Delta, a Drosophila neurogenic gene, is transcriptionally complex and encodes a protein related to blood coagulation factors and epidermal growth factor of vertebrates

Casey C. Kopczynski, Althea K. Alton,[1] Kim Fechtel, Pamela J. Kooh, and Marc A.T. Muskavitch

Programs in Genetics and Molecular, Cellular, and Developmental Biology, Department of Biology, Indiana University, Bloomington, Indiana 47405 USA

Delta (Dl) is required for normal segregation of the embryonic ectoderm into neural and epidermal cell lineages in *Drosophila melanogaster*. Loss-of-function mutations in *Dl* and other zygotic neurogenic loci lead to expansion of the neuroblast population at the expense of the dermoblast population within the ectoderm. Characterization of the transcriptional organization and maternal/embryonic expression within the chromosomal interval corresponding to *Dl* reveals that the locus encodes multiple transcripts: a minimum of two maternal transcripts, ~4.5 and 3.6 kb in length, and four zygotic transcripts, ~5.4 (two distinct species), 3.5, and 2.8 kb in length. These transcripts differ on the bases of differential splicing and differential polyadenylation site choice. The DNA sequence of a cDNA clone representing the predominant transcripts of the locus indicates that *Dl* encodes a transmembrane protein homologous to blood coagulation factors and epidermal growth factor. The relationship between coding sequences and transcript-specific exons within the locus suggests that *Dl* encodes multiple translational products.

[*Key Words: Drosophila*; neurogenic gene; neurogenesis; ectodermal differentiation; coagulation factor IX; epidermal growth factor]

Neurogenesis in *Drosophila* begins shortly after gastrulation with the inward segregation of ectodermal cells within the neurogenic regions of the embryo (Campos-Ortega and Hartenstein 1985). The internalized ectodermal cells become neuroblasts, the precursors of the neural cell lineage, and the ectodermal cells that remain on the peripheral surfaces of the embryo become dermoblasts, precursors of the epidermal cell lineage. Loss-of-function mutations that affect the zygotic neurogenic genes *Notch (N)*, *Delta (Dl)*, *Enhancer of split [E(spl)]*, *big brain (bib)*, *mastermind (mam)*, or *neuralized (neu)* lead to neural hypertrophy, reduction of the epidermis and embryonic lethality (Poulson 1937; Lehmann et al. 1983). The cellular basis of these defects is the failure to establish the epidermal lineage and the resultant overcommitment to the neural lineage within the developing embryo (Poulson 1937; Lehmann et al. 1983). The role of the neurogenic genes during the initial stages of neurogenesis is therefore to ensure correct partitioning of the ectoderm into neural and epidermal lineages.

Cell ablation studies conducted on the developing neuroectoderm in *Schistocerca* (Doe and Goodman

1985) indicate that cell–cell interactions are central to the establishment of the epidermal identity within the embryonic ectoderm. A similar conclusion was drawn in an elegant study in *Drosophila* that involved transplantation of single ectodermal cells from neurogenic mutant embryos into wild-type hosts (Technau and Campos-Ortega 1987). This study revealed that five of the zygotic neurogenic genes—*N, Dl, bib, mam*, and *neu*—encode products that can act nonautonomously at the level of single cells to affect the ability of ectodermal cells to adopt the epidermal identity.

Molecular analyses of the *N* locus provide evidence that implies a direct role for the *N* gene product in cell–cell interactions. The embryonic *N* product appears to be a transmembrane protein containing a tandem repetitive array comprised of sequences with significant similarity to vertebrate epidermal growth factor (EGF) within its putative extracellular domain (Wharton et al. 1985; Kidd et al. 1986). The putative intracellular domain includes sequences reminiscent of those found in proteins that interact with nucleoside triphosphates and exhibit protein kinase activity. The Notch protein is therefore

[1]Present address: Department of Biological Sciences, Western Illinois University, Macomb, Illinois 61455 USA.

structurally analogous to proteins known or believed to be involved in cell–cell interactions in a number of other organisms, often via mechanisms involving protein–protein contacts (Carpenter 1987; Ruoslahti 1988). The putative extracellular localization of a portion of the Notch protein is consistent with the known ability of this product to function at the cellular level in a non-automomous manner during development.

We have undertaken genetic and molecular analyses of *Dl* to achieve an understanding of the mechanism by which this locus participates in the establishment of cellular identity within the developing ectoderm. Analysis of the expression of the locus reveals that *Dl* produces multiple maternal and zygotic transcripts. We find that the predominant maternal and embryonic *Dl* transcripts appear to encode the same product. This protein exhibits a structure that is consistent with the participation of *Dl* in cell–cell interactions and contains sequence similarities to gene products known to participate in protein–protein contacts. Comparison of the location of coding sequences for the predominant *Dl* product in relation to alternative transcript-specific exonic regions implies that *Dl* may encode a number of distinct translational products maternally and zygotically.

Results

Physical definition of Dl

Our previous genetic and cytogenetic analyses of *Dl* (Alton et al. 1988) revealed that the distal breakpoint of the rearrangement *Df(3R)bxd^110^* (92A2) constitutes a telomere-proximal limit for *Dl*, while the distal breakpoint of the rearrangement Df(3R)ChaM1 (91F5-92A1) constitutes a centromere-proximal limit for *Dl*. Therefore, we isolated ~200 kbp of DNA encompassing this cytogenetic interval by chromosomal walking (Fig. 1; Bender et al. 1983).

Physical lesions within the walk were correlated with specific *Dl* mutations (Fig. 1) by comparative genomic DNA blot analysis of 69 independent *Dl* alleles induced on known parental backgrounds (Alton et al. 1988; S.B. Shepard, S.A. Broverman, and M.A.T. Muskovitch, in prep.). One class of mutations comprises deficiencies that remove portions (*bxd^110^*, *Cha^M9^*, *Cha^M1^*, *Dl^M2^*, *Dl^BX35^*, and *Dl^BX38^*) or the entirety (*Dl* alleles *BX2*, *BX6*, and *BX12*) of the chromosomal walk. Approximate end points have been defined for four deficiencies: *bxd^110^*, *Cha^M9^*, *Dl^M2^*, and *Cha^M1^*. The second class of mutations comprises nondeficiency alleles for which distinct molecular lesions can be defined. The distribution of the 15 independent nondeficiency alleles that map between 95 and 127 kbp strongly suggests that most or all of the DNA required for *Dl* function is within this interval, a hypothesis that is supported by the transcription analysis presented below.

Embryonic transcription of Dl

Subclones representing the entire walk were employed serially as probes to assess the accumulation of polyadenylated [poly(A)⁺] transcripts arising from this interval

of the genome during oogenesis and embryogenesis. This analysis, and further analyses presented below, revealed that *Dl* encodes one predominant maternal transcript (4.5M), one predominant zygotic transcript (5.4Z), one minor maternal transcript (3.6m), and three minor zygotic transcripts (5.4z, 3.5z, and 2.8z) (Figs. 2, 3, and 4). Each of the zygotic *Dl* transcripts exhibits maximum accumulation between 3 and 6 hr of embryogenesis (Fig. 2), immediately preceding and during the major portion of the neuroblast segregation phase (Campos-Ortega and Hartenstein 1985). The designation of 4.5M and 3.6m as maternal transcripts is based on the observation that these transcripts are found in unfertilized egg RNA (Fig. 2). We have not yet determined whether *Dl* transcripts 4.5 kb in length that are present throughout embryogenesis are synthesized during embryogenesis or represent stable, maternally loaded 4.5M.

Poly(A)⁺ RNA populations from unfertilized eggs and staged embryos were probed with a set of 12 adjacent subclones (fragments A–L, Fig. 4) to assess transcriptional structure within *Dl*. This analysis revealed that 5.4Z and 4.5M hybridize to an identical subset of these genomic fragments and that 3.5z, 2.8z, and 3.6m display different patterns of hybridization. Fragments D and F are specific for 3.5z, and fragment I is specific for 2.8z (Fig. 3a). Fragment B hybridizes to a transcript 5.4 kb in length (Fig. 3a). However, this transcript cannot be 5.4Z because fragment B does not hybridize with a cDNA insert that represents most or all of 5.4Z (Dl2, Fig. 3b). We designate this second 5.4-kb transcript 5.4z and infer that it is a minor zygotic *Dl* transcript. In situ hybridizations to embryonic tissue sections with fragment B and a probe that detects 5.4Z reveal that the subcellular localization of 5.4z is different from that of 5.4Z, further confirming that these transcripts are distinct (C.C. Kopczynski and M.A.T. Muskavitch, in prep.). Transcripts 3.6m and 3.5z are distinct because three genomic fragments (D, E, and F) that hybridize to 3.5z fail to hybridize to unfertilized egg transcripts (Fig. 4).

Hybridization at reduced stringency reveals a region of sequence similarity between *Dl* and *N* that maps between a *Pvu*II site 1.9 kb upstream of the junction between fragments J and K and the *Pst*I site between fragments K and L (Fig. 4). The probe employed to detect this similarity consists entirely of coding sequences from *N* that are related to vertebrate EGF (Wharton et al. 1985). This genomic region, which includes *Dl*-coding sequences (see below), is entirely excluded from 3.5z and partially excluded from 2.8z. It may also be excluded from 3.6m, given its proximity to the apparent 3' terminus of 3.6m.

