ubx* products expressed during embryogenesis. As of yet, we have not done detailed studies of other stages of development, and the possibility remains that at other times or in particular tissues, such as imaginal discs, additional products might be found. Our analysis has also neglected potential poly(A)$^-$ RNAs. One such RNA has been identified (Hogness *et al.*, 1985), but this 4.7 kb early embryonic form disappears before *Ubx* protein is detectable.

It is striking to note that the arrangement of the *bxd* and *Ubx* transcription units, as two differentially regulated promoters with alternately spliced products, resembles the architecture of the *Antp* transcriptional unit (Laughon *et al.*, 1986; Streoher *et al.*, 1986; Schneuwly *et al.*, 1986). In the latter case, however, transcription from the upstream promoter is thought to produce functional *Antp* products. Perhaps an evolutionarily early form of the *Ubx* domain was arranged similarly, such that the *bxd* and *Ubx* transcriptional units were fused, but what remains today are separate units in which the upstream *bxd* unit no longer produces meaningful products. Curiously, the alternate exons, which are spliced into the *Antp* transcript from the upstream promoter are noncoding, as are all the early *bxd* exons (Lipshitz *et al.*, 1987).

Ubx products show differential tissue and temporal specificity

Our examination provides evidence for temporal as well as tissue specificity in both splicing and polyadenylation. We believe that the temporal profile is most likely a reflection of the establishment of splicing and polyadenylation specificity in the development and differentiation of individual tissues during embryogenesis. Thus, in late embryos, the bulk of *Ubx* expression is confined to the nervous system and the splicing and polyadenylation profiles of RNA isolated from these embryos is similar to that of the isolated (N18) neural cultures.

While there is tissue specificity in splicing, it is not absolute in the cells which we have examined. In neural tissue (represented by the N18 cells) there is a strong preference for the 2S and 3S products, but other forms are also present at reduced levels. Similarly, myoblasts (M4) show predominantly the 1S and 1L forms, but the 2S species is also present. As development proceeds, these preferences may become more pronounced. Recently, Weinzierl *et al.* (1987) have shown that the mutation Ubx^{195} is caused by a stop codon in the second micro exon. The Ubx^{195} mutation was found to be unique among several *Ubx* mutations examined. While transformation of the larval CNS is observed, the adult CNS is not completely transformed. It was suggested that since Ubx^{195} should still make the 3L and 3S forms of the protein, these species of the *Ubx* protein family might be primarily responsible for segmental identity of the adult CNS. The larval CNS must require additional members of the *Ubx* family. Cultured neuronal cells, which are thought to be representative of larval neurons, express approximately equal levels of both the one-micro and no-micro species of *Ubx* transcripts. Since the larval CNS is transformed by the Ubx^{195} mutation, the 3L and 3S *Ubx* products must not be sufficient for its differentiation. We expect that a mutation in the first micro exon would have a more normal larval CNS, since the class 2 and class 3 *Ubx* proteins would be unaffected.

The tissue specificity in the choice of polyadenylation site is more dramatic than the splicing differences. We observe strong bias in myoblasts toward use of the first site, leading to production of the 3.2 kb message. In N18 neuronal cells, the 4.3 kb message is highly favored. The nervous system may in fact represent the major source of the 4.3 kb message. Akam and Martinez-Arias (1985) found by *in situ* hybridization that a probe specific for the 4.3 kb species showed preferential labeling in the nervous system. Since an equivalent 3.2 kb specific probe cannot be made, they were not able to determine unequivocally whether the 3.2 kb species was also present. The fact that another homeotic gene, *Antp*, shows a similar arrangement and tissue-specific utilization of alternate 3′ polyadenylation sites suggests some functional role for the different 3′ untranslated sequences. Since we see a correlation between the use of a particular polyadenylation site and the preference for certain splicing combinations, the two may be related. We know of no example, however, where splicing or polyadenylation specificity is controlled by sequences far from the relevant processing sites. The 3′ untranslated sequences might affect RNA stability or translation efficiency. There are two mutations, Cbx^{Twt} and Cbx^{3}, which interrupt the 3′ untranslated regions (Figure 2). Both produce a dominant transformation of anterior T2 to anterior T3 (Bender *et al.*, 1983), and Cbx^{3} has *Ubx* protein in the anterior wing disc, where it does not normally appear (White and Akam, 1985; Cabrera *et al.*, 1985). It is conceivable that *Ubx* RNA is normally made in the wing disc, but rapidly degraded, and that these mutations might stabilize the RNA. Other models for the dominant phenotype, such as aberrant segmental control of transcription initiation, seem more plausible.

