The K box, a conserved 3′ UTR sequence motif, negatively regulates accumulation of *Enhancer of split* Complex transcripts

Eric C. Lai, Christian Burks* and James W. Posakony‡

Department of Biology and Center for Molecular Genetics, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0349, USA

*Present address: Exelixis Pharmaceuticals, 260 Littlefield Avenue, South San Francisco, CA 94080, USA

‡Author for correspondence (e-mail: jposakony@ucsd.edu)

SUMMARY

Cell-cell interactions mediated by the Notch receptor play an essential role in the development of the *Drosophila* adult peripheral nervous system (PNS). Transcriptional activation of multiple genes of the *Enhancer of split* Complex [E(spl)-C] is a key intracellular response to Notch receptor activity. Here we report that most E(spl)-C genes contain a novel sequence motif, the K box (TGTGAT), in their 3′ untranslated regions (3′ UTRs). We present three lines of evidence that demonstrate the importance of this element in the post-transcriptional regulation of E(spl)-C genes. First, K box sequences are specifically conserved in the orthologs of two structurally distinct E(spl)-C genes (*m4* and *m8*) from a distantly related *Drosophila* species. Second, the wild-type *m8* 3′ UTR strongly reduces accumulation of heterologous transcripts in vivo, an activity that requires its K box sequences. Finally, *m8* genomic DNA transgenes lacking these motifs cause mild gain-of-function PNS defects and can partially phenocopy the genetic interaction of *E(spl)^D* with *Notch^spl*. Although E(spl)-C genes are expressed in temporally and spatially specific patterns, we find that K box-mediated regulation is ubiquitous, implying that other targets of this activity may exist. In support of this, we present sequence analyses that implicate genes of the *iroquois* Complex (Iro-C) and *engrailed* as additional targets of K box-mediated regulation.

Key words: Neurogenesis, Sensory organ development, 3′ untranslated region, RNA stability, Post-transcriptional regulation, CAAC motif

INTRODUCTION

The fruit fly *Drosophila melanogaster* is covered with a regular and stereotyped array of external mechanosensory organs, which constitute the majority of its peripheral nervous system (PNS). The determination and differentiation of these sensory organs takes place during late larval and early pupal development (Hartenstein and Posakony, 1989; Huang et al., 1991; Usui and Kimura, 1993), in a process that has been well characterized both genetically and molecularly. Briefly, the spatially patterned activity of the basic helix-loop-helix (bHLH) transcriptional activators achaete (ac) and scute (sc) define 'proneural clusters,' or groups of cells that are competent to adopt a sensory organ precursor (SOP) fate (Cabrera and Alonso, 1991; Cubas et al., 1991; Skeath and Carroll, 1991; Van Doren et al., 1992). Then, inhibitory cell-cell interactions requiring the function of the neurogenic genes restrict the SOP fate to a single cell in each proneural cluster (Dietrich and Campos-Ortega, 1984; Hartenstein and Posakony, 1990; Parks and Muskavitch, 1993; Schweisguth and Posakony, 1992; Tata and Hartley, 1995). This 'lateral inhibition' process is mediated largely by signaling through the Notch receptor. Notch signaling is further required for the cell

fate asymmetry of each of the three cell divisions of the SOP lineage, which yield the four distinct component cells of a mechanosensory organ (Hartenstein and Posakony, 1990; Posakony, 1994).

Recently (Lai and Posakony, 1998; Leviten et al., 1997), we described the Brd box and the GY box, sequence motifs located in the 3′ UTRs of many genes functioning or implicated in Notch signaling, including *Bearded* (*Brd*) (Leviten and Posakony, 1996) and multiple genes of the *Enhancer of split* complex [E(spl)-C]. The E(spl)-C encodes seven bHLH repressor proteins, which function as effectors of the lateral inhibitory signal (Klämbt et al., 1989; Schrons et al., 1992). This complex also includes a gene, *m4*, the protein product of which is novel but related to the protein encoded by *Brd* (Leviten et al., 1997). Another member of the E(spl)-C is *groucho* (*gro*), which encodes a ubiquitously expressed nuclear protein that functions as a transcriptional co-repressor for bHLH and other DNA-binding repressor proteins (Fisher and Caudy, 1998). Several E(spl)-C genes, including *m4*, are known to be directly activated in proneural clusters by the transcription factor Suppressor of Hairless in response to Notch receptor activity (Bailey and Posakony, 1995; Furukawa et al., 1995; Lecourtois and Schweisguth, 1995). *Brd* is likewise

expressed specifically in proneural clusters, under ac/sc control (Bailey and Posakony, 1995; Leviten et al., 1997; Singson et al., 1994). Moreover, gain-of-function alleles of *Brd* confer a bristle 'tufting' phenotype indistinguishable from that produced by loss-of-function mutations in genes of the Notch pathway (Leviten and Posakony, 1996), suggesting that *Brd* has a normal role in lateral inhibition. The Brd box motif shared by these genes negatively regulates protein and transcript levels, and mutation of Brd and GY boxes in a *Brd* transgene renders it hyperactive and capable of interfering with normal PNS development (Lai and Posakony, 1997). Thus, the Brd box appears to have an important function in controlling the activities of two families of genes involved in Notch signaling.

Here we report the identification of a novel hexamer motif, the K box (TGTGAT), that is present in the 3′ UTRs of seven of the nine genes of the E(spl)-C. We also describe an additional sequence, the CAAC motif, which is closely associated with Brd, GY and K boxes in these 3′ UTRs. Consistent with previous observations suggesting that the *E(spl)m8* 3′ UTR is associated with negative regulatory activity (Kramatschek and Campos-Ortega, 1994), we find using a heterologous reporter gene assay that the *m8* 3′ UTR mediates a spatially and temporally general form of negative regulation. This activity is largely mediated by K boxes, which act principally at the level of transcript accumulation. Using genomic *E(spl)m8* transgenes, we show that loss of K box-mediated regulation interferes with normal PNS development. In particular, *m8* transgenes lacking these motifs confer characteristic gain-of-function phenotypes and also exhibit genetic interactions with *Hairless* (*H*) and the *split* allele of *Notch*. Finally, we present sequence analyses that strongly suggest that *engrailed* and genes of the *iroquois* Complex (Iro-C) are additional targets of K box-mediated regulation.

MATERIALS AND METHODS

Drosophila stocks

w^1118^; P[w^+^, arm-lacZ-SV40 t] transgenic fly lines are described in Lai and Posakony (1997). *E(spl)^D^* is described in Lindsley and Zimm (1992), as are *H^E31^/TM6B* and *y w N^spl^*, utilized herein for genetic interaction studies involving *E(spl)m8* genomic DNA transgenes.

Molecular biology

General molecular biology techniques were performed as described (Ausubel et al., 1987). Preparation of RNA and protein extracts, and probing and quantitation of northern and western blots, were performed as described in Lai and Posakony (1997). The probe for detection of *E(spl)m8* transcripts (Fig. 8) was a 1.05 kb *Bgl*II fragment that includes nearly all of the *m8* transcription unit.

Plasmid construction

Mutant *E(spl)m8* 3′ UTRs

A 0.95 kb *Eco*RV-*Cla*I subclone containing the C-terminal region of *E(spl)m8*, its entire 3′ UTR and approximately 300 bp of downstream genomic DNA sequence was used as a template for site-directed mutagenesis. Primers (sequences available upon request) were used to generate the following mutant *E(spl)m8* 3′ UTRs, with novel diagnostic restriction sites introduced as indicated: *m8 K/CAAC del* (deletes all nucleotides between and including CAAC #1 and CAAC #5; nt 695-736 of the transcription unit); *m8 K1K2 mut* (changes K #1 from tc**TGTGAT**ag to tc**CTCGAG**ag [*Xho*I] and K #2 from ac**TGTGAT**gg to ac**CCCGGG**gg [*Sma*I]; and *m8 5x CAAC mut*

(changes CAAC #1-3 from ac**CAACCAACAAC**gc to ac**CACTTAAGTAC**gc [*Afl*II], CAAC #4 from cc**CAAC**tg to cc**GTAC**tg, and CAAC #5 from tgg**CAAC**aa to tgg**CCAG**aa [*Msc*I]). The specificity of all mutations was confirmed by sequencing.

arm-lacZ-m8 3′ UTR constructs

E(spl)m8 3′ UTRs were excised from the wild-type and three mutant constructs as *Nar*I-*Bgl*II fragments, which contain the last 9 nt of the *m8* coding region, the entire 3′ UTR and approximately 150 bp downstream of the polyadenylation signal. These fragments were filled in with Klenow polymerase and cloned into the *Hinc*II-*Eco*RV sites of *pBluescript* with the proximal region closer to the *Kpn*I site in the polylinker. These 3′ UTRs were then cloned into *pCaSpeR-ac-lacZ* (Lai and Posakony, 1997) as *Kpn*I-*Eco*RI fragments. The *ac* promoter was then replaced by the 1.8 kb *arm* promoter to create *arm-lacZ-m8 wt*, *arm-lacZ-m8 K/CAAC del*, *arm-lacZ-m8 K1K2 mut* and *arm-lacZ-m8 5x CAAC mut*.

E(spl)m8 genomic DNA constructs

A 2.6 kb *Eco*RI-*Cla*I wild-type *E(spl)m8* genomic DNA fragment was subcloned from a larger genomic fragment described by Knust et al. (1987). The 0.95 kb *Eco*RV-*Cla*I mutant 3′ UTR fragments replaced the corresponding wild-type fragment in the 2.6 kb subclone to create *m8 2.6 K/CAAC* and *m8 2.6 K1K2 mut*. The wild-type and mutant *E(spl)m8* genomic DNAs were then cloned into CaSpeR-K (Margolis et al., 1995) as *Kpn*I-*Xba*I fragments to yield *m8* wild-type, *m8 K/CAAC del* and *m8 K1K2 mut* P-element-transformation constructs.

Germline transformation

The P-element-transformation constructs described above were coinjected with a Δ2-3 helper plasmid into the recipient strain *w^1118^* (Rubin and Spradling, 1982). For each construct, 7-16 independent homozygous transgenic lines were analyzed.

