Copyright © 1992 by the Genetics Society of America

Seven Genes of the *Enhancer of split* Complex of *Drosophila melanogaster* Encode Helix-Loop-Helix Proteins

Elisabeth Knust, Herbert Schrons, Ferdi Grawe and José A. Campos-Ortega

Institut für Entwicklungsbiologie, Universität zu Köln, 5000 Köln 41, Germany

Manuscript received March 31, 1992
Accepted for publication June 19, 1992

ABSTRACT

Enhancer of split [*E(spl)*] is one of the neurogenic loci of *Drosophila* and, as such, is required for normal segregation of neural and epidermal cell progenitors. Genetic observations indicate that the *E(spl)* locus is in fact a gene complex comprising a cluster of related genes and that other genes of the region are also required for normal early neurogenesis. Three of the genes of the complex were known to encode helix-loop-helix (HLH) proteins and to be transcribed in nearly identical patterns. Here, we show that four other genes in the vicinity also encode HLH proteins and, during neuroblast segregation, three of them are expressed in the same pattern. We show by germ-line transformation that these three genes are also necessary to allow epidermal development of the neuroectodermal cells.

ENHANCER of split [*E(spl)*] is one of the so-called neurogenic loci of *Drosophila melanogaster* (LEHMANN *et al.* 1983; KNUST *et al.* 1987a). Whereas the other loci of this group are single genes, the *E(spl)* locus was found to be composed of a complex of several genes clustered together, called E(spl)-C, required to allow epidermal development of neuroectodermal cells (KNUST, TIETZE and CAMPOS-ORTEGA 1987; ZIEMER *et al.* 1988; KLÄMBT *et al.* 1989) [see CAMPOS-ORTEGA (1991) for a review]. Molecular data suggested that at least three transcription units, *m5*, *m7* and *m8*, are members of the gene complex, as they were found to encode highly conserved helix-loop-helix (HLH) proteins (KLÄMBT *et al.* 1989) and to be expressed in nearly identical patterns (KNUST, TIETZE and CAMPOS-ORTEGA 1987). Transformation experiments were used to show that transcription unit *m8* corresponds to the *E(spl)* gene (KLÄMBT *et al.* 1989; TIETZE, OELLERS and KNUST 1992). Another transcription unit in this region, *m9/m10*, which is also required for this process (ZIEMER *et al.* 1988; PREISS, HARTLEY and ARTAVANIS-TSAKONAS 1988), corresponds to the gene *groucho* (*gro*) (PREISS, HARTLEY and ARTAVANIS-TSAKONAS 1988). *gro* encodes an ubiquitously expressed protein with similarity to the β subunit of transducin (HARTLEY, PREISS and ARTAVANIS-TSAKONAS 1988; DELIDAKIS *et al.* 1991) and, therefore, differs structurally from those of the E(spl)-C. However, it was previously pointed out (KNUST, TIETZE and CAMPOS-ORTEGA 1987) that, beside these genes, others located further proximally are also required for early neurogenesis.

Here we show that, in addition to the three previously described HLH protein-encoding genes, four other genes in the neighborhood belong to the same family. Moreover, we show by germ-line transformation that three of them participate in mediating the decision between the neural and epidermal pathways. Finally, the transcription patterns of these three are very similar to that described for the RNA products of the other three, whereas the fourth (*m3*) is ubiquitously transcribed as previously shown (KNUST, TIETZE and CAMPOS-ORTEGA 1987). Transcription unit *m8* is *E(spl)*, but no visible phenotypes are known to be related to the other six HLH protein-encoding genes (SCHRONS, KNUST and CAMPOS-ORTEGA 1992). Thus, we have provisionally designated these genes *HLH-m3*, *HLH-m5*, *HLH-m7*, *HLH-m*β, *HLH-m*γ and *HLH-m*δ, based on the deduced structure of the encoded proteins, and proposed that these six genes, together with the *E(spl)* gene, form a gene complex which we have termed E(spl)-C.

MATERIALS AND METHODS

Drosophila stocks: Flies were grown on standard medium and crosses were performed either at room temperature or at 25°. We used the alleles *Df(3R)E(spl)*$^{R-A7.1}$ (deficient for the transcripts *m1* to *m9/m10*; KNUST *et al.* 1987a; KNUST, TIETZE and CAMPOS-ORTEGA 1987) and *Df(3R)gro*$^{ry78R-b32.2}$ (deficient for the transcripts *m*δ-*m9/m10*; SCHRONS, KNUST and CAMPOS-ORTEGA 1992), here referred to as *E(spl)*$^{R-A7.1}$ and *gro*$^{b32.2}$, respectively. Homozygous w^{1118} embryos were used as recipients for germ-line transformation, OregonR as wild-type stock.

Isolation of cDNA and genomic clones, subcloning and sequencing: The cDNA clones in λgt10 were isolated from cDNA libraries made from embryonic RNA of different developmental stages (0–3, 3–12 and 12–24 hr, respectively), kindly provided by L. KAUVAR (POOLE *et al.* 1985). The genomic library, derived from Oregon R and cloned into the EMBL-4 phage vector (PIRROTTA, HADFIELD and

G. H. J PRETORIUS 1983), was kindly provided by V. PIRROTTA. Screening of libraries, radioactive labeling of DNA probes, hybridizations, preparations of DNA and Southern blot analysis were essentially as described in SAMBROOK, FRITSCH and MANIATIS (1989). Screening under reduced stringency conditions was performed as described in KNUST *et al.* (1987b). For sequencing of subcloned fragments, we applied the chain termination method (SANGER, NICKLEN and COULSON 1977). Computer analysis was carried out on an IBM PC/AT with the DNA/protein sequence analysis software of J. M. PUSTELL/International Biotechnologies, Inc., New Haven, Connecticut. Computer homology searches were carried out using the program of LIPMAN and PEARSON (1985) and the NBRF protein bank.

Isolation of RNA, Northern blot analysis and *in situ* hybridization: Total RNA from staged embryos, third instar larvae, late pupae, male and female flies was isolated according to the method described by AUFFRAY and ROUGEON (1980) and enriched for poly(A^+) RNA by oligo(dT)-cellulose chromatography. The RNA was separated on agarose gels, blotted to nylon membranes and hybridized as described in VÄSSIN *et al.* (1987). *In situ* hybridizations to whole-mount embryos were performed with digoxigenin-labeled probes essentially as described by TAUTZ and PFEIFLE (1989). Antibody staining and cuticle preparations followed conventional protocols. Staging of embryos follows CAMPOS-ORTEGA and HARTENSTEIN (1985).

Germ-line transformation: A 13-kb *Sma*I-*Bam*HI genomic fragment containing transcription units *mβ*, *mγ* and *mδ* was isolated from a phage clone, using the *Sma*I site within the phage arm and the *Bam*HI site at map position −20.4 (see Figure 2B); another fragment containing only $m\delta^+$ was obtained as the distalmost 6-kb *Eco*RI fragment from the same phage clone, using the *Eco*RI cloning site of the phage and the genomic *Eco*RI site at map position at −27.5. These fragments were cloned into the pCaSpeR transformation vector (PIRROTTA 1988). Finally, a *Sal*I fragment containing gro^+ was obtained from map positions +12.5 to 21.7 and cloned in the pW8 vector (KLEMENZ, WEBER and GEHRING 1987). Germ-line transformation experiments were carried out essentially as described in SPRADLING (1986), transposase was supplied by coinjection of the Δ2–3 helper plasmid (LASKI, RIO and RUBIN 1986). Transgenic stocks were established over the appropriate balancers or kept as homozygotes.

RESULTS

Embryos homozygous for $Df(3R)E(spl)^{R\text{-}A7.1}$, thus lacking the genes *m1* to *gro*, develop a strong neurogenic phenotype (Figures 1 and 2); however, even more extreme phenotypes are observed if genes further proximal are also deleted (KNUST *et al.*, 1987a; ZIEMER *et al.*, 1988; Figure 1). Thus, we proposed that the deficiency $E(spl)^{R\text{-}A7.1}$ does not remove all the genes of the locus required for a normal separation of neural and epidermal cell lineages (KNUST, TIETZE and CAMPOS-ORTEGA 1987). Consequently, we screened genomic DNA proximal to the region defined by $E(spl)^{R\text{-}A7.1}$, up to map position −47 (Figure 2A) (KNUST, TIETZE and CAMPOS-ORTEGA 1987), with the coding region of *HLH-m5* under low stringency conditions and found three cross-hybridizing fragments (Figure 2B). Using several fragments from the region between map units −33.5 and −19.0, which included the cross-hybridizing fragments, to probe northern blots, we detected three transcripts, with sizes of 1.1, 1.0 and 1.1 kb, named *mβ*, *mγ* and *mδ*, respectively, from distal to proximal (Figure 2C). Whereas *mβ* is expressed at all developmental stages, transcripts from *mγ* and *mδ* can only be detected in poly(A^+) RNA from 0–10-hr embryos, but not at later stages. In addition to these three, a fourth transcript of 0.8 kb, called *mα*, is located between *mβ* and *m1* and can be detected in RNA prepared from 0–10-hr embryos, but not in later stages.

We sequenced cDNA and genomic clones corresponding to *mβ*, *mγ* and *mδ*; we also sequenced cDNA and genomic clones corresponding to *m3*, which also seems to participate in the control of the neural/epidermal dichotomy (SCHRONS, KNUST and CAMPOS-ORTEGA 1992), the sequence of which had not yet been determined. Conceptual translation of these four genes fails to uncover any intronic sequence and points to small protein products of 21.4, 23.2, 20.2 and 24.9 kD, respectively. A basic region is found at the amino terminus of all four proteins, immediately followed by two clusters of mainly hydrophobic amino acids (Figure 3), reminiscent of the basic domain and amphipathic helices of the so-called bHLH protein family. This family includes, among other members, the products of *myc* oncogenes (see LÜSCHER and EISENMANN 1990), proteins involved in myogenesis (see OLSON 1990), and several Drosophila proteins, some of them also involved in the neural development, such as daughterless (CAUDY *et al.*, 1988) and the four proteins encoded by the genes of the *achaete-scute* complex AS-C (VILLARES and CABRERA 1987; ALSONSO and CABRERA 1988; GONZÀLEZ *et al.* 1989). The four proteins presented here exhibit a high degree of sequence similarity to each other and to the proteins encoded by the genes *HLH-m5*, *HLH-m7* and *E(spl)* (KLÄMBT *et al.* 1989) (Figure 4A). Therefore, we propose to call these four genes *HLH-m3*, *HLH-mβ*, *HLH-mγ* and *HLH-mδ*. The homology includes, besides the bHLH motif, two further domains (helix III and helix IV), and a sequence of four amino acids, tryptophan-arginine-proline-tryptophan (W-R-P-W) which occurs at the carboxy terminus of all these proteins (Figure 4).

