The EMBO Journal vol.12 no.6 pp.2411-2418, 1993

The *Drosophila neuralized* gene encodes a C_3HC_4 zinc finger

B.Duane Price, Zhen Chang, Roger Smith, Sharri Bockheim and Allen Laughon[1]

Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

[1]Corresponding author

Communicated by V.Pirrotta

The neurogenic genes of *Drosophila* are required for cell-cell communication that determines the choice between neuronal and epidermal cell fate. Here we report the molecular characterization of the neurogenic gene *neuralized* (*neu*) and show that it encodes a protein containing a C_3HC_4 zinc finger DNA-binding motif. This motif has been previously characterized in a variety of regulatory proteins, including transcription factors, locus-specific *Drosophila* chromosomal proteins, and oncoproteins. These results suggest a nuclear function for *neu* in the cell-cell signalling process responsible for inhibiting neuronal determination.

Key words: C_3HC_4 zinc finger/*Drosophila*/*neuralized*/neurogenesis/neurogenic

Introduction

During neurogenesis in *Drosophila*, interactions between neighboring ectodermal cells are necessary to restrict the number of neuronal progenitors. Neurogenic mutations disrupt this process resulting in neuronal hypertrophy at the expense of epidermis (Artavanis-Tsakonis and Simpson, 1991; Campos-Ortega and Jan, 1991). Genetic and molecular analyses of the neurogenic loci have yielded clues to how this process of cell-cell communication takes place. Three neurogenic genes, *Notch*, *Delta* and *bigbrain*, encode transmembrane cell surface proteins (Wharton *et al.*, 1985; Kidd *et al.*, 1986; Vässin *et al.*, 1987; Kopczynski *et al.*, 1988; Rao *et al.*, 1990). In mosaic analysis, the autonomous neural development of *Notch* mutant cells (Hoppe and Greenspan, 1986, 1990; Heitzler and Simpson, 1991) suggests that *Notch* functions as a receptor for an inhibitory signal. Conversely, the nonautonomous epidermal development of *Delta* mutant cells along clone borders, and an increased frequency of neuronal commitment among neighboring wild-type cells, suggests that *Delta* functions as a signal that inhibits neuronal determination (Heitzler and Simpson, 1991). Cultured cells that express *Notch* adhere to cells expressing *Delta*, evidence that the two proteins interact directly (Fehon *et al.*, 1990).

A number of other neurogenic genes encode nuclear proteins. Seven genes of the *E*(*spl*) complex encode related helix-loop-helix proteins (Klämbt *et al.*, 1989; Knust *et al.*, 1992) which presumably function as transcriptional regulators (Murre *et al.*, 1989). Antibody staining has shown that the protein encoded by *mastermind* is nuclear; however, its only distinctive structural feature is the presence of homopolymeric runs of amino acids (Smoller *et al.*, 1990). Based upon a similarity to portions of the homeodomain and bacterial helix-turn-helix motifs, it has been proposed that the protein encoded by the neurogenic gene, *neuralized* (*neu*), is also a nuclear regulatory factor (Boulianne *et al.*, 1991). We have independently cloned and characterized *neu* and found that it encodes an additional DNA-binding domain found in a diverse group of regulatory proteins, suggesting that *neu* may affect gene expression critical for the choice of ectodermal cell fate.

Results

Cloning and mapping of the neu region

Molecular characterization of *neu* was initiated from a line, neu^{383}, in which a single *P*[lacW] enhancer trap transposon (Bier *et al.*, 1989), *P*[lacW](85C), disrupts *neu*. Nearly 2000 independent insertions of *P*[lacW] on the second and third chromosomes were generated in our laboratory and screened for β-galactosidase expression in imaginal discs. One recessive lethal line, #383, expressed the reporter construct predominantly in sensory mother cells, the progenitors of the adult peripheral nervous system. Plasmid rescue was used to clone DNA flanking the *P*[lacW] insert. The recovered genomic DNA mapped to 85C, the cytological position of *neu*. Complementation analysis revealed that this insertional mutation is allelic to the *neu* mutations neu^{9L119} and neu^{IF65}. A 32 kb genomic walk was initiated with the rescued DNA (Figure 1). Southern analysis was used to generate and confirm a restriction map of the region flanking the rescued genomic DNA.

Identification of the neu transcript

Transcription units flanking *P*[lacW](85C) insertion were characterized using DNA fragments from the region to probe Northern blots (Figure 2). Using a 450 bp *Asu*II-*Eco*RV genomic fragment (Figure 1) that lies 100 bp from *P*[lacW](85C) as a probe, three major transcripts of 4.1, 3.7 and 0.4 kb were detected on developmental Northern blots (Figure 2A). The proximity of the probe to *P*[lacW](85C) and its small size suggested that these were *neu* transcripts. The 4.1 kb *neu* transcript was detected in embryos, larvae, pupae and adults, with highest levels detected in 3-12 h embryos. The 3.7 kb transcript was detected in 0-12 h embryos and in adult females, with highest levels in 0-3 h embryos. Northern analysis with this probe, which represents a common exon, suggests that the 4.1 kb transcript is more abundant than the 3.7 kb transcript. Both transcripts are expressed during periods when lateral inhibition restricts neuronal determination (Brand and Campos-Ortega, 1988; Skeath and Carroll, 1992). A 4.5 kb genomic *Eco*RI fragment located 9 kb to the right of *P*[lacW](85C) (Figure 1) also detected the same

