The EMBO Journal vol.6 no.11 pp.3431-3440, 1987

# The neurogenic gene Delta of *Drosophila melanogaster* is expressed in neurogenic territories and encodes a putative transmembrane protein with EGF-like repeats

**Harald Vässin[1], Kirsten A.Bremer, Elisabeth Knust and Jose A.Campos-Ortega**

Institut für Entwicklungsphysiologie der Universität, Gyrhofstrasse 17, 5000 Köln 41, FRG

Communicated by J.Campos-Ortega

[1]Present address: Howard Hughes Medical Institute and Department of Physiology and Biochemistry, University of California, San Francisco, CA 94143, USA

**The decision of an ectodermal cell to take on a neural or an epidermal fate depends on its interactions with the neighbouring cells. In *Drosophila melanogaster*, the available evidence suggests that a regulatory signal necessary for epidermal commitment is built by the products of the so-called neurogenic genes. We have cloned 180 kb of genomic DNA surrounding the neurogenic gene Delta (*Dl*). Restriction fragment-length polymorphisms were mapped to a region of 25 kb. These 25 kb of DNA are assumed to contain essential parts, or all, of the *Dl* gene. Northern blots detect two developmentally regulated transcripts, of 5.4 and 4.6 kb, which are associated with the region where the mutants map. Several cDNA clones were recovered from embryonic cDNA libraries by homology to the 25 kb of genomic DNA. The complete sequence of a cDNA clone containing an insert of 4.73 kb was determined. The conceptual translation of the longest open reading frame yields a protein of 880 amino acids. This protein displays characteristics of a membrane protein, with intracellular, transmembrane and extracellular domains. The extracellular domain contains a tandem array of nine EGF-like repeats. In *in situ* hybridizations to tissue sections, transcripts homologous to *Dl* are detected in all territories with neurogenic abilities, e.g. the neurogenic ectoderm and the primordia of the sensory organs. Initially all cells of these neurogenic territories express *Dl*, but later on transcription of *Dl* becomes restricted to the cells that have adopted the neural fate. The topological specificity in the transcription of *Dl* corresponds to the one expected for a regulatory signal that mediates epidermal commitment.**

*Key words: Drosophila*/neurogenesis/Delta/expression/sequence

## Introduction

Neurogenesis, the segregation of the neuroblasts from the ectodermal germ layer to form the primordium of the central nervous system (CNS), is mediated by the neurogenic (NG) genes (Poulson, 1937; Wright, 1970; Lehmann *et al.*, 1981, 1983; Campos-Ortega, 1985). The NG-gene products are apparently required for processes of cell-cell interaction that lead to the segregation of two cell lineages from a common anlage of the ectoderm. In this anlage, neighbouring cells have to decide between an epidermal and a neural fate (Hartenstein and Campos-Ortega, 1984; Technau and Campos-Ortega, 1985, 1986). The NG genes seem to provide a regulatory signal that leads to epidermal commitment. Transplanting single cells lacking one of the NG genes into the NG ectoderm of the wild type has allowed a classification of the NG-gene functions as being either a source of, or a receptor for, the regulatory signal that mediates epidermal fate (Technau and Campos-Ortega, 1987).

Delta (*Dl*) is one of the NG genes (Lehmann *et al.*, 1981). It maps to chromosomal band 92A2 and exhibits a fairly complex genetic behaviour (Vässin *et al.*, 1985; Vässin and Campos-Ortega, 1987). *Dl* gene expression is essentially zygotic, with a phenocritical period at the time of neuroblast segregation (Lehmann *et al.*, 1983). However, there is also a maternal component of *Dl* gene expression (Vässin and Campos-Ortega, 1987) as well as a requirement of *Dl*-gene function for epidermal development of the imago (Dietrich and Campos-Ortega, 1984). Cell transplantations indicate that the *Dl* function is required as part of the source for the regulatory signal mediating epidermogenesis (Technau and Campos-Ortega, 1987). In order to elucidate its function during neurogenesis, we have begun to study the molecular organization of *Dl*. Here we report on our cloning of the *Dl* gene. The spatial and temporal pattern of gene expression and the deduced structure of a protein encoded by *Dl* support the contention that *Dl* provides the specificity required for the regulatory signal.

## Results

### *Cloning of the genomic region containing the Delta locus*

The gene *Dl* has been mapped, on the basis of cytogenetic and genetic analyses, to the chromosomal band 92A2 (Lehmann *et al.*, 1983; Vässin and Campos-Ortega, 1987). In order to clone the *Dl* gene, we took advantage of a *P* element inserted at 92A3 in a wild-type *Q* strain caught in the neighbourhood of Cologne (H.Vässin, unpublished). A genomic fragment containing the *P* element in question was isolated from an EMBL3 library prepared from genomic DNA of this wild-type strain. This fragment provided us with a start for a genomic walk to the chromosomal band 92A2. A total of 180 kb genomic DNA were isolated during this walk, 140 kb in the proximal and 40 kb in the distal direction from the site of P-element insertion (Figure 1).

Genomic digests of a total of 26 *Dl* alleles, 14 of them induced with X-rays, 10 with EMS, and the other two recovered from dysgenic crosses, were probed with fragments of the chromosomal walk on Southern blots, in order to obtain an estimate of the size of the *Dl* locus. Only *Dl* alleles induced in a known genetic background were used for molecular characterization (Vässin and Campos-Ortega, 1987). Restriction-enzyme-fragment-length polymorphisms, as compared with the parental strains, were detected in nine X-ray-induced alleles, in both alleles recovered from dysgenic crosses and in one of the EMS-induced alleles. Polymorphisms were mapped to the genomic region +75 to +100 (Figure 1). In one case ($Dl^{KX2}$), a difference was found outside the 25-kb region, at +18 to +21, in addition to another difference at +87 to +90. In a few chromosomal aberrations, the results obtained with genomic Southern blots were verified by *in situ* hybridization of $^3$H-labelled nick-translated probes to polytene chromosomes ($Dl^{III3}$, $Dl^{KX15}$; see Figure 2). Finally,

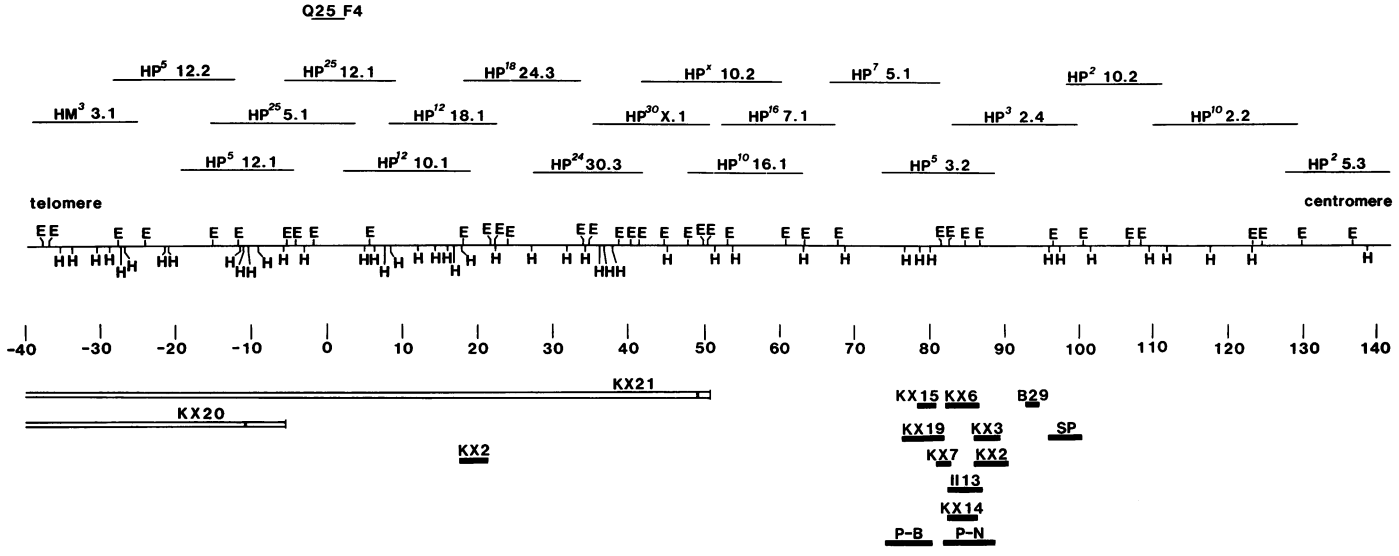



**Fig. 1.** Map of the cloned genomic region. Overlapping DNA inserts of recombinant phages, spanning the cloned region are indicated at the top. A restriction map of the corresponding genomic region is shown in the middle. Telomere is to the left, centromere to the right. Coordinates are indicated in kilobases. Zero corresponds to the insertion of a *P* element that served to start the chromosomal walk. In the lower part of the map, restriction fragment size differences found in 12 *Dl* alleles, as well as the distal breakpoints of two deletions that uncover the *Dl* locus, are indicated. Open bars indicate chromosomal material still present in deletions that uncover the *Dl* locus, solid bars indicate the regions to which restriction-fragment-length differences between *Dl* alleles and the respective parental strains were mapped. Lettering above the bars indicates the respective *Dl* allele designation. With the exception of $Dl^{KX2}$, in which a difference was found at map position +18 to +21 in addition to another one at +87 to +90, all other *Dl* alleles indicated showed only one restriction fragment length difference within the cloned region. *E*, *Eco*RI; *H*, *Hin*dIII.