***HLH-mβ*, *HLH-mγ* and *HLH-mδ* contribute to the neural/epidermal decision:** The striking similarity between the proteins encoded by *HLH-mβ*, *HLH-mγ*, *HLH-mδ* and *HLH-m3* and *HLH-m5*, *HLH-m7* and *E(spl)* suggests that these seven genes have common functions. We mentioned above that homozygosity for deficiencies extending further proximal to *m1* causes a more severe neurogenic phenotype than that of $E(spl)^{R\text{-}A7.1}$ (Figure 1). To test whether this is due to the lack of the genes *HLH-mβ*, *HLH-mγ* and *HLH-mδ* and, thus, whether these genes contribute to the

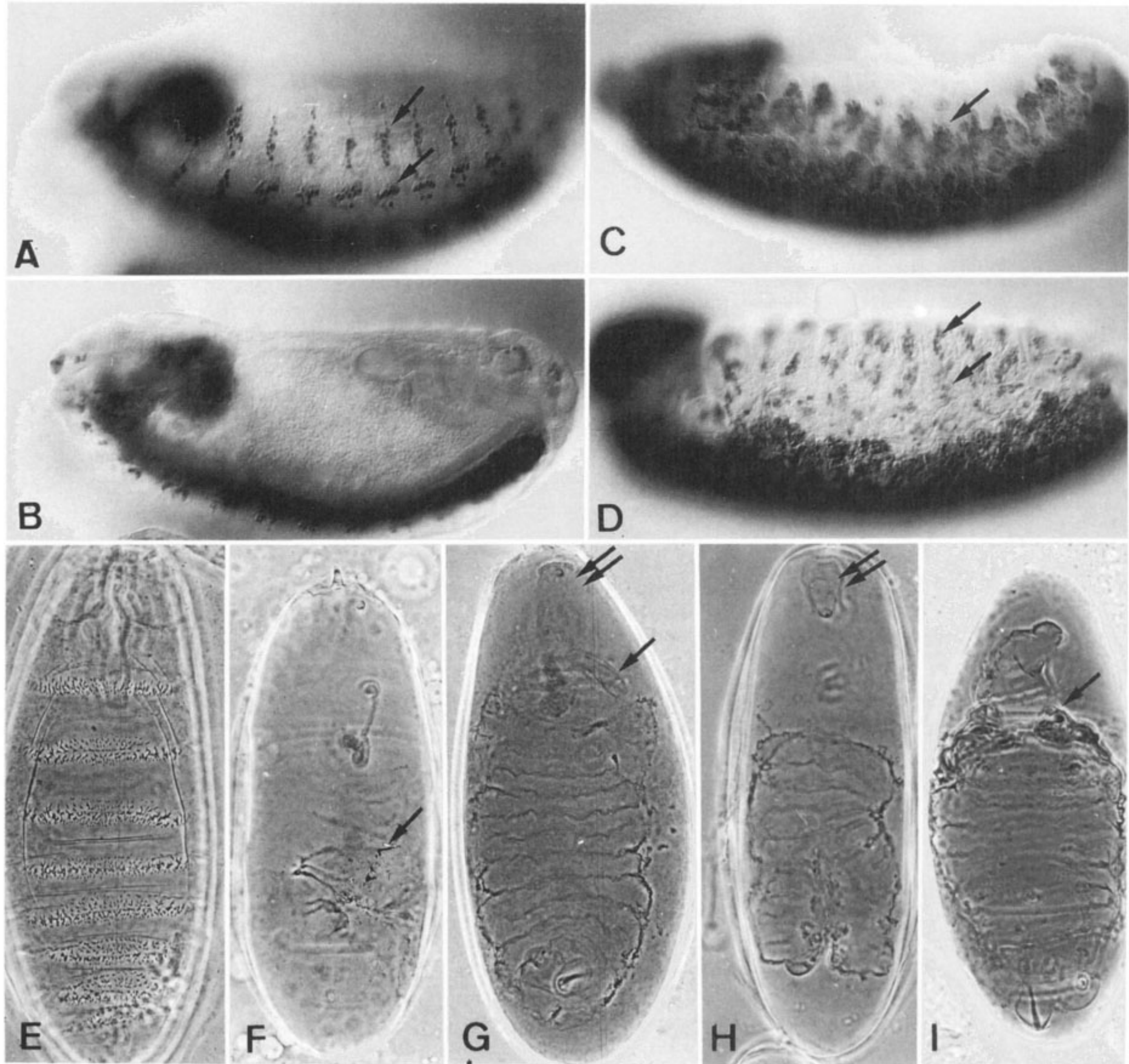


FIGURE 1.—**A** and **B** are lateral and medial focal planes of the same wild-type embryo, **C** shows an embryo homozygous for $gro^{b32.2}$, and **D** an embryo homozygous for $gro^{b32.2}$ and a copy of the transgenic *HLH*-[*mβ-mγ-mδ*]$^+$ fragment; all embryos are at late stage 15 and stained with the neural specific antibody 44c11 (kindly provided by Y. N. JAN). Notice the reduction in the severity of the neural hyperplasia of $gro^{b32.2}$ embryos caused by the *HLH*-[*mβ-mγ-mδ*]$^+$ fragment. **E–I** show cuticle preparations of fully developed embryos, wild-type (**E**), homozygous for $gro^{b32.2}$ (**F**), homozygous for $gro^{b32.2}$ and a copy of the transgenic *HLH*-[*mβ-mγ-mδ*]$^+$ fragment (**G**), homozygous for $E(spl)^{R\text{-}A7.1}$ (**H**), and homozygous for $E(spl)^{R\text{-}A7.1}$ with two copies of the *HLH*-[*mβ-mγ-mδ*]$^+$ fragment (**I**).

neural-epidermal lineage dichotomy, a genomic fragment of 13 kb including the coding regions of these three genes, henceforth referred to as the *HLH*-[*mβ-mγ-mδ]*$^+$ fragment, and another fragment of 5 kb, comprising the coding region of *HLH-mδ*$^+$ alone (see Figure 2), were subcloned into a transformation vector. Two transgenic lines were generated carrying the *HLH-mδ*$^+$ fragment and one carrying the *HLH*-[*mβ-mγ-mδ*]$^+$ fragment inserted on the third chromosome. These third chromosomal insertions were recombined into chromosomes carrying either the deficiency $E(spl)^{R\text{-}A7.1}$, which lacks the region extending from *m1* to *gro* (KNUST, TIETZE and CAMPOS-ORTEGA 1987), or the deficiency $gro^{b32.2}$, which is deficient for the region between a transcription unit proximal to *HLH-mδ* up to *E(spl)* (Figure 2A) and has a partially affected *gro* gene (SCHRONS, KNUST and CAMPOS-ORTEGA, 1992). In addition, we also tested another second chromosomal insertion of the *HLH*-[*mβ-mγ-mδ*]$^+$ fragment with the deficiency $gro^{b32.2}$. Embryos homozygous for a recombinant chromosome bearing $gro^{b32.2}$ and the *HLH*-[*mβ-mγ-mδ*]$^+$ fragment, *i.e.*, carrying two copies of the transgenic fragment, show a significantly reduced neurogenic phenotype (Figure 1, C, D and G). One copy of the transgenic fragment results in slightly less attenuation of the $gro^{b32.2}$ phenotype (not shown). The phenotype of these latter embryos is indistinguishable from that of $E(spl)^{R\text{-}A7.1}$ embryos in