© Oxford University Press

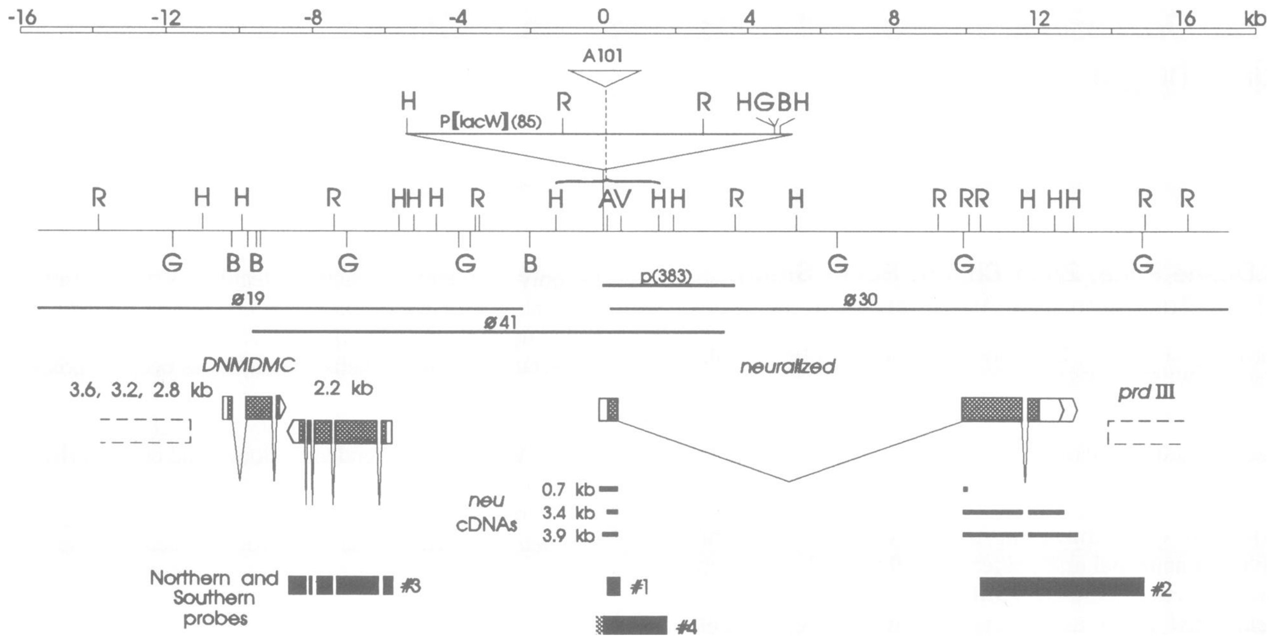

Fig. 1. Structure of the *neu* locus. The extent of the rescued genomic fragment p(383) and three genomic phage clones (ϕ19, ϕ30 and ϕ41) are shown as bars adjacent to a restriction map of the region surrounding *P*[lacW](85C). The positions of characterized transcription units are shown below as open boxes with coding regions shaded. *DNMDMC* and *neu* are transcribed from left to right; the gene encoding the 2.2 kb transcript is transcribed from right to left. The extent of characterized *neu* cDNAs of 0.7, 3.4 and 3.9 kb are indicated as bars adjacent to the *neu* transcription unit. The approximate location of the *prdIII* gene and a transcription unit encoding 3.6, 3.2 and 2.8 kb transcripts are represented as open dashed boxes (no cDNAs were obtained for these transcription units). Probes hybridized to the Northern blots in Figure 2 and Southern blots in Figure 6 are shown as numbered black boxes at the bottom. The abbreviations used for restriction sites are: B, *Bam*HI; E, *Eco*RI; G, *Bgl*II; and H, *Hin*dIII. A subset of the *Asu*II (A) and *Eco*RV (V) sites is also shown. A101 (not to scale) and *P*[lacW](85C) denote enhancer trap insertions. Scale is in kb from *P*[lacW](85C).

pattern of 4.1 and 3.7 kb transcripts (Figure 2B), indicating the presence of one or more additional *neu* exons.

Flanking transcription units discovered in our analysis serve to establish further the identity and outer molecular limits of the *neu* gene (Figure 1). Restriction mapping and DNA sequencing (data not shown) indicated that a gene located near the 3′ end of *neu* is *prdIII*, a locus previously identified at 85C on the basis of cross-hybridization to probes containing a paired repeat (Frigerio *et al.*, 1986). *prdIII* encodes 2.7 and 3.1 kb transcripts which exhibit temporal profiles of expression similar to that of *neu* (Figure 2B). A gene transcribed divergently beginning ~6 kb from *P*[lacW](85C) gives rise to a 2.2 kb transcript which is also expressed in a temporal pattern similar to that of *neu* (Figure 2C). cDNAs representing this 2.2 kb transcript were cloned and sequenced. Conceptual translation of this sequence predicts a 538 amino acid protein with no similarity to reported protein sequences (data not shown). 101 bp from the 3′ end of the 2.2 kb transcript lies the 3′ end of a gene encoding NAD-dependent 5,10-methylenetetrahydrofolate dehydrogenase-5,10-methenyltetrahydrofolate cyclohydrolase (DNMDMC) (Price and Laughon, 1993).

Using the 450 bp probe used in the Northern analysis, several overlapping *neu* cDNAs were recovered from imaginal disc and embryonic cDNA libraries. A composite *neu* cDNA sequence was generated from three of these cDNAs (Figure 3). The position of *P*[lacW](85C) was determined by comparing the sequence of a portion of the rescued plasmid with that of the *neu* cDNAs. This insertion lies within the untranslated leader of the *neu* transcript, suggesting that it may disrupt the synthesis or translation of *neu* mRNA (further evidence for this is described below). Two *neu* cDNAs contained poly(A)$^+$ tracts at sites

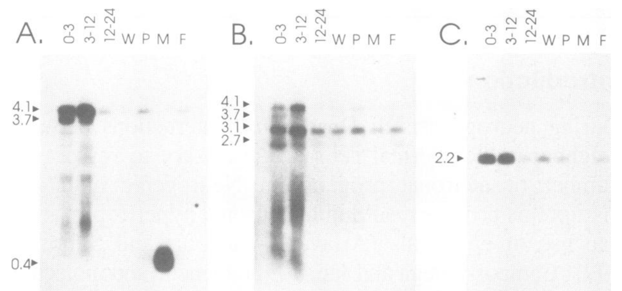

Fig. 2. Transcription units characterized by Northern analysis. 0–3 = 0–3 h embryos, 3–12 = 3–12 h embryos, 12–24 = 12–24 h embryos, W = wandering third larval instar, P = pupae, M = adult males and F = adult females. Size is indicated in kb. **(A)** An autoradiograph of a Northern blot probed with a 450 bp *Asu*II–*Eco*RV genomic fragment (probe #1, Figure 1), reveals three major transcripts of 4.1, 3.7 and 0.4 kb. The 4.1 kb transcript is expressed at all stages examined, with highest levels during hours 3–12 of embryogenesis. The 3.7 kb transcript is only expressed in adult females and during the first 12 h of embryogenesis. The 0.4 kb transcript is adult male-specific. Exposure of the Northern blot for 7 days. **(B)** An autoradiograph of a Northern blot probed with a 4.5 kb *Eco*RI genomic fragment (probe #2, Figure 1), reveals four major transcripts of 4.1, 3.7, 3.1 and 2.7 kb. The temporal expression of the 4.1 and 3.7 kb transcripts corresponds to the pattern of expression of the 4.1 and 3.7 kb *neu* transcripts shown in panel A. The 3.1 kb and 2.7 kb transcripts are derived from the *prdIII* locus. The 3.1 kb transcript is expressed at all stages examined with highest levels during hours 3–12 of embryogenesis. The 2.7 kb transcript is expressed only during the first 12 h of embryogenesis and in adult females. Exposure of the Northern blot was for 69 h. **(C)** An autoradiograph of a Northern blot probed with a 2.2 kb cDNA which maps from −6 kb to −9 kb (probe #3, Figure 1) and which detects a single transcript of 2.2 kb. This transcript is expressed at all stages except adult males, at highest levels during the first 12 h of embryogenesis. Exposure of the Northern blot was for 8 h.