Implications for models of Ubx function

Finally, we consider what purpose the alternate forms of the *Ubx* proteins might serve. The six alternative spliced variants do not have completely independent profiles of expression. We find that the longer form of each subgroup within the *Ubx* family is always the minor species and that its pattern of expression mirrors that of its equivalent first donor relative. Wilde and Akam (1987) have found that there is relatively high sequence variation between *D.melanogaster D.funebris* and *D.pseudoobscura* in the 27 bp region between the two donor sites. They also report that the second donor site is missing altogether in *Musca domestica*. Since

we find that the usage of the second donor site appears to be less frequent and that its profile is not independent of the first, it is unclear whether the products arising from its use are necessary for proper *Ubx* function. Differences in splice donor site usage might be more subtle than we could have detected in whole animal extracts or in the two tissues which we examined.

It seems likely that the larger variation resulting from the presence or absence of micro exons is significant, especially since these changes are regulated and show tissue specificity. The variable region is situated just upstream of the conserved homeobox element and just downstream of another small conserved region that encodes the peptide YPWM (Tyr-Pro-Trp-Met) (Mavilio *et al.*, 1986; Krumlauf *et al.*, 1987). The *Antp* gene shows a similar arrangement, with heterogeneity due to alternative splice donor sites occurring between the YPWM box and the homeobox (Schneuwly *et al.*, 1986; V.Stroeher and R.Garber, personal communication; J.Bermingham and M.Scott, personal communication). Curiously, they too find an optional exon, but in this case the variable exon (exon 6) is located upstream of the YPWM box. The variability produced by alternate splicing might influence the DNA binding specificity of the homeobox domain and/or alter the interactions of the *Ubx* protein with itself or other proteins. J.Theis and P.O'Farrell *et al.* (personal communication) recently demonstrated that the *engrailed* protein can associate with itself and other homeobox containing proteins. This association requires a 5′ portion of the homeobox encompassed in the 23 N-terminal amino acids. Since these sequences are immediately adjacent to variable regions in both the *Ubx* and *Antp* proteins, the variable sequences could influence the formation of multimers, either between the alternate forms of each family or between members of different homeobox-containing protein families. Different multimeric units could, in turn, have different DNA binding sites or regulatory interactions.

To evaluate the roles of these alternative *Ubx* RNAs, we would like to alter specific sequences within the transcription unit. Such experiments have not been possible because of the large size of the *Ubx* domain. We have recently devised a procedure for building up and manipulating the entire unit on low copy number plasmids. This should allow us to mutate the *Ubx* transcription unit *in vitro*, and to transform it into flies to assay the consequences of our changes.

Materials and methods

Isolation of Ubx cDNA

Libraries of cDNAs were obtained from L.Kauvar (Poole *et al.*, 1985), and N.Brown (unpublished). Initially some of the embryonic, larval and pupal gt10 libraries of L.Kauvar were screened with a 680 bp *Pst*I fragment from the 5′ *Ubx* region (see Figure 2). A total of 18 independent cDNAs were obtained from the Kauvar 6–9 h E6 embryonic library, three from the pupal Q4 library and four from the third instar H library. Subsequently, the 2.6 kb *Eco*RI fragment representing cDNA E6-16 was used as a probe to screen the E7,8 libraries. Twenty six additional clones were obtained. To look for putative '*abx*' clones, the genomic subclones 3136 and 3127 (Figure 1) were nick translated, pooled and used as probes to screen the Kauvar E6 library. Five were obtained. Five cDNAs homologous to the '*bx*' region were isolated by hybridization to a mixture of nick-translated subclones 3124, 3143 and 3142 (Figure 1).