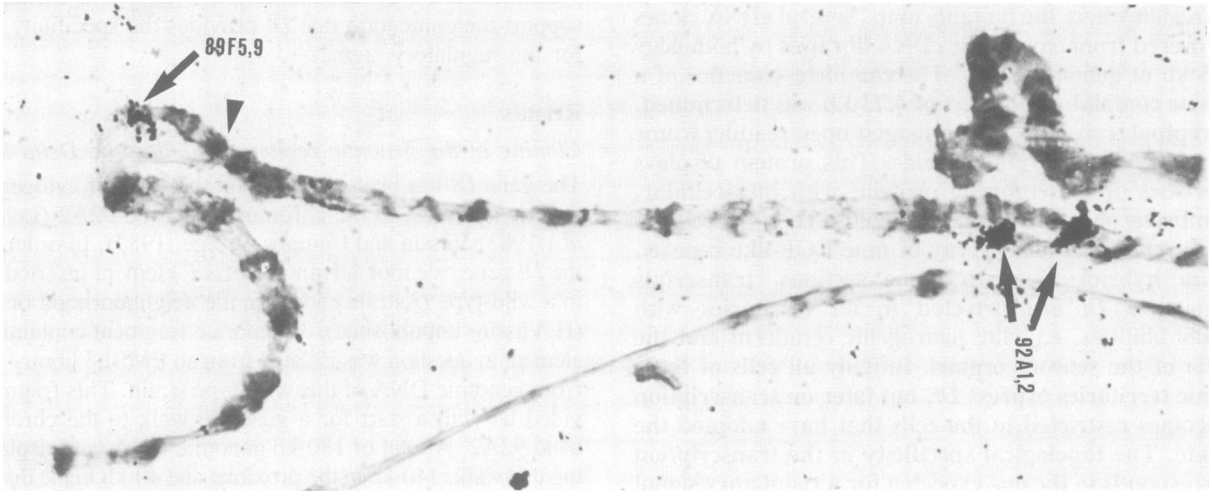



**Fig. 2.** *In situ* hybridization using $^3$H-labelled DNA of the recombinant phage clone $HP^2 2.4$ on salivary gland chromosomes of third-instar larvae heterozygous for the transposition $Tp(3;3)Dl^{III3}$. Hybridization signals (arrows) are visible in the chromosomal region 92A1-2 on both wild-type and mutant chromatides. Another hybridization signal can be seen at the proximal border between the transposed chromosomal segment, inserted in reversed orientation in 89F5-9. The arrowhead points to the distal insertion site of the transposed fragment. This hybridization pattern indicates that the molecular probe contains the wild-type equivalent of the chromosomal region where the breakpoint of $T(3;3)Dl^{III3}$ is located.

the distal breakpoints of two X-ray-induced deletions uncovering the *Dl* locus were mapped to −11 to −6 ($Df(3R)Dl^{KX20}$) and +49 to +51 ($Df(3R)Dl^{KX21}$). Based on these results, the region between +75 and +100 is assumed to represent essential parts, or most probably all, of the *Dl* locus.

### *Transcriptional activity of Dl*

Using different restriction fragments of the map region +80 to +100 as probes, three cDNA clones of different sizes were isolated from an oligo(dT)-primed cDNA library of 3−12-h-old embryos (kindly provided by L.Kauvar; see Poole *et al.*, 1985), the largest cDNA clone (*c3.2*) having an insert of 4.73 kb (Figure 3). In addition, two cDNA clones (*c4* and *c8*) were isolated on the basis of sequence homology to 'EGF-like repeats' (Knust *et al.*, 1987; and below). Based on restriction mapping and Southern blot analyses, four of the cDNA clones (*c3.2*, *c3*, *c4* and *c8*) had virtually identical 3′ ends, but differed mainly in the extent of 5′ sequences present. In addition, two restriction-enzyme sites are present in *c3.2* that are missing in *c3*, *c4* and *c8*. One of the cDNA clones (*c11*) differs from the others in that it misses ~1.7 kb of the 3′ end (Figure 3; see below).

Hybridization of small subfragments of the cDNA clones to the genomic walk and comparison of the restriction map allowed a rough alignment of the cDNAs and the homologous genomic

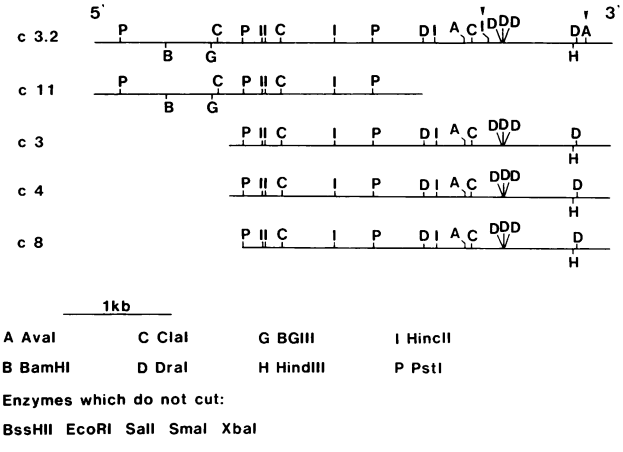



**Fig. 3.** Map of restriction sites of the *Dl* cDNA clones (*c3.2, c11, c3, c4* and *c8*) studied. The cDNA clones *c3, c4* and *c8* exhibit the same 3′ ends as *c3.2*, but they contain only smaller parts of the 5′ region. No differences have been detected in the pattern of restriction sites of *c3* and *c4*, which may therefore be identical. Two additional restriction cleavage sites are present in *c3.2* (indicated by arrowheads) that have not been found for the other cDNA clones.

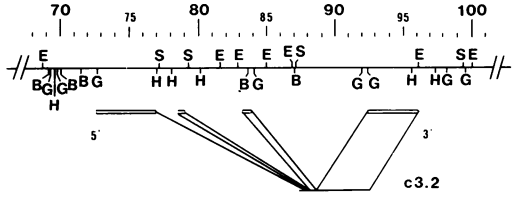



**Fig. 4.** Alignment of the cDNA clone *c3.2* with the genomic DNA. Subfragments showing homology, on the basis of Southern-blot analysis, are indicated. In the upper part a restriction map is shown of the genomic region between map positions +68 and +101. Open bars below the restriction map indicate the genomic fragments with homology, on the basis of Southern blot analysis, to the indicated regions of *c3.2*. The orientation of *c3.2* is indicated.

fragments (Figure 4). The transcript represented by the cDNA clones is thus composed of at least four exons together spanning the 25 kb of genomic DNA to which *Dl* mutations have been mapped.

To determine the size and developmental profile of the *Dl* transcript(s), Northern blots, either with poly(A)$^+$ or with total RNA from various developmental stages, were probed with fragments of the cDNA clone *c3.2*. Two major polyadenylated-stage-specific transcripts of 5.4 and 4.6 kb are present (Figure 5). Between 2 and 6 h of embryonic development the transcriptional activity of *Dl* is highest; this period overlaps with the phenocritical period of the *Dl* locus, as defined by Lehmann *et al.* (1983). During later stages of embryonic development, transcripts decrease in abundance more and more, but still remain visible up to the larval stage. During larval development, only a faint expression is detectable. Probing of Northern blots with progressively smaller probes from the 3′ end of *c3.2* (2.2-kb *Eco*RI/*Pst*I fragment, 1.3-kb *Eco*RI/*Cla*I fragment and 1.1-kb *Eco*RI/*Hin*cII fragment) results in stronger labelling of the 5.4-kb transcript throughout development, with the exception of 0−2-h-old embryos and adult females, in which the smaller transcript is more abundant (Figure 5). Thus, the transcripts detected in 0−2-h embryos most probably correspond to the maternal transcripts. In males only very small amounts of both transcripts can be detected.