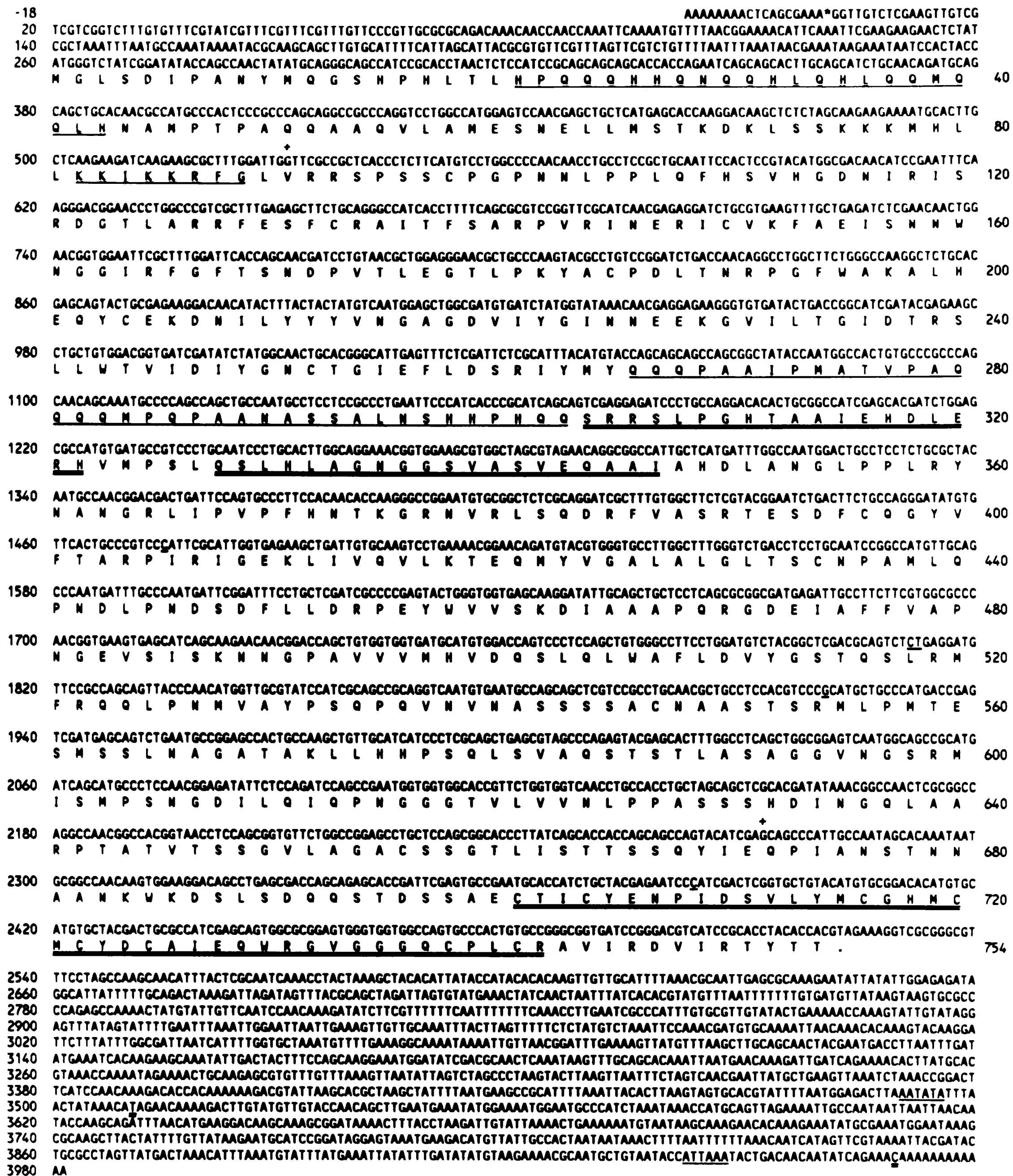

Fig. 3. Sequence analysis of *neu*. Both strands of 3999 bp of sequence were obtained from three overlapping *neu* cDNAs (Figure 1). A single strand of sequence from p(383) reveals that *P*[lacW](85C), indicated with an asterisk, falls within this cDNA sequence. Nucleotide position is indicated at the left; the *P*[lacW] insertion site is 0. The location of the neu^{383} insert suggests that these cDNAs correspond to *neu* transcripts. Sequence analysis also reveals that a 3.9 kb imaginal disc cDNA and a 3.4 kb embryonic cDNA, which probably represent the 4.1 kb and 3.7 kb *neu* transcripts, contain $poly(A)^+$ tracts following nucleotides 3510 and 3969 (double underlined) respectively. Potential $poly(A)^+$ addition signals are found at nucleotides 3490 and 3942 (underlined). Indicated by +, the splice donors follow nucleotides 530 and 2272. Both the 4.1 kb and 3.7 kb transcripts encode the same putative protein. The first ATG in the single long ORF is at nucleotide 260. Conceptual translation of the following ORF predicts a 754 amino acid protein of 82 kDa. Amino acid position is indicated at the right. The predicted protein contains short *opa* repeats at positions 21–43 and 256–303 (underlined), a putative nuclear localization signal at positions 82–89 (underlined), homology to helix–turn–helix domains at positions 304–347 (Boulianne *et al.*, 1991) (double underlined) and homology to the C_3HC_4 motif at positions 701–742 (double underlined). Nucleotides within the *neu* ORF at 1474, 1811, 1812, 1920 and 2383 (underlined) are not present in the sequence reported by Boulianne *et al.* (1991), which also contains the sequence TCCACGTCCCT in place of the G at nucleotide 1920.

separated by 458 bp, suggesting that the 4.1 and 3.7 kb *neu* transcripts arise via alternative polyadenylation. However, neither putative $poly(A)^+$ signal sequence is a perfect match to the consensus, AATAAA (Proudfoot and Brownlee, 1976). Comparison of genomic and *neu* cDNA restriction maps and Southern analysis indicated the presence of two introns. Genomic sequence was obtained for the splice donor and acceptor of each of these introns: both sets of splice junctions follow the GT/AG rule (Breathnach and Chambon, 1981). The 4.1 and 3.7 kb *neu* transcripts contain a single long open reading frame (ORF) encoding an 82 kDa protein of 754 amino acids. The first half of this ORF is virtually identical to the 411 codon *neu* ORF reported by Boulianne *et al.* (1991) but has an additional 343 codons at the 3′ end.

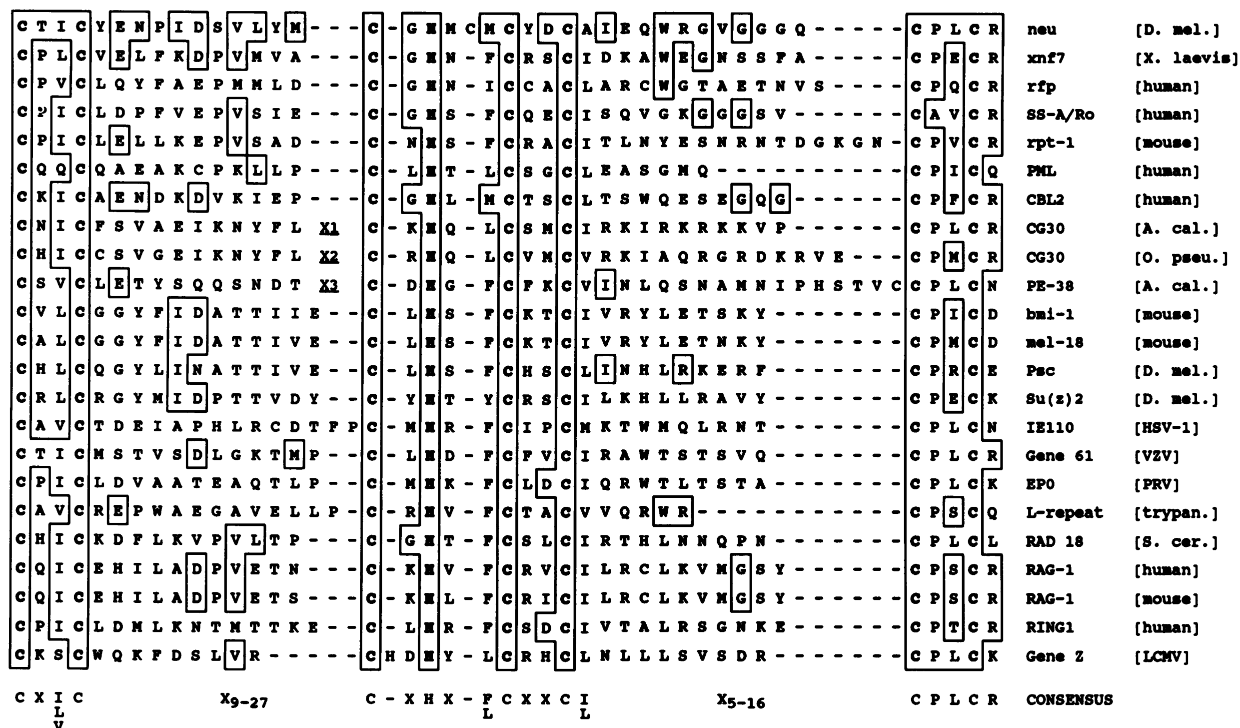

Fig. 4. Alignment of 23 proteins containing the C_3HC_4 motif (Blake *et al.*, 1991; Brunk *et al.*, 1991; Cheung, 1991; Haupt *et al.*, 1991; Kakizuka *et al.*, 1991; Reddy *et al.*, 1991; and references therein). Residues with identity to *neu* are boxed. The consensus can be represented as $CX[ILV]CX_{9-27}CXHX[FL]CXXC[IL]X_{5-16}CPLCR$. Both X_{9-27} and X_{5-16} are variable in terms of amino acid content and size, therefore no optimal alignment of these regions was attempted. X1, X2 and X3 represent amino acids omitted for brevity. X1 = QPIDRLTIIPVLELDT, X2 = QPVDAITILPIVELHT, X3 = CPFLIPTT.