β-galactosidase activity staining

Imaginal discs and other organs were dissected from late third-instar larvae and stained for β-galactosidase activity as described (Romani et al., 1989). Preparations were dehydrated in an ethanol series and mounted in Epon.

DNA sequence searches

Initial searches for sequence elements among E(spl)-C 3′ UTR sequences were carried out with the CONSENSUS program [(Hertz et al., 1990); Versions 4a of CONSENSUS and 3d of WCONSENSUS: G. Hertz, personal communication], which systematically – using a matrix consensus representation and information content analysis – scans a set of sequences for oligomers common to them.

RESULTS

Two new classes of 3′ UTR sequence motifs in genes of the E(spl)-C and *Brd*

Recently, we reported the identification of two novel classes of sequence motifs in the 3′ UTRs of *Brd* and most genes of the E(spl)-C, the Brd box (**AGCTTTA**) and the GY box (**GTCTTCC**) (Lai and Posakony, 1998; Leviten et al., 1997). Our investigation of the in vivo function of these elements has also been reported recently (Lai and Posakony, 1997). Here, we describe two additional classes of sequence elements in the 3′ UTRs of these genes, the K box and the CAAC motif.

Like the Brd box and the GY box, the K box is an exact sequence identity present in one or more copies in the 3′ UTRs of *Brd* and several genes of the E(spl)-C (Fig. 1A). A core

hexamer sequence (**TGTGAT**; rich in <u>k</u>eto nucleotides) defines the K box element, and there are additional strong biases in the 5′ and 3′ flanking nucleotides. While most of the genes listed in Fig. 1A contain single K boxes in their 3′ UTRs, two genes [*E(spl)m8* and *E(spl)mδ*] each contain a pair of closely spaced K boxes; this arrangement may be of particular consequence. With a single exception [in the coding region of *E(spl)m5*], all K boxes on the sense strand of these genes reside in their 3′ UTRs; moreover, within the 3′ UTRs, the K box sequence occurs only in the sense orientation. Thus, there is significant specificity to both the location and the orientation of the K box.

The K box sequence element appears not to have any consistent positional relationship with either Brd boxes or GY boxes, either in the linear sequence (Fig. 2) or in predicted secondary structures of these transcripts (data not shown). However, a fourth sequence element, the CAAC motif (**CAAC**) does link a subset of these other motifs. As shown in Fig. 1B, the tetranucleotide CAAC is tightly associated with one third of Brd, GY and K box elements. Moreover, the second Brd box in *E(spl)m4*, the third Brd box in *E(spl)m5*, and the first K box of *E(spl)m8* are all associated with multiple CAAC motifs. CAAC motifs are specifically concentrated upstream of Brd, GY and K boxes (Fig. 1B) and could conceivably influence their activities. Though there does not appear to be a strict spacing requirement between CAAC motifs and the Brd, GY or K boxes, CAAC sequences are found immediately upstream in more than half of the CAAC-Brd/GY/K pairs, with adjacent CAAC-K box pairs most prevalent (Fig. 1B). Overall, although this short motif occurs with essentially random frequency in the 3′ UTRs of E(spl)-C genes and *Brd*, its presence is highly enriched in the vicinity of Brd, GY and K boxes.

Evolutionary conservation of K boxes and CAAC motifs

The strict identity of the K box sequences in *Brd* and the genes of the E(spl)-C suggests that each nucleotide of this motif is important for its function. If this is the case, one might expect its precise sequence to be conserved over the course of *Drosophila* evolution. *D. hydei* is approximately 60 million years diverged from *D. melanogaster* (Beverley and Wilson, 1984), sufficiently distant for substantial sequence randomization to occur outside of coding and regulatory regions (Caccone and Powell, 1990). We analyzed the

sequences of the 3′ UTRs of two genes from *D. hydei*, *E(spl)m4* (a non-bHLH gene) and *E(spl)m8* (a bHLH gene), and found that K boxes and their flanking nucleotides are fully conserved in all three instances [Figs 1 and 3; see Lai and Posakony (1997)]. In particular, the most conserved region between the *m8* 3′ UTRs includes the pair of K boxes (Fig. 3), suggesting that this region is indeed of functional significance. Otherwise, these 3′ UTR sequences are quite divergent.

Although neither the exact location nor the number of CAAC motifs is completely conserved, this sequence is found upstream of *m4* B2, *m8* K1 and *m8* K2 in both species, while a single GY box in the 3′ UTR of each *m4* ortholog is also associated with a CAAC motif (Fig. 1). It is noteworthy that all eight CAAC motifs in the approximately 600 bp of combined *m8* 3′ UTR sequence are clustered around the K boxes (Fig. 3), further suggesting that the proximity of these motifs is not coincidental.

The K box mediates negative post-transcriptional regulation

To investigate the in vivo function of the K box, we made use of the *E(spl)m8* 3′ UTR, which includes two evolutionarily

A

K Box

Brd K1	TC<u>CAAC</u>**TGTGAT**CGAGGA
E(spl)m3 K1	TT<u>CAAC</u>**TGTGAT**ATAACT
E(spl)m4 K1	AAAAAT**TGTGAT**ACATTT
Dh m4 K1	CAAAAT**TGTGAT**AAATTT
E(spl)m5 K1	CTCCCG**TGTGAT**GGCCTG
E(spl)m7 K1	TAAGTC**TGTGAT**AGGGCA
E(spl)m8 K1	CGCATC**TGTGAT**AGCCCA
Dh m8 K1	CG<u>CAAC</u>**TGTGAT**AACCAA
E(spl)m8 K2	CC<u>CAAC</u>**TGTGAT**GGCAAC
Dh m8 K2	AC<u>CAAC</u>**TGTGAT**GGGAAC
E(spl)mγ K1	AGTTCA**TGTGAT**AACTAA
E(spl)mδ K1	TATAGA**TGTGAT**GTTCCT
E(spl)mδ K2	TGTTCC**TGTGAT**AGTTCT

A	343972	834335
C	537138	115452
G	040111	472321
T	523222	022335
Consensus	caac**TGTGAT**ag	

B

CAAC Motif

Brd B2	<u>CAAC</u>**AGCTTTA**ACTCT
Brd K1	<u>CAAC</u>**TGTGAT**CGAGGA
E(spl)m3 G1	<u>CAAC</u>AAGATCCGTT**GTCTTCC**AAAGT
E(spl)m3 K1	<u>CAAC</u>**TGTGAT**ATAACT
E(spl)m4 B1	TCTCC**AGCTTTA**AT<u>CAAC</u>
E(spl)m4 B2	<u>CAAC</u>T<u>CAAC</u>TGAAG**AGCTTTA**CAAAA
Dh m4 B2	<u>CAAC</u>ACTCAATCG**AGCTTTA**CAAAA
E(spl)m4 G2	<u>CAAC</u>**GTCTTCC**GTTAA
Dh m4 G1	<u>CAAC</u>AAT**GTCTTCC**ATTTT
E(spl)m5 B3	<u>CAAC</u>**AGCTTTA**ATAGT
E(spl)m5 G2	ATCCT**GTCTTCC**AAAAG<u>CAAC</u>
E(spl)m5 G3	<u>CAAC</u>GAATCATG**TCTTCC**AAAAT
E(spl)m8 K1	<u>CAAC</u><u>CAAC</u><u>CAAC</u>GCATC**TGTGAT**AGCC<u>CAAC</u>
Dh m8 K1	<u>CAAC</u>GC<u>AAC</u>**TGTGAT**AAC<u>CAAC</u>
E(spl)m8 K2	<u>CAAC</u>**TGTGAT**GG<u>CAAC</u>
Dh m8 K2	<u>CAAC</u>**TGTGAT**GGGAAC

Fig. 1. Two novel classes of sequence motifs in the 3′ UTRs of E(spl)-C genes and *Brd*. Genes labeled 'Dh' are the *Drosophila hydei* orthologs of the indicated loci. (A) Alignment of K box sequences found in the 3′ UTRs of E(spl)-C genes. A 6 nt sequence identity (bold) defines this motif, but strong biases are found in the flanking nucleotides as well. Underlined are instances in which CAAC motifs occur immediately 5′ to K boxes. *Brd* contains an atypically located K box that overlaps the C-terminal codon and stop codon (TGA); it also includes an adjacent CAAC motif. (B) Frequent occurrence of CAAC motifs in the immediate vicinity (≤ 10 bp) of Brd, GY and K boxes. The sequence CAAC (underlined) is found upstream of about one-third of the Brd, GY and K boxes in the E(spl)-C and *Brd*, and is found directly adjacent to the box sequence in half of these instances. Note that several box elements are associated with multiple CAAC motifs [*E(spl)m4* B2, *E(spl)m5* B3 and *E(spl)m8* K1]. Of the four *D. melanogaster* cases shown in which CAAC occurs downstream of a box sequence, two [*E(spl)m4* B1 and *E(spl)m8* K1] correspond to a CAAC motif located immediately 5′ to a nearby box [*E(spl)m4* G2 and *E(spl)m8* K2, respectively]; this feature is conserved in *D. hydei E(spl)m8*.
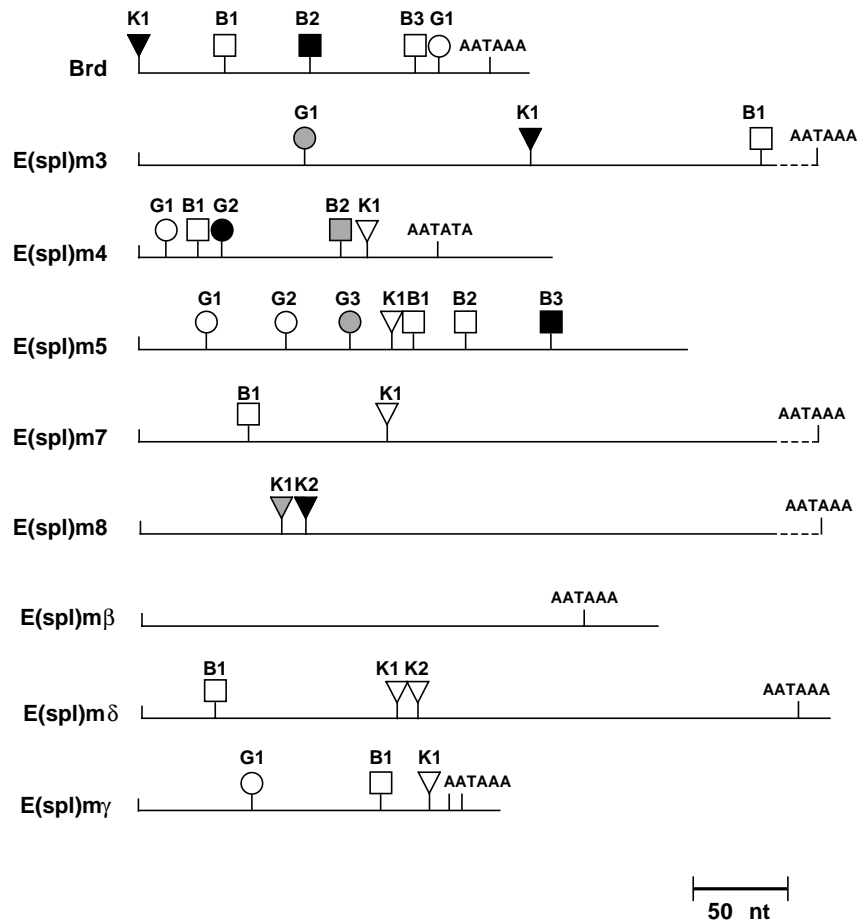
Fig. 2. Graphic map of 3′ UTR sequence motifs in E(spl)-C genes and *Brd*. Squares denote Brd boxes, circles GY boxes and triangles K boxes. Boxes with a CAAC motif immediately upstream are filled black; those associated with a more distant upstream CAAC motif are shaded light grey. Brd, GY and K boxes appear not to have any consistent positional relationship; however, there is a close pairing of K boxes in 3′ UTRs which contain two copies of this motif [8 bp spacing in *E(spl)m8* and 5 bp spacing in *E(spl)mδ*]. Note that *E(spl)m8* contains only K boxes, and that *E(spl)mβ* lacks Brd, GY and K boxes.