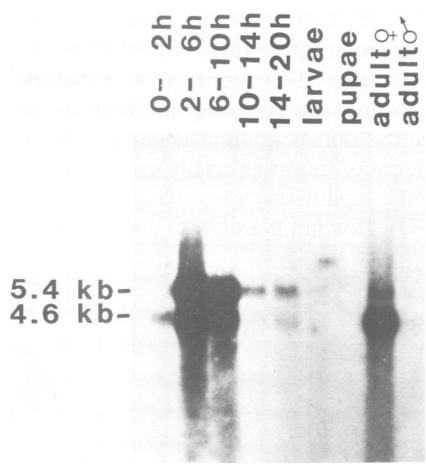



**Fig. 5.** Developmental Northern blot probed with a 0.5-kb *Hin*cII fragment from the 3′ end of the *c3.2* cDNA clone. The numbers above the lanes refer to hours of embryonic development at 25°C. Two major transcripts of 4.6 and 5.4 kb are visible throughout development, with the exception of the pupal stage, in which no transcripts are detectable. A weak signal can be detected in larvae and in adult males only after longer expression. With the probe used in this hybridization experiment, the 4.6-kb transcript is less represented than the 5.4-kb transcript; using probes from the 5′ end, both transcripts are equally represented.

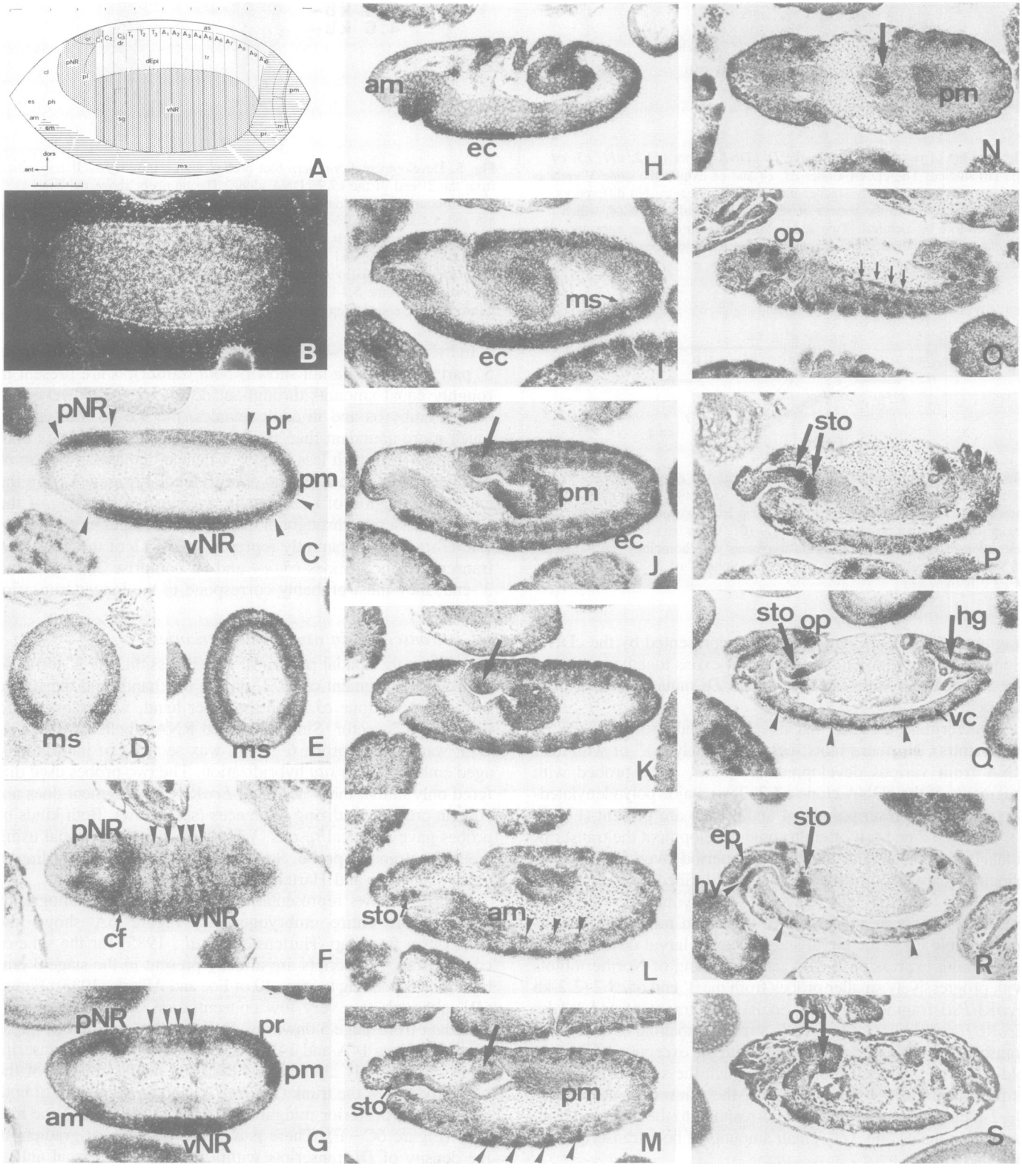
In hybridizations with a 265-bp *Eco*RI/*Pst*I fragment from the 5′ part of *c3.2* (data not shown) both transcripts are present in roughly equal amounts throughout development. However, in 0−2-h embyros and in adult females, the 4.6-kb transcript is much more abundant than the longer one; in adult males both transcripts are much less abundant than in females. Probing of Northern blots with a 333-bp *Hin*dIII/*Eco*RI fragment from the extreme 3′ region of *c3.2* results in exclusive labelling of the 5.4-kb transcript throughout development (data not shown). Therefore, *c3.2* apparently represents 4.73 kb of the 5.4-kb *Dl* transcript. Since clones *c3*, *c4* and *c8* comprise 333 bp of the 3′-end, they most probably correspond to the longer transcript as well.

### *Dl is transcribed in neurogenic territories*

To analyse the spatial pattern of *Dl* expression, the 3′ terminal *Eco*RI/*Cla*I fragment of *c3.2*, on the one hand, and the whole of the cDNA clone *c3.2*, on the other hand, were cloned into the Gemini 2 vector. Single-stranded RNA labelled with [$^{35}$S]-UTP was used to probe 6−8-μm wax sections of increasingly aged embryos by *in situ* hybridization. The two probes used differed only in that the 3′ terminal *Eco*RI/*Cla*I fragment does not contain cross-hybridizing sequences (see below). Both kinds of probes gave identical results. We failed to detect a signal using the sense strand as probe. Staging of embryos was according to Campos-Ortega and Hartenstein (1985).

Figure 6 shows representative examples of autoradiographs covering the entire embryogenesis; Figure 6A shows the blastoderm fate map (Hartenstein *et al.*, 1985) for the sake of comparison. Transcripts are already present in the stage 3 embryo [corresponding to stage 10 of Foe and Alberts (1983); Figure 6B], although at a very low concentration; they become very abundant from stage 5 onwards. During cellularization in stages 5 and 6 (stages 14A and 14B of Foe and Alberts), *Dl* transcription is particularly abundant in the cells of the anlagen of the NG ectoderm of the trunk (Figure 7A) and procephalon, of both anterior and posterior midgut and of the dorsal half of the hindgut (Figure 6C−E). There is a clear ventrodorsal gradient in the density of *Dl* transcripts within the NG ectodermal anlage

of the ventral cord (Figures 6D and 7A), corresponding to the gradient of neurogenic capabilities of the NG ectoderm (Technau and Campos-Ortega, 1986). Immediately before gastrulation transcription extends dorsally over the limits of the NG anlage to cross over the dorsal midline (Figure 6E). A conspicuous metameric pattern of transcriptionally active cells extending from the ventral NG region becomes evident in stage 7 embryos (Figure 6F–G). It is not possible to decide what kinds of cells form this banding pattern; they may perhaps correspond to the anlagen of the sensory organs, since the latter will become transcriptionally active during later stages. During germ band elongation, in stages 8–10, the entire ectoderm exhibits abundant transcriptional activity (Figure 6H–I). We have not been able to discern a banding pattern in these elongating embryos, probably due to folding of the dorsal regions of the ectoderm caused by germ-band elongation. However, since a metameric



distribution of transcripts again becomes apparent in late stage 10, we believe that the stripes continue to be present. Ventrally the metameric distribution of transcripts during late stage 10 (Figure 6J) is prominent in the mesectodermal cells and the primordia of the tracheal tree. Labelling in stage 10 is also apparent in the entire NG ectoderm, in anterior and posterior midgut primordia and in the cells of the posterior wall of the proctodeum, which after shortening will become the anterior wall of the hindgut.