The predicted neu protein contains a C_3HC_4 motif

Several features of the predicted *neu* protein indicate that it may be a DNA-binding regulatory protein (Figure 3), the most important of which is a C_3HC_4 zinc finger located near the carboxy-terminal end. The C_3HC_4 motif (Freemont *et al.*, 1991a) is found in more than 20 proteins (Figure 4) and its consensus sequence can be represented as $CX[ILV]CX_{9-27}CXHX[FL]CXXC[IL]X_{5-16}CPLCR$ (where X represents any amino acid). Only the histidine and cysteine residues are invariant. The *neu* C_3HC_4 sequence differs slightly from the motif consensus; the spacing between the histidine and the fourth cysteine is increased by one cysteine and there is an alanine at the [IL] position. Unlike other characterized C_3HC_4 proteins, the *neu* C_3HC_4 motif is found near the carboxy terminus of the protein. In other proteins, the C_3HC_4 motif has been shown to bind zinc (Salvato and Shimomaye, 1989) and to be necessary (Tagawa *et al.*, 1990) and sufficient (P.Freemont, R.Lovering and J.Trowsdale, unpublished, cited in Reddy and Etkin, 1991) for DNA binding. Although the best characterized C_3HC_4 protein, the herpes simplex virus type 1 immediate early protein 110 (Perry *et al.*, 1986), has been shown to function as a transcription factor (Everett, 1984; Gelman and Silverstein, 1985; O'Hare and Hayward, 1985a,b; Quinlan and Knipe, 1985), some C_3HC_4 proteins may have other functions that involve DNA contact. Other features of the predicted *neu* protein include homology to basic nuclear localization signals (Dang and Lee, 1989) and short runs of glutamine (Figure 3).

Expression pattern of neu transcripts

In situ hybridization was used to investigate the expression of *neu* transcripts during embryogenesis; our results are generally consistent with those reported by Boulianne *et al.* (1991). Before cellularization, low levels of *neu* transcripts are detected throughout the embryo (data not shown). After cellularization, *neu* expression is first detected along the ventral surface of the embryo. At first patchy (Figure 5A), this expression is soon uniform across the entire presumptive mesoderm (Figure 5B). Prior to and just after ventral furrow formation, transcripts are also weakly detected in transverse stripes (data not shown). As the ventral furrow seals, all cells expressing high levels of *neu* transcripts are internalized (Figure 5C). Transcript levels in the mesectoderm, the cells that lie along the ventral midline after gastrulation, are not detectably higher than in adjacent neurectodermal cells (data not shown). During germ band extension, the level of mesodermal expression rapidly decreased (data not shown).

At stage 9, *neu* expression is relatively uniform throughout the neurogenic ectoderm (data not shown). By stage 10, neuroblasts appear to express *neu* transcripts at slightly higher levels than surrounding cells (Figure 5D). Clusters of cells in the presumptive lateral epidermis express elevated levels of *neu* transcripts (Figure 5E). During stage 11, *neu* transcripts begin to decline in the ventral and lateral epidermis (data not shown). At late stages, expression appears to be maintained in at least some neural lineages (Figure 5F).

To test the assumption that *neu* is the gene encoding the 4.1 and 3.7 kb transcripts disrupted by *P*[lacW](85C), rather than the adjacent gene encoding the 2.2 kb transcript, we examined how *P*[lacW](85C) affects the expression of the two genes. *In situ* hybridization reveals that *P*[lacW](85C) dramatically reduces the level of the 4.1 and 3.7 kb transcripts (Figure 5G) but has no detectable effect on the fairly ubiquitous embryonic expression of the neighboring