conserved copies of this motif and lacks both Brd boxes and GY boxes. An *arm-lacZ-SV40 t* reporter transgene, utilizing 1.8 kb of the *armadillo* promoter and the 3′ UTR of the SV40 t antigen gene, is expressed at high levels in most cells throughout development (Lai and Posakony, 1997; Vincent et al., 1994). We compared the reporter activity of this transgene with that of a version that includes the *m8* 3′ UTR (Fig. 4) and found that the wild-type *m8* (*m8 wt*) 3′ UTR sharply reduces reporter gene activity relative to the *SV40 t* 3′ UTR (cf. Fig. 5A,F,K and Fig. 5B,G,L). Although *E(spl)m8*, as well as other genes of the E(spl)-C and *Brd*, are expressed in a spatially restricted pattern in both embryos and larvae, we observed a drastic and largely uniform reduction in reporter activity throughout larval and imaginal tissues of the third-instar larva as well as throughout embryonic development (Fig. 5 and data not shown). Thus, the *E(spl)m8* 3′ UTR appears to confer a ubiquitous and temporally general form of negative regulation.

To test the specific role of the K boxes in regulation by the *m8* 3′ UTR, we constructed mutant versions of the reporter transgene in which all K boxes and CAAC motifs were deleted (*m8 K/CAAC del*), or in which 4 or 5 point mutations were introduced into both K boxes (*m8 K1K2 mut*) (Fig. 4). As shown in Fig. 5 (C,H,M), deletion of the K/CAAC region restores moderate levels of reporter activity throughout all tissues. Interestingly, all transgenic lines carrying the *m8 K1K2 mut* construct exhibited significantly higher levels of reporter activity than those of the *m8 K/CAAC del* lines (Fig. 5D,I,N). Very comparable effects were observed with both of these K

box-mutant transgenes during embryonic development, indicating that a spatially and temporally general mode of negative regulation by the *m8* 3′ UTR requires the integrity of the K boxes (Fig. 5 and data not shown).

To investigate the possible role of the CAAC motifs, we constructed a mutant reporter gene containing 2 bp mutations in each of the five CAAC motifs surrounding the K boxes in the *m8* 3′ UTR (*m8 5x CAAC mut*). This mutant 3′ UTR retained strong negative regulatory activity, presumably due at least in part to its intact K boxes. Nevertheless, specific mutation of the CAAC motifs led to a detectable and reproducible increase in reporter activity relative to the *m8 wt* 3′ UTR (Fig. 5E,J,O). Therefore, CAAC motifs are required for the full regulatory activity of the *m8* 3′ UTR.

K boxes function primarily at the level of transcript accumulation

As our reporter gene results indicated that K boxes are essential for normal negative regulation by the *E(spl)m8* 3′ UTR, we were interested to determine at what level such regulation occurs. To address this, we performed northern and western blot analysis of RNA and protein accumulation for sets of five representative *arm-lacZ* transgenic lines for each 3′ UTR tested. Samples were prepared from 2-26 hour embryos, and reporter levels were quantitated relative to the ubiquitously expressed endogenous genes *rp49* (RNA) and *groucho* (protein).

We found that negative regulation mediated by K boxes is

D. melanogaster E(spl)m8   **K1**   **K2**

```
TAAAAACAAGGGGTTAAGTGGCAGGAAGGTAAGGG--TGGAGTGGCAATGGGAATGGAAGAACCAACCAACAACGCATCTGTGATAGCCCAACTGTGATGGCAACAAAAT----
|||||                |||||        |||  ||||       |||    |||||||||| ||||||||| |||||||||||||| ||||||
TAAAATGTCAGCAATA--------GCCGGTAAACACCAAGAG--GCAAAC-----AAAAG---------ACAACGCAACTGTGATAA-CCAACTGTGATGGGAACAAAACAATT
```
D. hydei E(spl)m8

D. m.
```
-----AGCAAAATTAAA-AAA-GCAATTAGCAAAACAAATATAAACCAATTTAAA------CAAATATC-----------AAAAC----AATTTAC-TTT----ATTGGGA
     |||||   |||| ||| ||||||||  |||| ||||| |||||        |||||       |||| ||||| ||| ||||||
AGTAAAGCAAT--TAAATAAATGCAATTAAG-AAAC----ATAAAATTATTTGACTTATGATTAATATGGGCTGGGCAAGGAAAAGGGTGACTTTACATTTCAGCATTGGGC
```
D. h.

D. m.
```
TGTTGGGAGGGGATCTAAGGGACCAATCCCTTAGTTGGTGGCAATTGGAATCCTTATTGGTAATTAAATTATTACCAATTAGGAAATATCAAATAATATATCTCTGCTCTGA
```


D. m.
```
CAATGGAAATCTATTTTGGCAATGACCGAAGTGAAATCTATTTTAGTTAATTAAAAATAATAAATTAAATATATTTCACTTAATTAAAATAAAAGTGTTTTAGAAAACTATC
```

Fig. 3. Alignment of the 3′ UTRs of *E(spl)m8* orthologs from *D. melanogaster* and *D. hydei*. The stop codons of both genes (TAA) are underlined. The 3′ UTR sequence of the *D. hydei* gene (Maier et al., 1993) may be incomplete, although a consensus polyadenylation signal (AATAAA) is underlined. The precise site of polyadenylation in the *D. melanogaster* transcript is likewise unknown, although usage of either of the underlined consensus signals would be consistent with the sizes of transcripts determined by northern blot analysis (Klämbt et al., 1989). Regions containing three or more consecutive nucleotide identities are indicated with vertical lines. The most highly conserved region between these homologs includes the pair of K boxes (**K1**, **K2**) and several of their associated CAAC motifs (shaded). In contrast, the CAAC sequence downstream of *D. melanogaster E(spl)m8* K2, a rare example of an isolated downstream CAAC motif in this gene set, is not conserved between *D. melanogaster* and *D. hydei*. These observations are consistent with the hypothesis that only 'upstream' CAACs are functionally constrained. This sequence alignment indicates that the two ATTTA motifs (underlined) noted previously in *D. melanogaster m8* (Kramatschek and Campos-Ortega, 1994) have not been conserved at the corresponding positions in *D. hydei m8*.

largely effected through decreased transcript levels. As shown in Fig. 6 and Table 1, mutation of the K boxes (*m8 K1K2 mut*) results in a large increase (approximately 4-fold) in the steady-state level of total *lacZ* transcripts, as well as a somewhat larger increase (6.5-fold) in the polyadenylated population. Quantitative western blot analysis using ^{125}I-Protein A showed that mutation of the K boxes has an additional effect on the steady-state protein level, leading to a nearly 11-fold difference between the β-galactosidase protein levels from wild-type and K box-mutant reporter gene constructs. Although we cannot rule out a potential effect of the K box on transcription, these data suggest that K boxes act primarily at the level of transcript stability and may secondarily reduce translational efficiency (approximately 2-fold) as well.

We also detected a distinct contribution of the CAAC motifs to regulation by the *E(spl)m8* 3′ UTR. In particular, specific mutation of the CAAC sequences (*m8 5x CAAC mut*) was found to result in a >2-fold increase in steady-state reporter

transcript level and a comparable increase in reporter protein level. These data suggest that the CAAC motifs also participate in negative regulation by the *m8* 3′ UTR, probably at the level of RNA stability. Despite this result, we observed that reporter transcript and protein levels from the *m8 K/CAAC del* construct were intermediate relative to those of the wild-type construct and the K box mutant, in agreement with the β-galactosidase activity staining data (cf. Figs 5 and 6). Although this deletion construct removes essentially only K boxes and CAAC motifs, nonspecific changes in RNA secondary structure or the

Table 1. Quantitation of RNA and protein accumulation from *arm-lacZ-m8 3′ UTR* transgenes

	K/CAAC del	K1K2 mut	5×CAAC mut
	m8 wt	m8 wt	m8 wt
2-26 hour embryo total RNA	2.39±0.55	4.05±1.83	1.91±0.70
2-26 hour embryo poly(A)+ RNA	2.15±0.71	6.58±1.86	2.45±0.68
2-26 hour embryo total protein	7.56±1.76	10.76±2.96	2.19±0.27

Accumulation of RNA and protein products from *arm-lacZ-m8 3′ UTR* reporter transgenes was quantitated relative to the endogenous controls *rp49* (RNA) and Groucho (protein) using a Molecular Dynamics phosphorimager (see Fig. 6). Values represent the ratio of the normalized (control-corrected) signal for each mutant construct to that of the fully wild-type construct. Three independent data sets were generated; means and standard errors are shown.