From late stage 11 onward, *Dl* transcripts become restricted to cells of both the central and the peripheral nervous systems, and they persist in the midgut and in the hindgut (Figures 6L–N and 7B–D). Particularly the neuroblasts, the median precursors, the primordia of the stomatogastric nervous system, of the optic lobes and of many recognizable, probably of all, sensory organs exhibit abundant transcripts (Figure 6M and R). During stage 13, transcripts disappear from the midgut and become prominent at the boundary between the proventriculus and the midgut, probably corresponding to the cells of the oesophageal ganglion of the stomatogastric nervous system (Figure 6P–R). In stage 16, transcriptional activity can only be detected in the primordium of the optic lobes, in some large sensory organs (e.g. hypophysis and epiphysis) and in the periventricular region of the foregut as well as in the anterior wall of the hindgut (Figure 6P–R). Transcripts are also very abundant within the trachea.

*Sequence analysis of a Dl transcript*

To analyse the structure of the *Dl* gene product(s), the sequence of the cDNA clone *c3.2* was determined (Figures 8 and 9). Computer analysis of this sequence revealed a large open reading frame (ORF) of 2640 bp, starting with an ATG preceded by the sequence AAAC in the −4 to −1 position. This sequence preceding ATG fulfills the requirements of a translation start for an eucaryotic gene (Kozak, 1984), and especially a *Drosophila* gene, as described by Cavener (1987). Preceding the ATG, three stop codons are present in frame; thus, we assume that the cDNA *c3.2* contains the entire translated part of the 5.4-kb transcript. Codon bias analysis, using a *Drosophila* codon bias table (J.Pustell/International Biotechnologies, Inc.), indicates that the large ORF is the only region within the sequence that shows a high probability to code for a *Drosophila* protein (Figure 8). The 3′ end of the sequence (Figure 9) contains a polyadenylation signal at position 4704 followed 20 bp downstream by a stretch of 13 adenosine residues that presumably correspond to the beginning of the poly(A) tail. There is a large untranslated region at the 3′ end of 1951 bp. A striking feature of this untranslated region are several copies of a ATTTA or a $A(T)_{n>3}A$ sequence. This sequence has been described to result in transcript instability when integrated in the 3′ region of an otherwise stable mRNA species. Since this motif has been found to be present in a number of transitionally expressed mRNA species, it may be involved in the regulation of specific degradation of those transcripts (Shaw and Kamen, 1986).

The deduced amino acid sequence of the large ORF (Figure 9) strongly suggests that *c3.2* encodes a transmembrane protein (Figure 10), with a structure similar to that proposed for the protein encoded by the NG gene *N* (Wharton *et al.*, 1985; Kidd *et al.*, 1986). The most conspicuous feature of the deduced *Dl* protein is a cystein-rich region, organized in nine complete and one incomplete repeats with homology to the so called 'EGF-like' repeats (Figure 11). Similar repeats are part of a number of mammalian proteins, as well as of the extracellular domain of the presumptive protein of the NG gene *N*, in which 36 of those repeats have been found (Wharton *et al.*, 1985; Kidd *et al.*, 1986), and in the proteins encoded by other *Drosophila* genes (Knust *et al.*, 1987). The N-terminal part of the putative *Dl* protein starts with a hydrophobic region (Figure 10), probably corresponding to a signal peptide (Watson, 1984; Sabatini *et al.*, 1982). In addition, a highly hydrophobic region followed by a stretch of six basic amino acids (KRKRKR), a characteristic feature of transmembrane regions of membrane-spanning proteins (Yarden *et al.*, 1986), is present. With the exception of the EGF-like repeats, computer analysis revealed no additional regions of significant homology, neither at the DNA nor at the amino acid-sequence level, between *N* and *Dl*. Small stretches of homology outside the region of the EGF-like repeats are due to the presence of an alanine stretch present in both sequences (data not shown).

## Discussion

The following arguments indicate that we have cloned and identified the genomic DNA of *Dl*. (i) The DNAs of 12 *Dl* alleles out of 26 alleles with known parental background exhibit restriction-fragment-length differences scattered around a region of 25 kb. It is highly improbable that the restriction fragment length

**Fig. 6.** Spatial pattern of transcription of *Dl* during embryogenesis. (**A**) shows a fate map of the blastoderm (Hartenstein *et al.*, 1985) for reference. The anlage of the neurogenic ectoderm (pNR, procephalic neurogenic region; vNR, ventral neurogenic region), and those invaginating at gastrulation (am, anterior midgut; ms, mesoderm; pr, proctodeum; mt, Malpighian tubules; pm, posterior midgut), are shaded in the figure. (**B**) is a dark field micrograph of a stage 3 embryo, before cell formation. (**C**) is a parasagittal section of a blastoderm, showing transcripts restricted to both the procephalic (pNR, arrowheads) and the ventral (vNR, arrowheads) subdivision of the neurogenic region, as well as in a dorsoposterior cap of the embryo comprising the anlage of the posterior midgut (pm) and the anlage of the posterior wall of the proctodeum (pr), which will give rise to the anterior wall of the hindgut (hg in Q). (**D** and **E**) are transverse sections of two embryos at the blastoderm stage [stage 6, corresponding to stage 14B of Foe and Alberts (1983)], showing that, initially, transcription is restricted to the ventral neurogenic region (**D**, shows ~50% egg length level); in the procephalic neurogenic region (at ~80% egg length, 0% at the posterior egg pole), transcripts extend across the dorsal midline (**E**). The mesodermal anlage (ms), as well as the mesoderm and its derivatives, are devoid of transcripts. (**F**) is a tangential and (**G**) a parasagittal section of the same late stage 6 embryo. Notice the conspicuous bands of *Dl* transcripts with a metameric pattern (arrowheads) in the dorsal epidermal anlage. Transcripts are also apparent in the anlage of the anterior midgut (am). (**H**) is a stage 8 embryo, during germ band elongation, (**I**) is a stage 10 embryo. The outer regions of the ectoderm (ec) contain abundant transcripts in both embryos. (**J**) is a late stage 10 embryo showing a metameric pattern of *Dl* transcripts in the region of the mesectodermal cells, from which the median neural cells will develop. The arrow in this figure, as well as in (**K**) (early stage 11), (**L**) (stage 11), (**M**) (late stage 11) and (**N**) (stage 12) point to a peculiar arrangement of *Dl*-transcribing cells at the posterior wall of the proctodeum. Abundant transcripts are also visible in the posterior midgut (pm). sto in (**L**) and (**M**) points to the three invaginations of the foregut that give rise to the stomatogastric nervous system. The evolution of the stomatogastric nervous system (sto) can be easily followed on these sections, from stage 11 (**L**) through stages 13–15 (**P–R**). Arrowheads in (**L**) point to individual labelled cells, topologically corresponding to median neural cells. Labelling within the epidermal primordium is also patchy, probably corresponding to the cells of the sensory organs. *Dl* transcripts in cells of developing sensory organs have been pointed to in a few cases, e.g. in the stage 12 embryo shown in (**O**) (small arrows) and in the stage 14 embryo in (**R**) (ep, epiphysis; hy, hypophysis). However, most sensory organs exhibit *Dl* transcripts. op in (**O**) corresponds to the anlage of the optic lobes. From stage 11 onwards, labelling is visible in the neuroblasts. In stage 14 (**Q**) neuroblasts cannot be distinguished any more; however, abundant transcripts are present over cells of the ventral cord (vc, arrowheads in **Q** and **R**), probably corresponding to the imaginal neuroblasts (unpublished observations; see Figure 7D). In the stage 16 embryo (**S**) transcription is still prominent in the anlage of the optic lobes. Other abbreviations: as, amnioserosa; cf, cephalic furrow; cl, clypeolabrum; dEpi, dorsal epidermis; dr, dorsal ridge; es, oesophagus; ph, pharynx; pl, procephalic lobe.

