**A Drosophila gene expressed in the embryonic CNS shares one conserved domain with the mammalian GAP-43**

SHI-CHUNG NG¹,², LIZABETH A. PERKINS¹,³, GARY CONBOY¹,², NORBERT PERRIMON¹,³ and MARK C. FISHMAN¹,²

¹Howard Hughes Medical Institute
²Developmental Biology Laboratory of the Massachusetts General Hospital and the MGH Cancer Center, Boston, MA 02114, USA
Departments of ³Medicine and ³Genetics, Harvard Medical School, Boston, MA 02115, USA

Summary

By cross hybridization with the mammalian growth-related protein, GAP-43, we have isolated several Drosophila cDNAs and genomic sequences. These sequences correspond to a single copy gene that encodes two developmentally regulated transcripts 2-4 and 2-0 kb in length. The predicted protein sequence from the cDNAs contains a stretch of 20 amino acids closely related to the mammalian GAP-43 protein. These residues are also highly conserved in a cDNA isolated from the nematode C. elegans. Prior to dorsal closure, expression of the Drosophila gene is observed in non-neuronal tissues, especially in the mesectoderm and presumptive epidermis, both in a metameric pattern. After dorsal closure, expression becomes restricted to sets of cells that are segmentally reiterated along the periphery of the nervous system. These cells appear to include at least one specific set of glia that may establish scaffolding for the development of the longitudinal neuropile.

Key words: Drosophila, nervous system, embryogenesis, GAP-43.

Introduction

Several approaches have been utilized to identify genes important to development of the nervous system of Drosophila. Mutations that cause phenotypic aberrations of the nervous system have identified specific genes. Among these are the neurogenic genes which result in hypertrophy of the CNS at the expense of the ventral epidermis (Lehmann et al. 1983) and affect the switch between epidermal and neural developmental pathways. Conversely, genes within the *achaete-scute* complex were identified by their hypotrophic affect on the CNS (Jimenez & Campos-Ortega, 1979). Also identified by this approach are the behavioural mutants passover and bendless (Thomas & Wyman, 1983) and disconnected (Stellar et al. 1987). Many other mutations identified by their effects on non-neuronal tissues have surprisingly been found to affect the nervous system. Among these genes are the segmentation genes fushi tarazu (Doe et al. 1988a) and even-skipped (Doe et al. 1988b), the sex-determination gene daughterless (Caudy et al. 1988), and other mutations with cuticular pattern defects including polyhomeotic (Smouse et al. 1988) and cut (Bodmer et al. 1987). Another powerful approach might be to seek Drosophila homologues of mammalian genes involved in neuronal development. These investigations were initiated as an attempt to identify genes related to the neuronal growth associated protein, GAP-43. The function of this neurone-specific gene is not known, although its expression correlates well with periods of neuronal growth during development and regeneration. It is a phosphoprotein enriched in the membranes of growth cones. Its phosphorylation state changes during long-term potentiation, and from this and several other observations has arisen the notion that GAP-43 may be involved in learning and memory (Skene & Willard, 1981a,b; Benowitz & Lewis, 1983; reviewed by Benowitz & Routtenberg, 1987). The sequences of GAP-43 cDNA have been identified from rat (Karns et al. 1987; Basi et al. 1987) human (Ng et al. 1988; Kosik et al. 1988), and mouse (Cimler et al. 1987) and are nearly identical.

By using low-stringency hybridization we identified several cDNAs from Drosophila that are related to mammalian GAP-43. This homology is restricted, however, to a short domain of 20 amino acids, which interestingly is also conserved in a predicted protein from the nematode C. elegans (Ng and Fishman, unpublished). Although this region may represent a conserved motif, it seems unlikely that the overall function of the Drosophila gene is equivalent to that of mammalian GAP-43. Despite this limited homology,
the cellular location of expression suggests that it is indeed important to development of the nervous system.