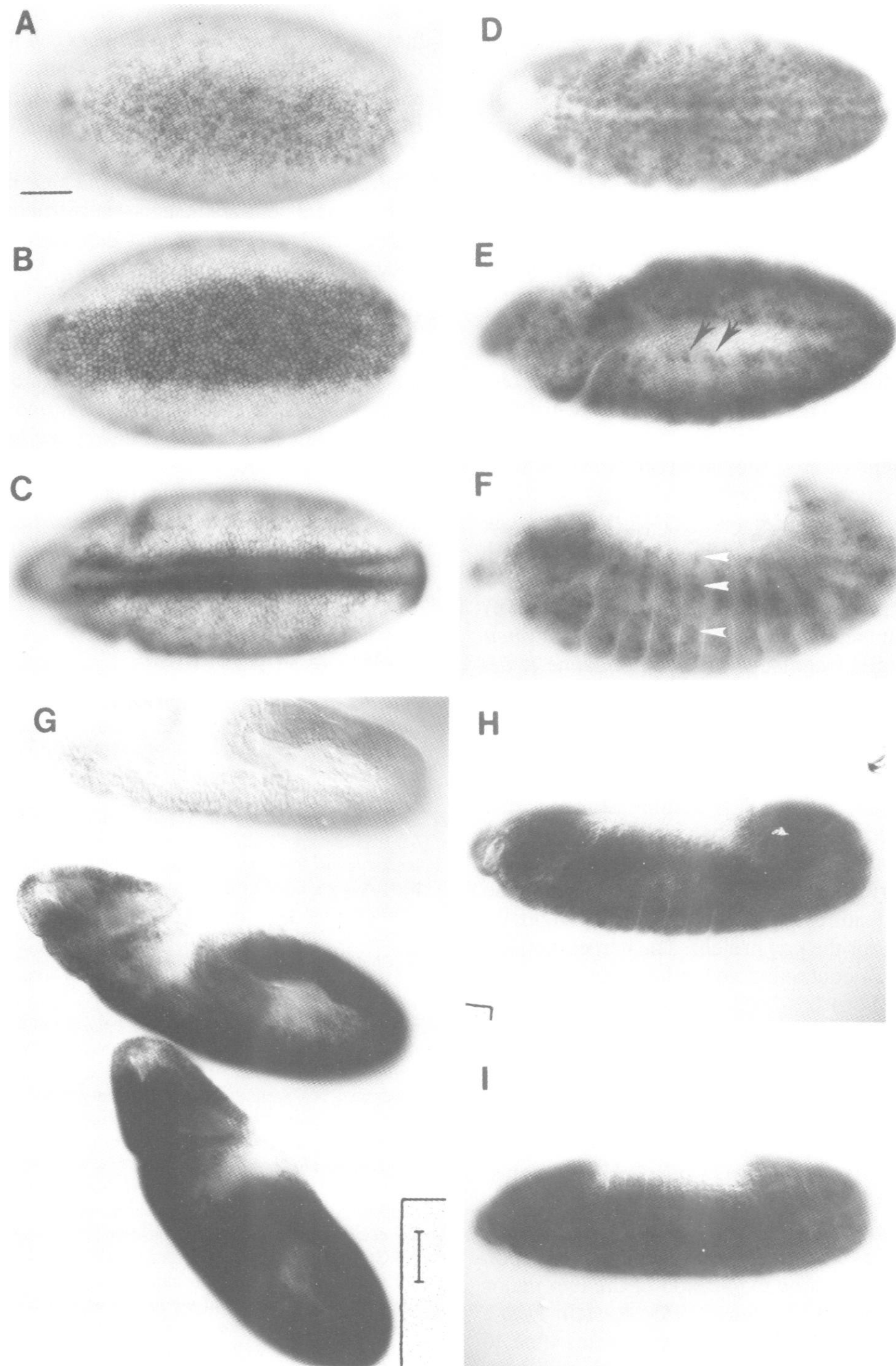

Fig. 5. Distribution of *neu* transcripts. Whole embryos hybridized with a digoxigenin-labelled riboprobe to the 5′ end of the third *neu* exon. All embryos are at the same magnification, scale bars in A and G = 63 μm, and are oriented with anterior to the left. (**A**) *neu* transcripts are first detected in a patchy pattern along the ventral surface of blastoderm embryos at about the time of cellularization. (**B**) Prior to ventral furrow formation, *neu* transcripts become evenly distributed throughout the presumptive mesoderm with sharply defined lateral boundaries. (**C**) At stage 6, the *neu*-expressing mesoderm invaginates during gastrulation. (**D**) At stage 10, *neu* is expressed in both neural and epidermal precursors along the ventral surface; however, the patchy pattern indicates that some cells express *neu* at higher levels. (**E**) Laterally, *neu* is also expressed in both the neural and epidermal precursors of the developing peripheral nervous system at stage 10. Again, patches of cells (arrows) exhibit high levels of *neu* expression. (**F**) At late stage 12, lateral ectodermal expression decreased with strong *neu* expression remaining in segmentally reiterated clusters of cells (arrows) which correspond in position to the developing peripheral nervous system. (**G**) Hybridization of the *neu* riboprobe to embryos collected from a $neu^{383}/TM3$ stock. Three classes of embryos were identified in a 1:2:1 Mendelian ratio. neu^{383}/neu^{383} (top; 49/192 total); $neu^{383}/TM3$ (middle) and *TM3/TM3* (bottom) embryos were identified on the basis of *neu* expression levels. Older neu^{383}/neu^{383} embryos can be identified on the basis of their neurogenic phenotype: they fail to form segments ventrally. All of these embryos showed dramatically reduced levels of *neu* transcripts (data not shown). (**H** and **I**) Hybridization of a riboprobe detecting the flanking 2.2 kb transcript to embryos collected from a $neu^{383}/TM3$ stock. A neu^{383}/neu^{383} mutant stage 14 embryo (I) shows no difference in the fairly ubiquitous pattern of 2.2 kb transcript expression compared with stage 14 $neu^{383}/TM3$ or *TM3/TM3* embryo (H). Similarly, no differences were observed at other embryonic stages (data not shown).

2.2 kb transcript (Figure 5H and I). These results confirm the identity of *neu*.

Discussion

While this work was in progress, another group reported the cloning and molecular characterization of *neu* (Boulianne *et al.*, 1991). That report predicted a *neu* protein of 411 amino acids, while our results predict a 754 amino acid *neu* protein. Comparison of DNA sequences from the two studies reveals a total of 27 differences. Four of these shift the reading frame resulting in the difference in predicted protein length. We sequenced these four regions in a corresponding genomic clone and found no differences compared with our cDNA sequence. We also note that our genomic restriction map differs somewhat from the previous report in the region surrounding the 5′ end of *neu*. We have confirmed the map by extensive restriction and Southern analysis of overlapping clones and of wild-type genomic DNA (see Materials and methods for details). Our mapping places the A101 enhancer trap insertion, a strong *neu* allele, within the same 2.7 kb *Hin*dIII fragment as *P*[lacW](85C) (data not shown) while in the previous report it was located further 5′ of *neu* in the adjacent *Hin*dIII fragment. Finally, Boulianne *et al.* (1991) reported a 2.5 kb transcript derived from the first *neu* intron. We have been unable to detect this transcript on Northern blots.

The most significant new information in this report is the identification of a potential DNA-binding domain, the C_3HC_4 motif, located near the carboxy terminus of the *neu* protein. Together with the potential helix–turn–helix DNA-binding motif described in the previous study (Boulianne *et al.*, 1991), this feature suggests a nuclear, potentially regulatory, function for the *neu* protein. The herpes simplex virus type 1, C_3HC_4-containing immediate early protein 110 (Perry *et al.*, 1986) has been shown to function as a transcriptional activator in co-transfection experiments (Everett, 1984; Gelman and Silverstein, 1985; O'Hare and Hayward, 1985a,b; Quinlan and Knipe, 1985). In other instances, the C_3HC_4 motif has been shown to be necessary (Tagawa *et al.*, 1990) or sufficient (Freemont *et al.*, 1991b) for DNA binding. Two related *Drosophila* C_3HC_4 proteins, *Posterior sex combs* (*Psc*) and *Suppressor two of zeste* [*Su(z)2*] (Brunk *et al.*, 1991; van Lohuizen *et al.*, 1991a), have been classified as Polycomb group genes, based upon the effect of *Psc* mutations on expression of the two major homeotic gene complexes (Jürgens, 1985; Rastelli *et al.*, 1993). These and other Polycomb group proteins bind to specific loci in the polytene chromosomes (Zink and Paro, 1989; Martin and Adler, 1993; Rastelli *et al.*, 1993) and are thought to play a role in the long range organization of chromatin (Paro, 1990). *Psc* and *Su(z)2* have also been implicated in chromosome pairing through their effects on the phenotype of *zeste* mutants (Wu *et al.*, 1988). These observations suggest that *neu* may affect ectodermal cell fate at the level of gene expression, either directly or as a transcription factor, or indirectly by affecting chromatin structure.

In addition to restricting neuronal determination, neurogenic genes have also been shown to function in muscle cell determination (Corbin *et al.*, 1991) and in controlling what appear to be cell fate choices in follicle cell lineages during oogenesis (Ruohola *et al.*, 1991; Xu *et al.*, 1992). Two proteins related in structure to *Notch* function in signalling that determines cell fate in the nematode *Caenorhabditis elegans* (Greenwald, 1985; Austin and Kimble, 1989; Yochem and Greenwald, 1989). *Notch* and *E(spl)* homologues have also been identified in vertebrates (Coffman *et al.*, 1990; Ellisen *et al.*, 1991; Weinmaster *et al.*, 1991; Sasai *et al.*, 1992). These results suggest that at least some neurogenic genes perform conserved functions in cell–cell interactions that determine cell fate among a wide variety of animals. Among the vertebrate genes containing the C_3HC_4 motif, *mel-18* (Tagawa *et al.*, 1990), *bmi-1* (Haupt *et al.*, 1991; van Lohuizen *et al.*, 1991b), *CBL2* (Blake *et al.*, 1991, *ret* (Takahashi *et al.*, 1988) and *PML-1* (de Thé *et al.*, 1991; Kakizuka *et al.*, 1991) are oncogenes or are associated with oncogenesis and therefore may be involved in regulating cell determination, while the *xnf7* gene is expressed predominantly in the *Xenopus* neural tube during embryogenesis (Reddy *et al.*, 1991). However, none of these genes exhibit extensive homology to *neu* and it therefore remains to be determined whether a true homolog exists in vertebrates.