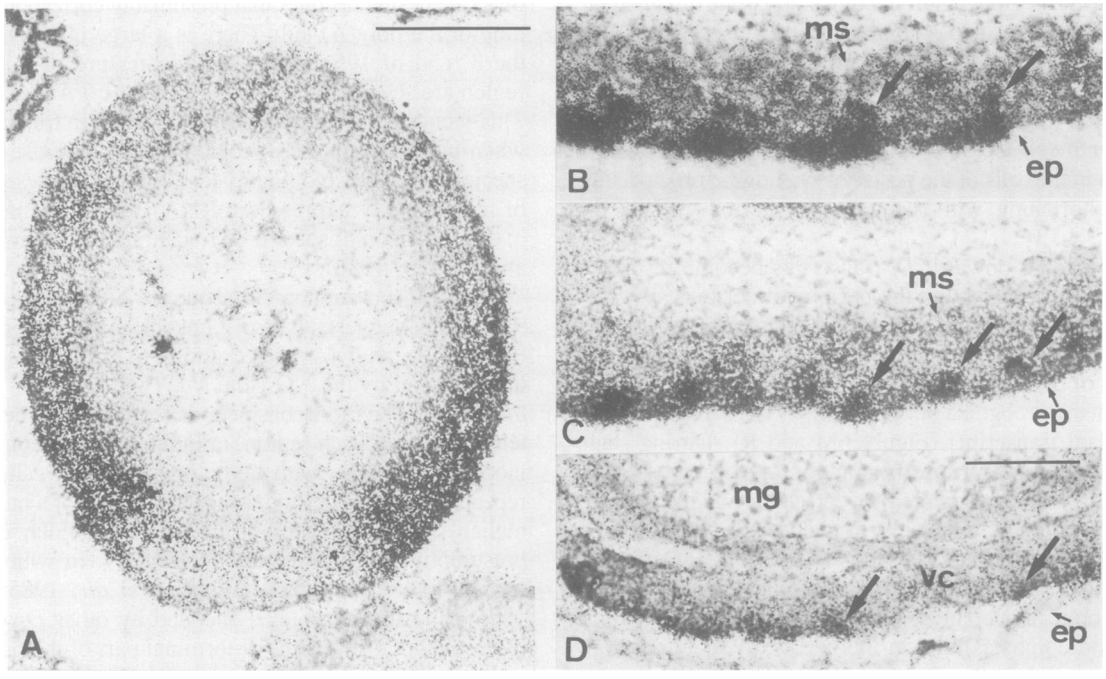



**Fig. 7.** Spatial pattern of *Dl* transcription. (**A**) shows a transversal section at ~50% egg length, to illustrate transcripts restricted to the ventral neurogenic region (cf. Figure 6A). Notice graded distribution of grains, with higher concentration at ventral levels. (**B–D**) show part of the germ band (sagittal section) of increasingly aged embryos; anterior is to the left. Notice in (**B**) (late stage 11) labelling is patchy in the epidermis (ep); arrows point to two neuroblasts (ms: mesoderm). Chiefly the neuroblasts (arrows) are labelled in (**C**) (stage 12), whereas the epidermis is virtually free of transcripts. In (**D**) (stage 14) labelling in the ventral cord (vc) is restricted to the ventralmost region (arrows), where the neuroblasts were visible shortly before this stage (mg, midgut). Bar = 50 μm.

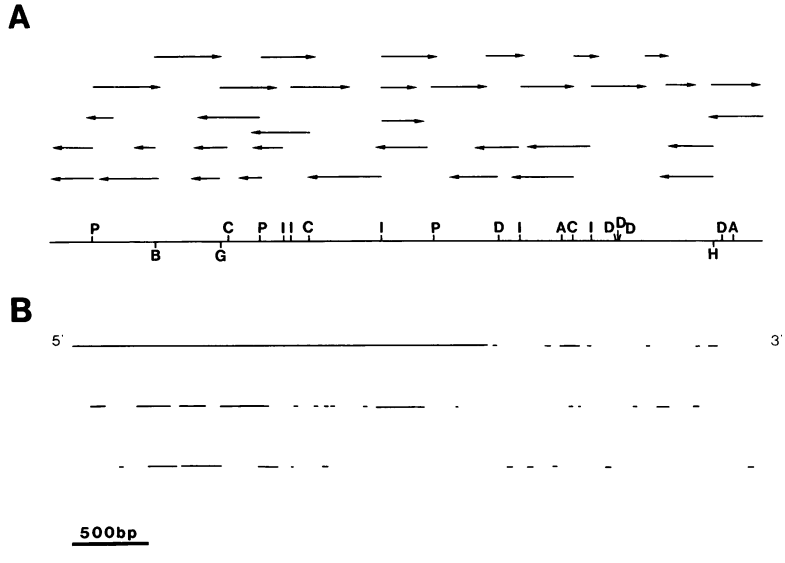



**Fig. 8.** Sequencing strategy of the cDNA clone *c3.2*. (**A**) shows the subsequences used to compose the complete sequence with respect to the restriction map of *c3.2*. In (**B**) the ORFs for three reading frames of the coding strand are displayed. Only those ORFs starting with an ATG and encoding more than five amino acid residues are shown. The largest ORF, located in the first reading frame, starts at base 127 and contains 2640 bp. This ORF was used for the deduction of the corresponding gene product. Symbols for restriction enzymes are as in Figure 3.

differences identified in the DNA of these *Dl* mutants is due to unspecific polymorphisms, since the genomic background of these mutations is known. One of the 12 alleles ($Dl^{B29}$) is EMS induced and carries a deletion of ~1 kb in this region. Another three alleles ($Dl^{III3}$, $Dl^{KX15}$ and $In(3R)Dl^{Sp}$) are chromosomal aberrations that have one breakpoint in the region indicated. Finally, two alleles ($Dl^{P\text{-}B55}$, $Dl^{P\text{-}N}$) have been isolated from dysgenic crosses and carry *P* element insertions. (ii) The *Dl* gene is uncovered by two deletions that behave phenotypically like complete deletions of the *Dl* gene (described in Vässin and Campos-Ortega, 1987). The distal breakpoints of these deletions have been mapped within the walk, several kilobases away from the 25-kb region, which is completely removed in these deletions. (iii) The 25-kb region gives rise to two transcripts with a temporal profile of expression in agreement with the phenocritical period of *Dl*. Northern blots probed with flanking