Sequence of the predominant product of Dl

Screening of cDNA libraries derived from 4- to 8-hr embryonic poly(A)⁺ RNA with *Dl* probes yielded a number of clones (Fig. 5). Hybridization of Dl1, Dl2, and Dl3 to subclones representing the entire *Dl* transcription unit (Fig. 5) demonstrates that each insert hybridizes exclusively to those genomic fragments within which exonic
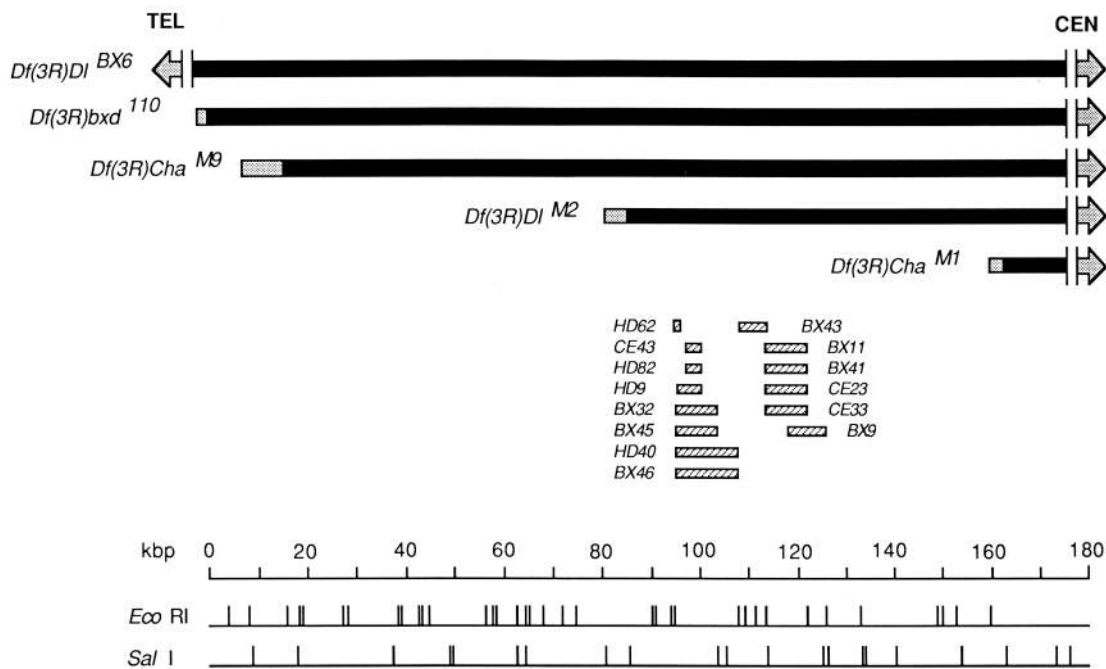
Figure 1. Physical organization of *Dl* and its flanking regions. Filled bars represent sequences removed in the rearrangements denoted to the left of each bar, respectively; stippled areas represent uncertainty in the localization of termini of rearrangements; arrowheads indicate that a deficiency extends beyond the interval depicted. Hatched bars above the coordinate scale represent the maximum extent of the molecular intervals within the walk affected by independent *Dl* mutations (described in Experimental procedures). Allele designations: (*BX*) Bloomington X-ray-induced; (*CE*) Cambridge EMS-induced; (*HD*) hybrid dysgenesis-induced. Coordinates are given in kbp of DNA relative to an arbitrary zero coordinate (*Bam*HI, 0 kbp). The positions of recognition sequences for restriction enzymes *Eco*RI and *Sal*I are depicted below the coordinate scale. Deficiencies bxd^{110} (Dl^-), Cha^{M9} (Dl^-), Dl^{M2} (Dl^-) and Cha^{M1} (Dl^+) have been described in Alton et al. (1988). Limits in our coordinate scheme for individual physical lesions depicted comprise bxd^{110}, between coordinate -1.5 kbp and the *Bam*HI site at 0 kbp; Cha^{M9}, 8.2 kbp *Eco*RI fragment, 7.4–15.6 kbp; Dl^{M2}, 4.9 kbp *Sal*I fragment, 81.9–87.0 kbp; Cha^{M1}, between an *Eco*RI site at 159.9 kbp and coordinate 162.9 kbp; *CS20* and *HD62*, 1.3-kbp *Sac*I fragment, 95.5–96.8 kbp; *CE43* and *HD82*, 3.0-kbp *Bam*HI–*Sac*I fragment, 98.6–101.6 kbp; *HD9*, 4.8-kbp *Sac*I fragment, 96.8–101.6 kbp; *BX32* and *BX45*, 7.7-kbp *Hin*dIII fragment, 96.0–103.7 kbp; *BX46* and *HD40*, 13.2-kbp *Eco*RI fragment, 95.7–108.9 kbp; *BX43*, three contiguous *Eco*RI fragments (1.3, 2.2, and 2.2 kbp), 108.9–114.6 kbp; *BX11*, *BX41*, *CE23*, and *CE33*, 8.5-kbp *Eco*RI fragment, 114.6–123.1 kbp; *BX9*, within 4 kbp of an *Eco*RI site at 123.1 kbp. (CEN) Centromere proximal; (TEL) telomere proximal.

sequences of 5.4Z and 4.5M have been mapped (Fig. 4). Inserts Dl2 and Dl3 appear to constitute full-length representations of 5.4Z and 4.5M, respectively. Analysis of a set of 11 additional cDNA inserts isolated from a library generated using 0- to 4-hr embryonic poly(A)$^+$ RNA, which is expected to be biased toward maternally loaded transcripts, supports the contention that 4.5M represents a maternal transcript (data not shown). Restriction map analysis indicates that Dl1 represents a transcript that is colinear with those represented by Dl2 and Dl3. DNA sequence data are consistent with the premise that Dl1 represents a partial copy of a *Dl* transcript because the 3' terminus of the insert does not contain sequences indicative of a poly(A)$^+$ tail.

The sequence of the 3' terminus of Dl3 indicates that the insert terminates at a site downstream of a polyadenylation consensus sequence (AATATA) identical to that utilized in the mouse α-amylase gene (Tosi et al. 1981). Restriction mapping and hybridization data place the 3' terminus of Dl2 at the same position as the 3' terminus of the zygotic transcript described previously by Vassin et al. (1987). The apparent polyadenylation

site for 5.4Z is ~800 bases downstream from that of 4.5M, and the 5' terminal extents of Dl2 and Dl3 are comparable (Fig. 5). These data suggest Dl2 and Dl3 represent transcripts that initiate at a common site and differ solely on the basis of polyadenylation site choice.

The sequence of Dl1 (Fig. 6) reveals that the transcript it represents has the potential to encode a putative transmembrane protein with significant sequence similarity to blood coagulation factor IX (FIX; Choo et al. 1982; Kurachi and Davie 1982) and vertebrate EGF (Gray et al. 1983; Sudhof et al. 1985). The deduced protein is 832 amino acids in length and contains a putative signal peptide (residues 1–25; Perlman and Halvorson 1983), extracellular domain (residues 26–595), transmembrane domain (residues 596–617; Kyte and Doolittle 1982); stop–transfer sequence (residues 620–625; Blobel 1980), and intracellular domain (residues 618–832). The sequence also contains five sites (residues 98, 137, 167, 421, and 649) exhibiting the consensus observed for positions of asparagine-linked glycosylation in proteins that are modified during passage through the rough endoplasmic reticulum and Golgi apparatus (Hirschberg
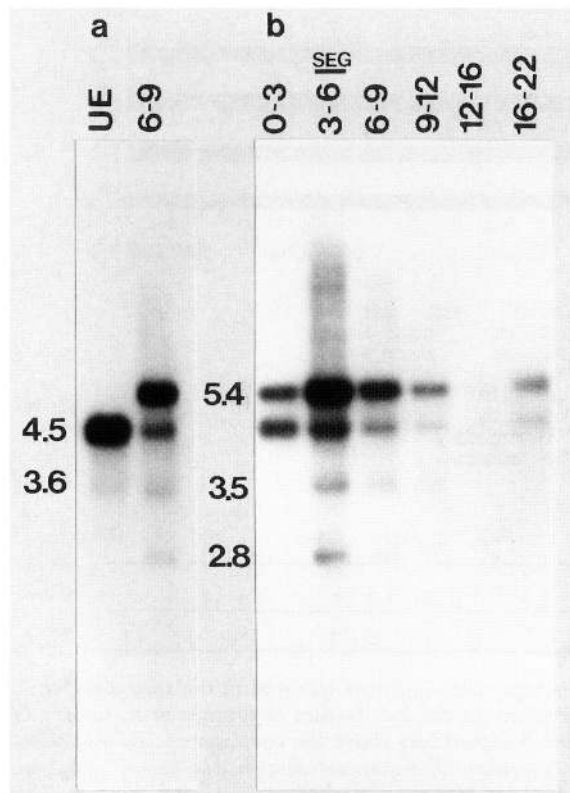
Figure 2. Temporal accumulation of *Dl* transcripts during oogenesis and embryogenesis. RNA blots were probed with fragment G (Fig. 4). Transcript lengths are given in kb of RNA. (SEG) Interval during which the majority of central nervous system neuroblast segregation occurs (Campos-Ortega and Hartenstein 1985). (*a*) Ten micrograms of poly(A)+ RNA from unfertilized eggs (UE), compared to 2 μg of poly(A)+ RNA from a mixed population of embryos, 6- to 9-hr postoviposition. (*b*) Each lane contains 3 μg poly(A)+ RNA prepared from mixed populations of embryos of the designated ages in hours postoviposition. Probing of same blot for *rp49* RNA (O'Connell and Rosbash 1984) suggests that the apparent increase in transcript levels in the 16- to 22-hr sample results from a higher proportion of poly(A)+ RNA in this sample, relative to poly(A)− contamination, than in other samples (data not shown).

and Snider 1987).