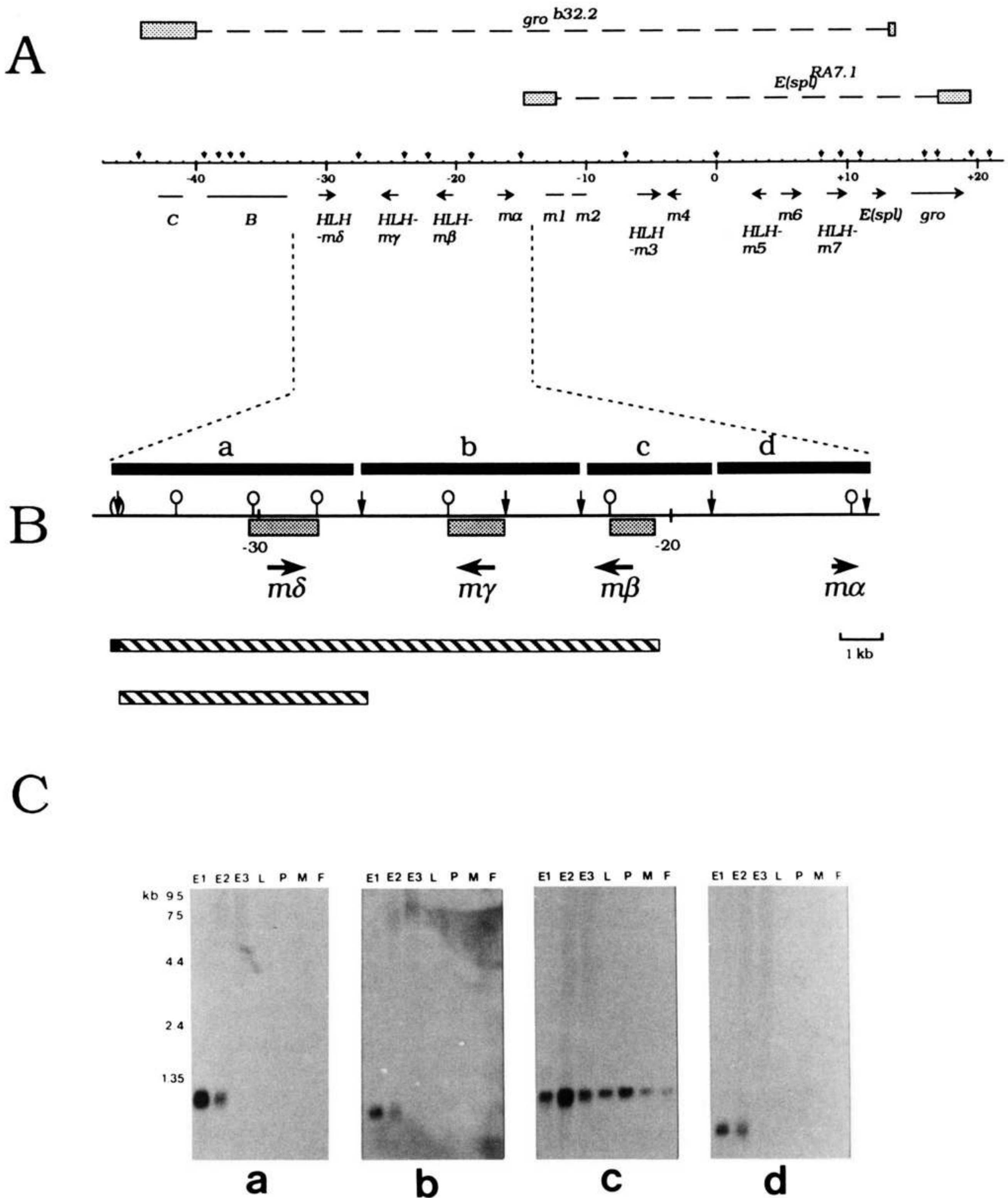


FIGURE 2.—**A** is a schematic representation of the genomic organization of the region of the E(spl)-C. Map units are given in kb (numbering according to KNUST, TIETZE and CAMPOS-ORTEGA 1987), *Eco*RI sites are indicated as arrowheads. Arrows represent the different transcripts and the direction of transcription, insofar as it is known. The position of transcripts C and B is from HART *et al.* (1990). The centromere is to the left. The broken lines indicate genomic regions missing in $E(spl)^{R\text{-}A7.1}$ and $gro^{b32.2}$, stippled bars the fragments to which the breakpoints of the deficiencies were mapped (KNUST, TIETZE and CAMPOS-ORTEGA 1987; SCHRONS, KNUST and CAMPOS-ORTEGA 1992). **B** shows a detailed map of the proximal genomic region of the E(spl)-C described in this paper. Arrowheads indicate *Eco*RI sites (the one in brackets is the site of the phage vector adjacent to the cloning site), open circles *Hin*dIII sites. The stippled boxes indicate the fragments that cross-hybridized, under low stringency conditions, with a probe containing the *m5* transcription unit. The cross-hatched bars below represent the fragments, which were used for germ line transformations. The upper bar (map unit −20.4 to −33.5), which includes a small fragment of the phage vector (black box), contains the transcription units $m\beta$, $m\gamma$ and $m\delta$, the lower one (map units −27.7 to −33.5) only that of $m\delta$. The black bars (a, b, c, d) indicate the genomic *Eco*RI fragments used to probe northern blots (see **C**). **C** Transcriptional activity of the genomic region shown in **B**. ^{32}P-Labeled genomic *Eco*RI fragments a, b, c and d were hybridized to poly(A^+) RNA isolated from different developmental stages. E1: 0–10-hr embryos; E2: 10–14-hr embryos; E3: 14–20-hr embryos; L: third instar larvae; P: pupae; M: adult males; F: adult females. In a–d, the same Northern blot was used.

```
HLH-m3
1      CTGGGAGCGTGTGTGAGTGATCATGCGGGAGCTTTGTGGCCCGGGAGAGCCTGATCCCCGTTTGCCCGGGAGAGCGTGGGCTGTCATATA
91     AAAGTAGGGACTGGGAACTAGTTTGCCACTCATTTCGTAAAAGTGCCGCCACTCGAACGCACCGCAATCGGTGGAACGCGCAACGCACGC
181    GCAACAAACTCGTCAAAAAGTAACCAATCTATGAAACCAACCAAAATCAGTAGTGATAATGTGCAGTTGAGAGACTAAATACAAAAGAAC
271    AATCCTTAAAATACACAACATAAAACAACCCAATCGATCATGGTCATGGAGATGTCCAAGACGTATCAGTACCGCAAGGTGATGAAGCCG
                                                 M  V  M  E  M  S  K  T  Y  Q  Y  R  K  V  M  K  P     17
361    CTGCTGGAGCGCAAACGGCGAGCCAGGATCAACAAGTGTCTGGACGATCTCAAGGATCTGATGGTGGAGTGCCTGCAGCAGGAGGGCGAG
        L  L  E  R  K  R  R  A  R  I  N  K  C  L  D  D  L  K  D  L  M  V  E  C  L  Q  Q  E  G  E     47
451.   CATGTCACCCGGCTGGAGAAGGCGGATATCCTGGAGTTGACCGTGGATCACATGAGGAAACTCAAGCAGCGGGGTGGTCTTTCGCTGCAG
        H  V  T  R  L  E  K  A  D  I  L  E  L  T  V  D  H  M  R  K  L  K  Q  R  G  G  L  S  L  Q     77
541    GGAGTAGTGGCTGGTGTTGGCAGCCCACCCACCTCAACCAGTACCGCCCACGTGGAGTCCTTTCGCTCCGGCTATGTGCATGCTGCCGAT
        G  V  V  A  G  V  G  S  P  P  T  S  T  S  T  A  H  V  E  S  F  R  S  G  Y  V  H  A  A  D    107
631    CAGATTACCCAGGTCCTGCTGCAGACACAGCAAACCGATGAGATTGGGCGCAAGATAATGAAATTCCTATCGACGCGACTAATTGAGCTG
.       Q  I  T  Q  V  L  L  Q  T  Q  Q  T  D  E  I  G  R  K  I  M  K  F  L  S  T  R  L  I  E  L    137
721    CAGACGCAGTTGCTCCAGCAGCAGCAACAGCAACAACACCAGCAGCAGCAAATACCGCAATCATCGGGCCGCTTAGCCTTCCCGCTGCTG
        Q  T  Q  L  L  Q  Q  Q  Q  Q  Q  H  Q  Q  Q  Q  I  P  Q  S  S  G  R  L  A  F  P  L  L    167
811    GGAGGATATGGACCGAGCCGCCGTGCCGCCATTAGTTACAGCTCCTTCCTGACCAGCAAGGACGAACTGATCGATGTGACATCCGTCGAT
        G  G  Y  G  P  S  R  R  A  A  I  S  Y  S  S  F  L  T  S  K  D  E  L  I  D  V  T  S  V  D    197
901    GGCAACGCGTTATCCGAGACGGCGTCGGTTAGCTCGCAGGAGTCCGGCGCCTCTGAGCCCGTCTGGAGGCCCTGGTAATTGGGCGTGCCC
        G  N  A  L  S  E  T  A  S  V  S  S  Q  E  S  G  A  S  E  P  V  W  R  P  W                   222
991    AATATGGATCTTATCCCTAATATTATAATCGCAACCCAAAAAAATGGCTATTGACCCTTCTGCAACAAGATCCGTTGTCTTCCAAAGTAA
1081   ATCATATCAACCTAAGTTAATTGAGGTCTAATCGATTTCTGAATGTTTTGAAATACTTTTAGTTTTAGTTCCGATAATGGTAATTTGTAA
1171   ATTAATTGTTTAATTTCAACTGTGATATAACTAAAGTACGTTGATCAATCAATCAGTGAATATATATTCGAGTGCGCTATTTTCCTACTC
1261   CAAAAGTTCCCTAAGTTTAATATCTCGTTGCGCTCTCTTTATTGTTTAGCCTATATTAAGCTTTAAATAATTTATAAATGTTAATTTATG
1351   ATTAGTTGAGAAAGAATAATCAAAAATACACTACCCAAATAAAATAAATGATTTAAATTGAAAAAAAAAAAA
```

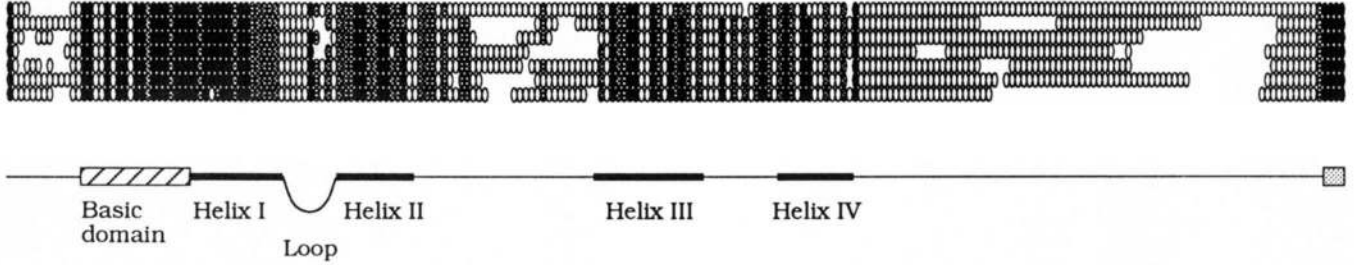
```
HLH-mβ
1      AAAAAAACAAAACCAAACTACAACAAAATGGTTCTGGAAATGGAGATGTCCAAGACCTATCAGTACCGCAAGGTGATGAAGCCCATGCTG
                                     M  V  L  E  M  E  M  S  K  T  Y  Q  Y  R  K  V  M  K  P  M  L     21
91     GAGCGTAAGCGCCGTGCCAGGATTAACAAGTGCCTGGACGAGCTCAAGGACATCATGGTGGAGTGCCTGACCCAGGAGGGCGAGCACATC
        E  R  K  R  R  A  R  I  N  K  C  L  D  E  L  K  D  I  M  V  E  C  L  T  Q  E  G  E  H  I     51
181    ACCCGGCTGGAGAAGGCCGACATCCTGGAGCTGACCGTGGAGCACATGAAGAAGCTGCGTGCCCAGAAGCAGCTGCGTCTGTCCAGCGTC
        T  R  L  E  K  A  D  I  L  E  L  T  V  E  H  M  K  K  L  R  A  Q  K  Q  L  R  L  S  S  V     81
271    ACCGGTGGAGTGTCTCCGAGTGCCGATCCCAAGCTTAGCATCGCTGAGAGTTTCCGTGCGGGCTACGTTCATGCTGCCAATGAAGTGTCC
        T  G  G  V  S  P  S  A  D  P  K  L  S  I  A  E  S  F  R  A  G  Y  V  H  A  A  N  E  V  S    111
361    AAAACCTTGGCCGCTGTACCCGGAGTGAGTGTGGATCTCGGCACCCAGCTGATGAGTCACCTTGGCCATCGTCTCAACTACCTGCAAGTG
        K  T  L  A  A  V  P  G  V  S  V  D  L  G  T  Q  L  M  S  H  L  G  H  R  L  N  Y  L  Q  V    141
451    GTGGTTCCCTCGCTGCCCATCGGAGTGCCTCTTCAAGCTCCAGTAGAGGATCAAGCTATGGTCACTCCACCACCCTCTGAATGCGGCAGC
        V  V  P  S  L  P  I  G  V  P  L  Q  A  P  V  E  D  Q  A  M  V  T  P  P  P  S  E  C  G  S    171
541    TTGGAGAGCGGAGCCTGCAGCCCGGCGCCCAGCGAGGCCAGCTCCACCTCTGGCCCCATGTGGCGTCCCTGGTGATCAACTGGACTCCGG
        L  E  S  G  A  C  S  P  A  P  S  E  A  S  S  T  S  G  P  M  W  R  P  W                      195
631    ATGATATTCACCTGAATTCGCCTTCAAGCGGAAGGCTTATA
```

```
HLH-mγ
1       AAGAAACACACAAAATGTCGTCGCTACAAATGTCCGAGATGTCCAAGACCTATCAGTACCGCAAGGTGATGAAGCCCATGCTGGAGAGG
                       M  S  S  L  Q  M  S  E  M  S  K  T  Y  Q  Y  R  K  V  M  K  P  M  L  E  R     25
90     AAGCGTCGTGCCAGAATCAACAAGTGCCTGGACGAGCTAAAGGATCTTATGGTTGCCACCTTGGAGTCCGAGGGCGAGCACGTCACCCGT
        K  R  R  A  R  I  N  K  C  L  D  E  L  K  D  L  M  V  A  T  L  E  S  E  G  E  H  V  T  R     55
180    TTGGAGAAAGCCGATATCCTGGAACTTACCGTCACCCATTTGCAAAAGATGAAGCAGCAGCGTCAGCACAAGCGGGCTTCCGGCGATGAG
        L  E  K  A  D  I  L  E  L  T  V  T  H  L  Q  K  M  K  Q  Q  R  Q  H  K  R  A  S  G  D  E     85
270    AGCTTGACTCCGGCGGAGGGCTTCCGTTCCGGCTACATCCATGCCGTCAATGAGGTCTCCCGTTCACTCTCCCAGCTGCCCGGCATGAAT
        S  L  T  P  A  E  G  F  R  S  G  Y  I  H  A  V  N  E  V  S  R  S  L  S  Q  L  P  G  M  N    115
360    GTCAGCCTAGGCACTCAGCTAATGACCCACCTGGGACAGCGCCTCAATCAGATCCAGCCAGCAGAAAAGGAAGTCTTGCCGGTGACCGCT
        V  S  L  G  T  Q  L  M  T  H  L  G  Q  R  L  N  Q  I  Q  P  A  E  K  E  V  L  P  V  T  A    145
450    CCACTCTCCGTGCACATCGCCAATCGGGATGCCTACTCCGTGCCCATTTCGCCAATCTCAACGTACGCCGGCAGTCCCAACTCCAATACC
        P  L  S  V  H  I  A  N  R  D  A  Y  S  V  P  I  S  P  I  S  T  Y  A  G  S  P  N  S  N  T    175
540    AGTTCGACCAGCCATTCCCTGTTGACCACCATCGATGTGACCAAGATGGAGGACGACAGCGAGGACGAGGAGAACGTCTGGCGTCCCTGG
        S  S  T  S  H  S  L  L  T  T  I  D  V  T  K  M  E  D  D  S  E  D  E  E  N  V  W  R  P  W    205
630    TAGATAGGATTACTTCACACCACAACTGAAAAAAAACGAAGGAGAGATCATAGAAAGTAAAAGTCTTCCGAGTCACTGGATTTCCACCAG
720    TCGACTCCCACCCTAGTTATAAGTAAACACCATAAACACACAAAGCTTATCTACTCCTGCGGCGTCAAGTACACTATTGCCTTAAG
```

```
HLH-mδ
1      GTATGTTTTAAAGTGACTACCGTGCAGTGCGGACGGGCACGAGCGACATCGCAAACATCGCAACTTTATTTACAAACAATCACCACACCA
91     AATCAAAACCCATTATACACAATGGCCGTTCAGGGTCAGAGATTTATGACCAAGACTCAGCATTACCGCAAGGTGACCAAACCCCTCCTG
                             M  A  V  Q  G  Q  R  F  M  T  K  T  Q  H  Y  R  K  V  T  K  P  L  L     23
181    GAGCGCAAAAGACGCGCCCGGATGAACCTGTATCTGGACGAACTCAAGGATCTCATCGTGGACACCATGGATGCGCAGGGCGAACAGGTC
        E  R  K  R  R  A  R  M  N  L  Y  L  D  E  L  K  D  L  I  V  D  T  M  D  A  Q  G  E  Q  V     53
271    AGCAAGCTGGAGAAAGCTGACATCTTAGAGCTCACCGTCAACTATCTGAAAGCGCAACAGCAGCAGCGAGTGGCCAATCCTCAATGGCCA
        S  K  L  E  K  A  D  I  L  E  L  T  V  N  Y  L  K  A  Q  Q  Q  Q  R  V  A  N  P  Q  W  P     83
361    CCACCCGACCAAGTTAACCTGGACAAATTTCGGGCCGGATACACCCAGGCTGCCTACGAGGTCTCGCACATATTCTCTACGGTTCCCGGC
        P  P  D  Q  V  N  L  D  K  F  R  A  G  Y  T  Q  A  A  Y  E  V  S  H  I  F  S  T  V  P  G    113
451    TTGGATCTCAAATTCGGCACCCACCTGATGAAGCAATTGGGTCACCAGCTCAAGGACATGAAACAGGAAGAAGAAATCATCGATATGGCC
        L  D  L  K  F  G  T  H  L  M  K  Q  L  G  H  Q  L  K  D  M  K  Q  E  E  E  I  I  D  M  A    143
541    GAAGAACCAGTTAATCTAGCCGACCAAAAACGTTCAAAGTCTCCGCGAGAAGAGGATATTCATCATGGCGAAGAAGTCTGGCGCCCCTGG
        E  E  P  V  N  L  A  D  Q  K  R  S  K  S  P  R  E  E  D  I  H  H  G  E  E  V  W  R  P  W    173
631    TGAAGGGCATTCATATATAACGCAACCAATTAACATTGATAGCTTTACATCTCAGTTTCATTAAGAAACAATCCAAGTTTTCCAAGGTCT
721    AGTGTAAACATTGCCTCAGAATCGTATTAGTTGGTAGTTGATTTGTATAGATGTGATGTTCCTGTGATAGTTCTTAAGTTCGTATTTTGT
811    ATTTGATCTCCCATAGGCTTACTGATTCATGTATTTAATTCATTGTAATTTATTTCATCAACTTTGCAAGAGCTTCCCCACCGAGCTTCC
901    ACTTAGCCTCATTAAGAACAAATTTTTGTGTATGTTTTAAAATTATATTTTTCGCTTTACACCATTAAGTATAAGCCATGAACATTTGAA
991    TAATAAAATGAACGTAAACTGAAAAAAA
```