More recently three embryonic libraries (0–8, 8–12 and 12–24 h) from N. Brown were screened with either a labeled 2.8 kb *Bam*HI fragment from subclone 3144, or a ~260 bp *Ava*I fragment from cDNA E6-7 (Figure 2). Seven cDNAs homologous to the *Ava*I probe were obtained from the 12–24 h library while 13 were obtained from the 0–8 and 8–12 h libraries which were homologous to the 2.8 kb *Bam*HI fragment.

Preparation of DNA and RNA hybridization probes and hybridization conditions

Purified DNA fragments were labeled with ^{32}P by nick translation to a specific activity of 600 Ci/mmol. To make single stranded DNA probes, fragments to be labeled were first cloned into either phages pMP18 or -19 (Yanisch-Perron *et al.*, 1985), or the plasmids pEMBL8,9 (Dente *et al.*, 1983) which permit the recovery of single stranded DNA after infection with the phage FI. Radioactive probes were produced by Klenow extension of a reverse hybridization primer or sequencing primer (New England Biolabs) in the presence of [^{32}P]dCTP (600 Ci/mmol). To prepare high specific activity RNA probes for hybridization to blots containing S1 nuclease digestion products, fragments of interest were subcloned into pGEM vectors from Promega Biotec. Templates were then linearized with appropriate restriction endonucleases and transcribed with T7 RNA polymerase in the presence of [^{32}P]ATP (>3000Ci/mmol) according to the specifications supplied by Promega Biotec.

Hybridizations and washing conditions were as previously described (Bender *et al.*, 1983a), except that when RNA probes were used, the wash temperature was increased to 70°C.

DNA Sequencing

Fragments were first cloned into MP18,19 or pEMBL vectors and then sequenced using standard dideoxy chain termination procedures (Sanger *et al.*, 1977). To sequence across the splice junctions, deletions of individual pEMBL subclones containing cDNA inserts were constructed using known restriction sites.

RNA preparation and Northern hybridizations

Total RNA was prepared from wild type Canton S *D.melanogaster* of embryonic, larval, pupal and adult stages by the phenol/chloroform/urea method essentially as described by Wold *et al.* (1978). Polyadenylated RNA was selected by passage of these extracts through columns containing oligo deoxythymidylic acid–cellulose (T3: Collaborative Research) as described by Bantle *et al.* (1976). In typical Northern hybridization experiments, 5 μg of poly(A)$^+$ RNA was separated on 1.0% agarose, 2.2 M formaldehyde gels in 10 mM MOPS buffer as described by Maniatis *et al.* (1982). The RNA was then transferred directly by capillary action to either nitrocellulose or Nytran (Schliecher and Schell). The membranes were then baked for 2 h at 80°C under vacuum and hybridized with various probes as described.

Cultured cells from *Drosophila* gastrulae were prepared essentially as described by Furst and Mahowald (1985). Total RNA from cultured undifferentiated neuroblasts (N4), myoblasts (M4) and differentiated neurons (N18) was prepared by the guanidinium/cesium chloride method (Maniatis *et al.*, 1982; Perkins and Mahowald, in preparation). Briefly, 70–80 μg of total RNA can be recovered from 4.5×10^7 cultured neuroblasts. This same number of neuroblasts, when allowed to differentiate overnight, can yield up to 200 μg of total RNA. Myoblast cultures routinely yielded one third the amount of total RNA as neuroblast cultures (L.Perkins, unpublished observations).

S1 nuclease protection studies

A variation on the typical S1 mapping procedure (Berk and Sharp, 1977) was employed for the experiments described in this report. Single phage strands (pMP18) carrying various inserts were hybridized overnight with either 5 μg of poly(A)$^+$ RNA from staged animals or 10 μg of total RNA in the case of neuroblast and myoblast experiments. The hybrids were then treated with 400 U of S1 (Pharmacia or Boehringer Mannheim) at 42°C for 2.5 h. The digestion products were electrophoresed through 6% polyacrylamide 6 M urea sequencing gels. ^{32}P end-labeled *Msp*I digests of pBR322 served as mol. wt markers. The fragments were transferred onto Nytran membranes by electroblotting. The RNA was lightly crosslinked to the nylon membrane by UV irradiation (1 min 30 s at a distance of 20 cm from an unfiltered UV bulb). The blots were then hybridized with labeled RNA transcripts generated by T7 transcription of linearized pGEM vectors containing various inserts. In most instances the probe was specific for detecting the RNA strand of the hybrid. In order to vary the cleavage efficiency across from a looped out region of nucleic acid, the temperature was varied. At 37°C, cleavage was minimal (<5%), while at 55°C, cutting was >80%. In most experiments described here, a temperature of 42–45°C was used, where cleavage was in the 20–50% range.