Hybridization of the 5′ terminal *Eco*RI–*Alu*I fragment from Dl1 to genomic subclones places the start codon and most or all of the signal peptide in fragment A. Genomic DNA sequence analysis places the stop codon in fragment L (Figs. 4 and 5). Comparison of the positions of the putative start (AUG) and stop (TAA) codons for this product in relation to the structures of Dl1, Dl2, and Dl3 implies that the predominant maternal and zygotic *Dl* transcripts share the same coding capacity with Dl1 (Figs. 5 and 6). We refer to this putative translational product of *Dl* as Dl5.4Z/4.5M (DlZM). The sequence we have determined is substantially, although not entirely, in agreement with that reported by Vassin et al. (1987) for a product of the *Dl* locus (see Table 1 and Discussion).

The most striking feature of the sequence of DlZM is the presence of an array of cysteine-rich repeats that ex-

tends from residue 217–566 within the putative extracellular domain of the protein (Figs. 6, 7, and 8). This internally repetitive array (Figs. 6 and 7) is detectable in a DOTPLOT self-comparison of DlZM (data not shown). Use of the FASTP sequence comparison programs (Lippmann and Pearson 1985) to analyze the similarity between the DlZM cysteine-rich array and protein sequences entered in the Protein Information Resource (PIR) data base yields a number of significant matches. The best match is the previously described EGF-related array of the N gene product (Figs. 7 and 8; Wharton et al. 1985; Kidd et al. 1986). This result is consistent with the observed cross-hybridization between *Dl* and *N* (Fig. 4). Another significant match is detected with the EGF-related array of the *lin-12* gene product described in the nematode *Caenorhabditis elegans* (Greenwald 1985). Surprisingly, significant matches are also detected with FIX of the vertebrate blood coagulation pathway (Fig. 7), as well as a number of other proteins in the pathway, including factor X, factor VII, factor XII, protein S, protein C, and von Willebrand factor (Furie and Furie 1988). Remaining significant matches are observed with EGF precursor (Gray et al. 1983; Scott et al. 1983), low-density lipoprotein receptor (Sudhof et al. 1985), β-integrin (Fig. 8; Takmun et al. 1986), leukocyte adhesion protein β-chain (Kishimoto et al. 1987), thrombospondin (Lawler and Hynes 1986), and plasminogen activators (Ny et al. 1984; Holmes et al. 1985). No highly significant matches are detected between other portions of DlZM and sequences currently compiled in the PIR data base.

Further scrutiny of repetitive sequences within the DlZM and Notch proteins reveals that consensus sequences that yield ungapped alignments with a region of FIX can be derived from DlZM and Notch repeat arrays (Fig. 7). The previously noted similarity to EGF (Wharton et al. 1985; Kidd et al. 1986; Vassin et al. 1987), based on the repetitive motif containing six cysteines, which is the hallmark of EGF-related proteins, is also embedded within these alignments. The DlZM, Notch, and FIX consensus sequences contain another motif (Fig. 7) that has been correlated with the β-hydroxylation of aspartate and asparagine in a number of EGF-related proteins (Stenflo et al. 1987). The biological significance of this modification is not currently understood.

Discussion

Multiple alternative transcripts arise from *Dl* during oogenesis and embryogenesis

The temporal pattern of transcript accumulation from the *Dl* locus during embryogenesis is consistent with that expected for a zygotic neurogenic gene based on developmental, genetic, and comparative criteria. The predominant embryonic transcript 5.4Z and the minor embryonic transcripts 5.4z, 3.5z, and 2.8z exhibit maximum accumulation between 3 and 6 hr of embryogenesis and cannot be detected in unfertilized eggs. Their accumulation therefore precedes and encompasses the period of neuroblast segregation and establishment of the epidermal lineage within embryonic
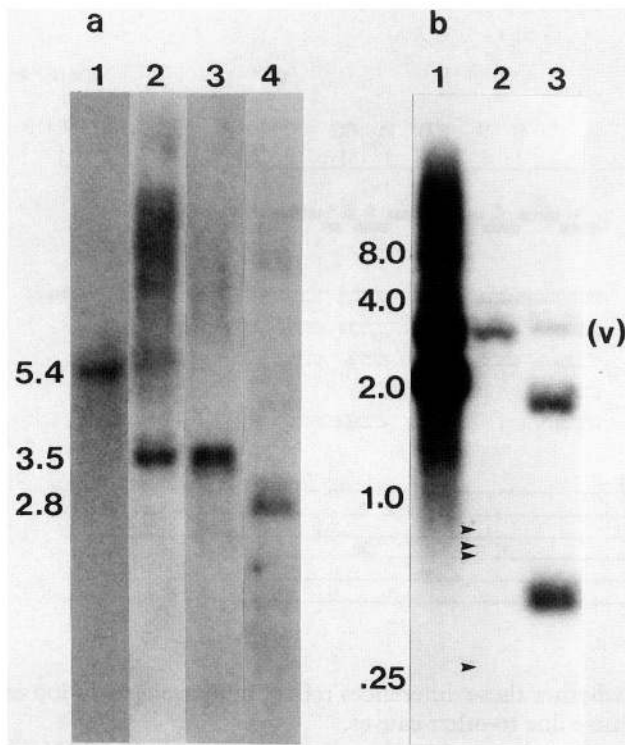
Figure 3. Identification of genomic regions specific for minor *Dl* zygotic transcripts. (*a*) RNA blots probed with *Dl* genomic fragments (Fig. 4). (Lane *1*) Fragment B; (lane *2*) fragment D; (lane *3*) fragment F; (lane *4*) 1.7-kbp *SalI* fragment that consists of fragment I and the adjacent 0.5 kbp of fragment J. Lanes contain 3 μg of 3- to 6-hr embryonic poly(A)+ RNA (lanes *1–3*) or 3- to 4-hr embryonic poly(A)+ RNA (lane *4*). Transcript lengths are given in kb of RNA. (*b*) DNA blot of genomic subclones probed with D12. (Lane *1*) Fragment A subclone digested with *Bam*HI, *Sac*I, and *Ava*I; (lane *2*) fragment B subclone digested with *Sac*I and *Hinc*II; (lane *3*) fragment C digested with *Sal*I and *Pst*I. Arrowheads in lane *2* mark positions of ethidium-bromide-stained bands representing the entirety of fragment B. A scale for DNA fragment length is given in kbp. (v) Hybridization to vector DNA due to contamination of the probe with vector sequences.

neurogenic regions (Campos-Ortega and Hartenstein 1985). This maximum also overlaps with the embryonic phenocritical period for *Dl* function, from 1 to 6 hr post-oviposition (Lehmann et al. 1983). Investigations of the expression of two other neurogenic loci, *N* (Grimwade et al. 1985) and *E(spl)* (Preiss et al. 1988), indicate that these loci encode transcripts that exhibit maximum accumulation during the first third of embryogenesis (2–7 hr postoviposition). The embryonic pattern of *Dl* transcript accumulation is consistent with the hypothesis that some or all of these transcripts encode products required for the regulation of the segregation of the ectoderm into neural and epidermal lineages.

The accumulation of the maternal *Dl* transcripts is consistent with observations suggesting that *Dl* encodes a maternal component (Dietrich and Campos-Ortega

1984). The minor maternal transcript, 3.6m, appears to accumulate in unfertilized eggs, but not in embryos. The predominant maternal transcript, 4.5M, is present in embryos long after the disappearance of other transcripts within the walk that appear to accumulate in unfertilized eggs, but not in early embryos (Alton et al. 1989). Although it is possible that 4.5M is an extremely stable transcript, this transcript is probably also synthesized during embryogenesis.

Transcription of *Dl* yields a family of alternatively processed maternal and embryonic transcripts. We infer that genomic fragments that specifically cross-hybridize to different minor zygotic transcripts contain exons specific to each of these transcripts, respectively (Fig. 4). This implies that alternative splicing contributes to the transcriptional complexity of *Dl*. We also note that the 3'-terminal extent of hybridization differs for different transcripts, consistent with the premise that alternative polyadenylation site choice also contributes to structural variation among transcripts. We know this to be the case for 5.4Z and 4.5M (Fig. 5). Although we cannot yet eliminate the possibility that minor transcripts include sequences downstream of the 3' terminus of 5.4Z, we do not detect any of the minor transcripts with probes covering an interval extending 30 kb downstream of this point. It is noteworthy that *Dl* encodes distinct maternal and embryonic transcripts because few *Drosophila* genes are known to produce transcripts that differ in structure during these two phases of the life cycle (Tautz et al. 1987; Preiss et al. 1988; Alton et al. 1989).