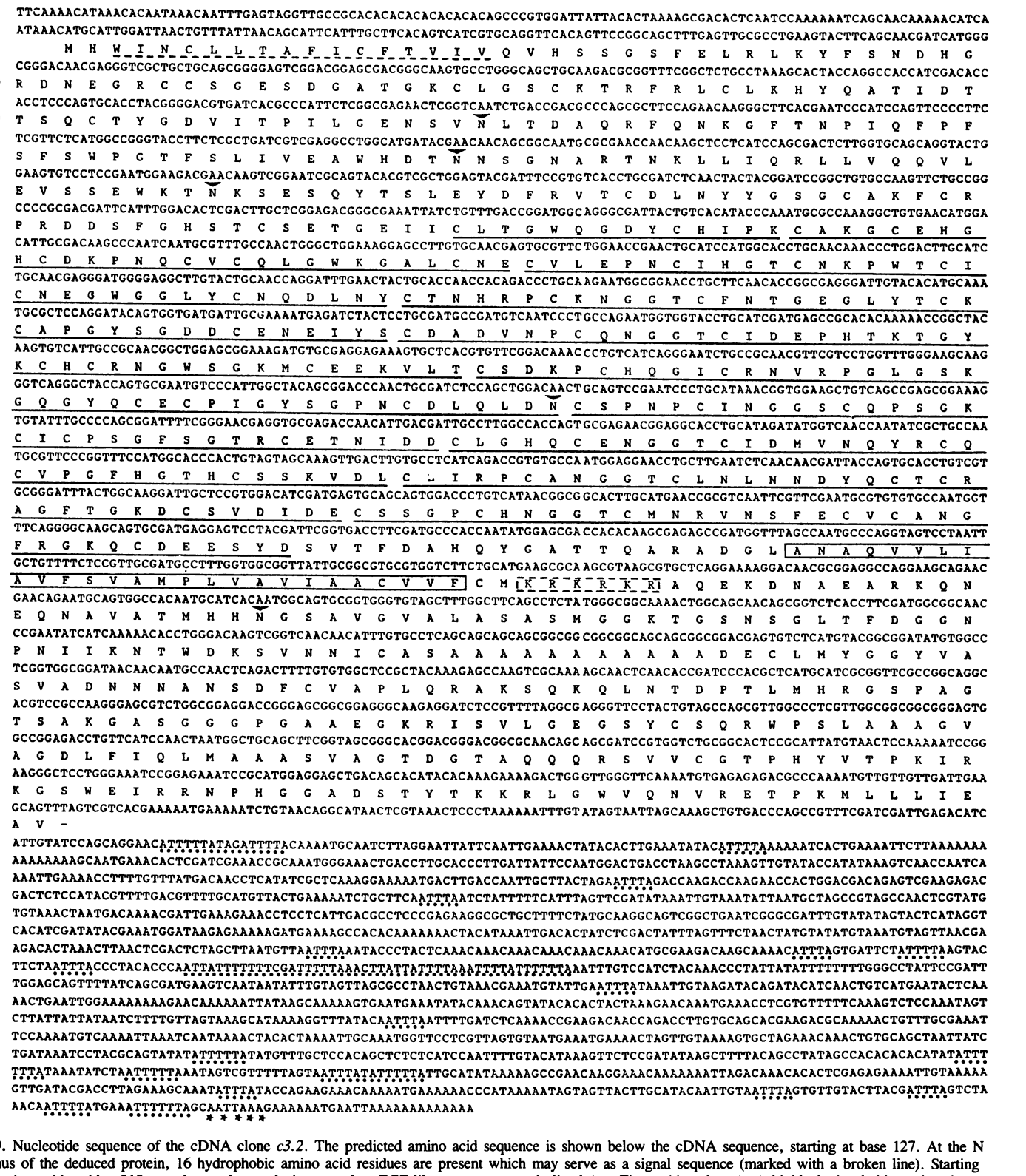


**Fig. 9.** Nucleotide sequence of the cDNA clone *c3.2*. The predicted amino acid sequence is shown below the cDNA sequence, starting at base 127. At the N terminus of the deduced protein, 16 hydrophobic amino acid residues are present which may serve as a signal sequence (marked with a broken line). Starting with amino acid residue 218, one incomplete and nine complete EGF-like repeats are underlined (see Figure 11 and text). A highly hydrophobic region that may correspond to the transmembrane sequence is boxed. This presumptive membrane-spanning region is followed by a short stretch of six basic amino acids (boxed with broken lines). Potential glycosylation sites are indicated with arrowheads. Within the untranslated 3′ region, several copies of the sequence motif ATTTA, or $AT_{n>3}A$, are marked with a dotted line. The polyadenylation signal is indicated by asterisks.

sequences do not uncover any additional transcript which could be related to the periods of *Dl* gene activity. (iv) The spatial pattern of transcription during embryogenesis, as revealed by *in situ* hybridization with cDNA homologous to these 25 kb of genomic DNA, shows a correlation to the process of neurogenesis.

Therefore, these 25 kb of genomic DNA most probably contain all coding sequences of the *Dl* gene. This contention is particularly well supported by the spatial pattern of expression shown by the *c3.2* cDNA, of which we would like to discuss its possible functional significance. During embryogenesis, *Dl* is transcribed ex-

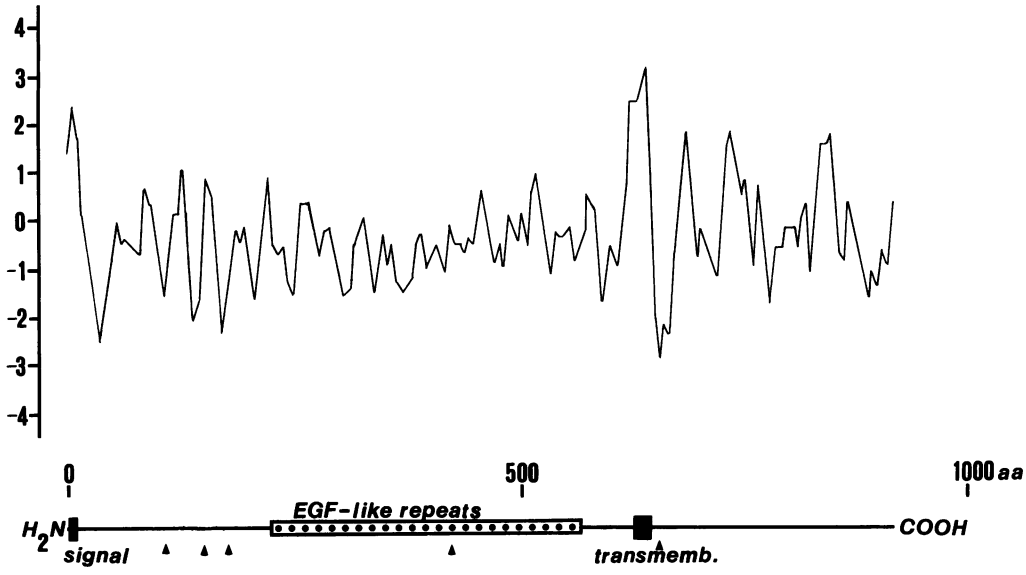



**Fig. 10.** Hydropathy blot and primary structure of one *Dl* gene product as deduced from the sequence of the cDNA clone *c3.2*. The putative signal peptide and transmembrane sequences are indicated by black bars. The open bar with small dots corresponds to the nine complete and one incomplete EGF-like repeats. Potential glycosylation sites are indicated by small arrowheads. The scale is in amino acid residues.

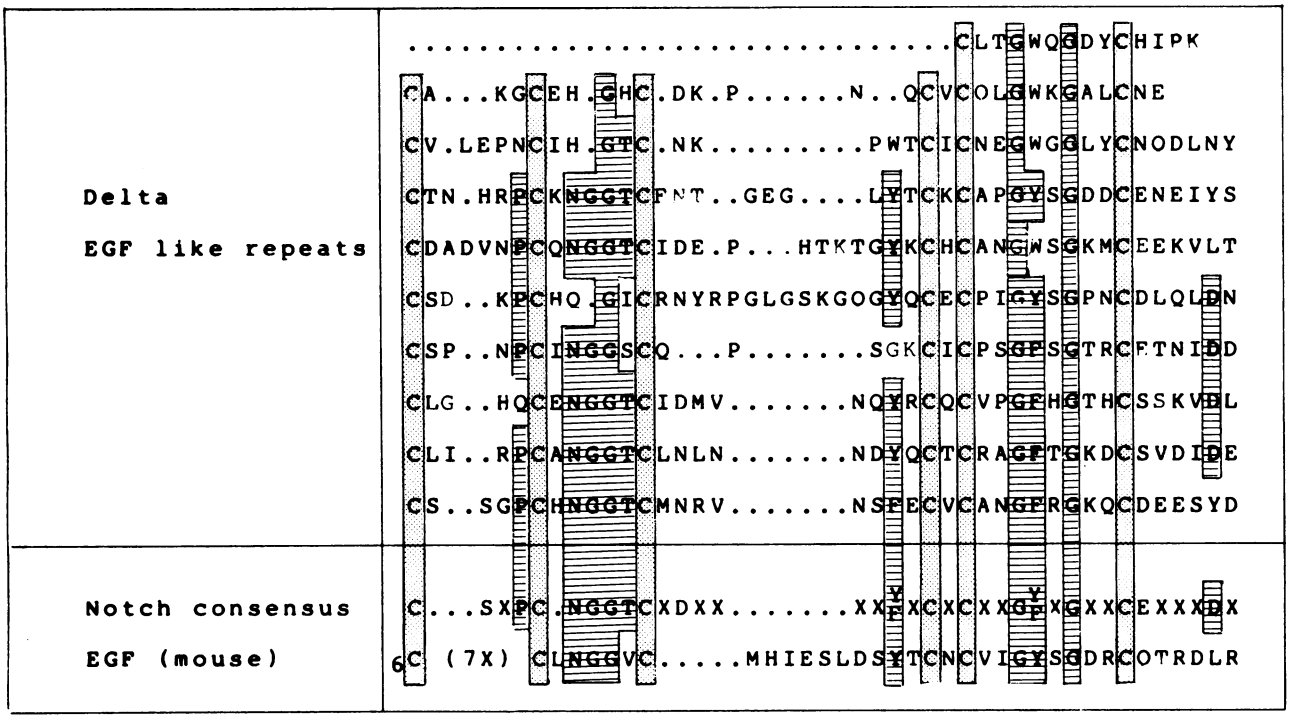



**Fig. 11.** Alignment of the nine complete EGF-like repeats, as well as of one incomplete repeat preceding the other nine, with the consensus sequence of the 36 EGF-like repeats proposed by Wharton *et al.* (1985), and of the mouse EGF (Gray *et al.*, 1983). Four of these repeats were already published by Knust *et al.* (1987). The present sequence differs from the published one in four amino acid residues, most probably reflecting a technical error. Cystein and other amino acid residues occupying corresponding positions are shaded.