FIGURE 3.—The nucleotide sequence of *HLH-m3*, *HLH-mβ*, *HLH-mγ* and *HLH-mδ* and the deduced amino acid sequences of the corresponding proteins. The sequence of *HLH-m3* was deduced from genomic DNA and one cDNA (*cE36.3*), extending from nucleotide 498 to 1411 and including a poly(A)-stretch. The conceptual translation of the longest open reading frame of 666 nucleotides indicates a protein of 222 amino acids with a calculated molecular mass of 24.9 kd. Although there are several in frame ATG codons at the start of the putative protein, the first fits the translation start consensus sequence best. Two canonical polyadenylation signals (AATAAA) (Birnstiel, Busslinger and STRUB 1985) are underlined. *HLH-mβ*: This is the sequence of the genomic DNA, for no cDNA could be isolated for *HLH-mβ*. Therefore, no information on the putative transcriptional start and termination sites is available. There is one open reading frame of 585 nucleotides, which is potentially coding (STADEN 1984) and encodes a protein of 195 amino acids with a calculated molecular mass of 21.4 kD. The conceptual translation starts at an ATG that fits the consensus sequence for Drosophila start sites (C/AAAA/CATG) (CAVENER 1987). *HLH-mγ*: The genomic DNA and three cDNAs were used for sequence analysis, the longest one (*cX4*), which consists of 805 nucleotides, does not cover the entire transcribed region, as the size of the transcript was shown to be 1.0 kb in length by northern blot analysis (see Figure 2C). cDNA *cE521* stretches from nucleotide 77 to 685 and the third, *cE132* extends from nucleotide 1 to 685. There is an open reading frame of 615 nucleotides with a high probability of being coding. The first ATG is likely to be used as translational start codon, as it fits the consensus sequence. The putative protein consists of 205 amino acids and has a calculated molecular mass of 23.2 kD. *HLH-mδ*: Three cDNAs and a genomic fragment of *HLH-mδ* were sequenced. These cDNAs consist of 1083 (*cE2*; 29–1111), 938 (*cA2*; 1–938) and 260 (*cE3*; 416–674) nucleotides, respectively, and thus seem to represent the entire transcript (1.1 kb according to Northern blot analysis; Figure 1C). There is a long open reading frame of 519 nucleotides, which is probably coding. The deduced protein of 173 amino acids has a calculated molecular mass of 20.2 kD. The translational start site nearly fits the consensus sequence. A canonical polyadenylation signal (AATAAA) is underlined. The EMBL Data Bank accession numbers for the four proteins are: X67046, X67047, X67048 and X67049, respectively.

A



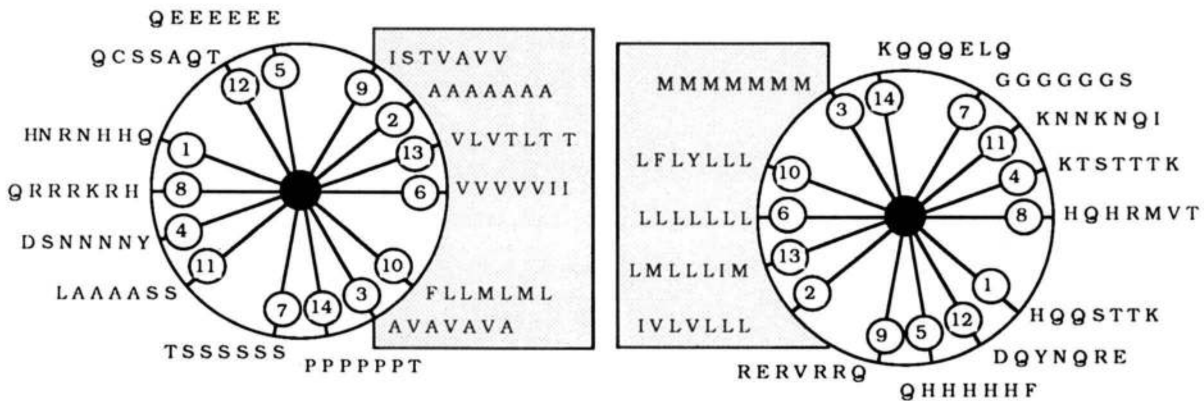
B

```
                  basic       helix I                 helix II

m3       ( 7-69)  KTYQYRKVMKPLLERKRRARINKCLDDLKDLMVECLQQEGEHVTRLEKADILELTVDHMRKLK
m5       (14-74)  KTQHYLKVKKPLLERQRRARMNKCLDTLKTLVAEFQGDDA--ILRMDKAEMLEAALVFMRKQV
m7       ( 9-69)  KTYQYRKVMKPLLERKRRARINKCLDELKDLMAECVAQTGD--AKFEKADILEVTVQHLRKLK
E(spl)   ( 6-66)  KTQIYQKVKKPMLERQRRARMNKCLDNLKTLVAELRGDDG--ILRMDKAEMLESAVIFMRQQK
mβ       ( 9-71)  KTYQYRKVMKPMLERKRRARINKCLDELKDIMVECLTQEGEHITRLEKADILELTVEHMKKLR
mγ       (11-73)  KTYQYRKVMKPMLERKRRARINKCLDELKDLMVATLESEGEHVTRLEKADILELTVTHLQKMK
mδ       (11-73)  KTQHYRKVTKPLLERKRRARMNLYLDELKDLIVDTMDAQGEQVSKLEKADILELTVNYLKAQQ
                  **  * ** **:***:****:*****:**:::::     :: : :::**::** ::  ::: :
consensus         KT  Y KV KP LER RRAR NKCLD LK                  KA  LE
                        :  **::*::****:* :* :**:::::          :::**::** ::  ::  :
hairy    (27-89)  PLKSDRRSNKPIMEKRRRARINNCLNELKTLILDATKKDPARHSKLEKADILEKTVKHLQELQ
```

C

```
                      helix III      helix IV

m3     (98-137) FRSGYVHAADQITQVLLQTQQTD-EIGRKIMKFLSTRLIEL
m5     (89-129) FKNGYMNAVSEISRVMACTPAMSVDVGKTVMTHLGVEFQRM
m7     (83-123) FRAGYIRAANEVSRALASLPRVDVAFGTTLMTHLGMRLNQL
E(spl) (83-123) FKNGYMNAVNEVSRVMASTPGMSVDLGKSVMTHLGRVYKNL
mβ     (99-139) FRAGYVHAANEVSKTLAAVPGVSVDLGTQLMSHLGHRLNYL
mγ     (93-133) FRSGYIHAVNEVSRSLSQLPGMNVSLGTQLMTHLGQRLNQI
mδ     (93-133) FRAGYTQAAYEVSHIFSTVPGLDLKFGTHLMKQLGHQLKDM
                *: **: *: *:*: :   * ::: :*: :*::**  :  :
consensus       F  GY  A  E S      P      G   M  LG
                *: *::  : *:*:     *         :: :*   :  :
hairy (107-143) FKAGFADCVNEVSRFPGIEPAQR----RRLLQHLSNCINGV
```

D



heterozygosity with $gro^{b32.2}$. These results demonstrate that the three genes contribute to the neural-epidermal lineage decision during early neurogenesis, and that their absence is responsible for the stronger neurogenic phenotype of $gro^{b32.2}$ as compared with that of $E(spl)^{R\text{-}A7.1}$ (Figure 1).