Fig. 4. Graphic depiction of the *E(spl)m8* 3′ UTRs used in this study. The positions of the stop codon (TAA), K boxes (K), CAAC motifs (c) and polyadenylation signal (AATAAA) are marked. *E(spl)m8* 3′ UTRs studied are wild type (*m8 wt*), a mutant bearing a deletion of the K box/CAAC motif region (*m8 K/CAAC del*), a mutant bearing a cluster of 4 or 5 point mutations in the two K boxes (*m8 K1K2 mut*), and a mutant bearing 2 point mutations in each of the five CAAC motifs (*m8 5x CAAC mut*).

Fig. 5. Negative regulation mediated by K boxes is spatially general. β-galactosidase activity staining of late third-instar (A-E) wing imaginal discs, (F-J) eye-antenna discs and (K-O) salivary glands. (A,F,K) An *arm-lacZ-SV40 t* transgene is expressed at high levels throughout all imaginal discs and larval tissues. (B,G,L) Reporter activity from the *arm-lacZ-m8 wt* construct is strongly repressed in all cells. (C,H,M) Deletion of the K and CAAC boxes (*arm-lacZ-m8 K/CAAC del* transgene) results in a general increase in reporter activity throughout these tissues, relative to *arm-lacZ-m8 wt*. (D,I,N) Specific mutation of the two K boxes (*arm-lacZ-m8 K1K2 mut* transgene) results in even higher levels of reporter activity throughout these tissues. (E,J,O) Specific mutation of the 5 CAAC motifs (*arm-lacZ-m8 5x CAAC mut* transgene) results in a very modest increase in reporter activity levels relative to the *m8 wt* construct. These stained preparations are representative of results for 7-10 transgenic lines for each construct; stainings were carried out for identical times (35 minutes).
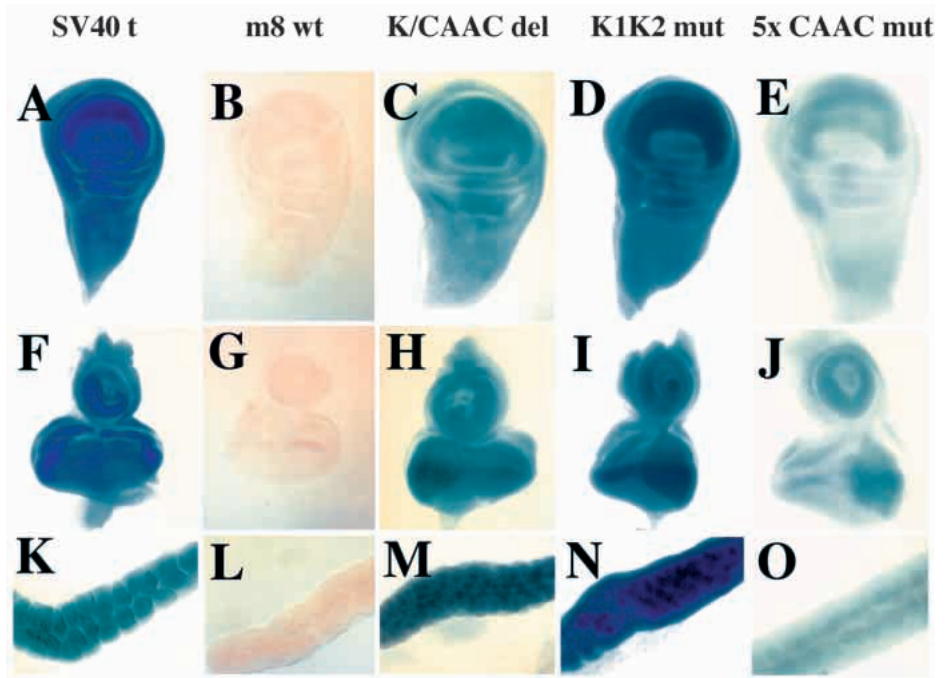
shortening of the *m8* 3′ UTR might indirectly affect its metabolism. Thus, the activity of the *K/CAAC del* 3′ UTR may not be directly predictable from the results with the point mutant constructs. Nevertheless, this deletion does cause a strong increase in steady-state transcript and protein levels (Fig. 6; Table 1), consistent with the loss of K box and CAAC motif regulation from this construct.

Loss of K box-mediated regulation interferes with normal adult PNS development

A previous study has shown that the hyperactivity of the original *E(spl)^D* allele of *E(spl)m8* is due to a small deletion at the 3′ end of the gene, including C-terminal codons and the 3′ UTR (Tietze et al., 1992). Since this deletion removes both of the K box elements that we have identified, we were interested to investigate whether loss of K box-mediated regulation is sufficient to allow *m8* transgenes to interefere with PNS development. While two copies of a wild-type *m8* transgene are not sufficient to perturb sensory organ development (Tietze et al., 1992), overexpression of *E(spl)m8* causes pronounced phenotypic defects in the PNS (Nakao and Campos-Ortega, 1996; Tata and Hartley, 1995).

We subcloned a 2.6 kb *E(spl)m8* genomic DNA fragment containing 1.17 kb of upstream promoter sequence as our wild-type transgene. This fragment contains all known transcriptional activation sequences for this gene (Bailey and Posakony, 1995; Kramatschek and Campos-Ortega, 1994; Singson et al., 1994), but is approximately 100 bp shorter than the transgene described by Tietze et al. (1992). We also constructed mutant *m8* transgenes containing the mutations in the two K boxes (*m8 K1K2 mut*) or a deletion of the *K/CAAC* region (*m8 K/CAAC del*) as illustrated in Fig. 4.

As expected, none of 12 homozygous wild-type transgenes conferred any detectable mutant phenotype in the otherwise wild-type background of the recipient strain. In contrast, 4/14 lines homozygous for the *m8 K1K2 mut* transgene exhibited

highly (>80%) penetrant PNS and/or wing venation defects, while an additional 3 lines exhibited low-penetrance (>20%) defects (Fig. 7A-D). Collectively, these defects included loss
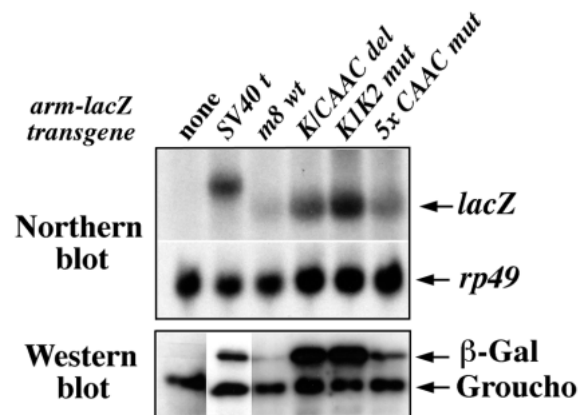
Fig. 6. K boxes negatively regulate steady-state levels of both transcript and protein (see Table 1). RNA and protein extracts of 2-26 hour embryos were prepared from representative sets of 5 transgenic lines for each *arm-lacZ* construct and analyzed by northern and western blot. (Top) Total RNA probed for *lacZ* (reporter transgene) and *rp49* (endogenous control) transcripts. The transformation recipient strain *w^1118* was used as the negative control (none). The *SV40 t* 3′ UTR is ~500 bp longer than the *E(spl)m8* 3′ UTR; thus, a larger transcript is detected in the former case. The *m8 wt* 3′ UTR strongly decreases steady-state *lacZ* transcript levels by comparison to the *SV40 t* 3′ UTR. Relative to this, *lacZ* RNA levels are elevated by the *m8 K/CAAC del* mutant, greatly elevated by the *m8 K1K2 mut* construct, and modestly elevated by the *5x CAAC mut* construct. (Bottom) Total protein probed simultaneously for β-galactosidase (reporter transgene) and Groucho (endogenous control). The same qualitative differences are observed in both the total RNA and total protein populations, but the quantitative effects of mutations that affect the K boxes are greater in the protein population (see Table 1).

of anterior orbital, postvertical and scutellar macrochaetes, loss of interocellar microchaetes, and extra/missing crossvein material. A smaller proportion of the *m8 K/CAAC del* transgenes exhibited such defects, with only one out of sixteen lines exhibiting a fully penetrant mutant phenotype and three other lines displaying low-penetrance defects. Although these effects are quite mild, they are qualitatively similar to those produced by *E(spl)m8* overexpression constructs, indicating that they are likely to be gain-of-function phenotypes (Nakao and Campos-Ortega, 1996; Tata and Hartley, 1995). Therefore, we conclude that K box-mediated regulation is necessary to restrict levels of *E(spl)m8* activity in vivo.

K box-mutant *E(spl)m8* transgenes display phenotypic interactions with *H* and *spl*

To characterize further the in vivo activity of K box-mutant transgenes, we examined their behavior in sensitized genetic backgrounds. Loss-of-function mutations in *Hairless* (*H*) increase the activity of Su(H), a transcription factor that directly links Notch receptor activity with transcriptional activation of E(spl)-C genes, including *E(spl)m8* (Bailey and Posakony, 1995; Schweisguth and Posakony, 1994). Flies heterozygous for the null allele H^{E31} display PNS defects consistent with excess Notch signaling, including missing or double-socket sensory organs (Bang et al., 1991). This is particularly evident on the head of the fly (Fig. 7E). Ten independent insertions of each *m8* construct were tested in a single copy in a $H^{E31}/+$ background. The wild-type transgenes caused only a minimal reduction (<5%) in the numbers of wild-type head macrochaetae (Fig. 7F; Table 2). In contrast, the *m8 K/CAAC del* and *m8 K1K2 mut* transgenes had pronounced effects in this genetic background, causing approximately 30% and 40% reduction in head macrochaetes, respectively

(Fig. 7G; Table 2). Enhancement of the *H* phenotype is consistent with increased *m8* activity conferred by these mutant transgenes.