Materials and methods

Isolation of Drosophila GAP-43 related cDNA and genomic clones

Overlapping cDNA clones were isolated at reduced stringency from both a Drosophila 3-12 h embryonic library (Poole et al. 1985) and from a size-selected 9-12 h embryonic library (a generous gift of Kai Zinn and Corey Goodman) using as probes both the complete rat GAP-43 cDNA (Karns et al. 1987) and a 300 bp subcloned fragment, KZ30 was isolated from an embryonic Agt10 (9-12 h) library. The genomic piece, 3L, was isolated from the Maniatis Charon 4A genomic library. Genomic clones were isolated independently from both strands by the chain-termination method (Feinberg & Vogelstein, 1983). Filters were washed in 2 x SSC, 0.1% SDS at 50°C before autoradiography.

Total RNA from staged Drosophila collections reared at 25°C in population cages was obtained by the guanidinium/caesium chloride method (Maniatis et al. 1982). Poly (A)^+ RNA was affinity-purified on oligo(dT)-cellulose (Type III, Collaborative Research, Inc.) and synthetic oligonucleotide primers. DNA sequences were analyzed by the Sequenase enzyme (T7 DNA polymerase, United States Biochemical Co.) and synthetic oligonucleotide primers. DNA sequences were analysed by UWGCG (University of Wisconsin Genetics Computer Group) and Beckman Microgenie software packages. Database searching were done by the Fast-p and Fast-n algorithms (Lipman & Pearson, 1985).

Northern and Southern Analyses

Fig. 1. Map of Drosophila GAP-43-related cDNA clones aligned with the genomic DNA. cDNA clones 10a4 and 9c1 were independently isolated from a Drosophila embryonic (3-12 h) Agt10 library, and clone KZ30 was isolated from an embryonic Agt11 (9-12 h) library. The genomic piece, 3L, was isolated from the Maniatis Charon 4A genomic library. The method of isolation of these clones is described in the Materials and methods. They are aligned to show the region of overlap. The arrows in 3L delineate the exon that is colinear with the cDNA. A, Aval; K, Kpn1, An, poly A tail.
Fig. 2. Nucleotide and predicted amino acid sequence of *Drosophila* GAP-43-related cDNA. The nucleotide and deduced amino acid sequences of a *Drosophila* GAP-related cDNA are shown. There is one potential glycosylation site at amino acid position 81 and the serine at position 137 is a potential phosphorylation site for protein kinases A and C. The ‘GAP motif’ is boxed and the direct repeats at the C-terminal are overlined. The two polyadenylation signals, separated by 300 base pairs, are underlined.

**Results**

*Drosophila* and *C. elegans* cDNAs contain a predicted domain related to GAP-43

Using both the complete rat GAP-43 cDNA (Karns et al. 1987) and a 300 bp *HpaII* subcloned fragment (which contains only the coding region) as probes in duplicate screening at reduced stringency, we isolated three overlapping cDNA clones from *Drosophila* embryonic *Agt10* and *Agt11* cDNA libraries and a genomic clone from the Maniatis *Drosophila* library (Maniatis et al. 1978), as shown in Fig. 1. The longest cDNA clone, KZ30, is 2.2 kb and is close to the full length of the largest and more abundant transcript (2.4 kb) as judged by Northern blot analysis (see below).

The sequence and longest predicted open reading frame of the cDNA is shown in Fig. 2. The entire predicted open reading frame and 3′-untranslated region are contained within a single long exon, with an intron–exon boundary in the 5′-untranslated region. The genomic sequence diverges from the cDNA after the polyadenylation signal (AATAAA), which is followed by a poly A stretch at the end of one cDNA clone. One other potential polyadenylation signal is located 300 bp upstream. We have no direct evidence for its use, although there is a smaller transcript (2.0 kb) that is less abundant than the 2.4-kb one (see below).

The longest open reading frame that is contained in the KZ30 clone can encode a protein of 441 amino acids. It is not related to protein or nucleotide sequences identified in the NBRF data bases.