The structure of the predominant Dl *maternal/embryonic product implies participation in protein–protein interactions*

The predominant zygotic and maternal *Dl* transcripts appear to encode the same polypeptide, which we have designated DlZM in light of its apparent dual-stage expression. The deduced amino acid sequence of this protein implies that the predominant product of *Dl* is a transmembrane protein that possesses a putative extracellular domain with extensive similarity to EGF and FIX of vertebrates. This structure is, in large part, analogous to that deduced for the *Notch* protein on the basis of DNA sequence analyses (Fig. 8; Wharton et al. 1985; Kidd et al. 1986). This extensive structural analogy and sequence similarity implies that *Dl* and *N* are homologous genes, derived by descent from a common ancestor.

Comparison of the sequence we have determined for DlZM with that previously published by Vassin et al. (1987) yields unanticipated, as well as expected, differences (Table 1). We detect instances of neutral third base pair variation in 14 codons within the coding sequences that overlap in the two sequences, as would be expected on the basis of interstrain variation. We also detect two relative missense alterations, one of which is conservative (A ↔ V), the other of which is not (K ↔ N). The amino acid assignment we have determined at one of these positions (K) appears to be more consistent with the putative signal peptide context within which this

Figure 4. Transcriptional structure of *Dl*. The interval of the chromosomal walk from coordinate 91–133 kbp is depicted. Each arrow represents a composite of those restriction fragments that cross-hybridize to a transcript(s) of the indicated length; vertical lines within arrows represent termini of restriction fragments analyzed (solid bars A–L). Hatching in the box underlying fragment B, documenting hybridization to a 5.4-kb RNA, indicates that this hybridization is specific to a minor zygotic transcript (5.4z). The stippled box represents an interval within which EGF-related sequences from within the *Notch* locus cross hybridize (see Experimental procedures). (kb) Kilobases of RNA; (M/m) accumulates in eggs during oogenesis; (Z/z) accumulation predominantly zygotic. The positions of recognition sequences for restriction enzymes are represented by vertical strokes on appropriate lines; (R) *Eco*RI; (S) *Sal*I. (CEN) Centromere proximal; (TEL) telomere proximal.
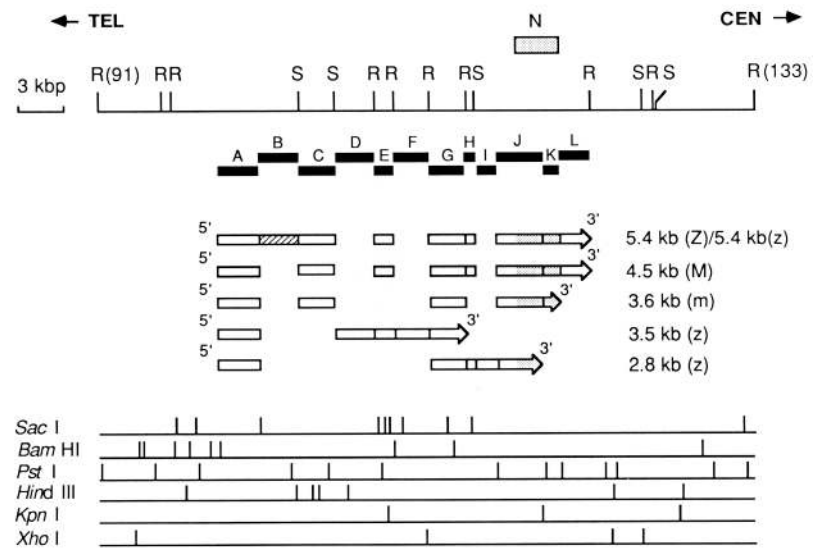

residue resides. Dramatic differences arise as the result of a number of apparent frameshift alterations near the carboxyl termini of the two deduced sequences. We deduce the presence of a terminal methionine residue (position 832) in place of an internal leucine residue at the corresponding position in the previously published sequence. Thus, our sequence implies a protein of 832 amino acids, whereas that of Vassin et al. (1987) implies an 880-residue polypeptide. We also define an internal tetrapeptide (ACSS, positions 801–804) that differs from the corresponding interval of the previously published sequence. We are currently attempting to determine

whether these differences reflect interstrain variation or arise due to other causes.

We propose that the primary structure of DlZM reflects the participation of this molecule in protein–protein interactions that are essential for establishment of the epidermal lineage within embryonic neurogenic regions. This hypothesis is based largely on the observed similarity of the tandemly repeated sequences within the extracellular domain to FIX and EGF. The optimum alignment between DlZM and FIX involves the ninth complete repeat of the DlZM array (Fig. 7), a repeat that exhibits 47% identity (18 of 38 residues) with the

Figure 5. Selected cDNA inserts representing *Dl* transcripts and their relationship to the genome. Thick bars represent cDNA inserts; the line in the lower portion represents a part of the chromosomal walk (coordinates 95–133 kbp; Fig. 1). The correlations of restriction sites in cDNA segments and the genome, indicated by connecting lines, are based on a combination of cDNA/genomic DNA hybridization analysis and cDNA sequence data (not shown). Open boxes below the genomic map represent cross-hybridization between genomic fragments and transcripts 5.4Z and 4.5M, as in Fig. 4, with two exceptions. The 5'-terminal extent of these two transcripts has been delimited to a 1.4-kbp *Xba*I–*Ava*I fragment within fragment A (Fig. 4). The 3' terminus of 4.5(M) has been localized to a position ~800 bp upstream of the *Eco*RI site defining the 3' terminus of fragment L (Fig. 4) on the basis of RNA blot analyses and sequence comparisons with the data of Vassin et al. (1987). Positions of recognition sequences for restriction enzymes are represented by vertical strokes on appropriate lines: (B) *Bam*HI; (G) *Bgl*II; (H) *Hin*cIII; (K) *Kpn*I; (P) *Pst*I; (R) *Eco*RI; (S), *Sal*I; (R), *Eco*RI

linker introduced during cloning of Dl1. ATG and TAA represent putative start and stop codons, respectively, as defined in Dl1 by DNA sequence analysis (Fig. 6) and in fragments A and L by cDNA/genomic cross hybridization analysis (Fig. 4). (A)$_n$ represents the site of polyadenylation based on our sequence data (for Dl3) or those of Vassin et al. (1987) (for a cDNA that apparently represents a partial copy of Dl2).
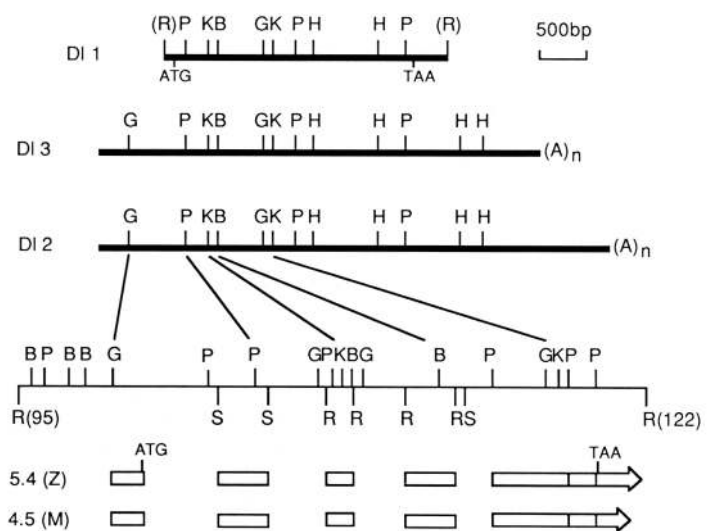
aligned consensus derived from human and bovine FIX sequences (residues 47–84; Choo et al. 1982; Kouichi et al. 1979; Kurachi and Davie 1982). The ninth repeat also yields the optimum alignments defined by the FASTP comparison programs between DlZM and both factor X and factor VII. Rees et al. (1988) have shown that site-directed mutagenesis of any one of the aligned residues within FIX corresponding to residues 529, 531, or 546 of DlZM (rep9, Fig. 7) abolishes the stimulation of an enzymatic activity of FIX by other clotting factors in vitro. These investigators infer that this region of FIX participates in a protein–protein interaction that involves members of the coagulation pathway (Rees et al. 1988). The repeats of the extracellular domain are also similar in primary structure to EGF. The EGF peptide hormone is involved in a protein–protein interaction with the EGF receptor (Carpenter and Cohen 1979), and the EGF-like motif has been detected in a number of other proteins known or believed to participate in protein–protein interactions (Fig. 8; Gray et al. 1983; Scott et al. 1983; Sudhof et al. 1985; Wharton et al. 1985; Kidd et al. 1986; Furie and Furie 1988; Jones et al. 1988; Montell and Goodman 1988). The extracellular domain of DlZM therefore exhibits similarity to structural motifs that have been implicated in protein–protein interactions in vertebrates.

Given the possibility that the DlZM extracellular domain participates in protein–protein interactions, the structure of DlZM provides a molecular rationale for the observed nonautonomous action of *Dl* in cell transplantation experiments (Technau and Campos-Ortega 1987). The accessibility of this domain of DlZM to the extracellular compartment, in either a free or membrane associated state, would permit it to act extrinsically to those cells in which it is synthesized. Indeed, DlZM exhibits structural analogy to members of the peptide hormone receptor family, as well as to proteins that are found in or interact with the extracellular matrix (Fig. 8). Our results are consistent with the premise, previously set forth by Vassin et al. (1987), that the participation of DlZM in protein–protein interactions is essential for cell–cell interactions required for acquisition of epidermal identity within neurogenic regions of the embryo.