Homozygous $gro^{b32.2}$ or $E(spl)^{R\text{-}A7.1}$ recombinants carrying only the *HLH-mδ*$^+$ fragment do not show any modification of their neurogenic phenotypes. However, the *HLH-mδ*$^+$ fragment together with two copies of a gro^+ transgene (SCHRONS, KNUST and CAMPOS-ORTEGA 1992) achieves a detectable reduction of the $E(spl)^{R\text{-}A7.1}$ phenotype (to the same extent as the phenotype shown in Figure 1G). It is remarkable that, when present separately, two copies of a gro^+ or two copies of the *HLH-mδ*$^+$ transgene do not seem to modify the $E(spl)^{R\text{-}A7.1}$ phenotype (not shown). Finally, homozygous $E(spl)^{R\text{-}A7.1}$ recombinants bearing the *HLH-[mβ-mγ-mδ]*$^+$ fragment show a reduction in their neurogenic phenotypes (Figure 1, H and I). The deficiency $E(spl)^{R\text{-}A7.1}$ itself does not affect *HLH-mβ*, *HLH-mγ* or *HLH-mδ*, that is to say, these latter embryos have a total of four copies of the wild-type alleles of the three genes.

With respect to the participation of *HLH-m3* in early neurogenesis, the reader is referred to SCHRONS, KNUST and CAMPOS-ORTEGA (1992).

Transcripts from *m4*, *HLH-m5*, *HLH-m7*, *E(spl)*, *HLH-mβ*, *HLH-mγ* and *HLH-mδ* are restricted to epidermoblasts: Digoxigenin-labeled probes have been used to study the distribution of RNAs of *HLH-mβ*, *HLH-mγ* and *HLH-mδ* in embryonic whole mounts. *HLH-m3* transcripts have previously been shown to be distributed ubiquitously during embryogenesis (KNUST, TIETZE and CAMPOS-ORTEGA 1987); this has been confirmed by digoxigenin-labeled probes (not shown). For the purposes of comparison with the observations on *HLH-mβ*, *HLH-mγ* and *HLH-mδ*, we also studied using digoxigenin-labeled probes the distribution of transcription unit *m4*, and of *HLH-m5*, *HLH-m7* and *E(spl)* RNA, previously described on the basis of *in situ* hybridization of radioactively labeled probes to tissue sections (KNUST, TIETZE and CAMPOS-ORTEGA 1987).

Whereas transcripts of *m4*, *HLH-m5*, *HLH-m7* and *E(spl)* exhibit virtually identical spatial distributions during early embryonic stages up to stage 9 (KNUST, TIETZE and CAMPOS-ORTEGA 1987) (Figure 5, A–D) (staging follows CAMPOS-ORTEGA and HARTENSTEIN 1985), transcripts of *HLH-mβ*, *HLH-mγ* and *HLH-mδ* are not detected until stage 7. Transcription of these latter genes at stage 7 follows the same spatial pattern as *m4*, *HLH-m5*, *HLH-m7* and *E(spl)*. When the tip of the elongating germ band has reached about 20% egg length, the domain of distribution of RNA from each of the seven genes expands from the midline laterally to invade the neuroectoderm (Figure 5D), in such a way that, as germ band elongation proceeds, most of the cells of the neuroectoderm begin transcription of all six genes of the E(spl)-C and of *m4* (Figure 5, D and E). It is remarkable that during these initial stages of germ band elongation, RNA from the six E(spl)-C genes and *m4* is found exclusively within the neuroectoderm, both in the territory of the trunk and in cephalic regions (Figures 6 and 7). The distribution of these seven RNAs within the neuroectoderm changes rapidly during the period of fast germ band elongation (Figure 5, D and E). Immediately preceding and following SI neuroblast segregation, RNA is present in two longitudinal stripes on each side of the ventral midline, one paramedial and the other lateral, each about three to four cells wide, and connected to each other by single RNA containing cells (Figures 5D, 6A and 7, A–F). The region between the two stripes which is almost devoid of transcripts during this stage, coincides with the zone of mitotic activity

FIGURE 4.—**A**, Schematic representation of the bHLH proteins of the E(spl)-C, aligning, from top to bottom, HLH-m3 (this work), HLH-m5, HLH-m7, E(spl) (taken from KLÄMBT *et al.* 1989), HLH-mβ, HLH-mγ, HLH-mδ (this work), using the program CLUSTAL (HIGGINS and SHARP 1988). Identical amino acids are represented as black ovals, conserved amino acids as grey ovals and divergent amino acids as open ovals. Putative protein domains are indicated below. Note the high degree of conservation in the basic and HLH domains and in two additional regions (helix III and IV). All seven proteins terminate with the same tetrapeptide, W-R-P-W. **B** and **C** show amino acid sequence similarities of different domains of HLH-m3, HLH-m5, HLH-m7, E(spl), HLH-mβ, HLH-mγ, HLH-mδ and hairy [taken from KLÄMBT *et al.* (1989), RUSHLOW *et al.* (1989), and this work]. Identical amino acids (asterisks) and conserved amino acids (colons) are indicated as such whenever six or seven amino acids at a given position are identical or conserved, respectively. The consensus sequences are derived from the seven proteins of the E(spl)-C. **B**, Amino acid sequence similarities in the basic and the HLH domains. The degree of conservation is highest in the basic region and in the two amphipathic helices. Note particularly the proline residue in the basic domains of all seven proteins, which is unique among the members of the bHLH-protein familiy, but also present in the hairy protein. **C**, Helix III/IV. The alignment shows a region of 41 amino acids present in all seven bHLH proteins of the E(spl)-C. The consensus sequence derived from the seven proteins is indicated. Within the first 25 amino acids, hydrophobic amino acids are arranged in a three-four repeat, which is reminiscent of α-helical segments, *e.g.*, those described for lamins or in the amphipatic helices of the bHLH proteins [see MURRE, SCHONLEBER MCCAW and BALTIMORE (1989) for discussion]. Within the region shown, two putative amphipathic helices, each of 14 amino acids, can be formed, with hydrophobic amino acids on one side and charged residues on the other side. This becomes evident when the amino acids are arranged on a helical wheel, aligning the amino acids of HLH-m3, HLH-m5, HLH-m7, E(SPL), HLH-mβ, HLH-mγ and HLH-mδ, from distal to central (**D**). Despite the sequence similarity between the HLH proteins of the E(spl)-C and part of hairy (amino acids 107–143), including conservation of the positions of some of the hydrophobic residues, the putative α-helix III is not as obvious in the hairy protein, whereas the putative helix IV seems to be conserved.

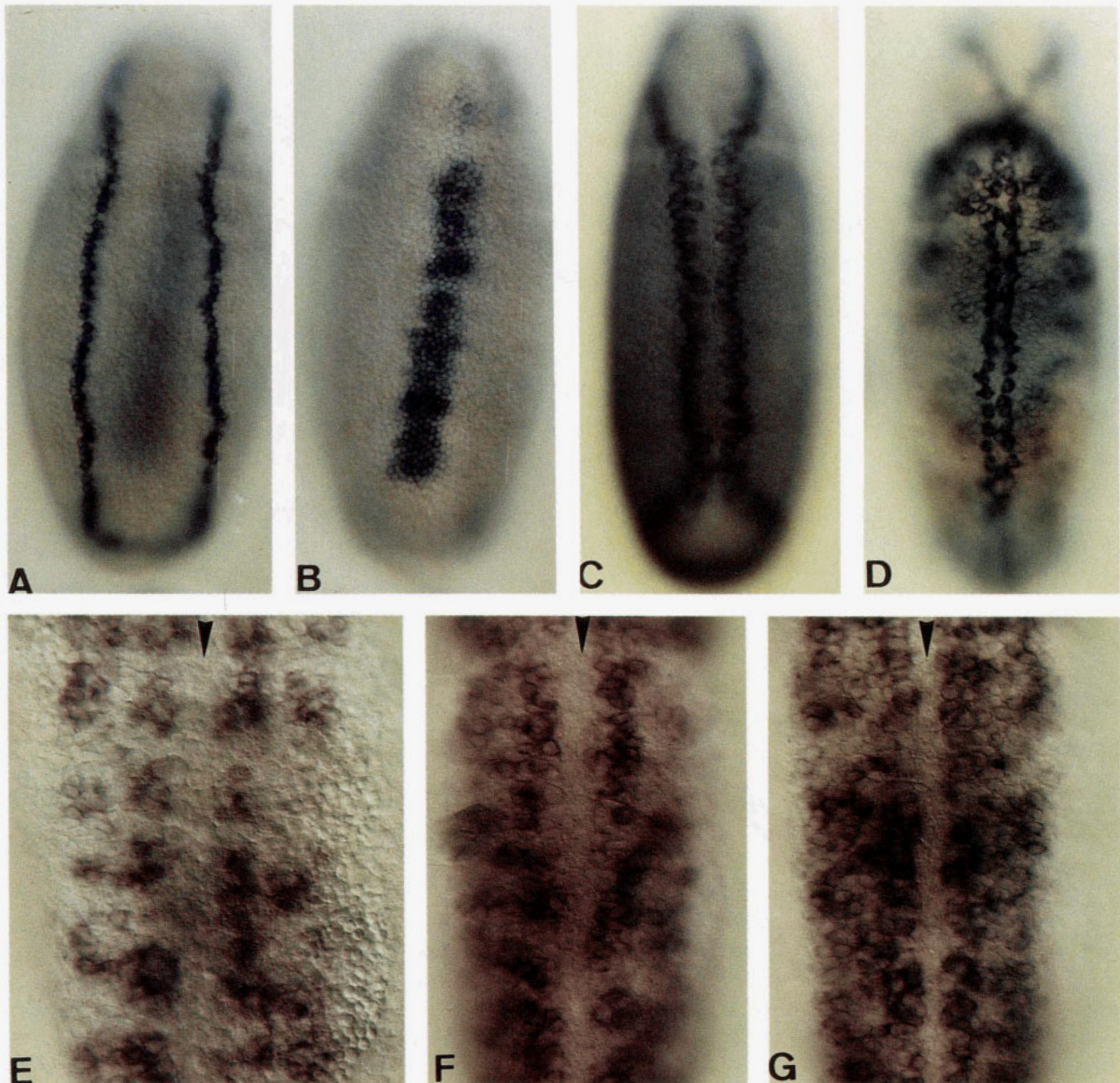


FIGURE 5.—**A–D** show *in situ* hybridizations with a digoxigenin-labeled *HLH-m5* probe to embryonic whole-mounts. **A** and **B** show ventral and dorsal focal planes, respectively, of the same stage 5 embryo, **C** and **D** are ventral focus planes of two different, stage 7 and 8 embryos (CAMPOS-ORTEGA and HARTENSTEIN 1985). From stage 5 to 7, transcripts from *m4*, *HLH-m5*, *HLH-m7* and *E(spl)* exhibit a virtually identical spatial distributions. All accumulate in rows one cell wide on either side of the mesodermal anlage. In addition to these rows, transcripts from *HLH-m5*, *HLH-m7* and *E(spl)* are expressed during stages 5 and 6 in a dorsomedial band two to three cells in width (**B**) which precisely coincides with the anlage of the amnioserosa (HARTENSTEIN, TECHNAU and CAMPOS-ORTEGA 1985); RNA from *m4* is also present in these same cells, but at a much lower concentration (not shown). During closure of the ventral furrow, the two ventrolateral rows meet at the midline and expression persists for some time (**C** and **D**). *HLH-mβ*, *HLH-mγ* and *HLH-mδ* are not expressed in any detectable pattern during these early developmental stages. During elongation of the germ band, RNA from all seven genes accumulates exclusively in cells of the neuroectoderm in similar, if not identical, patterns. **E** and **F** show, for example, expression of *E(spl)* in part of the germ band in three stage 8 embryos. The distribution is irregular during stage 7 and early stage 8, but in late stage 8 (**E**) two clusters of labeled cells per hemisegment are detected. This arrangement develops further (**F**) and within a few minutes evolves into a distinct striped pattern at the transition between stages 8 and 9, when RNA from all seven genes accumulates in a ladder-like arrangement, in two continuous cell rows on either side of the midline (arrowhead), one paramedial and the other lateral, with a few labeled cells in between (**G**).