Comparison of the sequence of the *Drosophila* cDNA with that of GAP-43 shows that the overall nucleotide and predicted amino acid homology is low (about 40% at the nucleotide level and 14% at the amino acid level), except for one domain (see below). The probes, however, do have predicted similarities. Both are charged and hydrophilic. Rat GAP-43 is associated with the internal face of the plasma membrane; however, it has no membrane-spanning hydrophobic domain (Karns et al. 1987) and the mechanism of attachment is unknown. The predicted *Drosophila* protein is also hydrophilic, except for a hydrophobic stretch of about 20 amino acids near the amino terminus. We were curious as to whether this might be a signal sequence, especially since there is one potential N-linked glycosylation site (Fig. 2), but the predictive indices of von Heijne (1984) suggest that the protein is more likely not secreted, although the hydrophobic domain may indicate that it is membrane-associated.

GAP-43 is a substrate for protein kinase C, and the predicted *Drosophila* protein contains serines that are in an appropriate context to fulfill this function (Kishi-moto et al. 1985), as shown in Fig. 2. GAP-43 binds calmodulin, and the sequence of the *Drosophila* protein between amino acids 145 and 158 forms a reasonably good consensus calmodulin-binding region, i.e. a basic amphiphilic a-helix (Erickson-Viitanen & DeGrado, 1987; Alexander et al. 1988).

There is one region of 57 nucleotides in the *Drosophila* clone that is 76% identical at the nucleotide level to GAP-43 and is the likely source of the cross hybridization between the species as demonstrated by Southern blot analysis of different restriction fragments. It is enclosed in the box in Fig. 2. The predicted amino acid sequence that spans this region is quite similar between the predicted *Drosophila* protein and GAP-43, as shown in Fig. 3. To investigate this apparent evolutionary conservation further, we isolated and sequenced cDNA clones from *C. elegans* that hybridized to the *Drosophila* probe.

**Fig. 3.** The ‘GAP motif.’ Portions of the protein deduced from KZ30 of *Drosophila*, GAP-43 of rat and CN2 of *C. elegans* (Ng and Fishman, unpublished) are aligned to emphasize their relatedness. Identical amino acids are shaded darkly; related amino acids are shaded lightly. Each sequence is continuous and the boxes serve to outline the most conserved regions. The numbers refer to the positions of the amino acids in the predicted open reading frames.
same rat GAP-43 probes under the same conditions that we used to identify the Drosophila clones (Ng & Fishman, unpublished). The longest open reading frame of the C. elegans clone is not clearly related over its entire length, but does contain one region similar to rat GAP-43. This region is the same as that conserved between Drosophila and rat. For convenience we refer to this conserved domain as the 'GAP motif' and, until a genetic lesion is identified, we refer to the Drosophila gene as KZ30. The predicted amino acid sequence of this region is not closely related to others in the data banks.

KZ30 is a single copy X-linked gene

Southern blot analysis of BamHI, HindIII, and EcoRI digests all gave rise to patterns indicative of a single copy gene (data not shown). To determine the cytological position of the KZ30 locus, we labelled subcloned fragments in pGem3Z with biotin and used them as probes for in situ hybridization to the polytene chromosomes of larval salivary glands. All cDNA clones hybridize at the same polytene chromosomal location, in region 17E on the X chromosome (data not shown). There are 9 bands within this region (Bridges, 1938). Bands 17E1–2 are clearly visible just distal to the area of hybridization that we will assign KZ30 within chromosomal bands 17E3–9.

KZ30 expression is developmentally regulated

Northern analysis reveals that the gene encodes two developmentally regulated poly(A)+ transcripts which are 2-4 and 2-0 kb, respectively (Fig. 4). Both transcripts are present throughout all developmental stages but increase in abundance between 4 and 13 h of embryonic development and again during the first 24 h of pupation. Both of these time periods are characterized by major morphogenetic movements and structuring (4 to 13 h) or restructuring (pupation) of the basic body plan (Campos-Ortega & Hartenstein, 1985). The abundance of the 2-4 kb transcript is very low in adult flies compared to that observed in 0 to 1 h embryos, a time prior to zygotic transcription and therefore representing stored maternal RNAs (Fig. 4). The RNA from adult flies comes from a population composed of both males and females; therefore, the weak signal detected in adult flies most probably corresponds to an ovarian message stored in the developing eggs. However, expression in non-ovarian tissues of the adult cannot be excluded at this time.