Materials and methods

Enhancer trap screen and plasmid rescue

Approximately 2000 independent insertions of the enhancer trap construct *P*[lacW] were obtained on second and third chromosomes by mobilization of an X-linked copy of *P*[lacW] (kindly provided by E.Bier) as previously described (Bier *et al.*, 1989). Initial chromosomal assignment was based on segregation of the *P*[lacW] mini-*white* gene from dominantly marked balancer chromosomes. Homozygous viable insertions were maintained as homozygous lines and recessive lethal insertions were maintained as heterozygous lines balanced by *CyO* or *TM3*. To identify lines bearing insertions that expressed *lacZ* in imaginal disc complexes, five to ten wandering third instar larvae were dissected from each line and stained with X-gal following standard protocols (Ashburner, 1989).

To recover genomic DNA flanking the neu^{383} insert, total genomic DNA was isolated (Ashburner, 1989) and 0.2 μg digested with *Eco*RI. The DNA was precipitated, resuspended and ligated following standard protocols (Sambrook *et al.*, 1989). Circular DNA bearing the bacterial origin of replication and the ampicillin resistance gene of *P*[lacW] and flanking genomic DNA were selected by transformation of Epicurian Coli AG1 competent cells (Stratagene). The recovered neu^{383} DNA, p(383), contains pUC and *P* element sequences derived from *P*[lacW] and 3.5 kb of genomic DNA.

Mapping and isolation of genomic clones

To determine the cytological position of the neu^{383} *P*[lacW] insertion, p(383) was labelled with biotin and used for *in situ* hybridization to wild-type polytene chromosomes following standard protocols (Ashburner, 1989). To recover genomic clones of the *neu* locus, the 1.7 kb *Hin*dIII fragment of p(383) was used to probe a Canton S wild-type genomic library in Charon 4A (Maniatis *et al.*, 1978) following standard protocols (Sambrook *et al.*, 1989). A single genomic clone was obtained and characterized. Two flanking genomic clones were obtained by using the ends of the insert to re-screen the genomic library. An *Eco*RI, *Bam*HI, *Bgl*II and *Hin*dIII restriction map (Figure 1) was generated from analysis of ethidium bromide stained gels and Southern analysis of genomic clones recovered from the Canton S–Charon 4A phage library following standard protocols (Sambrook *et al.*, 1989).

This restriction map and the location of *P*[lacW](85C) were confirmed by Southern analysis of Oregon R and 383/*TM3*, *Ser* DNA (Figure 6). Genomic DNAs were isolated following standard protocols (Ashburner, 1989), digested, electrophoresed, transferred to nitrocellulose (Schleicher and Schuell) and probed at 65°C in aqueous solution following standard protocols (Sambrook *et al.*, 1989). In this analysis, two *Hin*dIII fragments from the rescued plasmid (probe #4, Figure 1), reveal that *P*[lacW](85C) resides on 6.9 kb *Eco*RI, >18 kb *Bam*HI, 10.4 kb *Bgl*II and 2.8 kb *Hin*dIII fragments (Figure 6A). A full-length cDNA representing the flanking 2.2 kb transcription unit (probe #3, Figure 1) detects *Eco*RI, *Bam*HI, *Bgl*II and *Hin*dIII fragments which do not overlap those detected by probing with the

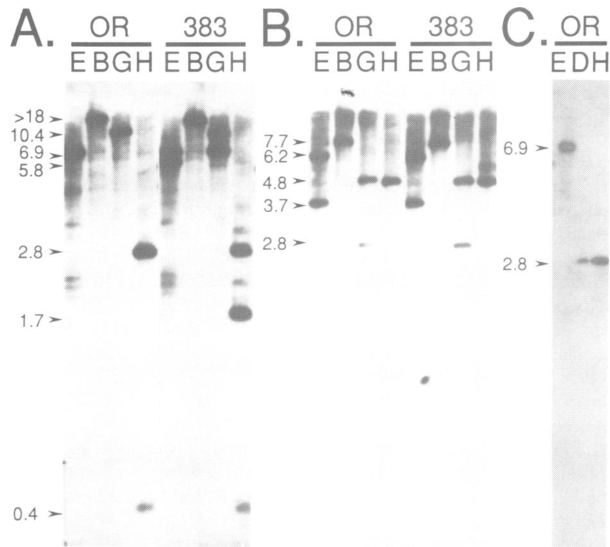

Fig. 6. Verification of restriction map by genomic Southern analysis. E = *Eco*RI, B = *Bam*HI, G = *Bgl*II, H = *Hin*dIII, D = *Eco*RI and *Hin*dIII double digest. OR = Oregon R genomic DNA, 383 = neu^{383}/*TM3 Ser* genomic DNA. All sizes are indicated in kb. (**A**) An autoradiograph of genomic Southern blot probed with two *Hin*dIII fragments from p(383) (probe #4, Figure 1), reveals novel 383 fragments. These fragments, 5.8 kb *Eco*RI, 6.9 kb *Bgl*II and 1.7 kb *Hin*dIII, are consistent with the restriction map in Figure 1. Note that the neu^{383} genomic DNA was incompletely digested by *Hin*dIII. (**B**) An autoradiograph of the same genomic blot as shown in (A), stripped and reprobed with a cDNA derived from the 2.2 kb flanking transcription unit (probe #3, Figure 1). All bands are different from those in (A) and are consistent with the restriction map shown in Figure 1. (**C**) An autoradiograph of a genomic Southern blot probed with a 450 bp *Asu*II–*Eco*RV fragment containing the 5′-most *neu* exon (probe #1, Figure 1), reveals that this exon lies within an *Eco*RI-insensitive 2.8 kb *Hin*dIII fragment. The relatively lower intensity of the band in the *Eco*RI–*Hin*dIII lane appears to have resulted from uneven transfer of DNA to the membrane.

rescued plasmid (Figure 6B). Both results are consistent with our phage insert-derived map.

This map differs from that reported by Boulianne *et al.* (1991) in the region flanking the 5′-most *neu* exon. We therefore used genomic Southern analysis to examine a critical difference between the two maps. The previously reported map predicts that the 5′-most *neu* exon resides on a 2.0 kb *Hin*dIII–*Eco*RI fragment. Our map predicts that this exon lies on a 2.8 kb *Hin*dIII fragment which does not contain an *Eco*RI site. When used to probe a Southern blot of *Eco*RI–*Hin*dIII digested Oregon R DNA, the 450 bp *Asu*II–*Eco*RV fragment containing the 5′-most *neu* exon (probe #1, Figure 1) reveals that this exon lies on a 2.8 kb *Hin*dIII fragment which does not contain an *Eco*RI site (Figure 6C). This result is consistent with our map. In addition, extrapolation from the previously reported map predicts that plasmid rescue of *P*[lacW](85C) using *Eco*RI should have resulted in the recovery of 500 bp of genomic DNA, rather than the observed 2.8 kb.

To map the A101 enhancer trap insertion (Huang *et al.*, 1991), a 2.7 kb *Hin*dIII fragment containing the 5′-most exon of *neu* was used to probe Southern blots of digested Oregon R wild-type and heterozygous A101 genomic DNAs following standard protocols (Sambrook *et al.*, 1989). Comparison of the A101 and wild-type lanes revealed that the A101 insertion disrupts the 2.8 kb *Hin*dIII fragment (data not shown).