Dl *may encode multiple translational products*

We have shown that transcript-specific exons for the minor zygotic *Dl* transcripts map between the initiation and termination codons of DlZM (Figs. 4 and 5). If such an exon contains coding sequences, any polypeptide encoded by the appropriate transcript must contain amino acids not present in DlZM. If such an exon does not contain coding sequences, then any polypeptide encoded by the appropriate transcript must initiate or terminate at a position that differs from that of DlZM. The exclusion of portions of the DlZM-coding sequences corresponding to the FIX/EGF-like repeats from different minor zygotic transcripts also implies that these transcripts must differ in coding potential from 5.4Z and 4.5M (Fig. 4). Therefore, *Dl* may encode alternative translational products in addition to DlZM, one or more of which may also be essential for regulation of the proper partitioning of ectodermal cells into neural and epidermal lineages during *Drosophila* embryogenesis.

Experimental procedures

Drosophila *stocks*

The *Dl* alleles employed in this study are described in Yedvobnick et al. (1985); Alton et al. (1988, 1989) and S.B Shepard, S.A. Broverman, and M.A.T. Muskavitch (in prep.). Markers and chromosomes are described in Lindsley and Grell (1968), except for *TM6B, Hu e Tb ca (TM6B)*, which is described in Craymer (1984).

Chromosomal walking

Chromosomal walking was performed by standard techniques (Bender et al. 1983) using a genomic library generated (Maniatis et al. 1982) in the bacteriophage λEMBL3 vector from chromosomal DNA of a *Drosophila* strain isogenic for a single third chromosome (*sse⁴ ro*). Selected DNA fragments along the walk were subcloned into either Bluescribe or Bluescript plasmid vectors (Stratagene, Inc.), as appropriate. The initial probe for the walk within 92A1-2 was obtained by isolating a genomic fusion fragment resulting from the rearrangement *Tp(3)bxd¹¹⁰* in which the chromosomal interval extending from 91D1 to 92A2 is inserted into the *bithoraxoid (bxd)* domain of the Bithorax complex (BX-C) (Lindsley and Grell 1968). The *bxd* genomic segment Dm3106 was employed to isolate a 16.2-kb *Sal*I fusion fragment from a bacteriophage λEMBL3 genomic library generated using DNA prepared from adult flies heterozygous for the transposition. In situ hybridization of the fusion fragment to wild-type polytene chromosomes demonstrated that it hybridized to 89E1-4 (BX-C) and 92A1-2 (*Dl*).

DNA blot analysis

Molecular lesions associated with specific mutations were localized by comparative genomic DNA blot analysis, using labeled genomic subclones as probes. *Drosophila* genomic DNA was isolated from six to eight flies, as described previously (Lis et al. 1983), except for elimination of organic extractions and RNase treatment, and resuspended in distilled water overnight at 4°C. DNA was digested with appropriate enzymes and size-fractionated on 0.8% (wt/vol) agarose gels. Fragments were transferred from both sides of the gel (Maniatis et al. 1982) to two sheets of nylon membrane, using buffers described for alkaline transfer to GeneScreen Plus (New England Nuclear, Inc.). Filters were dried overnight and pretreated at 37°C for 2–6 hr in 50% (vol/vol) formamide, 5× SSC, 50 mM Tris-HCl (pH 8.0), 10× Denhardt's, and 1% (wt/vol) SDS (mix A). Hybridizations were performed in a fresh aliquot of mix A with the addition of 0.25 mg/ml sheared denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate (Pharmacia; 500,000), and ^{32}P-labeled nick-translated probe (final concentration of 10^7 cpm/ml) at 37°C for 36 hr. Filters were washed at 60°C with three changes of 0.1× SSC and 0.1% (wt/vol) SDS prior to autoradiography. Cross-hybridization between *N* and *Dl* was detected by probing subclones of genomic fragments within *Dl* at reduced stringency (Knust et al. 1987), using a 3.0-kbp *Bgl*II fragment from *N* composed entirely of EGF-related coding sequences (Wharton et al. 1985).

Table 1. *Comparison of sequences for cDNA inserts representing* Dl *transcripts*

Position/Sequence[a]		Alteration[b]	Coding impact[c]	Verification[d]		
ref. A	ref. B			DSC	7dG	SSG
44-5(−)	35(G)	deletion	none	+	+	
150(A)	141(C)	missense	K↔N	+		
246(C)	237(T)	tpbc	none	+		
549(T)	540(C)	tpbc	none	+		
612(G)	603(A)	tpbc	none	+		
861(A)	852(G)	tpbc	none	+		
894(G)	885(A)	tpbc	none	+		
981(T)	972(A)	tpbc	none	+		
1009(C)	1000(T)	tpbc	none	+		
1413(A)	1404(C)	tpbc	none	+		
1482(A)	1473(G)	tpbc	none	+		
1506(T)	1497(C)	tpbc	none	+		
2391(C)[e]	2382(T)	tpbc	none	+		
2415(A)	2406(C)	tpbc	none	+		
2469(T)	2460(C)	tpbc	none	+		
2550(G)	2541(A)	tpbc	none	+		
2536-7(−)[f,g]	2528(A)	deletion	frameshift	+	+	+
2548(C)[g]	2539-40(−)	addition	frameshift	+	+	+
2570(C)	2561(T)	missense	A↔V	+		
2628-9(−)[h]	2619(T)	deletion	frameshift	+	+	+
2661(T)[i]	2653(G)	transition	none	+	+	+
2746-7(−)	2739(C)	deletion	none	+	+	+

[a] Nucleotide positions are presented as designated in the respective publications. (Ref. A) C.C. Kopczynski et al., this work (Ref. B) Vassin et al. 1987. The identity of the nucleotide of interest is given in parentheses. (−) Nucleotide absent from one sequence relative to the other sequence.

[b] Deletions and additions noted are for our sequence relative to that of Vassin et al. (1987). (tpbc) Third position base change in codon.

[c] When differences occur in the identity of a given amino acid residue, the amino acid encoded by the sequence presented in this work is given to the left in the column.

[d] (DSC) Double-strand sequencing of the Dl1 cDNA insert, as described herein; (7dG) single-strand or double-strand sequencing of the Dl1 cDNA insert in the presence of 7-deazaguanine nucleoside triphosphate, as described in Mizusawa et al. (1986); (SSG) single-strand sequencing of genomic fragments K and L (Fig. 4), as described herein; (+) this method was employed to confirm the sequence difference noted.

[e] Sequence presented in this work also verified by restriction mapping of a *Sac*II recognition site specific to our sequence within the Dl1 cDNA insert and genomic fragment L (Fig. 4).

[f] Sequence presented in this work also verified by restriction mapping of an *Nla*IV recognition site specific to our sequence within the Dl1 cDNA insert.

[g] Sequence presented in this work also verified by single-strand sequencing within genomic fragment K (Fig. 4), as described by Mizusawa et al. (1986).

[h] Sequence presented in this work also verified by restriction mapping of an *Nde*I recognition site specific to our sequence within the Dl1 cDNA insert and genomic fragment L (Fig. 4).

[i] Our sequence data indicate the presence of a G at this position in genomic fragment L (Fig. 4).