Quantitation of the S1 nuclease protected fragments

In order to determine more accurately the relative amounts of each digestion product, autoradiographs were scanned by an LKB laser densitometer (LKB Instruments), and the peaks were quantitated by the Gel scan program supplied with the instrument. To ensure that hybridization was performed under condition of DNA excess, serial 2-fold dilutions of the MP18 single strands were hybridized to a constant (5 μg) amount of poly(A)$^+$

RNA. The point at which a diminution in signal was first observable was judged to be just limiting in DNA, and four times more than this amount was used for subsequent experiments.

To determine the linear response range of the autoradiographic film, serial 2-fold dilutions of poly(A)$^+$ RNA were hybridized to a constant amount of the appropriate mp18 derivative DNA, and the samples were prepared as described above. The blots were then exposed for various lengths of time and the films scanned by densitometry.

To normalize for the differences in the actual amounts of RNA contained in samples from different stages, a Northern blot containing 5 μg of poly(A)$^+$ from each of the stages was hybridized to a probe specific for the ribosomal gene *RP49*. We assume that this transcript is expressed uniformly throughout development (O'Connell and Rosbash, 1984), and we therefore normalized each lane for the differences exhibited by the *RP49* hybridization.

Acknowledgements

We are grateful to Larry Kauvar and Nick Brown for their generous gifts of cDNA libraries. We thank Tony Mahowald and Norbert Perrimon in whose lab some of this work was carried out and Tony Mahowald for generously supplying samples of N18 RNA, and K.Kornfeld, V.Stroeher, J.Bermingham and P.O'Farrell for communication of results prior to publication. We are much indebted to Mark Peifer and Jeff Simon for critical comments on the manuscript, and for many stimulating discussions. The research was supported by a Damon Runyon–Walter Winchell postdoctoral fellowship to M.B.O. and a National Institutes of Health grant to W.W.B.