Northern analysis

Total RNA was isolated from staged 0–3 h, 3–12 h and 12–24 h *Drosophila* embryos, wandering third instar larvae, pupae and adult males and females using a guanidinium thiocyanate method (Chomczynski and Sacchi, 1987). Poly(A)$^+$ mRNA was selected on an oligo(dT) cellulose column using a batch binding protocol (Sambrook *et al.*, 1989). A total of 10 μg poly(A)$^+$ mRNA was loaded per lane on a 0.7% agarose, 6.7% formaldehyde gel (Sambrook *et al.*, 1989). A 1 μg aliquot of RNA standards (Gibco BRL) was loaded for calibration of the gels. Nucleic acid was transferred to a Zeta Probe membrane (Bio-Rad) following the manufacturer's alkaline blotting protocol.

The autoradiographs shown in Figure 2 are from blots probed with a 450 bp *Asu*II–*Eco*RV genomic fragment, a 4.5 kb *Eco*RI genomic fragment and a 2.2 kb cDNA (see text). In addition, a 5.5 kb genomic *Eco*RI fragment completely within the first *neu* intron was used to probe Northern blots. The molecular map reported by Boulianne *et al.* (1991) shows the 3′ half of the 2.5 kb transcription unit on a similar fragment. In two separate experiments we could not detect transcripts with this 5.5 kb genomic *Eco*RI fragment. All DNAs were labelled with [α-^{32}P]dATP (DuPont) by extension of random hexanucleotides (Boehringer Mannheim) with the Klenow fragment of *Escherichia coli* DNA polymerase I (Ausubel *et al.*, 1987) to a specific activity of 10^9 c.p.m./μg. Pre-hybridization, hybridization of the blots with $>10^6$ c.p.m./ml probe at 65°C for >12 h, and washes at 65°C were performed following the manufacturer's dextran sulfate protocol. Blots were exposed to film for 8–168 h (Figure 2) at −70°C with intensification screen. Hybridized DNA probes were removed from the blots following the manufacturer's protocol. The blots were reprobed with the ribosomal protein 49 gene (O'Connell and Rosbash, 1984) to control for lane-to-lane variation in loading of RNA (data not shown).

Isolation of cDNAs

To isolate *neu* cDNAs, a 450 bp *Asu*II–*Eco*RV genomic fragment containing the 3′ end of the first *neu* exon, probe #1 (Figure 1), was used to screen 2×10^5 plaques of a *Drosophila* imaginal disc cDNA library in λgt10 (kindly provided by G.Rubin) and a 3–12 h Oregon R embryonic cDNA library in λgt10 (Poole *et al.*, 1985) following standard protocols (Sambrook *et al.*, 1989). Four independent larval and five independent embryonic pure phage stocks were recovered. Restriction and Southern analyses suggest that, except for their termini, these cDNAs are contiguous.

cDNAs representing the 2.2 kb transcript were identified using a 2.1 kb genomic *Eco*RI fragment (one artificial *Eco*RI site from the end of a genomic phage clone) to screen 2×10^5 plaques of the imaginal disc library. Four independent phage stocks representing the 2.2 kb transcript were recovered.

Sequence analysis

All sequence analysis was from DNA subcloned into Bluescript KS+ (Stratagene). The majority of *neu* sequence was obtained from a 0.7 kb cDNA fragment representing the 5′ end of *neu*, a 2.8 kb *Eco*RI fragment of a 3.9 kb cDNA representing the 3′ end of the 4.1 kb *neu* transcript (Figure 1), and a 1.2 kb *Eco*RI–*Hin*dIII genomic fragment containing the 3′ end of the second *neu* exon and the splice donor for the second *neu* intron. Nested deletions of the cDNA fragments in both directions and of the genomic fragment in one direction were generated with an Erase-a-Base kit (Promega) following the manufacturer's protocol. Additional *neu* sequence was obtained from a 2.4 kb *Eco*RI fragment of a cDNA representing the 3′ end of the 3.7 kb transcript (Figure 1), a 400 bp genomic *Eco*RI fragment representing the center of the *neu* transcript, a 1 kb genomic *Eco*RI fragment containing the first *neu* intron acceptor, a 1 kb genomic *Hin*dIII fragment containing the second *neu* intron acceptor, and a 450 bp *Asu*II–*Eco*RV genomic fragment containing the first *neu* intron donor. In addition, part of a 1.7 kb *Hin*dIII fragment of p(383) was sequenced to determine the point of insertion of *P*[lacW](85C).

Dideoxy DNA sequencing was done using Sequenase (USB) according to the manufacturer's protocol. The sequencing reactions were run on 60 cm long 6% polyacrylamide, 50% urea, 0.5 × TBE gels following standard protocols (Sambrook *et al.*, 1989). Dried gels were exposed to film at room temperature, usually for 24 h.

In situ hybridization

A digoxigenin-labelled riboprobe specific to the 4.1 and 3.7 kb *neu* transcripts was generated from a 1 kb genomic *Hin*dIII fragment and a riboprobe specific to the 2.2 kb transcript was generated from a cDNA representing this gene using the Genius labelling kit (Boehringer Mannheim). The riboprobes were synthesized with T3 RNA polymerase (Boehringer Mannheim) and partially hydrolyzed for 40 min at 60°C in 100 μl of 100 mM $NaCO_3$ according to the manufacturer's protocol. The labelled RNA was precipitated with 100 μg of glycogen (Boehringer Mannheim), and resuspended in 100 μl of hybridization buffer (50% formamide, 5 × SSC, 0.1% Tween, 100 μg/ml herring sperm DNA, 50 μg/ml heparin). *In situ* hybridization was performed according to Tautz and Pfeifle (1989) as modified by Jiang *et al.* (1991). Prior to mounting for microscopy, tissues were washed with PBS plus 0.1% Tween and equilibrated in 60% glycerol.

Acknowledgements

We thank E.Bier for providing us with the *P*[lacW] enhancer trap construct; M.Boedigheimer, N.Brown and J.Skeath for help with the enhancer trap screen; G.Rubin for providing us with the imaginal disc cDNA library;

D.Miller for help with the initial molecular analysis; an anonymous reviewer, S.Blair, W.Dove, W.Engels, C.Luongo, J.Skeath and members of our laboratory for valuable comments on this manuscript; and V.Pirrotta and P.Adler for sharing unpublished results. This work was supported by grants to A.L. from the National Science Foundation and The Milwaukee Foundation Shaw Scholars program. B.D.P. received support from NIH predoctoral training grant GM07133. This is publication #3339 from the Laboratory of Genetics, University of Wisconsin-Madison.