Figure 6. DNA sequence of the cDNA insert Dl1 and the deduced primary amino acid sequence of DlZM. The coordinate of the first nucleotide in each row is listed at *left*; the coordinate of the last deduced amino acid in each row is listed at *right*. The dashed box indicates the core of a putative signal sequence (Perlman and Halvorson 1983); the solid box indicates a putative transmembrane domain (Kyte and Doolittle 1982). Bold underlining underscores cysteine-rich internal repeat elements with sequence similarity to FIX and EGF (see Fig. 7). Circles encompass asparagine residues within contexts defined previously for potential N-linked glycosylation sites (Hirschberg and Snider 1987). (*) The putative stop codon. Additions and deletions between this sequence and that of Vassin et al. (1987), differences altering coding capacity and codon third position changes with no impact on coding capacity are described in Table 1.

```
   1  GGAGGAATTATTCAAAACATAAACACAATAAACAATTTGAGTAGTTGCCGCACACACACACACACAGCCCGTGGATTATTACACTAAAAGCGACACTCAATCCAAAAAATCAGCAACA
 121  AAAACATCAATAAACATGCATTGGATTAAATGTTTATTAACAGCATTCATTTGCTTCACAGTCATCGTGCAGGTTCACAGTTCCGGCAGCTTTGAGTTGCGCCTGAAGTACTTCAGCAAC
                          M   H   W   I   K  [C   L   L   T   A   F   I   C   F   T   V   I   V   Q   V] H   S   S   G   S   F   E   L   R   L   K   Y   F   S   N  35

 241  GATCACGGGCGGGACAACGAGGGTCGCTGCTGCAGCGGGGAGTCGGACGGAGCGACGGGCAAGTGCCTGGGCAGCTGCAAGACGCGGTTCGGCTCTGCCTAAAGCACTACCAGGCCACC
       D   H   G   R   D   N   E   G   R   C   C   S   G   E   S   D   G   A   T   G   K   C   L   G   S   C   K   T   R   F   R   L   C   L   K   H   Y   Q   A   T  75

 361  ATCGACACCACCTCCCAGTGCACCTACGGGGACGTGATCACGCCCATTCTCGGCGAGAACTCGGTCAATCTGACCGACGCCCAGCGCTTCCAGAACAAGGGCTTCACGAATCCCATCCAG
       I   D   T   T   S   Q   C   T   Y   G   D   V   I   T   P   I   L   G   E   N   S   V  (N)  L   T   D   A   Q   R   F   Q   N   K   G   F   T   N   P   I   Q 115

 481  TTCCCCTTCTCGTTCTCATGGCCGGGTACCTTCTCGCTGATCGTCGAGGCCTGGCATGATACGAACAATAGCGGCAATGCGCGAACCAACAAGCTCCTCATCCAGCGACTCTTGGTGCAG
       F   P   F   S   F   S   W   P   G   T   F   S   L   I   V   E   A   W   H   D   T  (N)  N   S   G   N   A   R   T   N   K   L   L   I   Q   R   L   L   V   Q 155

 601  CAGGTACTGGAGGTGTCCTCCGAATGGAAGACGAACAAGTCGGAATCGCAGTACACGTCGCTGGAGTACGATTTCCGTGTCACCTGCGATCTCAACTACTACGGATCCGGCTGTGCCAAG
       Q   V   L   E   V   S   S   E   W   K   T  (N)  K   S   E   S   Q   Y   T   S   L   E   Y   D   F   R   V   T   C   D   L   N   Y   Y   G   S   G   C   A   K 195

 721  TTCTGCCGGCCCCGCGACGATTCATTTGGACACTCGACTTGCTCGGAGACGGGCGAAATTATCTGTTTGACCGGATGGCAGGGCGATTACTGTCACATACCCAAATGCGCCAAAGGCTGT
       F   C   R   P   R   D   D   S   F   G   H   S   T   C   S   E   T   G   E   I   I   C   L   T   G   W   Q   G   D   Y   C   H   I   P   K   C   A   K   G   C 235

 841  GAACATGGACATTGCGACAAACCCAATCAATGCGTTTGCCAACTGGGCTGGAAGGGAGCCTTGTGCAACGAGTGCGTTCTGGAACCGAACTGCATCCATGGCACCTGCAACAAACCCTGG
       E   H   G   H   C   D   K   P   N   Q   C   V   C   Q   L   G   W   K   G   A   L   C   N   E   C   V   L   E   P   N   C   I   H   G   T   C   N   K   P   W 275

 961  ACTTGCATCTGCAACGAGGGTTGGGGAGGCTTGTACTGCAACCAGGATCTGAACTACTGCACCAACCACAGACCCTGCAAGAATGGCGGAACCTGCTTCAACACCGGCGAGGGATTGTAC
       T   C   I   C   N   E   G   W   G   G   L   Y   C   N   Q   D   L   N   Y   C   T   N   H   R   P   C   K   N   G   G   T   C   F   N   T   G   E   G   L   Y 315

1081  ACATGCAAATGCGCTCCAGGATACAGTGGTGATGATTGCGAAAATGAGATCTACTCCTGCGATGCCGATGTCAATCCCTGCCAGAATGGTGGTACCTGCATCGATGAGCCGCACACAAAA
       T   C   K   C   A   P   G   Y   S   G   D   D   C   E   N   E   I   Y   S   C   D   A   D   V   N   P   C   Q   N   G   G   T   C   I   D   E   P   H   T   K 355

1201  ACCGGCTACAAGTGTCATTGCCGCAACGGCTGGAGCGGAAAGATGTGCGAGGAGAAAGTGCTCACGTGTTCGGACAAACCCTGTCATCAGGGAATCTGCCGCAACGTTCGTCCTGGTTTG
       T   G   Y   K   C   H   C   R   N   G   W   S   G   K   M   C   E   E   K   V   L   T   C   S   D   K   P   C   H   Q   G   I   C   R   N   V   R   P   G   L 395

1321  GGAAGCAAGGGTCAGGGCTACCAGTGCGAATGTCCCATTGGCTACAGCGGACCCAACTGCGATCTCCAGCTGGACAACTGCAGTCCGAATCCATGCATAAACGGTGGAAGCTGTCAGCCG
       G   S   K   G   Q   G   Y   Q   C   E   C   P   I   G   Y   S   G   P   N   C   D   L   Q   L   D  (N)  C   S   P   N   P   C   I   N   G   G   S   C   Q   P 435

1441  AGCGGAAAGTGTATTTGCCCCAGCGGATTTTCGGGAACGAGATGCGAGACCAACATTGACGATTGTCTTGGCCACCAGTGCGAGAACGGAGGCACCTGCATAGATATGGTCAACCAATAT
       S   G   K   C   I   C   P   S   G   F   S   G   T   R   C   E   T   N   I   D   D   C   L   G   H   Q   C   E   N   G   G   T   C   I   D   M   V   N   Q   Y 475

1561  CGCTGCCAATGCGTTCCCGGTTTCCATGGCACCCACTGTAGTAGCAAAGTTGACTTGTGCCTCATCAGACCGTGTGCCAATGGAGGAACCTGCTTGAATCTCAACAACGATTACCAGTGC
       R   C   Q   C   V   P   G   F   H   G   T   H   C   S   S   K   V   D   L   C   L   I   R   P   C   A   N   G   G   T   C   L   N   L   N   N   D   Y   Q   C 515

1681  ACCTGTCGTGCGGGATTACTGGCAAGGATTGCTCCGTGGACATCGATGAGTGCAGCAGTGGACCCTGTCATAACGGCGGCACTTGCATGAACCGCGTCAATTCGTTCGAATGCGTGTGT
       T   C   R   A   G   F   T   G   K   D   C   S   V   D   I   D   E   C   S   S   G   P   C   H   N   G   G   T   C   M   N   R   V   N   S   F   E   C   V   C 555

1801  GCCAATGGTTTCAGGGGCAAGCAGTGCGATGAGGAGTCCTACGATTCGGTGACCTTCGATGCCCACCAATATGGAGCGACCACACAAGCGAGAGCCGATGGTTTAGCCAATGCCCAGGTA
       A   N   G   F   R   G   K   Q   C   D   E   E   S   Y   D   S   V   T   F   D   A   H   Q   Y   G   A   T   T   Q   A   R   A   D   G   L   A   N   A   Q   V 595

1921  GTCCTAATTGCTGTTTTCTCCGTTGCGATGCCTTTGGTGGCGGTTATTGCGGCGTGCGTGGTCTTCTGCATGAAGCGCAAGCGTAAGCGTGCTCAGGAAAAGGACAACGCGGAGGCCAGG
      [V   L   I   A   V   F   S   V   A   M   P   L   V   A   V   I   A   A   C   V   V   F] C   M   K   R   K   R   K   R   A   Q   E   K   D   N   A   E   A   R 635

2041  AAGCAGAACGAACAGAATGCAGTGGCCACAATGCATCACAATGGCAGTGCGGTGGGTGTAGCTTTGGCTTCAGCCTCTATGGGCGGCAAAACTGGCAGCAACAGCGGTCTCACCTTCGAT
       K   Q   N   E   Q   N   A   V   A   T   M   H   H  (N)  G   S   A   V   G   V   A   L   A   S   A   S   M   G   G   K   T   G   S   N   S   G   L   T   F   D 675

2161  GGCGGCAACCCGAATATCATCAAAAACACCTGGGACAAGTCGGTCAACAACATTTGTGCCTCAGCAGCAGCAGCGGCGGCGGCGGCAGCAGCGGCGGACGAGTGTCTCATGTACGGCGGA
       G   G   N   P   N   I   I   K   N   T   W   D   K   S   V   N   N   I   C   A   S   A   A   A   A   A   A   A   A   A   A   A   D   E   C   L   M   Y   G   G 715

2281  TATGTGGCCTCGGTGGCGGATAACAACAATGCCAACTCAGACTTTTGTGTGGCTCCGCTACAAAGAGCCAAGTCGCAAAAGCAACTCAACACCGATCCCACGCTCATGCACCGCGGTTCG
       Y   V   A   S   V   A   D   N   N   N   A   N   S   D   F   C   V   A   P   L   Q   R   A   K   S   Q   K   Q   L   N   T   D   P   T   L   M   H   R   G   S 755

2401  CCGGCAGGCACGTCAGCCAAGGGAGCGTCTGGCGGAGGACCGGGAGCGGCGGAGGGCAAGAGGATCTCTGTTTTAGGCGAGGGTTCCTACTGTAGCCAGCGTTGGCCCTCGTTGGCGGCG
       P   A   G   T   S   A   K   G   A   S   G   G   G   P   G   A   A   E   G   K   R   I   S   V   L   G   E   G   S   Y   C   S   Q   R   W   P   S   L   A   A 795

2521  GCGGGAGTGGCCGGAGCCTGTTCATCCCAGCTAATGGCTGCAGCTTCGGCAGCGGGCACGGACGGGACGGCGCAACAGCAGCGATCCGTGGTCTGCGGCACTCCGCATATGTAACTCCAA
       A   G   V   A   G   A   C   S   S   Q   L   M   A   A   A   S   A   A   G   T   D   G   T   A   Q   Q   Q   R   S   V   V   C   G   T   P   H   M   *           832

2641  AAATCCGGAAGGGCTCCTGGTAAATCCGGAGAAATCCGCATGGAGGAGCTGACAGCACACATACACAAAGAAAGACTGGGTTGGGTTCAAAATGTGAGAGAGACGCCAAAATGTTGTTGTT
2761  GATTGAAGCAGTTTAGTCGTCACGAAAAATGAAAAATCTGTAACAGGCATAACTCGTAAATCCCTAAAAAATTTGTATAGTAATTAGCAAAGCTGTGACCCAGCCGTTTCGATC  2874
```

Figure 6. (*See facing page for legend.*)

Figure 7. Repeats within the DlZM primary amino acid sequence and their relationships to other proteins. (*Top*) Alignment of cysteine-rich repeats within the deduced sequence of DlZM. This alignment is based on that published by Rees et al. (1988) for FIX and Notch. The coordinate of the last amino acid residue in each row is listed at *right*. (*Bottom*) Consensus sequences for the nine full-length repeats of the DlZM cysteine-rich array (Fig. 6); the 36 full-length repeats of the Notch cysteine-rich array (Wharton et al. 1985; Kidd et al. 1986); human and bovine FIX cysteine-rich motif (Kouichi et al. 1979; Kurachi and Davies 1982); human, mouse, and rat EGF (Gregory and Preston 1977; Gray et al. 1983; Scott et al. 1983; Simpson et al. 1985); and β-hydroxylation of aspartate or asparagine (Stenflo et al. 1987). Consensus sequences include single residues, alternate residues (D/N, F/Y), or any amino acid (X) present in 50% or more of the repeat elements (DlZM and Notch); residues present in two or three of the sequences (EGF); residues present in both sequences (FIX); and a previously published consensus (β-hydroxylation). Visual means were employed to maximize the alignment of the DlZM repeats and the FIX and EGF consensus sequences. Boxes encompass identical amino acid residues or conservative alternative residues (D/N, F/Y, or S/T) at single positions.