clusively in cells of territories with neurogenic abilities, with two exceptions. These neurogenic territories include the NG region of the ectoderm, the anlagen of the stomatogastric nervous system, of the optic lobes and of many sensory organs. Although it is difficult to map the positions of all known sensory organs (Campos-Ortega and Hartenstein, 1985; Ghysen *et al.*, 1986) on *in situ* hybridizations, we have detected transcripts in several of the primordial sensilla. Therefore, we believe that the cells of all sensory organs do indeed transcribe *Dl*. Abundant transcripts were also detected in cells of the anterior and posterior midgut primordia, where neural cells are known to originate during embryonic development (Y.N.Jan, personal communication; J.A.Campos-Ortega, unpublished). There are two apparent exceptions: the anlage of the anterior half of the hindgut and of the trachea. We do not know whether these cells are neurogenically active. Neural cells are known to occur in the hindgut of other insects, e.g. in the moth *Manduca sexta* (Taghert, personal communication), but there is as yet no indication of neurons in the

hindgut of *Drosophila*. Transcription in the tracheal primordia and in the tracheal tree has probably no relation to neurogenesis. Transcripts have also been detected in the tracheal tree while probing with several other different genes that encode EGF-like repeated polypeptides (Knust *et al.*, 1987; U.Dietrich and E.Knust, unpublished); *Dl* transcripts in the trachea may well be related to other functions of the EGF-like repeats independent of neurogenesis.

Although initially all cells of these various primordia express *Dl*, transcripts become progressively restricted to the cells that have taken on the neural fate (Figure 7B–D). This behaviour is very striking, and we assume that it is causally related to the neurogenic decision taken by these cells. On the basis of laser ablations in the NG ectoderm of grasshoppers, Doe and Goodman (1985) have proposed that the neuroblasts prevent the remaining ectodermal cells from entering neurogenesis. One is therefore tempted to speculate about a possible role for *Dl* in mediating this inhibitory signal (see below).

Sequencing of the cDNA clone *c3.2* allowed us to deduce a possible molecular structure for (one of) the *Dl* product(s). The presence of a N-terminal signal sequence, as well as a second hydrophobic region followed by six basic amino acids, a characteristic structure of a transmembrane region, strongly suggests that at least one *Dl* gene product is a membrane-spanning protein. The most prominent structures of this deduced protein are nine EGF-like repeats in the putative extracellular protein domain. This finding is rather interesting, since 36 EGF-like repeats have been reported for the presumptive extracellular part of the protein encoded by another of the NG genes, namely *N* (Wharton *et al.*, 1985; Kidd *et al.*, 1986). Other genes of the *Drosophila* genome are known to encode such EGF-like repeats (Knust *et al.*, 1987; U.Dietrich and E.Knust, personal communication). In *Caenorhabditis elegans*, EGF-like repeats have been described for the product of *lin-12*, a gene known to be involved in the regulation of cell fates (Sternberg and Horvitz, 1984; Greenwald, 1985). Although the presence of EGF-like repeats in the products of all these genes suggests an important role for these repeats in the respective regulation processes, the functional significance of the repeats in the different proteins is not understood. Experimental evidence has been reported by Graf *et al.* (1987) showing that a short nine amino acid peptide present in the laminin B1 chain, which has homology to a part of the EGF-like repeats (Sasaki *et al.*, 1987), seems to be involved in processes that are based on protein–protein interaction, such as cell attachment, migration and cell surface-receptor binding. In addition, it has been reported by Davis *et al.* (1987) that EGF-like repeats in the LDL receptor are essential for its ability to bind one of two receptor ligands.

It is not clear from our data what the structural differences are between the translation products of the 5.4- and 4.6-kb transcripts. Northern blot analysis shows that both transcripts differ in that at least 333 bp of the 3′ end of the 5.4-kb transcript are missing in the 4.6-kb one. However, hybridization experiments with different subfragments of the translated region of *c3.2* show this region to be present in both transcripts without obvious differences. This, of course, does not rule out the possibility of different splicing patterns for both transcripts in the translated region. Due to the lack of a full length cDNA clone for either of the two transcripts, we cannot rule out the possible occurrence of differential splicing in the 5′ region. In fact, none of the cDNA clones studied in our work can be correlated with the smaller transcript. Four clones (*c3.2, c3, c4* and *c8*) contain the *Hin*dIII/*Eco*RI fragment that is specific to the 5.4-kb transcript, whereas the fifth clone (*c11*) is most likely the result of priming from one of the A-rich sequences in the 3′ untranslated region.

Given that the 4.6-kb transcript lacks at least some of the ATTTA sequence motifs implicated in RNA degradation in its 3′ untranslated region, this transcript may be more stable than the longer one. The 4.6-kb transcript is the main maternal transcript; this may explain the need for a more stable transcript. However, this is not sufficient to explain the simultaneous expression of both transcripts during later development.

The NG gene products are apparently responsible for a regulatory signal that passes between cells of the NG ectoderm and leads to epidermal commitment of the receiving cell. On the basis of genetic analyses (Vässin *et al.*, 1985), *N* and *Dl* have been shown to interact with *E(spl)*. Technau and Campos-Ortega (1987) have presented evidence from transplantation experiments that *E(spl)* is the only NG gene that is located on the receptor side, whereas both *N* and *Dl* belong to the signal side of this communication pathway. The putative proteins encoded by *N* and *Dl* exhibit extracellular domains with strikingly similar structure. We propose that the proteins of *N* and *Dl*, either together or separately, act as ligands for the hypothetical receptor protein, perhaps provided by *E(spl)*. Whereas transcription of *N* has been shown to be ubiquitous during early neurogenesis (Yedvobnik *et al.*, 1985), *Dl* expression exhibits the necessary topological specificity required to restrict the epidermogenic signal to the regions where it is needed. We hope to obtain experimental evidence for these latter speculations from the study of the various NG-gene products.

## Materials and methods

*Genomic DNA preparations*

Genomic DNA was prepared from adult flies as described by Weigel *et al.* (1987). Digested genomic DNA was blotted onto GeneScreenPlus membrane (NEN/Dupont) according to Reed and Mann (1985). Transfers to nitrocelullose filters were carried out according to Maniatis *et al.* (1982). Hybridization was as described by Weigel *et al.* (1987) but hybridization under conditions of low stringency was according to Knust *et al.* (1987).

*Screening of recombinant phage libraries*

Preparation of genomic libraries and genomic walking was as described by Weigel *et al.* (1987). Wild-type genomic libraries of Pirotta *et al.* (1983) and R.Rudloff (unpublished) cloned in EMBL-4 were used. Phage DNA preparation and screening of genomic as well as cDNA libraries was performed as described by Maniatis *et al.* (1982). Embryonic cDNA libraries were obtained from L.Kauvar (Poole *et al.*, 1985).

*Northern blot analysis*

Total RNA was isolated by the hot phenol method (Jowett, 1986). Staged embryos, larvae, pupae and adult flies were ground in liquid nitrogen prior to homogenization in phenol. Electrophoresis of RNA was in 1.2 or 0.7% agarose gels in 0.2 M MOPS, 50 mM Na-acetate, 1 mM EDTA (pH 7.0), 2.2 M formaldehyde. Aliquots of 50–80 μg of total RNA were denatured in 6 M glyoxal, 0.2 M MOPS, 50 mM Na-acetate, 1 mM EDTA (pH 7.0) for 15 min at 65°C. The gel was run at 30 V for 12–15 h. As mol. wt markers an RNA ladder (BRL) was used. The RNA was transferred without any pretreatment of the gels in 20 × SSC on to nytran membranes (Schleicher and Schuell). The filter was baked at 80°C for 2 h and prehybridized in 50% formamide, 5 × SSC, 1 × Denhardt, 1% SDS, 20 mM Na-phosphate buffer, 100 μg/ml yeast RNA. Hybridization was performed in the same solution with nick-translated $^{32}$P-labelled DNA at $0.5–1 \times 10^6$ c.p.m./ml for 40–48 h at 42°C. The filters were washed twice in 1 × SSC, 1% SDS for 20 min at 42°C and once in 0.1 × SSC, 1% SDS at 50°C.