References

Artavanis-Tsakonis,S. and Simpson,P. (1991) *Trends Genet.*, **7**, 403-408.
Ashburner,M. (1989) *Drosophila: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Austin,J. and Kimble,J. (1989) *Cell*, **58**, 565-571.
Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Smith,J.A., Seidman,J.G. and Struhl,K. (1987) *Current Protocols in Molecular Biology.* John Wiley and Sons, New York.
Bier,E., Vaessin,H., Shepherd,S., Lee,K., McCall,K., Barbel,S., Ackerman,L., Carretto,R., Uemura,T., Grell,E., Jan,L.Y. and Jan,Y.N. (1989) *Genes Dev.*, **3**, 1273-1287.
Blake,T.J., Shapiro,M., Morse,H.C.,III and Langdon,W.Y. (1991) *Oncogene*, **6**, 653-657.
Boulianne,G.L., de la Concha,A., Campos-Ortega,J.A., Jan,L.Y. and Jan,Y.N. (1991) *EMBO J.*, **10**, 2975-2983.
Brand,M. and Campos-Ortega,J.A. (1988) *Roux's Arch. Dev. Biol.*, **197**, 457-470.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Brunk,B.P., Martin,E.C. and Adler,P.N. (1991) *Nature*, **353**, 351-353.
Campos-Ortega,J.A. and Jan,Y.N. (1991) *Annu. Rev. Neurosci.*, **14**, 399-420.
Cheung,A.K. (1991) *J. Virol.*, **65**, 5260-5271.
Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.*, **162**, 156-159.
Coffman,C., Harris,W. and Kintner,C. (1990) *Science*, **249**, 1438-1441.
Corbin,V., Michelson,A.M., Abmayr,S.M., Neel,V., Alcamo,E., Maniatis,T. and Young,M.W. (1991) *Cell*, **67**, 311-323.
Dang,C.V. and Lee,W.M.F. (1989) *J. Biol. Chem.*, **264**, 18019-18023.
de Thé,H., Lavau,C., Marchio,A., Chomienne,C., Degos,L. and Dejean,A. (1991) *Cell*, **66**, 675-684.
Ellisen,L.W., Bird,J., West,D.C., Soreng,A.L., Reynolds,T.C., Smith,S.D. and Sklar,J. (1991) *Cell*, **66**, 649-661.
Everett,R.D. (1984) *EMBO J.*, **3**, 3135-3141.
Fehon,R.G., Kooh,P.J., Rebay,I., Regan,C.L., Xu,T., Muskavitch,M.A.T. and Artavanis-Tsakonas,S. (1990) *Cell*, **61**, 523-534.
Freemont,P.S., Hanson,I.M. and Trowsdale,J. (1991) *Cell*, **64**, 483-484.
Frigerio,G., Burri,M., Bopp,D., Baumgartner,S. and Noll,M. (1986) *Cell*, **47**, 735-746.
Gelman,I.H. and Silverstein,S. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 5265-5269.
Greenwald,I. (1985) *Cell*, **43**, 583-590.
Haupt,Y., Alexander,W.S., Barri,G., Klinken,S.P. and Adams,J.M. (1991) *Cell*, **65**, 753-763.
Heitzler,P. and Simpson,P. (1991) *Cell*, **64**, 1083-1092.
Hoppe,P.E. and Greenspan,R.J. (1986) *Cell*, **46**, 773-783.
Hoppe,P.E. and Greenspan,R.J. (1990) *Development*, **109**, 875-885.
Huang,F., Dambly-Chaudière,C. and Ghysen,A. (1991) *Development*, **111**, 1087-1095.
Jiang,J., Kosman,D., Ip,Y.T. and Levine,M. (1991) *Genes Dev.*, **5**, 1881-1891.
Jürgens,G. (1985) *Nature*, **316**, 153-155.
Kakizuka,A., Miller,W.H., Jr, Umesono,K., Warrell,R.P.,Jr, Frankel,S.R., Murty,V.V.V.S., Dmitrovsky,E. and Evans,R.M. (1991) *Cell*, **66**, 663-674.
Kidd,S., Kelly,M.R. and Young,M.W. (1986) *Mol. Cell. Biol.*, **6**, 3094-3108.
Klämbt,C., Knust,E., Tietze,K. and Campos-Ortega,J.A. (1989) *EMBO J.*, **8**, 203-210.
Knust,E., Schrons,H., Grawe,F. and Campos-Ortega,J.A. (1992) *Genetics*, **132**, 505-518.
Kopczynski,C.C., Alton,A.K., Fechtel,K., Kooh,P.J. and Muskavitch, M.A.T. (1988) *Genes Dev.*, **2**, 1723-1735.
Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connell,C., Quon,D., Sim,G.K. and Efstratiadis,A. (1978) *Cell*, **15**, 687-701.
Martin,E.C. and Adler,P.N. (1993) *Development*, **117**, 641-655.
Murre,C., McCaw,P.S. and Baltimore,D. (1989) *Cell*, **56**, 777-783.
O'Connell,P. and Rosbash,M. (1984) *Nucleic Acids Res.*, **12**, 5495-5513.
O'Hare,P. and Hayward,G.S. (1985a) *J. Virol.*, **53**, 751-760.
O'Hare,P. and Hayward,G.S. (1985b) *J. Virol.*, **56**, 723-733.
Paro,R. (1990) *Trends Genet.*, **6**, 416-421.
Perry,L.J., Rixon,F.J., Everett,R.D., Frame,M.C. and McGeoch,D.J. (1986) *J. Gen. Virol.*, **67**, 2365-2380.
Poole,S.J., Kauvar,L.M., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37-43.
Price,B.D. and Laughon,A. (1993) *Biochim. Biophys. Acta*, in press.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211-214.
Quinlan,M.P. and Knipe,D.M. (1985) *Mol. Cell. Biol.*, **5**, 957-963.
Rao,Y., Jan,L.Y. and Jan,Y.N. (1990) *Nature*, **345**, 163-167.
Rastelli,L., Chan,C.S. and Pirrotta,V. (1993) *EMBO J.*, **12**, 1513-1522.
Reddy,B.A. and Etkin,L.D. (1991) *Nucleic Acids Res.*, **19**, 6330.
Reddy,B.A., Kloc,M. and Etkin,L. (1991) *Dev. Biol.*, **148**, 107-116.
Ruohola,H., Bremer,K., Baker,D., Swedlow,J.R., Jan,L.Y. and Jan,Y.N. (1991) *Cell*, **66**, 433-449.
Salvato,M.S. and Shimomaye,E.M. (1989) *Virology*, **173**, 1-10.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual.* 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sasai,Y., Kageyama,R., Tagawa,Y., Shigemoto,R. and Nakanishi,S. (1992) *Genes Dev.*, **6**, 2620-2634.
Skeath,J.B. and Carroll,S.B. (1992) *Development*, **114**, 939-946.
Smoller,D., Friedel,C., Schmid,A., Bettler,D., Lam,L. and Yedvobnick,B. (1990) *Genes Dev.*, **4**, 1688-1700.
Tagawa,M., Sakamoto,T., Shigemoto,K., Matsubara,H., Tamura,Y., Ito,T., Nakamura,I., Okitsu,A., Imai,K. and Taniguchi,M. (1990) *J. Biol. Chem.*, **265**, 20021-20026.
Takahashi,M., Inaguma,Y., Hiai,H. and Hirose,F. (1988) *Mol. Cell. Biol.*, **8**, 1853-1856.
Tautz,D. and Pfeifle,C. (1989) *Chromosoma*, **98**, 81-85.
van Lohuizen,M., Frasch,M., Wientjens,E. and Berns,A. (1991a) *Nature*, **353**, 353-355.
van Lohuizen,M., Verbeek,S., Scheijen,B., Wientjens,E., van der Gulden,H. and Berns,A. (1991b) *Cell*, **65**, 737-752.
Vässin,H., Bremer,K.A., Knust,E. and Campos-Ortega,J.A. (1987) *EMBO J.*, **6**, 3431-3440.
Weinmaster,G., Roberts,V.J. and Lemke,G. (1991) *Development*, **113**, 199-205.
Wharton,K.A., Johansen,K.M., Xu,T. and Artavanis-Tsakonas,S. (1985) *Cell*, **43**, 567-581.
Wu,C.-T., Jones,R.S., Lasko,P.F. and Gelbart,W.M. (1988) *Mol. Gen. Genet.*, **218**, 559-564.
Xu,T., Caron,L.A., Fehon,R.G. and Artavanis-Tsakonas,S. (1992) *Development*, **115**, 913-922.
Yochem,J. and Greenwald,I. (1989) *Cell*, **58**, 553-564.
Zink,B. and Paro,R. (1989) *Nature*, **337**, 468-471.

Received on January 8, 1993; revised on March 1, 1993

Note added in proof

The sequence reported in this paper has been deposited in the GenBank Data Library under the accession number L12218.