References

Akam,M.E. and Martinez-Arias,A. (1985) *EMBO J.*, **4**, 1689–1700.
Bantle,J.A., Maxell,I.H. and Hahn,W.E. (1976) *Anal. Biochem.*, **72**, 413–427.
Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721–732.
Beachy,P.A., Helfand,S.L. and Hogness,D.S. (1985) *Nature*, **313**, 545–551.
Bender,W., Spierer,P. and Hogness,D.S. (1983a) *J. Mol. Biol.*, **168**, 17–33.
Bender,W., Akam,M., Karch,F., Beachy,P.A., Peifer,M., Spierer,P., Lewis,E.B. and Hogness,D.S. (1983b) *Science*, **221**, 23–29.
Bender,W., Weiffenbach,B., Karch,F. and Peifer,M. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 173–180.
Cabrera,C.V., Botas,J. and Garcia-Bellido,A. (1985) *Nature*, **318**, 569–571.
Casanova,J., Sánchez-Herrero,E. and Morata,G. (1985) *Cell*, **42**, 663–669.
Dambly-Chaudiere,C. and Ghysen,A. (1987) *Genes Dev.*, **1**, 297–306.
Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645–1655.
Duncan,I. (1987) *Annu. Rev. Genetics*, **21**, 285–319.
Furst,A. and Mahowald,A.P. (1985) *Dev. Biol.*, **109**, 184–192.
Ghysen,A., Jan,L.Y. and Jan,Y.N. (1985) *Cell*, **40**, 943–48.
Hayes,P.H., Sato,S. and Denell,R.E. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 545–549.
Haynes,S.R., Rebbert,M.L., Mozer,B.A., Forquignon,F. and Dawid,I.B. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1819–1823.
Hogness,D.S., Lipshitz,H.D., Beachy,P.A., Peattie,D.A., Saint,R.B., Goldschmidt-Clermont,M., Harte,P.J., Gaviz,E.R. and Helfand,S.L. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 181–194.
Hooper,J.E. (1986) *EMBO J.*, **5**, 2321–2329.
Kaufman,T.C., Lewis,R. and Wakimoto,B. (1980) *Genetics*, **94**, 115–133.
Kerridge,S. and Dura,J.M. (1982) *Dev. Genet.*, **3**, 207–214.
Krumlauf,R., Holland,P.W.H., McVey,J.H. and Hogan B.L.M. (1987) *Development*, **99**, 603–617.
Laughon,A., Boulet,A.M., Bermingham,J.R., Laymon,R.A. and Scott,M.P. (1986) *Mol. Cell. Biol.*, **6**, 4676–4689.
Lewis,E.B. (1978) *Nature*, **276**, 565–570.
Lewis,E.B. (1981) In Brown,D. and Fox,F. (eds), *Developmental Biology Using Purified Genes. ICN–UCLA Symposium on Molecular and Cellular Biology*. Academic Press, New York, Vol. XXIII, pp. 189–208.
Lewis,E.B. (1982) In Burger,M. (ed.), *Embryonic Development: Genes and Cells*. Proceedings of the IX Congress of the International Society of Developmental Biology. Alan R. Liss, New York, pp. 269–289.
Lipshitz,H.D., Peattie,D.A. and Hogness,D.S. (1987) *Genes Dev.*, **1**, 307–322.
Martinez-Arias,A. and Lawrence,P. (1985) *Nature*, **313**, 639–642.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Mavilio,F., Simeone,A., Giampaolo,A., Faiella,A., Zappavigna,V., Acampora,D., Poiana,G., Russo,G., Peschle,C. and Boncinelli,E. (1986) *Nature*, **324**, 664–670.
Nüsslein-Volhard,C. (1979) *Symp. Soc. Dev. Biol.*, **37**, 185–211.
Nüsslein-Volhard,C. and Wieschaus,E. (1980) *Nature*, **287**, 795–801.
O'Connell,P. and Rosbash,M. (1984) *Nucleic Acids Res.*, **12**, 5495–5513.
Peifer,P. and Bender,W. (1986) *EMBO J.*, **5**, 2293–2303.
Poole,S.J., Kauvar,L., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37–43.
Saari,G. and Bienz,M. (1987) *EMBO J.*, **6**, 1775–1780.
Sanchez-Herrero,E., Vernos,I., Marco,R. and Morata,G. (1985) *Nature*, **313**, 108–113.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schneuwly,S., Kuroiwa,A., Baumgartner,P. and Gehring,W.J. (1986) *EMBO J.*, **5**, 733–739.
Stroeher,V.L., Jorgensen,E.M. and Garber,R.L. (1986) *Mol. Cell. Biol.*, **6**, 4667–4675.
Struhl,G. and Akam,M. (1985) *EMBO J.*, **4**, 3259–3264.
Tiong,S., Bone,L.M. and Whittle,J.R.S. (1985) *Mol. Gen. Genet.*, **200**, 335–342.
Weinzierl,R., Axton,J., Ghysen,A. and Akam,M. (1987) *Genes Dev.*, **1**, 386–397.
White,R.A.H. and Akam,M.E. (1985) *Nature*, **318**, 567–569.
White,R.A.H. and Wilcox,M. (1984) *Cell*, **39**, 163–171.
White,R.A.H. and Wilcox,M. (1985a) *EMBO J.*, **4**, 2035–2043.
White,R.A.H. and Wilcox,M. (1985b) *Nature*, **318**, 563–567.
Wilde,C.D. and Akam,M. (1987) *EMBO J.*, **6**, 1393–1401.
Wold,B.J., Klein,W.H., Hough-Evans,B.R., Britten,R.J. and Davidson,E.H. (1978) *Cell*, **14**, 941–950.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.

Received on November 10, 1987