that has been found in the intermediate subdivision of the neuroectoderm immediately before neuroblast segregation (HARTENSTEIN and CAMPOS-ORTEGA 1984; FOE 1989). The anteroposterior extents of the lateral and medial stripes differ. Caudally, at the level of parasegment 14, the lateral stripe extends two- to three-cell diameters further than the medial stripe (Figure 7, B, D and F), whereas orally, at the level of parasegment 1, the medial stripe extends two-cell diameters further into the cephalic furrow than the lateral stripe. Whereas epidermoblasts exhibit high levels of RNA derived from these genes, the subjacent SI neuroblasts are clearly devoid of these molecules (Figure 6, E and F), with the exception of E(spl): a

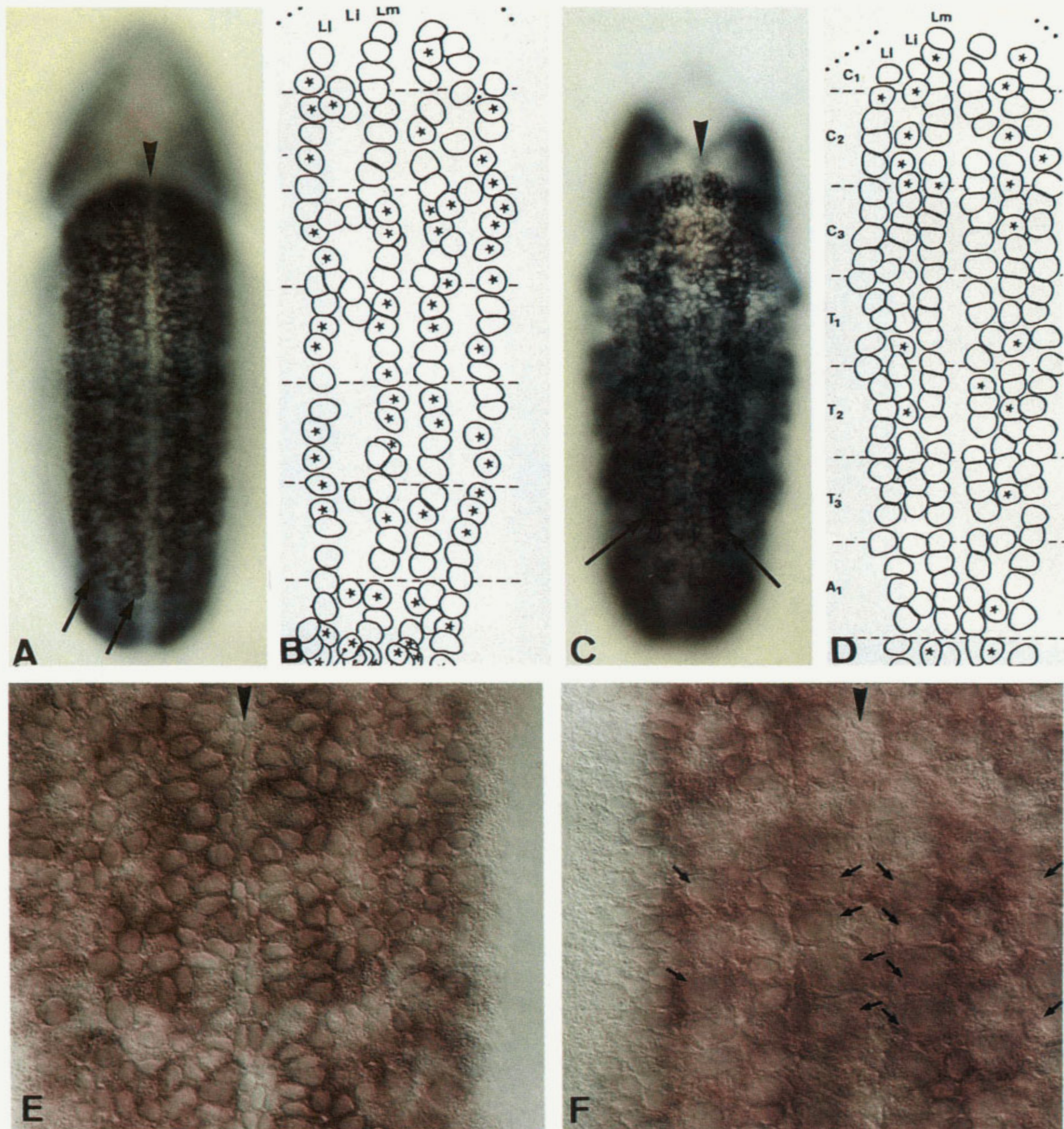


FIGURE 6.—**A** and **C** show *in situ* hybridizations with an *HLH-m5* probe to stage 9 and stage 10 embryos, **B** and **D** [taken from HARTENSTEIN and CAMPOS-ORTEGA (1984) with permission of the editor] are schematic drawings of the arrangement of SI and SII neuroblasts in embryos of the same ages. In stage 9 embryos, transcripts from *HLH-mβ*, *HLH-mγ*, *HLH-mδ*, *HLH-m5*, *HLH-m7* and *E(spl)*, as well as *m4*, are arranged in a ladder-like manner, in two rows of cells on either side connected by stripes of labeled cells (the arrows point to the two rows, the arrowhead to the midline). SI neuroblasts are located immediately subjacent to these rows and the connecting stripes, suggesting that they have segregated from the region in which labeled cells occur. Compare to the pattern of SI neuroblasts shown in **B**. During stages 9 and 10, RNA from the seven genes becomes diffusely distributed in all cells of the neuroectoderm (**C**), as well as in some cells of the dorsal epidermal primordium. At the beginning of stage 10, transcripts from the seven genes are particularly abundant in a row about three cells wide located on either side in an intermediate position with respect to the ladder-like arrangement of the previous stage (arrows, the arrowhead points to the midline). SII neuroblasts segregate from this region during stage 10 to complete the intermediate row of neuroblasts (HARTENSTEIN and CAMPOS-ORTEGA 1984) (compare the neuroblast pattern at this stage, shown in **D**). Transcripts from the seven genes become restricted to epidermoblasts. **E** and **F** show two planes of focus through part of the germ band of a stage 9 embryo hybridized with *HLH-m5*, to show labeling in the epidermoblasts but not in SI neuroblasts (several are indicated by small arrows). The arrowhead points to the midline.

small amount of RNA from this gene can occasionally, *i.e.*, in some embryos, be detected in the neuroblasts.

We wish to emphasize that, prior to the lineage segregation, the array of RNA-containing cells correlates remarkably well with the zone from which the SI neuroblasts segregate and that, after segregation, the transcripts become restricted to epidermoblasts (compare the distribution of transcripts in the neu-

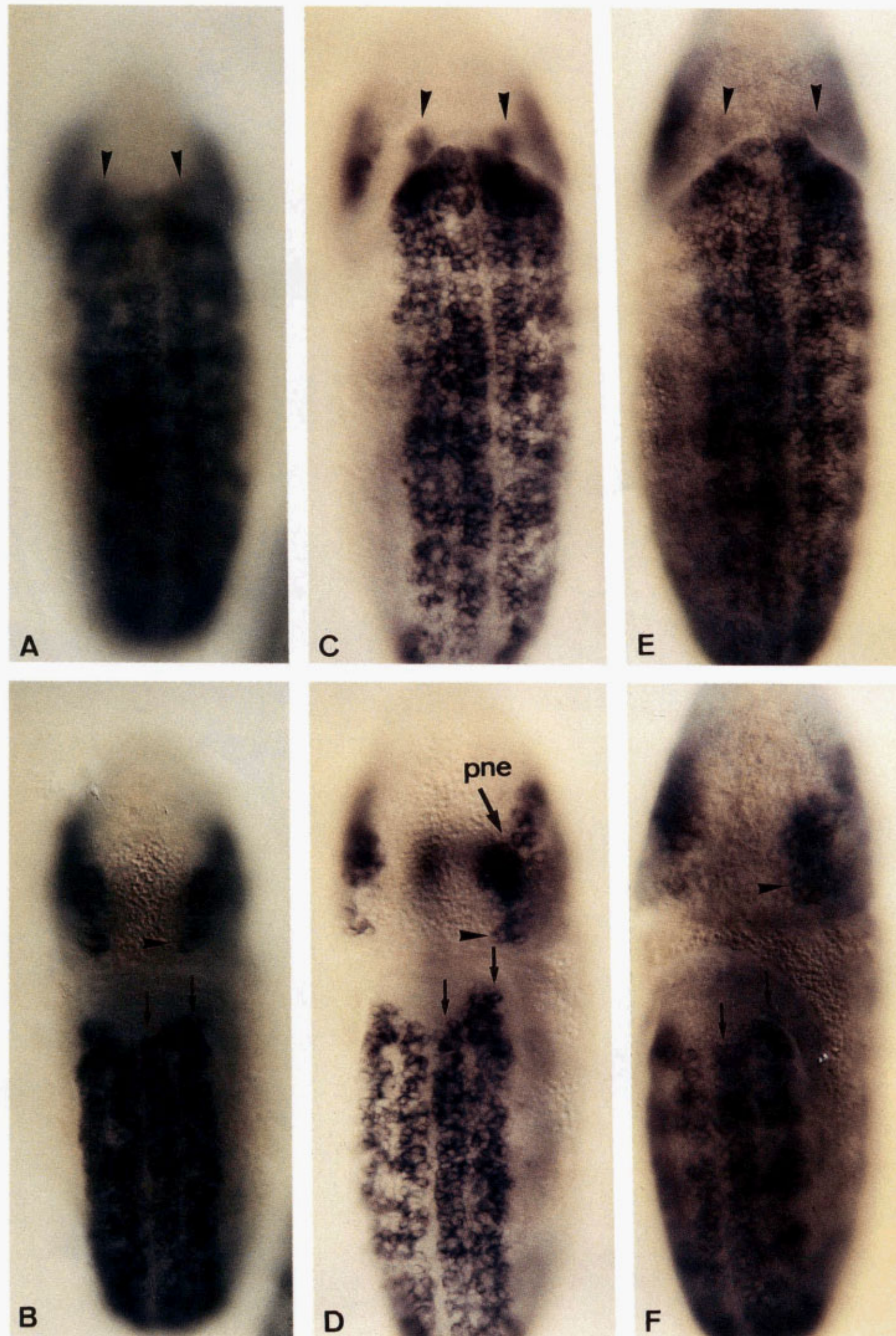


FIGURE 7.—Ventral (**A**, **C** and **E**) and dorsal (**B**, **D** and **F**) planes of focusing through stage 9 embryos which have been hybridized *in situ* with probes of *HLH-mβ* (**A-B**), *HLH-mγ* (**C-D**) and *HLH-mδ* (**E-F**). Notice that the three genes are transcribed in the procephalic (pne) and the ventral neuroectoderm, and that the distribution of RNA is very similar, or even identical, for all three. The arrowheads (two in **A**, **C** and **E**, and one in **B**, **D** and **E**) point to groups of cells which contain RNA from the three genes; the two arrows point to the longitudinal rows of ventral neuroectodermal cells from which SI neuroblasts have segregated (see Figure 5).

roectoderm with the map of SI neuroblasts, Figure 6, A–B) (see HARTENSTEIN and CAMPOS-ORTEGA 1984). This arrangement suggests a causal relationship between the presence of transcripts from these genes in a given cell within these regions and its development as a epidermoblast. Unfortunately, the high density of transcripts does not allow us to distinguish on whether the cells that become committed as neuroblasts lack transcripts prior to segregation, or whether they first stop transcribing these genes after having segregated.