Maximal expression of the 2-0 kb transcript correlates with that of the 2-4 kb transcript, with the highest RNA levels between 4 and 6 h of embryonic development and the first 24 h of pupation. This transcript, however, can be detected at a low level in all of the remaining developmental stages tested.

Expression during the first half of embryogenesis is outside the nervous system

Transcripts from the KZ30 gene are expressed in a complex spatial and temporal pattern during embryogenesis. From fertilization through germ band elongation expression is uniformly distributed in all embryonic tissues. Subsequently, several regions of the embryo express the gene transiently. Among the tissues that hybridize intensely is the mesectoderm (also referred to as midline cells) (Fig. 5A,B,E), which gives rise to components of the CNS, both glial and neural, along the ventral midline (Thomas et al. 1988).

Intense labelling of the presumptive epidermis and amnioserosa is observed from germ band elongation through germ band retraction and to the end of dorsal closure. This labelling is not uniform in that the
hybridization is always most intense in the dorsalmost cells of the epidermis (Fig. 5C,D,F,G). Along with this dorsoventral polarity is a prominent segmental, anteroposterior polarity such that the highest level of expression is just anterior to the segment borders (Fig. 5G,H). Interestingly, no expression is observed in the CNS when neuroblasts are dividing and when the major axonal tracts are established (Fig. 5C,D,F).

Expression in the second half of embryogenesis is confined to the central nervous system
Following dorsal closure, expression of the *KZ30* gene appears exclusively restricted to the central nervous system (Fig. 6, 7, and schematized in Fig. 8). This expression begins as foci aligned in a linear array just dorsal to the longitudinal neuropile (Fig. 6A,B). Subsequently, other foci become visible. Serial transverse
Fig. 6. Expression in the CNS. All panels are sagittal or parasagittal sections with anterior to the left and ventral at the bottom. This figure illustrates KZ30 expression during the last 3 embryonic stages. Following dorsal closure, (early stage 15 (A and B)), expression in the dorsal epidermis (de) rapidly decreases and disappears. At this time, expression in the CNS first begins in a subset of cells lying just dorsal to the longitudinal neuropile (ln) of the ventral nerve cord (vnc). As development proceeds, more intense expression appears at particular foci within the ventral nerve cord (C,D,E), the suboesophageal ganglia (sbg) (F) and the supraoesophageal ganglia (spg) (see Fig. 7C). These foci appear in a segmental pattern along the ventral nerve cord with favourable sections showing 12 foci lying dorsal to the longitudinal neuropile (arrowheads in C), 12 foci lying ventral to the longitudinal neuropile, and 12 foci lying along the ventral side of the ventral nerve cord (arrowheads in E). Most likely condensation of the ventral nerve cord in the fully mature embryo makes it difficult to visualize these as segmentally reiterated foci (not shown). Scale bar: (A–F) 100 μm.

sections suggest that the foci have a bilateral symmetry with each hemisegment having at least four foci in addition to the dorsal one: there is one just ventral to the longitudinal neuropile, and three are at the periphery of the ventral nerve cord (Fig. 7A,B). Extensions of hybridization from the foci form processes that encircle the longitudinal neuropile, and extend dorsoventrally adjacent to the midline (Fig. 7C and Fig. 8). These foci are reiterated in a segmental pattern, 12 times along the ventral nerve cord (Fig. 6C,D,E), with the dorsal foci positioned between the posterior and anterior commissures (Fig. 8). This position coincides with that found for a subset of glial cells (Bastiani & Goodman, 1986; Jacobs & Goodman, unpublished). Foci of expression
Drosophila homologue of mammalian GAP-43