We also examined the effect of *m8* transgenes on the phenotype of $N^{spl}/+$ females, whose eyes are largely wild-type. All seven wild-type *m8* transgenes failed to enhance *spl*, even when two copies of the transgene were present [cf. Fig. 7H,L
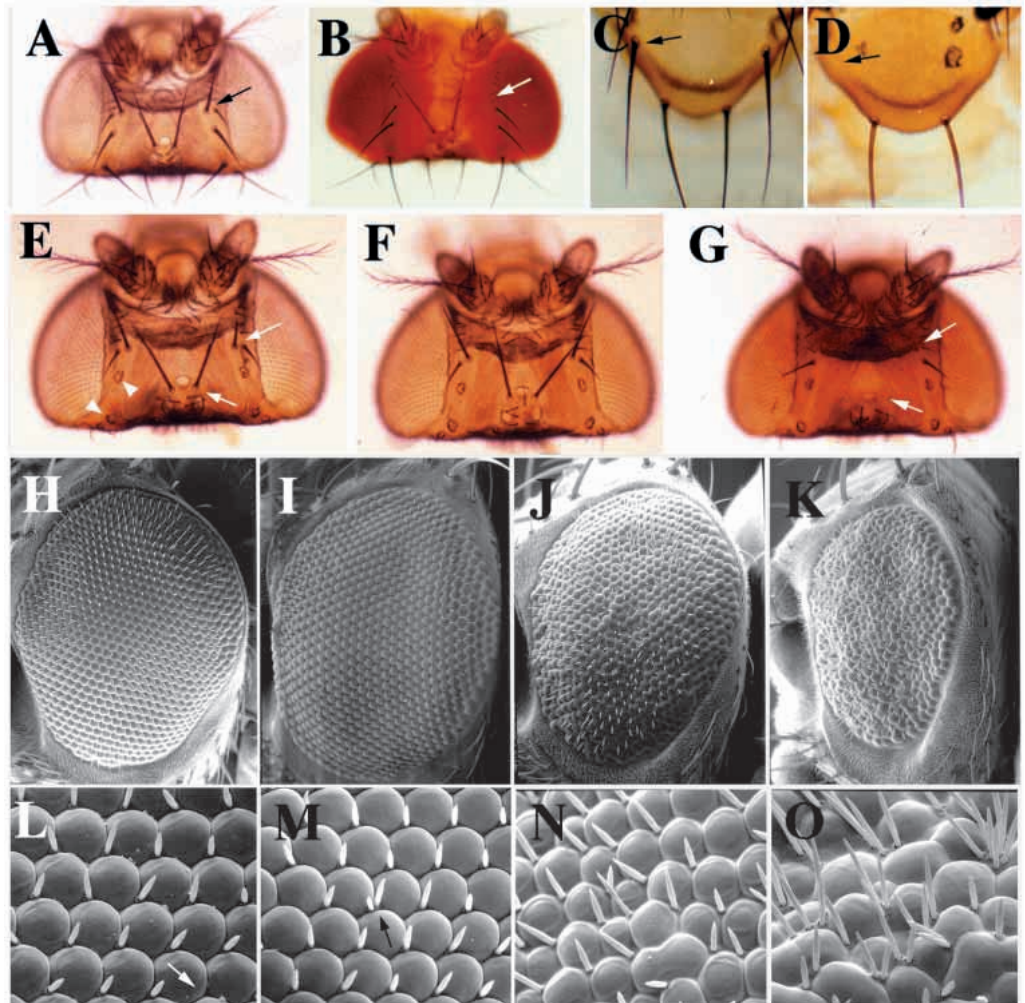
Fig. 7. *E(spl)m8* genomic DNA transgenes bearing mutations of the K boxes are hyperactive. (A,B) Heads and (C,D) scutella dissected from adult flies. (A,C) Wild-type flies and (B,D) flies homozygous for an *m8 K1K2 mut* genomic DNA transgene. Note missing orbital macrochaetae (B) and missing scutellar macrochaetae (D) (arrows). Flies bearing wild-type *m8* transgenes display no visible mutant phenotypes (not shown). (E-G) $H^{E31}/+$ heterozygous flies bearing a single copy of the following transgenes: (E) none, (F) *m8 wt*, (G) *m8 K1K2 mut*. $H^{E31}/+$ flies have several missing or double-socket sensory organs (E, arrowheads). This phenotype is enhanced by a single copy of a K box-mutant *m8* transgene (G, arrows), but not by a wild-type *m8* transgene (F). 6/10 *m8 K1K2 mut* lines were left with the only the two medial orbital bristles (G) in one-third or more of the flies examined (see Table 2). (H-K, L-O) Scanning electron micrographs of the compound eyes of $N^{spl}/+$ (*spl/+*) heterozygous females that are homozygous for the following transgenes: (H,L) none, (I,M) *m8 wt*, (J,N) *m8 K/CAAC del*, and (K,O) *m8 K1K2 mut*. Note that *spl/+* females (H,L) as well as *spl/+* females that are homozygous for an *m8 wt* transgene (I,M) have wild-type eyes with regular ommatidial patterning; missing or double interommatidial bristles are occasionally observed (arrows in L,M). In contrast, *spl/+* females homozygous for an *m8 K/CAAC del* transgene often display a slight roughening of the eyes (J), while such flies homozygous for an *m8 K1K2 mut* transgene exhibit a much stronger roughening and reduction of the eye field (K). Higher-magnification micrographs illustrate the high degree of ommatidial fusion and bristle multiplication observed with the *m8 K/CAAC del* (N) and *m8 K1K2 mut* (O) transgenes in the *spl/+* background.

Table 2. Interaction of *E(spl)m8* transgenes with *H*

Genotype	No. of wild-type head macrochaetae	No. of lines that enhance *H*
H/+	6.31±0.15	–
H/+; m8wt/+	6.00±0.17	0/10
H/+; m8K/Cdel/+	4.42±0.15	2/10
H/+; m8KKmut/+	3.86±0.16	6/10

10 independent transformant lines for each *E(spl)m8* transgene construct were tested in single copy in a *H^E31^/+* background. The number of wild-type macrochaetae on the heads of 5 males and 5 females were counted for each line tested (10 females were counted if the transgene insert is on the X chromosome); means and standard errors for the 100 flies examined for each construct are shown. For reference, wild-type flies have 14 head macrochaetae. We define enhancement of the *H/+* phenotype as the presence of only the two medial orbital bristles (see Fig. 7G) in one-third or more of the flies bearing a given transgene insert.

and 7I,M; see also Tietze et al. (1992)]. However, K box-mutant transgenes (four out of six *m8 K/CAAC del* lines and five out of seven *m8 K1K2 mut* lines) clearly enhanced *spl*, producing roughened eyes in *spl/+* females homozygous for the transgene (Fig. 7J,N and K,O). In accord with our other assays of *m8* 3′ UTR activity, the degree of enhancement was greater for the specific *m8 K1K2 mut* transgene than for the *m8 K/CAAC del*etion transgene, as the former caused a greater degree of eye roughening and loss of ommatidia than the latter.

The *E(spl)^D* mutation is associated with an elevated steady-state level of transcript

The *spl*-enhancing effects of K box-mutant *E(spl)m8* transgenes (Fig. 7) are reminiscent of those characteristic of the original *E(spl)^D* mutant. In view of our finding that mutation of K box sequences leads to elevated levels of transcript accumulation from *lacZ* reporter transgenes (Fig. 6; Table 1), we tested the hypothesis that the hyperactivity of the *E(spl)^D* allele might likewise be reflected in increased transcript accumulation. As shown in Fig. 8, we did indeed observe that, in *E(spl)^D/E(spl)^D* embryos and third-instar larvae, the truncated mutant mRNA is present at much higher levels than the normal *E(spl)m8* transcript in wild-type animals. In contrast to a previous report (Knust et al., 1987), we fail to detect any normal-sized *E(spl)m8* mRNA in *E(spl)^D* homozygotes (Fig. 8). Our results are consistent with the interpretation that deletion of K box sequence elements in the *E(spl)^D* mutation contributes to elevated steady-state transcript levels and hence to the hyperactivity of this allele.

DISCUSSION

Multiple modes of post-transcriptional regulation of E(spl)-C genes

The 3′ UTRs of eukaryotic mRNAs, once thought to be unimportant trailers following protein coding regions, often contain sequences that link RNA transcripts to a wide variety of regulatory pathways. 3′ UTRs have now been found to influence polyadenylation efficiency, transcript localization, transcript stability, and translational efficiency [reviewed by Chen and Shyu (1995); Curtis et al. (1995); St. Johnston (1995)]. Multiple regulatory functions are often localized within a single 3′ UTR; indeed, the known complexity of *cis*-

regulatory activities resident in 3′ UTRs is increasingly reminiscent of that long associated with transcriptional regulatory DNA.

We have identified multiple classes of shared sequence motifs in the 3′ UTRs of *Brd* and the genes of the E(spl)-C, including the Brd box (**AGCTTTA**), the GY box (**GTCTTCC**), and the K box (**TGTGAT**). These motifs subject these genes to multiple modes of post-transcriptional regulation [Lai and Posakony (1997); this work]. In addition, we have identified a fourth sequence, the **CAAC** motif, that is located upstream of many Brd, GY and K boxes, and may modulate their activity.

In the present study, we have shown that the integrity of the K box sequence is essential for the strong, negative post-transcriptional regulation conferred on a heterologous *lacZ* reporter gene by the *E(spl)m8* 3′ UTR, and that the CAAC motif may also participate in such regulation. Within the context of the *m8* gene itself, we found that K box elements are required to limit gene activity and that lack of such regulation has phenotypic consequences in the development of the adult PNS. Whereas wild-type *m8* transgenes did not detectably affect development and failed to show interactions with *H* and *spl* mutations, specific mutation of the two K boxes in an otherwise identical *m8* transgene caused mild gain-of-function PNS defects and yielded strong genetic interactions with both *H* and *spl*. Taken together, these data strongly support our proposal that the K box sequence mediates negative regulation in vivo, and demonstrate the developmental importance of this regulation.