Figure 8. Comparison of the inferred one-dimensional structure of the predominant product of the *Dl* locus with the structures of proteins known or believed to be involved in cell–cell interactions. The schematic depictions represent the amino acid sequences of known or putative mature forms of vertebrate EGF receptor (Ullrich et al. 1984), Notch (Wharton et al. 1985; Kidd et al. 1986), DlZM (this paper; Vassin et al. 1987), nerve growth factor receptor (Johnson et al. 1986), vertebrate β-integrin (Takmun et al. 1986), and *Drosophila* laminin B1 chain (Montell and Goodman 1988). Cysteine-rich domains are symbolized by fine hatching; FIX/EGF-like domains are symbolized by heavy hatching.

RNA blot analysis

Total RNA was prepared from staged embryos (Meyerowitz and Hogness 1982), and poly(A)+ RNA was selected by oligo(dT)-cellulose chromatography (Maniatis et al. 1982). Poly(A)+ RNA samples were denatured with glyoxal and size-fractionated on 1.2% agarose gels (Maniatis et al. 1982), then transferred to Biotrans nylon membranes according to manufacturer's instructions (ICN Biomedicals, Inc.). Filters were UV-irradiated to fix RNA (Church and Gilbert 1984) and pretreated at 37°C for 6 hr in mix A, including 5% (wt/vol) SDS, 0.1% (wt/vol) NaPP$_i$, 0.5 mg/ml sheared denatured salmon sperm DNA, and 10% (wt/vol) dextran sulfate. Hybridizations were performed in a fresh aliquot of the same solution with the addition of ^{32}P-labeled nick-translated probe (final concentration of 10^7 cpm/ml) at 37°C for 12–16 hr. Filters were washed at 70°C with three changes of 2× SSC, 0.1% (wt/vol) NaPP$_i$, and 1% (wt/vol) SDS prior to autoradiography. Transcription was oriented by hybridization to embryonic RNA blots, using ^{32}P-labeled single-stranded RNA probes generated by in vitro transcription of genomic subclones (fragments A, E, and G; Fig. 4).

Isolation of cDNA clones

The cDNA clone containing insert D11 was isolated by screening a recombinant phage library made using 4- to 7-hr embryonic poly(A)+ RNA (provided by Barry Yedvobnick, Emory University; Yedvobnick et al. 1985) with a ^{32}P-labeled nick-translated probe that hybridizes to 5.4Z and 4.5 M. The cDNA clones containing inserts D12 and D13 were isolated from recombinant plasmid libraries made using either 0- to 4-hr or 4- to 8-hr embryonic poly(A)+ RNA (provided by Nicholas Brown, Harvard University; Brown and Kafatos 1988) by screening with ^{32}P-oligo-labeled (Feinberg and Vogelstein 1984) probes specific for 5.4Z or that hybridize to 5.4Z and 4.5M. Differential hybridization behavior with appropriate probes was assessed to identify candidate clones representing 5.4Z and 4.5M, based on RNA blot data (Figs. 2–4).

DNA sequencing and data analysis

Fragments from genomic and cDNA clones were subcloned into Bluescript vectors (KS, +, and − versions; Stratagene, Inc.). Single-stranded DNA was prepared from each subclone, according to Mead et al. (1986), using either JM109 or NM522 as hosts strains and the phage M13K07 as helper. Subcloned fragments were sequenced by the dideoxynucleotide-termination method, using a modified T7 DNA polymerase (Sequenase; U.S. Biochemicals, Inc.) and [α-^{35}S]dATP (New England Nuclear, Inc; 110 Ci/mmole). The sequencing strategy employed to determine the complete sequence of both strands of D11 (Fig. 6) is presented in Figure 9. DNA sequences were analyzed using programs from each of three software packages: the University of Wisconsin Genetics Computer Group (UWGCG) programs; the FASTP programs (Lipmann and Pearson 1985); and the IBI-Pustell programs (International Biotechnologies, Inc.).

General procedures

Phage (Cameron et al. 1977) and plasmid (Maniatis et al. 1982) DNAs were isolated using standard techniques. Autoradiography was performed with Kodak XAR film; intensifying screens were used for ^{32}P autoradiography.
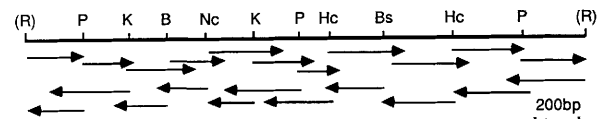

Figure 9. Sequencing strategy for cDNA insert Dll. The thick bar represents the Dll cDNA insert. Positions of recognition sequences for selected restriction sites within Dll are represented by vertical strokes. Abbreviations are the same as those in Fig. 5, except (Nc) *Nco*I; (Hc) *Hinc*II; (Bs) *Bst*EII. Arrows represent extent of DNA sequence data obtained from single subclones: Rightward pointing arrows indicate sequence from mRNA strand of cDNA; leftward pointing arrows indicate sequence from template strand of cDNA.

Acknowledgments

Note

Sequence data described in this paper have been submitted to the EMBL/GenBank Data Libraries under accession number Y00222.

References

Alton, A.K., K. Fechtel, A.L. Terry, S.B. Meikle, and M.A.T. Muskavitch. 1988. Cytogenetic definition and morphogenetic analysis of *Delta*, a gene affecting neurogenesis in *Drosophila melanogaster. Genetics* **118**: 235–245.

Alton, A.K., K. Fetchel, C.C. Kopczynski, S.B. Shepard, P.J. Kooh, and M.A.T. Muskavitch. 1989. Molecular genetics of *Delta*, a locus required for ectodermal differentiation in *Drosophila. Dev. Genet.* (in press).

Blobel, G. 1980. Intracellular protein topogenesis. *Proc. Natl. Acad. Sci.* **77**: 1496–1500.

Brown, N.H. and F.C. Kafatos. 1988. Functional cDNA libraries from *Drosophila* embryos. *J. Mol. Biol.* **203**: 425–437.

Bender, W., P. Spierer, and D.S. Hogness. 1983. Chromosomal walking and jumping to isolate DNA from the *Ace* and *rosy*

loci and the *bithorax* complex in *Drosophila melanogaster* *J. Mol. Biol.* **168**: 17–33.

Cameron, J.R., P. Philippsen, and R.W. Davis. 1977. Analysis of chromosomal integration and deletions of yeast plasmids. *Nucleic Acids Res.* **4**: 1429–1448.

Campos-Ortega, J.A. and V. Hartenstein. 1985. The embryonic development of *Drosophila melanogaster*. Springer, Berlin.

Carpenter, G. 1987. Receptors for epidermal growth factor and other peptide mitogens. *Annu. Rev. Biochem.* **56**: 881–914.

Carpenter, G. and S.N. Cohen. 1979. Epidermal growth factor. *Annu. Rev. Biochem.* **48**: 193–216.

Choo, K.H., K.G. Gould, D.J.G. Rees, and G.G. Brownlee. 1982. Molecular cloning of the gene for human anti-haemophilic Factor IX. *Nature* **299**: 178–180.

Church, G.M. and W. Gilbert. 1984. Genomic sequencing. *Proc. Nalt. Acad. Sci.* **81**: 1991–1995.

Craymer, L. 1984. New mutants. *Drosophila Inf. Serv.* **60**: 234.

Dietrich, U. and J.A. Campos–Ortega. 1984. The expression of neurogenic loci in imaginal epidermal cells of *Drosophila melanogaster*. *J. Neurogenet.* **1**: 315–332.

Doe, C.Q. and C.S. Goodman. 1985. Early events in insect neurogenesis. II. The role of cell interactions and cell lineage in the determination of neuronal precursor cells. *Dev. Biol.* **111**: 206–219.

Feinberg, A.P. and B. Vogelstein. 1984. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**: 266–267.

Furie, B. and B.C. Furie. 1988. The molecular basis of blood coagulation. *Cell* **53**: 505–518.

Gray, A., T.J. Dull, and A. Ullrich. 1983. Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000–molecular weight protein precursor. *Nature* **303**: 722–725.