*In situ hybridization*

*In situ* hybridization to salivary gland chromosomes was carried out according to Pardue (1985) modified as described by Weigel *et al.* (1987). The protocol used for *in situ* hybridization to embryonic tissue sections is essentially the same as described by Ingham *et al.* (1985), with laboratory modifications. For hybridization, $^{35}$S-labelled RNA probes, synthesized with the Riboprobe Gemini-2 system

(Promega), were used. Hybridization solutions with a final activity of $7 \times 10^5$ c.p.m./μl were used. Slides were exposed between 3 and 8 days.

*Sequencing*

Sequencing was carried out essentially as described by Sanger *et al.* (1980). Most of the sequence was obtained by sequencing restriction fragments, subcloned into M13mp18 or M13mp9 vectors. Part of the sequence was obtained by plasmid sequencing using commercially available primers (BRL and Stratagene) for the Gemini (Promega) or Bluescript (Stratagene) system. Some of the fragments sequenced with the plasmid sequencing method were shortened by exonuclease III/mung-bean nuclease digestion in the Bluescript vector (Stratagene, exo/mung DNA sequencing system, Instruction Manual). In addition three oligonucleotides were synthesized for use as a primer in the M13 sequencing protocol, to close small gaps in the sequence or to obtain a sufficient overlap between adjacent sequenced fragments. The primers were synthesized on a DNA-synthesizer (Applied Biosystems).

*Computer analysis*

Computer analysis of the nucleotide and the predicted amino acid sequences was performed on a IBM PC/AT with the DNA/Protein Sequence Analysis Software of J.M.Pustell/International Biotechnologies, Inc., New Haven. For a computer homology search the program of Lipmann and Pearson (1985) and the NBRF Protein Bank was used.

*Other procedures*

Restriction-digested phage DNA or gel-purified fragments were subcloned into pGem2 (Promega) or Bluescript (Stratagene). Preparation of plasmid DNA was performed according to Willimzig (1985) for small-scale preparation and Holmes and Quigley (1981) for large-scale preparation.

## Acknowledgements

We are grateful to Sigrid Baars for expert technical assistance, our colleagues in Köln for discussions, Michael Brand for critical reading of the manuscript, and the Deutsche Forschungsgemeinschaft (DFG, SFB 74) for financial support.

## References

Campos-Ortega,J.A. (1985) *Trends Neurosci.*, **8**, 245–250.
Campos-Ortega,J.A. and Hartenstein,V. (1985) *The Embryonic Development of Drosophila melanogaster*. Springer Verlag, Berlin, Vol. VIII, pp. 1–227.
Cavener,D.R. (1987) *Nucleic Acids Res.*, **15**, 1353–1361.
Davis,C.G., Goldstein,J.L., Südhof,T.C., Anderson,R.G.W., Russel,D.W. and Brown,M.S. (1987) *Nature*, **326**, 760–765.
Dietrich,U. and Campos-Ortega,J.A. (1984) *J. Neurogenet.*, **1**, 315–332.
Doe,C.Q. and Goodman,C.S. (1985) *Dev. Biol.*, **111**, 193–205.
Foe,V.A. and Alberts,B.M. (1983) *J. Cell Sci.*, **61**, 31–70.
Ghysen,A., Dambly-Chaudiere,C., Aceves,E., Jan,L.Y. and Jan,Y.N. (1986) *Wilhelm Roux's Arch. Dev. Biol.*, **195**, 281–289.
Graf,J., Iwamoto,Y., Sasaki,M., Martin,G.R., Kleinman,H.K., Robey,F.A. and Yamada,Y. (1987) *Cell*, **48**, 989–996.
Gray,A., Dull,T.J. and Ullrich,A. (1983) *Nature*, **303**, 722–725.
Greenwald,I. (1985) *Cell*, **43**, 583–590.
Hartenstein,V. and Campos-Ortega,J.A. (1984) *Wilhelm Roux Arch. Dev. Biol.*, **193**, 308–325.
Hartenstein,V., Technau,G.M. and Campos-Ortega,J.A. (1985) *Wilhelm Roux Arch. Dev. Biol.*, **194**, 213–216.
Holmes,D.S. and Quigley,M. (1981) *Anal. Bichem.*, **14**, 193–197.
Ingham,P.W., Howard,K.R. and Ish-Horowicz,D. (1985) *Nature*, **318**, 439–445.
Jowett,T. (1986) In Roberts,D.B. (ed.), *Drosophila*. IRL Press, Oxford and Washington DC, pp. 275–286.
Kidd,S., Kelley,M.R. and Young,M.W. (1986) *Mol. Cell. Biol.*, **6**, 3094–3108.
Knust,E., Dietrich,U., Tepass,U., Bremer,K.A., Weigel,D., Vässin,H. and Campos-Ortega,J.A. (1987) *EMBO J.*, **6**, 761–766.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
Lehmann,R., Dietrich,U., Jimenez,F. and Campos-Ortega,J.A. (1981) *Wilhelm Roux's Arch. Dev. Biol.*, **190**, 226–229.
Lehmann,R., Jimenez,F., Dietrich,U. and Campos-Ortega,J.A. (1983) *Wilhelm Roux's Arch. Dev. Biol.*, **192**, 62–74.
Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435–1441.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
Pardue,M.L. (1985) In Hames,B.D. and Higgins,S.J. (eds), *Nucleic Acid Hybridization*. IRL Press, Oxford, pp. 179–202.
Pirotta,V., Hadfield,C. and Pretorius,G.H.J. (1983) *EMBO J.*, **2**, 927–934.
Poulson,D.F. (1937) *Proc. Natl. Acad. Sci. USA*, **23**, 133–137.
Poole,S.J., Kauvar,L.M., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37–43.
Reed,K.C. and Mann,D.A. (1985) *Nucleic Acids Res.*, **13**, 7207–7221.
Sabatini,D.D., Kreibich,G., Morimoto,T. and Adesnik,M. (1982) *J. Cell Biol.*, **92**, 1–22.
Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161–178.
Sasaki,M., Kato,S., Kohno,K., Martin,G.R. and Yamada,Y. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 935–939.
Shaw,G. and Kamen,R. (1986) *Cell*, **46**, 659–667.
Sternberg,P.W. and Horvitz,H.R. (1984) *Annu. Rev. Genet.*, **18**, 489–524.
Technau,G.M. and Campos-Ortega,J.A. (1985) *Wilhelm Roux's Arch. Dev. Biol.*, **194**, 196–212.
Technau,G.M. and Campos-Ortega,J.A. (1986) *Wilhelm Roux's Arch. Dev. Biol.*, **195**, 445–454.
Technau,G.M. and Campos-Ortega,J.A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4500–4504.
Vässin,H. and Campos-Ortega,J.A. (1987) *Genetics*, **116**, 433–445.
Vässin,H., Vielmetter,J. and Campos-Ortega,J.A. (1985) *J. Neurogenet.*, **2**, 291–308.
Weigel,D., Knust,E. and Campos-Ortega,J.A. (1987) *Mol. Gen. Genet.*, **207**, 374–384.
Watson,M.E.E. (1984) *Nucleic Acids Res.*, **12**, 5145–5164.
Wharton,K.A., Johansen,K.M., Xu,T. and Artavanis-Tsakonas,S. (1985) *Cell*, **43**, 567–581.
Willimzig,M. (1985)*Trends Genet.*, **1**, 158.
Wright,T.R.F. (1970) *Adv. Genet.*, **15**, 261–395.
Yarden,Y., Escobedo,J.A., Kuang,W.-J., Yang-Feng,T.L., Daniel,T.O., Tremble,P.M., Chen,E.Y., Ando,M.E., Harkins,R.N., Franke,U., Fried,V.A., Ullrich,A. and Wiliams,L.T. (1986) *Nature*, **323**, 226–232.
Yedvobnick,B., Muskavitch,M.A.T., Wharton,K.A., Halpern,M.E., Paul,E., Grinwade,B.G. and Artavanis-Tsakonas,S. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 841–854.

*Received on July 10, 1987; revised on August 4, 1987*