The cell fate decisions controlled by Notch signaling are sensitive to small changes in gene dosage (Ashburner, 1982; Bang et al., 1991; Heitzler and Simpson, 1991; Schweisguth and Posakony, 1994), implying the importance of precise control over the expression levels of the genes involved, such as the E(spl)-C. Negative post-transcriptional regulation by K box and Brd box elements, which principally affect transcript accumulation and translational efficiency, respectively, are two means by which such fine control may be obtained [Lai and
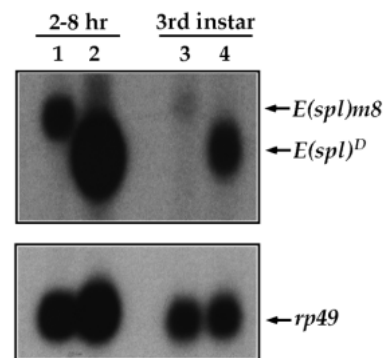

Fig. 8. Elevated steady-state levels of transcript in the *E(spl)^D* mutant. Each lane contains 5 μg of total RNA from *w^1118* (1, 3) or *E(spl)^D* homozygous (2, 4) animals; RNA was isolated from 2-8 hour embryos (1, 2) or third-instar larvae (3, 4). Northern blots were probed to detect *E(spl)m8* and *rp49* (loading control) transcripts. The *E(spl)^D* transcript, which bears a deletion of the normal *E(spl)m8* C-terminus and proximal 3′ UTR (Klämbt et al., 1989), is truncated and present at greatly elevated levels relative to the wild-type transcript during both embryonic and larval development.
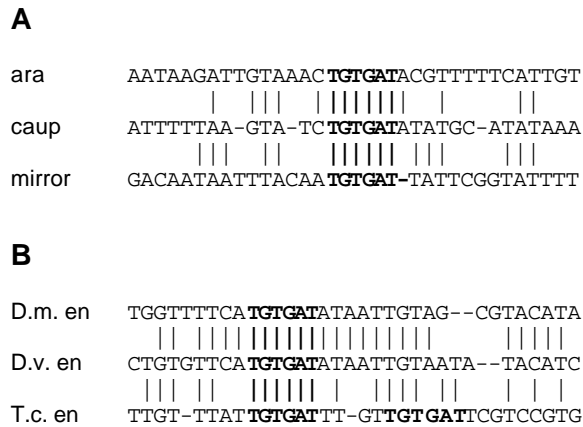
A

```
ara     AATAAGATTGTAAACTGTGATACGTTTTTCATTGT
             |  |||  ||||||||  |      ||
caup    ATTTTTAA-GTA-TCTGTGATATATGC-ATATAAA
             ||| ||   |||||| |||     |||
mirror  GACAATAATTTACAATGTGAT-TATTCGGTATTTT
```

B

```
D.m. en  TGGTTTTCATGTGATATAATTGTAG--CGTACATA
              || ||||||||||||||||||||     |||||
D.v. en  CTGTGTTCATGTGATATAATTGTAATA--TACATC
              ||| ||  ||||||  |  |||| ||   | | |
T.c. en  TTGT-TTATTGTGATTT-GTTGTGATTCGTCCGTG
```

Fig. 9. K boxes in the 3′ UTRs of Iro-C genes and *engrailed*. (A) Alignment of the K box regions from the three Iro-C genes. The K boxes in *ara* and *caup* match a more restricted 8 nt consensus sequence (cTGTGATa) derived from K boxes in E(spl)-C genes (see Fig. 1). (B) An alignment of a portion of the 3′ UTRs of *engrailed* orthologs from *D. melanogaster*, *D. virilis*, and *T. castaneum*. The K box region has clearly been conserved among these distantly related insect species. Note that *T. castaneum en* contains two closely spaced K boxes, similar to K box pairs found in *D. melanogaster E(spl)m8* and *E(spl)mδ* (see Figs 1 and 2).

Posakony (1997); this study]. As most genes of the E(spl)-C are apparently subject to negative regulation mediated by both Brd boxes and K boxes (Fig. 2), accumulation of E(spl)-C proteins should be quantitatively and temporally quite responsive to transcriptional regulation.

Alone among the genes of the E(spl)-C, *E(spl)mβ* appears to lack Brd, GY and K boxes in its 3′ UTR (Fig. 2). Its pattern of expression in wing imaginal discs is also unique: *mβ* transcripts accumulate in a complex pattern that covers much of the wing imaginal disc and is not coincident with proneural clusters (de Celis et al., 1996), the sites of co-expression of most of the other E(spl)-C genes, including *m4*, *m7*, *m8*, *mδ* and *mγ* (de Celis et al., 1996; Singson et al., 1994). Moreover, it is likely that *mβ* functions principally in Notch signaling during the development of the pupal wing, where it is expressed in intervein regions (de Celis et al., 1997). Thus, the unique expression pattern and likely functional specialization of *mβ* are correlated with its lack of 3′ UTR motifs that are otherwise widely shared within the E(spl)-C. In contrast to the other genes, it may be unnecessary or even deleterious to regulate levels of *mβ* expression via Brd boxes or K boxes.

Regulatory effects of the *E(spl)D* mutation

The E(spl)-C was originally identified via the mutant $E(spl)^D$, which displays an allele-specific interaction with N^{spl} (*spl*) (Lindsley and Zimm, 1992). In particular, $E(spl)^D$ renders the *spl* eye phenotype dominant in females heterozygous for *spl*. *spl*-enhancing activity was localized to a small deletion in the 3′ region of *E(spl)m8* that removes the C-terminal 56 amino acid residues, adds 9 novel amino acids and removes 315 nucleotides of the 3′ UTR (Klämbt et al., 1989; Tietze et al., 1992). The regulatory basis of $E(spl)^D$'s effects has been difficult to establish, partly because the behavior of this allele has been inconsistent in different phenotypic assays. For

example, enhancement of the *spl* phenotype behaves as a gain-of-function activity, while high-level overexpression of the $E(spl)^D$ protein results in dominant-negative phenotypes. In spite of these apparently contradictory effects, a consistent finding has been that the deletion itself appears to increase the activity of the gene. Thus, a construct that includes only a premature stop codon [$E(spl)^{+, stop}$] has less *spl*-enhancing activity as well as less dominant-negative activity in an overexpression assay than comparable constructs that include the deletion [$E(spl)^{D, stop}$], even though both of these constructs encode the same protein (Giebel and Campos-Ortega, 1997; Tietze et al., 1992).

We have presented several lines of evidence from multiple in vivo phenotypic assays that strongly indicate that a K box-mutant *E(spl)m8* genomic DNA transgene behaves as a hypermorph. As mutations were introduced only into untranslated regions, all E(spl)m8 proteins expressed from our constructs are expected to be completely wild type. Notably, we found that K box-mutant *m8* transgenes can largely phenocopy the *spl*-enhancing activity of $E(spl)^D$. This verifies that *spl* enhancement is a gain-of-function activity and suggests that the hyperactivity of the $E(spl)^D$ allele is due in part to its lack of K boxes. Consistent with this interpretation, we have shown here that $E(spl)^D$ mutants exhibit elevated levels of accumulated transcript. The observed levels (Fig. 8) are high enough, however, that they cannot be accounted for solely by relief from K box-mediated regulation.

$E(spl)^D$ proteins also lack the C-terminal WRPW motif, which is required to recruit the co-repressor Groucho by means of a direct protein-protein interaction (Paroush et al., 1994). Since bHLH repressor proteins function as dimers, it is possible that the single Groucho molecule recruited by an $E(spl)^D$/E(spl)+ heterodimer may be functionally sufficient. In this situation, moderately increased levels of truncated $E(spl)^D$ monomers might serve to elevate overall E(spl) protein function, yielding a hypermorphic effect. On the contrary, when the mutant protein is strongly overexpressed, DNA-bound complexes of $E(spl)^D$/$E(spl)^D$ homodimers might predominate; these would be predicted to function in a dominant-negative fashion by virtue of their complete inability to recruit Groucho. Loss of K box-mediated regulation under these circumstances would be predicted to result in an exacerbation of dominant-negative activity.

Genes of the Iro-C and *engrailed* are likely targets of K box-mediated regulation

Since K box-mediated regulation appears to operate in most cells, it is probable that some genes expressed outside of proneural clusters will also prove to be subject to this mode of control. We have identified a number of *Drosophila* genes with K box motifs in their 3′ UTRs, which represent candidate targets of K box regulation. Those of particular interest include the homeobox genes of the *iroquois* Complex (Iro-C) and *engrailed* (*en*).

The Iro-C, consisting of the genes *araucan* (*ara*), *caupolican* (*caup*) and *mirror*, encodes a family of three related homeodomain transcription factors (Gomez-Skarmeta et al., 1996; McNeill et al., 1997). All three genes appear to contribute to a 'prepattern' for sensory organs in the wing imaginal disc, and the ara and caup proteins have been shown to directly activate expression of *ac* and *sc* (Gomez-Skarmeta

et al., 1996; Kehl et al., 1998). Each of the three Iro-C genes contains a single K box in its 3′ UTR (Fig. 9A). Moreover, the K boxes in *ara* and *caup* conform to a much more restricted consensus sequence found in many E(spl)-C K boxes (cTGTGATa) (see Fig. 1). The presence of the K box motif in all three Iro-C genes resembles its widespread distribution in the 3′ UTRs of genes in the E(spl)-C. Thus, genes that activate (Iro-C) and repress [E(spl)-C] proneural gene expression may be regulated by the same post-transcriptional mechanism.

The *en* homeobox gene controls posterior compartment identity in *Drosophila* and has a fundamental role in metazoan patterning. We have found that the *en* 3′ UTR contains a K box located within a large block of sequence conserved between the distantly related species *D. melanogaster* and *D. virilis* (Fig. 9B) (Kassis et al., 1986; Poole et al., 1985). Furthermore, the *en* K box-containing region has been conserved between *Drosophila* and the flour beetle *Tribolium castaneum*, which are at least three hundred million years diverged (Fig. 9B) (Brown et al., 1994). Notably, the 3′ UTR of *Tribolium en* contains a pair of K boxes similar to the K box pairs found in *Drosophila E(spl)m8* and *E(spl)mδ*, suggesting that this paired arrangement is of particular significance. The sequence of the novel second K box in *Tribolium en* is related to the corresponding sequence from *Drosophila*, suggesting that either *Tribolium* has acquired a second K box or the dipteran lineage has lost a K box. Overall, the evolutionary conservation of the *en* K box region among these insects strongly suggests that it is subject to functional constraint.