Greenwald, I. 1985. *lin-12*, a nematode homeotic gene, is homologous to a set of mammalian proteins that includes epidermal growth factor. *Cell* **43**: 583–590.

Gregory, H. and B.M. Preston. 1977. The primary structure of human urogastrone. *Int. J. Pept. Protein Res.* **9**: 107–118.

Grimwade, B.G., M.A.T. Muskavitch, W.J Welshons, B. Yedvobnick, and S. Artavanis–Tsakonas. 1985. The molecular genetics of the *Notch* locus in *Drosophila melanogaster*. Dev. Biol. **107**:503–519.

Hirschberg, C.B. and M.D. Snider. 1987. Topography of glycosylation in the rough endplasmic reticulum and Golgi apparatus. *Annu. Rev. Biochem.* **56**: 63–88.

Holmes, W.E., D. Pennica, M. Blaber, M.W. Rey, W.A. Guenzler, G.J. Steffens, and H.L. Heyneker. 1985. Cloning and expression of the gene for prourokinase in *E. coli*. *BioTechnology* **3**: 923–929.

Johnson, D., A. Lanahan, C.R. Buck, A. Sehgal, C. Morgan, E. Mercer, M. Bothwell, and M. Chao. 1986. Expression and structure of the human NGF receptor. *Cell* **47**: 545–554.

Jones, F.S., M.P. Burgoon, S. Hoffmann, K.L. Crossin, B.A. Cunningham, and G.M. Edelman. 1988. A cDNA clone for cytotactin contains sequences similar to epidermal growth factor-like repeats and segments of fibronectin and fibrinogen. *Proc. Natl. Acad. Sci.* **85**: 2186–2190.

Kidd, S., M.W. Kelley, and M.W. Young. 1986. Sequence of the *Notch* locus of *Drosophila*: relationship of the encoded protein to mammalian clotting and growth factors. *Mol. Cell. Biol.* **6**: 3094–3108.

Kishimoto, T.K., K. O'Connor, A. Lee, T.M. Roberts, and T.A. Springer. 1987. Cloning of the *beta*-subunit of the leukocyte adhesion proteins: homology to an extracellular matrix receptor defines a novel supergene family. *Cell* **48**: 681–690.

Knust, E., U. Dietrich, U. Tepass, K.A. Bremer, D. Weigel, H. Vassin, and J.A. Campos-Ortega. 1987. EGF homologous sequences encoded in the genome of *Drosophila melanogaster*, and their relation to neurogenic genes. *EMBO J.* **6**: 761–766.

Kouichi, K., L.H. Ericsson, D.L. Enfield, K.A. Walsh, H. Neurath, E.W. Davie, and K. Titani. 1979. Comparison of the amino acid sequence of bovine coagulation factor IX (Christmas factor) with that of other vitamin K-dependent plasma proteins. *Proc. Natl. Acad. Sci.* **76**: 4990–4994.

Kurachi, K. and E.W. Davie. 1982. Isolation and characterization of a cDNA coding for human factor IX. *Proc. Natl. Acad. Sci.* **79**: 6461–6464.

Kyte, J. and R.F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**: 105–132.

Lawler, J. and R.O. Hynes. 1986. The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. *J. Cell Biol.* **103**: 1635–1648.

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*. *Wilhelm Roux's Arch. Dev. Biol.* **192**: 62–74.

Lindsley, D.H. and E.H. Grell. 1968. Genetic variations of *Drosophila melanogaster*. *Carnegie Inst. Wash. Publ.* **627**.

Lis, J.T., J.A. Simon, and C.A. Sutton. 1983. New heat shock puffs and *beta*-galactosidase activity resulting from transformation of *Drosophila* with an *hsp70-lacZ* hybrid gene. *Cell* **35**: 403–410.

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

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.

Mead, D.A., E. Skorupa-Szczesna, and B. Kemper. 1986. Single-stranded DNA 'blue' T and promotor plasmids: A versatile tandem promotor system for cloning and protein engineering. *Prot. Eng.* **1**: 67–74.

Meyerowitz, E.M. and D.S. Hogness. 1982. Molecular organization of a *Drosophila* puff site that responds to ecdysone. *Cell* **28**: 165–176.

Mizusawa, S., S. Nishimura, and F. Seela. 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucleic Acids Res.* **14**: 1319–1324.

Montell, D.J. and C.S. Goodman. 1988. *Drosophila* substrate adhesion molecule: sequence of laminin B1 chain reveals domains of homology with mouse. *Cell* **53**: 463–473.

Ny, T., F. Elgh, and B. Lund. 1984. The structure of the human tissue-type plasminogen activator gene: correlation of intron and exon structures to functional and structural domains. *Proc. Natl. Acad. Sci.* **81**: 5355–5359.

O'Connell, P. and M. Rosbash. 1984. Sequence, structure and codon preference of the *Drosophila* ribosomal protein 49 gene. *Nucleic Acids Res.* **12**: 5495–5513.

Perlman, D. and H.O. Halvorson. 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J. Mol. Biol.* **167**: 391–409.

Poulson, D.F. 1937. Chromosomal deficiencies and embryonic development of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **23**: 133–137.

Preiss, A., D.A. Hartley, and S. Artavanis-Tsakonas. 1988. Enhancer of split, a gene required for embryonic neural development in *Drosophila*. *EMBO J.* (in press).

Rees, D.J.G, I.M. Jones, P.A. Hanford, S.J. Walter, M.P. Snouf, K.J. Smith, and G.G. Brownlee. 1988. The role of *beta*-hydroxyaspartate and adjacent carboxylate residues in the first

EGF domain of human factor IX. *EMBO J.* **7**: 2053–2061.

Ruoslahti, E. 1988. Fibronectin and its receptors. *Annu. Rev. Biochem.* **57**: 375–414.

Scott, J., M. Urdea, M. Quiroga, R. Sanchez-Pescador, N. Fong, M. Selby, W.J. Rutter, and G.I. Bell. 1983. Structure of a mouse submaxillary messenger RNA encoding eidermal growth factor and seven related proteins. *Science* **221**: 236–240.

Simpson, R.J., J.A. Smith, R.L. Moritz, M.J. O'Hare, P.S. Rudland, J.R. Morrison, C.J. Lloyd, B. Grego, A.W. Burgess, and E.C. Nice. 1985. Rat epidermal growth factor: complete amino acid sequence. *Eur. J. Biochem.* **153**: 29–637.

Stenflo, J., A. Lundwall, and B. Dahlback. 1987. *beta*-Hydroxyasparagine in domains homologous to epidermal growth factor precursor in vitamin K-dependent protein S. *Proc. Natl. Acad. Sci.* **84**: 368–372.

Sudhof, T.C., J.L Goldstein, M.S. Brown, and D.W. Russell. 1985. The LDL receptor gene: a mosaic of exons shared with different proteins. *Science* **228**: 815–822.

Takmun, J.W., D.W. DeSimone, D. Fonda, R.S. Patel, C. Buck, A.F. Horwitz, and R.O. Hynes. 1986. Structure of integrin, a glycoprotein involved in the transmembrane linkage between fibronectin and actin. *Cell* **46**: 271–282.

Tautz, D., R. Lehmann, H. Schnurch, R. Schuh, E. Seifert, A. Keinlin, K. Jones, and H. Jackle. 1987. Finger protein of novel structure encoded by hunchback, a second member of the gap class of *Drosophila* segmentation genes. *Nature* **327**: 383–389.

Technau, G.M. and J.A. Campos-Ortega. 1987. Cell autonomy of expression of neurogenic genes of *Drosophila melanogaster. Proc. Natl. Acad. Sci.* **84**: 4500–4504.

Tosi, M., R. Young, O. Hagenbuchle, and U. Schibler. 1981. Multiple polyadenylation sites in mouse *alpha*-amylase gene. *Nucleic Acids Res.* **9**: 2313–2323.

Ullrich, A., L. Coussens, J.S. Hayflick, T.J. Dull, A. Gray, A.W. Downward, E.L.V. Mayes, N. Whittle, M.D. Waterfield, and P.H. Seeburg. 1984. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* **309**: 418–425.

Vassin, H., K.A. Bremer, E. Knust and J.A. Campos-Ortega. 1987. The neurogenic gene *Delta* of *Drosophila melanogaster* is expressed in neurogenic territories and encodes a putative transmembrane protein with EGF-like repeats. *EMBO J.* **6**: 3431–3440.

Wharton, K.A., K.M. Johansen, T. Xu, and S.Artavanis-Tsakonas. 1985. Nucleotide sequence from the neurogenic locus *Notch* implies a gene product that shares homology with proteins containing EGF-like repeats. *Cell* **43**: 567–581.

Yedvobnick, B., M.A.T. Muskavitch, K.A. Wharton, M.E. Halpern, E. Paul, B. Grimwade, and S. Artavanis-Tsakonas. 1985. Molecular genetics of *Drosophila* neurogenesis. *Cold Spring Harbor Symp. Quant. Biol.* **50**: 841–854.

Delta, a Drosophila neurogenic gene, is transcriptionally complex and encodes a protein related to blood coagulation factors and epidermal growth factor of vertebrates.

C C Kopczynski, A K Alton, K Fechtel, et al.

References	This article cites 54 articles, 16 of which can be accessed free at: **http://genesdev.cshlp.org/content/2/12b/1723.full.html#ref-list-1**
License	
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**