During stage 9, the distribution pattern of RNA

from all seven genes changes considerably. Labeling within the procephalic lobe becomes diffusely distributed and nearly reaches the cephalic furrow (not shown). The stripe pattern found previously in the trunk is no longer visible and RNA is found in all cells of the neuroectoderm, as well as in some cells of the dorsal epidermal primordium. A further striking feature of the pattern of RNA distribution is detectable at the beginning of stage 10, at the time of segregation of the SII neuroblasts (Figure 6C). At this stage, although the entire neuroectodermal region contains RNA derived from these genes, transcripts are particularly abundant in segmentally arranged cell patches that are located along the entire germ band in an intermediate position with respect to the banded arrangement seen in the previous stage. SII neuroblasts segregate during stage 10 from this intermediate region to complete the intermediate row of neuroblasts; after segregation, SII neuroblasts are devoid of RNA from any of the genes, whereas the epidermoblasts that remain in the intermediate region contain high levels of these transcripts (Figure 6C).

We should mention that differences in the distribution of the seven transcripts become apparent during stage 11 and persist until stage 15, when transcription vanishes. These differences have not been characterized in detail.

DISCUSSION

In this report, we describe three new genes, *HLH-mβ*, *HLH-mγ* and *HLH-mδ*, which are involved in mediating the neuroepidermal lineage dichotomy. This claim is supported by the observation that a genomic fragment containing the coding sequences of the three genes reduces the neurogenic phenotype of $gro^{b32.2}$ embryos to a level comparable to that of $E(spl)^{R\text{-}A7.1}$ embryos and virtually identical to that of embryos heterozygous for $gro^{b32.2}$ with $E(spl)^{R\text{-}A7.1}$. Hence, the more severe neurogenic phenotype found normally in embryos homozygous for the former deletion is due to the lack of *HLH-mβ*, *HLH-mγ* and *HLH-mδ*. This observation also indicates that the transgenic $HLH\text{-}[m\beta\text{-}m\gamma\text{-}m\delta]^+$ fragment contains indeed three functional genes which are capable of providing a comparable level of genetic activity to that of the endogenous *HLH-mβ*, *HLH-mγ* and *HLH-mδ* genes. Two copies of the transgenic $HLH\text{-}[m\beta\text{-}m\gamma\text{-}m\delta]^+$ fragment cause an even stronger reduction of the neurogenic phenotype of $gro^{b32.2}$ embryos. This phenotype is weaker than that produced by homozygosity for $E(spl)^{R\text{-}A7.1}$, which is due to the fact that *gro* is not completely impaired in $gro^{b32.2}$, whereas it is absent in $E(spl)^{R\text{-}A7.1}$ (SCHRONS, KNUST and CAMPOS-ORTEGA, 1992). In addition, either two copies of the $HLH\text{-}m\delta^+$ fragment in the presence of two transgenic copies of gro^+, or two copies of the $HLH\text{-}[m\beta\text{-}m\gamma\text{-}m\delta]^+$ fragment alone, cause a similar attenuation of the phenotype of $E(spl)^{R\text{-}A7.1}$ in both cases. Since $E(spl)^{R\text{-}A7.1}$ is itself $HLH\text{-}m\beta^+$, $HLH\text{-}m\gamma^+$ and $HLH\text{-}m\delta^+$ (KNUST, TIETZE and CAMPOS-ORTEGA 1987), the transgenic animals in fact have four copies of either $HLH\text{-}m\delta^+$, in the first case, or of $HLH\text{-}m\beta^+$, $HLH\text{-}m\gamma^+$ and $HLH\text{-}m\delta^+$, in the second. Therefore, these observations suggest that the number of genes encoding HLH proteins, and therefore the amount of these protein molecules, is important. This points to functional redundancy of these genes.

The main conclusion of our work is that the E(spl)-C comprises at least seven genes encoding bHLH proteins. The proteins encoded by *HLH-m3*, *HLH-mβ*, *HLH-mγ* and *HLH-mδ*, described in this paper, exhibit a high degree of amino acid sequence similarity to each other as well as to the proteins encoded by *HLH-m5*, *HLH-m7* and *E(spl)* (KLÄMBT *et al.* 1989). The amino acid sequences of the seven putative proteins are particularly conserved within the basic domain and the two amphipathic helices, a motif which has been described to be necessary for DNA binding and protein dimerization (MURRE, SCHONLEBER MCCAW and BALTIMORE 1989; MURRE *et al.* 1989; DAVIS *et al.* 1990). The seven proteins of the E(spl)-C are unusual, in that they contain a proline residue in the basic domain, which is thought to interfere with DNA binding. The MyoD protein of the mouse, for example, acts as dominant negative regulator after introduction of a proline residue into its basic domain (DAVIS *et al.* 1990). In the case of the E(spl) protein, however, *in vitro* as well as *in vivo* experiments have demonstrated that, in spite of the presence of the proline residue, the basic domain is required for DNA binding and for enhancement of the *split* phenotype (TIETZE, OELLERS and KNUST 1992). In addition, all seven proteins terminate with the tetrapeptide W-R-P-W. Finally, we found that a region of 41 amino acids within the C-terminal halves of the proteins exhibits a high degree of sequence conservation among all seven (Figures 3 and 4B). The regular spacing of hydrophobic residues in a four-three repeat suggests the presence of a coiled-coil structure, similar to the structure proposed for lamins (MCKEON, KIRSCHNER and CAPUT 1986) or leucine zippers (LANDSCHULZ, JOHNSON and MCKNIGHT 1988). Helical wheel analysis of this region predicts an α-helical structure, which can be organized as two amphipathic helices (see Figure 4D). One side of these hypothetical helices is composed predominantly of hydrophobic amino acids, while the other side consists mainly of charged amino acids or residues with uncharged polar chains. Other members of the bHLH family contain, in addition to the bHLH motif, one or two leucine zippers (*e.g.*, *myc* and AP-4) [see LÜSCHER and EISENMANN (1990) and HU *et al.* (1990) for reviews]. In the

AP-4 protein, only one of the leucine zipper motifs has the three-four repeat of hydrophobic amino acids characteristic for a coiled-coil structure. The two leucine zippers are necessary for homodimerization of the protein and their elimination allows the formation of heterodimers with other HLH proteins, which otherwise does not take place, but which does not interfere with DNA binding to the specific target site (HU *et al.* 1990). Although the motif found in the bHLH proteins of the E(spl)-C does not fulfill the criteria for a leucine zipper (LANDSCHULZ, JOHNSON and MCKNIGHT 1988), its conservation in all seven proteins and its putative coil-coiled structure is suggestive of involvement in selective protein-protein interactions; these interactions would be different from those mediated by the HLH motif, and could involve either the same partner, thus stabilizing homodimers, as described for AP-4, or a variety of other proteins, and thus increasing the potential regulatory complexity of these gene products.

Among the bHLH proteins described so far, the seven proteins of the E(spl)-C are most similar to the hairy protein of *D. melanogaster* (RUSHLOW *et al.* 1989). Besides the great similarity in the basic domain, including the proline residue mentioned above, other domains are also conserved, such as the HLH domain and the C-terminal tetrapeptide W-R-P-W (Figure 4B). Although there is some amino acid sequence conservation in the region of the hypothetical helix III/IV (Figure 4C), the ability to form an α-helix is only apparent in helix IV of the hairy protein, and less obvious in helix III. *hairy* is involved in the regulation of segmentation during embryogenesis and the establishment of bristle pattern formation during larval and pupal development (INGHAM *et al.* 1985). Whether the structural similarity between its product and the seven proteins described here has any functional implications, awaits further experiments.

Among the seven proteins of the E(spl)-C, HLH-m3 slightly differs from the others. With respect to the sequence, it carries several nonconserved amino acids at positions, where the others are perfectly matched. Furthermore, it contains a polyglutamine stretch distally to helix IV, comparable to the hairy protein, which might be involved in transcriptional activation. The most dramatic difference, however, concerns its expression pattern. In contrast to the other six, *HLH-m3* RNA is supplied maternally, and later on, it is expressed ubiquitously during all stages of embryonic development, whereas the other six, as well as *m4*, exhibit a distinct and very similar expression pattern, at the time when separation of neural and epidermal precursor cells occurs. Although some minor differences exist, as for example in the mesectodermal cell stripes, which express only *HLH-m5*, *HLH-m7* and *E(spl)* RNAs (as well as *m4*), but not *HLH-mβ*, *HLH-mγ* or *HLH-mδ*, the patterns of *HLH-mβ*, *HLH-mγ*, *HLH-mδ*, *HLH-m5*, *HLH-m7*, *E(spl)* and *m4*, but not that of *HLH-m3*, are indistinguishable from each other during the stages of SI and SII neuroblast segregation (Figures 6 and 7). All seven transcripts accumulate within the regions from which the SI and SII neuroblasts segregate and correlate with epidermoblast fate. We assume that the same correlation holds true for the SIII neuroblasts; however, the SIII segregation pattern is not as obvious as that of the other two subpopulations and, consequently, the correlation cannot be established as easily. The presence of these transcripts in cells which have chosen the epidermal pathway and their absence from the neuroblasts, supports a function in the commitment of epidermal cells, as had been deduced from the analysis of the phenotype of loss-of-function mutations (LEHMANN *et al.* 1983; KNUST *et al.* 1987a). We have mentioned above that, apart from their presence in epidermoblasts, *E(spl)* transcripts have occasionally been detected in the neuroblasts after separation of the lineages. This could be due either to higher stability of the *E(spl)* transcripts, as compared with the other transcripts of the complex, which would allow them to be detected in the neuroblasts for some time after segregation, or to other unknown reasons.

In *D. melanogaster*, a few gene complexes comprising several structurally and functionally related genes are known. The best known are two gene complexes comprising multiple homeobox-containing genes, the *Antennapedia* complex (Antp-C) and the *Bithorax* complex (BX-C) [see PEIFER, KARCH and BENDER (1987) and SCOTT and CARROLL (1987) for reviews]. Genes within these two complexes control the segmental identity during embryogenesis and later development. Since mutations in individual genes result in homeotic transformation of part of the body into another part, each of the genes has a specific function. The E(spl)-C differs from the BX-C and Antp-C in that at least two of the genes encoding HLH proteins [*E(spl)* itself and *HLH-m7*] are dispensable for the viability of the fly and the loss of their function does not cause any obvious phenotype (SCHRONS, KNUST and CAMPOS-ORTEGA 1992).

Another well known gene complex, the AS-C, contains at least four genes, which are involved in the development of both CNS and PNS (JIMÉNEZ and CAMPOS-ORTEGA 1979, 1987; DAMBLY-CHAUDIÈRE and GHYSEN 1987; GHYSEN and DAMBLY-CHAUDIÈRE 1988) and also encode HLH proteins, each of which has a characteristic expression pattern (VILLARES and CABRERA 1987; ALONSO and CABRERA 1988; GONZÀLEZ *et al.* 1989). The AS-C exhibits a similar genetic organization and its constituent genes are to some extent functionally redundant (JIMÉNEZ and CAMPOS-ORTEGA 1979). It has been shown that the AS-C

products are necessary for the development of neuroectodermal cells as neuroblasts (JIMÉNEZ and CAMPOS-ORTEGA 1990). Reciprocal regulation of the activity of the genes of the E(spl)-C and the AS-C, either at the transcriptional or at post-transcriptional level, is an attractive possibility (CAMPOS-ORTEGA 1991).