Gain-of-function alleles of *extramacrochaetae* and *Serrate* are associated with loss of K box motifs

We have shown here that loss of K box-mediated regulation from the *E(spl)m8* 3′ UTR strongly elevates reporter gene transcript and protein levels, and confers readily detectable gain-of-function activities on an *E(spl)m8* genomic DNA transgene, including the capacity to substantially phenocopy the *spl*-enhancing effects of *E(spl)^D*. These findings prompted us to consider whether known gain-of-function alleles of other genes may likewise be associated with loss of K box elements from their 3′ UTRs. A survey of extant hypermorphic mutants identified two such alleles, *Achaetous* (*Ach*) and *Serrate^D* (*Ser^D*).

extramacrochaetae (*emc*) encodes an HLH protein that antagonizes the activity of bHLH proneural activators by forming 'poisoned' heterodimers unable to bind DNA (Ellis et al., 1990; Garrell and Modolell, 1990; Van Doren et al., 1991). Curiously, the *emc* 3′ UTR resembles many E(spl)-C 3′ UTRs, as it contains a Brd box, four copies of a GY box-like sequence (GTTTTCC) and a K box. The only known gain-of-function allele of *emc*, *Ach*, is associated with a transposon insertion, which deletes the codons for the C-terminal 42 amino acids as well as the entire normal 3′ UTR (Garrell and Modolell, 1990). *Ser* encodes one of two known ligands for the Notch receptor and functions in pattern formation and cell proliferation in the developing wing (Fleming et al., 1990; Rebay et al., 1991). The hypermorphic allele *Ser^D* is associated with a transposon insertion into the 3′ UTR of the gene; the mutant transcript is truncated but encodes a wild-type protein (Thomas et al., 1995). A single K box motif is lost from the 3′ UTR in *Ser^D*.

Both *Ach* and *Ser^D* are associated with insertion of the *Tirant* transposon (Garrell and Modolell, 1990; Thomas et al., 1995).

However, the gain-of-function nature of these mutants is unlikely to be due to a strong transcriptional activation effect of this transposon, as transcription of *Ser* is not increased in *Ser^D* (Thomas et al., 1995). Instead, the transposon insertion is thought to result in increased stability of mutant transcripts. We suggest that loss of K box-mediated regulation in *Ach* and *Ser^D* may contribute to their hypermorphic character.

Is there a link between gene duplication and negative regulation mediated by Brd boxes and K boxes?

An unknown but probably very significant fraction of the genes in a typical metazoan genome can be mutated to an inactive state without conferring a mutant phenotype that is readily detectable under laboratory conditions (Miklos and Rubin, 1996). Members of paralogous multigene families are perhaps especially likely to fall into this category. Such genes are frequently referred to as 'redundant' because of their apparent functional overlap with other genes, but we may be confident that their evolutionary retention in an active state indicates that they are not truly redundant or otherwise dispensable. Nevertheless, in the absence of a recognizable loss-of-function phenotype, the existence of a particular gene of this type is frequently first revealed by the dominant phenotypic effects of a gain-of-function allele. Such is the case for the genes of the E(spl)-C and for *Brd* (Leviten and Posakony, 1996; Lindsley and Zimm, 1992).

Negative regulatory mechanisms that limit gene activity are obvious targets for mutations that will yield gain-of-function alleles. We have previously presented evidence that loss of Brd box sequences from the 3′ UTR and the consequent loss of the negative regulation mediated by this motif, accounts at least in part for the gain-of-function properties of the original *Brd^1* allele (Lai and Posakony, 1997; Leviten et al., 1997; Leviten and Posakony, 1996). In this paper, we have similarly shown that mutation of K box elements in the *E(spl)m8* 3′ UTR confers a partial phenocopy of *E(spl)^D*.

We are struck by the apparent prevalence of Brd box and K box motifs among genes that are members of paralogous gene families. The 3′ UTRs of five of the seven bHLH repressor-encoding genes of the E(spl)-C (*m3*, *m5*, *m7*, *mγ* and *mδ*) contain at least one copy of both of these classes of sequence element. We have suggested (Leviten et al., 1997) that *Brd* and *E(spl)m4* constitute a small gene family. *m4* also contains both Brd boxes and K boxes in its 3′ UTR, while *Brd* contains Brd boxes and possibly a K box (see legend to Fig. 1). Three other examples of duplicated gene sets in *Drosophila* appear to reinforce this association. First, *en* and *invected*, homeobox genes with overlapping functions in controlling posterior compartment identity, contain a K box and a Brd box, respectively, in their 3′ UTRs (Coleman et al., 1987; Poole et al., 1985). Morever, the phylogenetic analysis presented above (Fig. 9) strongly suggests that the K box is important for *en* regulation. Second, all three genes in the Iro-C contain K boxes in their 3′ UTRs, are co-expressed in the wing imaginal disc and function to activate expression of *ac* and *sc* (Gomez-Skarmeta et al., 1996; Kehl et al., 1998). Finally, *PKC53E* and *inaD*, two protein kinase C genes in the 53E cytological region, each contain single Brd boxes in their 3′ UTRs (Rosenthal et al., 1987; Schaeffer et al., 1989). Although the normal role of *PKC53E* is unknown, its expression pattern overlaps that of

inaD in the eye and thus it may have a related function. Like the bHLH genes of the E(spl)-C and the *Brd/E(spl)m4* family, all of these are examples of protein sets related by primary sequence, expression pattern and putative or known function.

With these considerations in mind, it is intriguing to consider the possible evolutionary significance of negative regulatory motifs and mechanisms such as those exemplified by the Brd box and K box. Whether these elements were present in the ancestral single-copy genes prior to duplication and were then retained by the duplicated gene copies, or whether they were independently acquired by the members of the gene set after duplication, is unclear. In either case, it may be that the presence of these negative regulatory motifs made the gene duplication events more tolerable, by limiting the possibly deleterious gain-of-function effects that could result from the corresponding increase in gene activity. Thus, it is possible that negative post-transcriptional regulation may have not only specific *cis*-acting effects on gene expression, but also more general consequences for genome evolution.

REFERENCES

Ashburner, M. (1982). The genetics of a small autosomal region of *Drosophila melanogaster* containing the structural gene for alcohol dehydrogenase. III. Hypomorphic and hypermorphic mutations affecting the expression of *Hairless*. *Genetics* **101**, 447-459.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1987). *Current Protocols in Molecular Biology*. New York: John Wiley & Sons.

Bailey, A. M. and Posakony, J. W. (1995). Suppressor of Hairless directly activates transcription of *Enhancer of split* Complex genes in response to Notch receptor activity. *Genes Dev.* **9**, 2609-2622.

Bang, A. G., Hartenstein, V. and Posakony, J. W. (1991). *Hairless* is required for the development of adult sensory organ precursor cells in *Drosophila*. *Development* **111**, 89-104.

Beverley, S. M. and Wilson, A. C. (1984). Molecular evolution in *Drosophila* and the higher Diptera II. A time scale for fly evolution. *J. Mol. Evol.* **21**, 1-13.

Brown, S. J., Patel, N. H. and Denell, R. E. (1994). Embryonic expression of the single *Tribolium engrailed* homolog. *Dev. Genet.* **15**, 7-18.

Cabrera, C. V. and Alonso, M. C. (1991). Transcriptional activation by heterodimers of the *achaete-scute* and *daughterless* gene products of *Drosophila*. *EMBO J.* **10**, 2965-2973.

Caccone, A. and Powell, J. R. (1990). Extreme rates and heterogeneity in insect DNA evolution. *J. Mol. Evol.* **30**, 273-280.

Chen, C. Y. and Shyu, A.-B. (1995). AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem. Sci.* **20**, 465-470.

Coleman, K. G., Poole, S. J., Weir, M. P., Soeller, W. C. and Kornberg, T. (1987). The *invected* gene of *Drosophila*: Sequence analysis and expression studies reveal a close kinship to the *engrailed* gene. *Genes Dev.* **1**, 19-28.

Cubas, P., de Celis, J.-F., Campuzano, S. and Modolell, J. (1991). Proneural

clusters of *achaete-scute* expression and the generation of sensory organs in the *Drosophila* imaginal wing disc. *Genes Dev.* **5**, 996-1008.

Curtis, D., Lehmann, R. and Zamore, P. D. (1995). Translational regulation in development. *Cell* **81**, 171-178.

de Celis, J. F., Bray, S. and Garcia-Bellido, A. (1997). Notch signalling regulates *veinlet* expression and establishes boundaries between veins and interveins in the *Drosophila* wing. *Development* **124**, 1919-1928.

de Celis, J. F., de Celis, J., Ligoxygakis, P., Preiss, A., Delidakis, C. and Bray, S. (1996). Functional relationships between *Notch*, *Su(H)* and the bHLH genes of the *E(spl)* complex: the *E(spl)* genes mediate only a subset of *Notch* activities during imaginal development. *Development* **122**, 2719-2728.

Dietrich, U. and Campos-Ortega, J. A. (1984). The expression of neurogenic loci in the imaginal epidermal cells of *Drosophila melanogaster*. *J. Neurogenetics* **1**, 315-332.

Ellis, H. M., Spann, D. R. and Posakony, J. W. (1990). *extramacrochaetae*, a negative regulator of sensory organ development in Drosophila, defines a new class of helix-loop-helix proteins. *Cell* **61**, 27-38.

Fisher, A. L. and Caudy, M. (1998). Groucho proteins: transcriptional corepressors for specific subsets of DNA-binding transcription factors in vertebrates and invertebrates. *Genes Dev.* **12**, 1931-1940.

Fleming, R. J., Scottgale, N. T., Diederich, R. J. and Artavanis-Tsakonas, S. (1990). The gene *Serrate* encodes a putative EGF-like transmembrane protein essential for proper ectodermal development in *Drosophila melanogaster*. *Genes Dev.* **4**, 2188-2201.

Furukawa, T., Kobayakawa, Y., Tamura, K., Kimura, K., Kawaichi, M., Tanimura, T. and Honjo, T. (1995). Suppressor of Hairless, the *Drosophila* homologue of RBP-Jκ, transactivates the neurogenic gene *E(spl)m8*. *Jpn. J. Genet.* **70**, 505-524.

Garrell, J. and Modolell, J. (1990). The Drosophila *extramacrochaetae* locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. *Cell* **61**, 39-48.

Giebel, B. and Campos-Ortega, J. A. (1997). Functional dissection of the *Drosophila Enhancer of split* protein, a suppressor of neurogenesis. *Proc. Natl. Acad. Sci. USA* **94**, 6250-6254.

Gomez-Skarmeta, J. L., del Corral, R. D., de la Calle-Mustienes, E., Ferres-Marco, D. and Modolell, J. (1996). *araucan* and *caupolican*, two members of the novel Iroquois Complex, encode homeoproteins that control proneural and vein-forming genes. *Cell* **85**, 95-105.

Hartenstein, V. and Posakony, J. W. (1989). Development of adult sensilla on the wing and notum of *Drosophila melanogaster*. *Development* **107**, 389-405.

Hartenstein, V. and Posakony, J. W. (1990). A dual function of the *Notch* gene in *Drosophila* sensillum development. *Dev. Biol.* **142**, 13-30.

Heitzler, P. and Simpson, P. (1991). The choice of cell fate in the epidermis of Drosophila. *Cell* **64**, 1083-1092.

Hertz, G. Z., Hartzell, G. W. and Stormo, G. D. (1990). Identification of consensus patterns in unaligned DNA sequences known to be functionally related. *Comput. Appl. Biosci.* **6**, 81-92.

Huang, F., Dambly-Chaudière, C. and Ghysen, A. (1991). The emergence of sense organs in the wing disc of *Drosophila*. *Development* **111**, 1087-1095.

Kassis, J. A., Poole, S. J., Wright, D. K. and O'Farrell, P. H. (1986). Sequence conservation in the protein coding and intron regions of the *engrailed* transcription unit. *EMBO J.* **5**, 3583-3589.

Kehl, B. T., Cho, K.-O. and Choi, K.-W. (1998). *mirror*, a *Drosophila* homeobox gene in the *iroquois* complex, is required for sensory organ and alula formation. *Development* **125**, 1217-1227.

Klämbt, C., Knust, E., Tietze, K. and Campos-Ortega, J. (1989). Closely related transcripts encoded by the neurogenic gene complex *Enhancer of split* of *Drosophila melanogaster*. *EMBO J.* **8**, 203-210.

Knust, E., Tietze, K. and Campos-Ortega, J. A. (1987). Molecular analysis of the neurogenic locus Enhancer of split of *Drosophila melanogaster*. *EMBO J.* **6**, 4113-4123.

Kramatschek, B. and Campos-Ortega, J. A. (1994). Neuroectodermal transcription of the *Drosophila* neurogenic genes *E(spl)* and *HLH-m5* is regulated by proneural genes. *Development* **120**, 815-826.

Lai, E. C. and Posakony, J. W. (1997). The Bearded box, a novel 3′ UTR sequence motif, mediates negative post-transcriptional regulation of *Bearded* and *Enhancer of split* Complex gene expression. *Development* **124**, 4847-4856.

Lai, E. C. and Posakony, J. W. (1998). Regulation of *Drosophila* neurogenesis by RNA:RNA duplexes? *Cell* **93**, 1103-1104.

Lecourtois, M. and Schweisguth, F. (1995). The neurogenic Suppressor of

Hairless DNA-binding protein mediates the transcriptional activation of the *Enhancer of split* Complex genes triggered by Notch signaling. *Genes Dev.* **9**, 2598-2608.

Leviten, M. W., Lai, E. C. and Posakony, J. W. (1997). The *Drosophila* gene *Bearded* encodes a novel small protein and shares 3′ UTR sequence motifs with multiple *Enhancer of split* Complex genes. *Development* **124**, 4039-4051.

Leviten, M. W. and Posakony, J. W. (1996). Gain-of-function alleles of *Bearded* interfere with alternative cell fate decisions in *Drosophila* adult sensory organ development. *Dev. Biol.* **176**, 264-283.

Lindsley, D. L. and Zimm, G. G. (1992). *The Genome of Drosophila melanogaster*. San Diego: Academic Press, Inc.

Maier, D., Marte, B. M., Schafer, W., Yu, Y. and Preiss, A. (1993). Drosophila evolution challenges postulated redundancy in the E(spl) gene complex. *Proc. Natl. Acad. Sci. USA* **90**, 5464-5468.

Margolis, J. S., Borowsky, M. L., Steingrimsson, E., Shim, C. W., Lengyel, J. A. and Posakony, J. W. (1995). Posterior stripe expression of *hunchback* is driven from two promoters by a common enhancer element. *Development* **121**, 3067-3077.

McNeill, H., Yang, C. H., Brodsky, M., Ungos, J. and Simon, M. A. (1997). *mirror* encodes a novel PBX-class homeoprotein that functions in the definition of the dorsal-ventral border in the *Drosophila* eye. *Genes Dev.* **11**, 1073-1082.

Miklos, G. L. and Rubin, G. M. (1996). The role of the genome project in determining gene function: Insights from model organisms. *Cell* **86**, 521-529.

Nakao, K. and Campos-Ortega, J. A. (1996). Persistent expression of genes of the *Enhancer of split* complex suppresses neural development in *Drosophila*. *Neuron* **16**, 275-286.

Parks, A. L. and Muskavitch, M. A. (1993). *Delta* function is required for bristle organ determination and morphogenesis in *Drosophila*. *Dev. Biol.* **157**, 484-496.

Paroush, Z., Finley, R. L., Jr., Kidd, T., Wainwright, S. M., Ingham, P. W., Brent, R. and Ish-Horowicz, D. (1994). Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* **79**, 805-815.

Poole, S. J., Kauvar, L. M., Drees, B. and Kornberg, T. (1985). The *engrailed* locus of *Drosophila*: Structural analysis of an embryonic transcript. *Cell* **40**, 37-43.

Posakony, J. W. (1994). Nature versus nurture: asymmetric cell divisions in Drosophila bristle development. *Cell* **76**, 415-418.

Rebay, I., Fleming, R. J., Fehon, R. G., Cherbas, L., Cherbas, P. and Artavanis-Tsakonas, S. (1991). Specific EGF repeats of Notch mediate interactions with Delta and Serrate: Implications for Notch as a multifunctional receptor. *Cell* **67**, 687-699.

Romani, S., Campuzano, S., Macagno, E. R. and Modolell, J. (1989). Expression of *achaete* and *scute* genes in *Drosophila* imaginal discs and their function in sensory organ development. *Genes Dev.* **3**, 997-1007.

Rosenthal, A., Rhee, L., Yadegari, R., Paro, R., Ullrich, A. and Goeddel, D. V. (1987). Structure and nucleotide sequence of a *Drosophila melanogaster* protein kinase C gene. *EMBO J.* **6**, 433-441.

Rubin, G. M. and Spradling, A. C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**, 348-353.

Schaeffer, E., Smith, D., Mardon, G., Quinn, W. and Zuker, C. (1989). Isolation and characterization of two new drosophila protein kinase C genes, including one specifically expressed in photoreceptor cells. *Cell* **57**, 403-12.

Schrons, H., Knust, E. and Campos-Ortega, J. A. (1992). The *Enhancer of split* complex and adjacent genes in the 96F region of *Drosophila melanogaste*r are required for segregation of neural and epidermal progenitor cells. *Genetics* **132**, 481-503.

Schweisguth, F. and Posakony, J. W. (1992). *Suppressor of Hairless*, the Drosophila homolog of the mouse recombination signal-binding protein gene, controls sensory organ cell fates. *Cell* **69**, 1199-1212.

Schweisguth, F. and Posakony, J. W. (1994). Antagonistic activities of *Suppressor of Hairless* and *Hairless* control alternative cell fates in the *Drosophila* adult epidermis. *Development* **120**, 1433-1441.

Singson, A., Leviten, M. W., Bang, A. G., Hua, X. H. and Posakony, J. W. (1994). Direct downstream targets of proneural activators in the imaginal disc include genes involved in lateral inhibitory signaling. *Genes Dev.* **8**, 2058-2071.

Skeath, J. B. and Carroll, S. B. (1991). Regulation of *achaete-scute* gene expression and sensory organ pattern formation in the *Drosophila* wing. *Genes Dev.* **5**, 984-995.

St. Johnston, D. (1995). The intracellular localization of messenger RNAs. *Cell* **81**, 161-170.

Tata, F. and Hartley, D. A. (1995). Inhibition of cell fate in *Drosophila* by *Enhancer of split* genes. *Mech. Dev.* **51**, 305-315.

Thomas, U., Jonsson, F., Speicher, S. A. and Knust, E. (1995). Phenotypic and molecular characterization of *Ser^D*, a dominant allele of the *Drosophila* gene *Serrate*. *Genetics* **139**, 203-213.

Tietze, K., Oellers, N. and Knust, E. (1992). *Enhancer of split^D*, a dominant mutation of *Drosophila*, and its use in the study of functional domains of a helix-loop-helix protein. *Proc. Natl. Acad. Sci. USA* **89**, 6152-6156.

Usui, K. and Kimura, K. (1993). Sequential emergence of the evenly spaced microchaetes on the notum of *Drosophila*. *Roux's Arch. Dev. Biol.* **203**, 151-158.

Van Doren, M., Ellis, H. M. and Posakony, J. W. (1991). The *Drosophila extramacrochaetae* protein antagonizes sequence-specific DNA binding by *daughterless/achaete-scute* protein complexes. *Development* **113**, 245-255.

Van Doren, M., Powell, P. A., Pasternak, D., Singson, A. and Posakony, J. W. (1992). Spatial regulation of proneural gene activity: auto- and cross-activation of *achaete* is antagonized by *extramacrochaetae*. *Genes Dev.* **6**, 2592-2605.

Vincent, J. P., Girdham, C. H. and O'Farrell, P. H. (1994). A cell-autonomous, ubiquitous marker for the analysis of *Drosophila* genetic mosaics. *Dev. Biol.* **164**, 328-331.