We would like to thank LARRY ZIPURSKY for the gift of cDNA clones of *HLH-mγ* and *HLH-mδ*, PAUL HARDY for constructive criticisms on the manuscript, and CARIN THERES and EVA THEILENBERG for expert technical assistance. The research reported here was supported by grants from the Deutsche Forschungsgemeinschaft (DFG, SFB 243).

LITERATURE CITED

ALONSO, M. C., and C. V. CABRERA, 1988 The *achaete-scute* gene complex of *Drosophila melanogaster* comprises four homologous genes. EMBO J. **7:** 2585–2591.

AUFFRAY, C., and F. ROUGEON, 1980 Purification of mouse immunoglobulin heavy chain messenger RNAs from total myeloma tumor RNA. Eur. J. Biochem. **107:** 303–314.

BIRNSTIEL, M. L., M. BUSSLINGER and K. STRUB, 1985 Transcription termination and 3′ processing: the end is in site! Cell **41:** 349–359.

CAMPOS-ORTEGA, J. A., 1991 Mechanisms of early neurogenesis in *Drosophila melanogaster*. Comments Dev. Neurobiol. **1:** 251–266.

CAMPOS-ORTEGA, J. A., and V. HARTENSTEIN, 1985 *The Embryonic Development of Drosophila melanogaster*. Springer Verlag, New York.

CAUDY, M., H. VÄSSIN, M. BRAND, R. TUMA, L. Y. JAN and Y. N. JAN, 1988 *daughterless*, a gene essential for both neurogenesis and sex determination in *Drosophila*, has sequence similarities to *myc* and the *achaete-scute* complex. Cell **55:** 1061–1067

CAVENER, D., 1987 Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. Nucleic Acids Res. **15:** 1353–1361.

DAMBLY-CHAUDIÈRE, C., and A. GHYSEN, 1987 Independent subpatterns of sense organs require independent genes of the *achaete-scute* complex in *Drosophila* larvae. Genes Dev. **1:** 297–306.

DAVIS, R. L., P. F. CHENG, A. B. LASSAR and H. WEINTRAUB, 1990 The MyoD DNA binding domain contains a recognition code for muscle specific gene activation. Cell **60:** 733–746

DELIDAKIS, C., A. PREISS, D. HARTLEY and S. ARTAVANIS-TSAKONAS, 1991 Two genetically distinct functions reside within the Enhancer of split locus of of *Drosophila melanogaster*. Genetics **129:** 803–823.

FOE, V. E., 1989 Mitotic domains reveal early commitment of cells in *Drosophila* embryos. Development **107:** 1–22.

GHYSEN, A., and C. DAMBLY-CHAUDIÈRE, 1988 From DNA to form: the achaete-scute complex. Genes Dev. **2:** 495–501.

GONZÀLEZ, F., S. ROMANI, P. CUBAS, J. MODOLELL and S. CAMPUZANO, 1989 Molecular analysis of asense, a member of the *achaete-scute* complex of *Drosophila melanogaster*, and its novel role in optic lobe development. EMBO J. **8:** 3553–3562.

HART, A. C., H. KRÄMER, D. L. VAN VACTOR, M. PAIDHUNGAT and S. L. ZIPURSKY, 1990 Induction of cell fate in the *Drosophila* retina: the bride of sevenless protein is predicted to contain a large extracellular domain and seven transmembrane segments. Genes Dev. **4:** 1835–1847

HARTENSTEIN, V., and J. A. CAMPOS-ORTEGA, 1984 Early neurogenesis in wildtype *Drosophila melanogaster*. Roux's Arch. Dev. Biol. **193:** 308–325.

HARTENSTEIN, V., G. M TECHNAU and J. A. CAMPOS-ORTEGA, 1985 Fate-mapping in wild-type *Drosophila melanogaster*. III. A fate map of the blastoderm. Roux's Arch. Dev. Biol. **194:** 213–216.

HARTLEY, D. A., A. PREISS and S. ARTAVANIS-TSAKONAS, 1988 A deduced gene product from the Drosophila neurogenic locus *Enhancer of split* shows homology to mammalian G-protein *β* subunit. Cell **55:** 785–795.

HIGGINS, D. G., and P. M. SHARP, 1988 CLUSTAL; a package for performing multiple sequence alignment on a microcomputer. Gene **73:** 237–244.

HU, Y. F., B. LASCHER, A. ADMON, N. MERMOD and R. TJIAN, 1990 Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. Genes Dev. **4:** 1741–1752.

INGHAM, P. W., S. M. PINCHIN, K. R. HOWARD and D. ISH-HOROWICZ, 1985 Genetic analysis of the hairy locus in *Drososphila melanogaster*. Genetics **111:** 463–486.

JIMÉNEZ, F., and J. A. CAMPOS-ORTEGA, 1979 A region of the *Drosophila* genome necessary for CNS development. Nature **282:** 310–312.

JIMÉNEZ, F., and J. A. CAMPOS-ORTEGA, 1987 Genes in subdivision 1B of the *Drosophila melanogaster* X-chromosome and their influence on neural development. J. Neurogenet. **4:** 179–200.

JIMÉNEZ, F., and J. A. CAMPOS-ORTEGA, 1990 Defective neuroblast commitment in mutants of the achaete-scute complex and adjacent genes of *Drosophila melanogaster*. Neuron **5:** 81–89.

KLÄMBT, C., E. KNUST, K. TIETZE and J. A. CAMPOS-ORTEGA, 1989 Closely related transcripts encoded by the neurogenic gene complex *Enhancer of split* of *Drosophila melanogaster*. EMBO J. **8:** 203–210.

KLEMENZ, R., U. WEBER and W. J. GEHRING, 1987 The *white* gene as marker in a new P-element vector for gene trasfer in *Drosphila*. Nucleic Acids Res. **15:** 3947–3959.

KNUST, E., K. TIETZE and J. A. CAMPOS-ORTEGA, 1987 Molecular analysis of the neurogenic locus Enhancer of split of *Drosophila melanogaster*. EMBO J. **6:** 4113–4123.

KNUST, E., K. A. BREMER, H. VÄSSIN, A. ZIEMER, U. TEPAβ and J. A. CAMPOS-ORTEGA, 1987a The Enhancer of split locus and neurogenesis in *Drosophila melanogaster*. Dev. Biol. **122:** 262–273.

KNUST, E., U. DIETRICH, U. TEPAβ, K. A. BREMER, D. WEIGEL, H. VÄSSIN and J. A. CAMPOS-ORTEGA, 1987b EGF-homologous sequences encoded in the genome of *Drosophila melanogaster* and their relation to neurogenic genes. EMBO J. **6:** 761–766.

LANDSCHULZ, W. H., P. F. JOHNSON and S. L. MCKNIGHT, 1988 The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240:** 1759–1764.

LASKI, F. A., D. C. RIO and G. M. RUBIN, 1986 Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing. Cell **44:** 7–19.

LEHMANN, R., F. JIMENEZ, U. DIETRICH and J. A. CAMPOS-ORTEGA, 1983 On the phenotype and development of mutants of early neurogenesis in *Drosophila melanogaster*. Roux's Arch. Dev. Biol. **192:** 62–74.

LIPMAN, D. J., and W. R. PEARSON, 1985 Rapid and sensitive protein similarity searches. Science **227:** 1435–1441.

LÜSCHER, B., and R. N. EISENMANN, 1990 New light on Myc and Myb. Part I. Myc. Genes Dev. **4:** 2025–2035.

MCKEON, F. D., M. KIRSCHNER and D. CAPUT, 1986 Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins. Nature **319:** 463–468.

MURRE, C., P. SCHONLEBER MCCAW and D. BALTIMORE, 1989 The amphipathic helix-loop-helix: a new DNA-binding and dimerization motif in immunglobulin enhancer binding, daughterless, MyoD and myc proteins. Cell **56:** 777–783.

MURRE, C., P. SCHONLEBER MCCAW, H. VAESSIN, M. CAUDY, L. Y. JAN, Y. N. JAN, C. V. CABRERA, J. N. BUSKIN, S. D. HAUSCHKA,

A. B. LASSAR, H. WEINTRAUB and D. BALTIMORE, 1989 Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell **58:** 537–544.

OLSON, E. N., 1990 MyoD family: a paradigm for development? Genes Dev. **4:** 1454–1461.

PEIFER, M., F. KARCH and W. BENDER, 1987 The bithorax complex: control of segmental identity. Genes Dev. **1:** 891–898

Pirrota, V., 1988 pp. 437–456 in *A Survey of Molecular Cloning Vectors and Their Uses*, edited by R. L. RODRIGUEZ and T. TREINHARDT. Butterworths, Boston.

PIRROTTA, V., C. HADFIELD and G. H. J PRETORIUS, 1983 Microdissection and cloning of the white locus and the 3B1–3C2 region of the *Drosophila* X chromosome. EMBO J. **2:** 927–934.

POOLE, S. J., L. M. KAUVAR, B. DREES and T. KORNBERG, 1985 The *engrailed* locus of Drosophila: structural analysis of an embryonic transcript. Cell **40:** 37–43

PREISS, A., D. A. HARTLEY and S. ARTAVANIS-TSAKONAS, 1988 The molecular genetics of *Enhancer of split*, a gene required for embyonic neural development in *Drosophila*. EMBO J. **7:** 3917–3928.

RUSHLOW, C. A., A. HOGAN, S. M. PINCHIN, K. M. HOWE, M. LARDELLI and D. ISH-HOROWICZ, 1989 The *Drosophila* hairy protein acts in both segmentation and bristle patterning and shows homology to N-*myc*. EMBO J. **8:** 3095–3103.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHRONS, H., E. KNUST and J. A. CAMPOS-ORTEGA, 1992 The *Enhancer of split* complex and adjacent genes in the 96F region of *Drosophila melanogaster* are required for segregation of neural and epidermal progenitor cells. Genetics **132:** 481–503.

SCOTT, M. P., and S. B. CARROLL, 1987 The segmentation network and homeotic gene network in early *Drosophila* development. Cell **51:** 689–698.

SPRADLING, A. C., 1986 pp. 175–197 in *Drosophila, a Practical Approach*, edited by D. B. ROBERTS. IRL Press, Oxford.

STADEN, R., 1984 Graphic methods to determine the function of nucleic acid sequences. Nucleic Acids Res. **12:** 521–538.

TAUTZ, D., and C. PFEIFLE, 1989 A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals a translational control of the segmentation gene *hunchback*. Chromosoma **98:** 81–85.

TIETZE, K., OELLERS, N. and E. KNUST, 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.

VÄSSIN, H., K. A. BREMER, E. KNUST and J. A. CAMPOS-ORTEGA, 1987 The neurogenic locus *Delta* of *Drosophila melanogaster* is expressed in neurogenic territories and encodes a putative transmembrane protein with EGF-like repeats. EMBO J. **6:** 3431–3440.

VILLARES, R., and C. V. CABRERA, 1987 The *achaete-scute* gene complex of *Drosophila melanogaster*: conserved domains in a subset of genes required for neurogenesis and their homology to *myc*. Cell **50:** 415–424.

ZIEMER, A., K. TIETZE, E. KNUST and J. A. CAMPOS-ORTEGA, 1988 Genetic analysis of *Enhancer of split*, a locus involved in neurogenesis in *Drosophila melanogaster*. Genetics **119:** 63–74.

Communicating editor: T. SCHÜPBACH