Fig. 7. Expression in the CNS of transverse sectioned embryos. All panels are transverse sections with ventral at the bottom. This figure illustrates the expression during the last 3 embryonic stages when expression is confined to the CNS. Expression in the CNS is first seen at foci lying just dorsal to the longitudinal neuropile (In) in the ventral nerve cord (vnc) (see Fig. 6A and B). Later, when the other hybridizing foci are visible, serial transverse sections reveal these foci to have bilateral symmetry with each hemisegment having at least 5 foci (arrowheads in A). 1 each immediately dorsal and ventral to the longitudinal neuropile, 1 adjacent to the ventral midline, and 2 each along the lateral periphery of the ventral nerve cord. Processes of hybridization appear to originate from the foci dorsal to the longitudinal neuropile and extend around it both laterally and medially (A and B) and occasionally laterally along the periphery of the ventral nerve cord. The medial processes appear to extend ventrally through the ventral nerve cord (arrowheads in C) into the ventralmost foci. Processes from these ventralmost foci are occasionally seen to extend a short distance laterally along the periphery of the ventral nerve cord (not shown). Within the supra- and suboesophageal ganglia (spg and sbg, respectively) it is difficult to recognize a repetitive pattern of focal expression; however, favourable sections suggest a bilateral symmetry (C). Additionally, in mature embryos, like the longitudinal neuropile of the ventral nerve cord, the neuropile within the 'brain' appears outlined by KZ30 expression (D). Other abbreviations: cec, circumoesophageal commissure. Scale bar: (A–D) 100 μm.

Discussion

The ‘GAP motif’

We identified the KZ30 gene by its cross hybridization with the mammalian growth-related protein GAP-43. One region of the predicted Drosophila protein sequence, the ‘GAP motif’, is closely related to that of rat GAP-43, and is also found in the predicted protein from the homologous gene in C. elegans. Outside of this domain the degree of relatedness based on sequence analysis is low, which, combined with its absence of expression in at least most neurones, makes it unlikely that the invertebrate protein serves an identical function to GAP-43. Other similarities between the predicted KZ30 protein and GAP-43 do exist. For example, both are charged and hydrophilic, and consensus sequences that could potentially serve as phosphorylation sites and for calmodulin binding are recognizable in KZ30.

The ‘GAP motif’ region is conserved between arthropods (which include Drosophila), nematodes and mam-
mals, which diverged from chordates about 700 million years ago, in the Precambrian era (Wood, 1988). Thus this domain is likely to have emerged prior to the division of the ancestral bilateral metazoans into Protostomia and Deuterostomia, and may represent a conserved functional domain. Conserved domains of other proteins from these phyla have been found. For example, the vertebrate EGF motif is observed in the predicted lin-12 protein of *C. elegans*, and several proteins of *Drosophila* including Notch, Delta and laminin B1 (Greenwald, 1985; Wharton et al. 1985; Vassin et al. 1987; Montell & Goodman, 1988). The homeodomain, first described in several developmentally important *Drosophila* genes, is also present in several mammalian genes and mec-3 of *C. elegans* (McGinnis et al. 1984; Way & Chalfie, 1988). It has been suggested that these conserved blocks represent functional units that perhaps arose by 'exon shuffling' between genes (Doolittle, 1985; Sudhof et al. 1985). Hence the 'GAP motif is a natural target for mutagenesis studies designed to elucidate the function of the mammalian protein.

**KZ30 expression is developmentally regulated and spatially complex**

The overall pattern of embryonic expression of KZ30 does not resemble that for *Drosophila* genes so far described by in situ hybridization. Its early expression is characterized by the sequential and transient labelling of discrete non-neural domains, whereas its late expression appears restricted to segmentally reiterated foci in the central nervous system.

Three features of early expression are especially noteworthy. First, the mesectoderm, which gives rise to some glial and neural components of the CNS (Thomas et al. 1988; Crews et al. 1988), labels intensely. The relative intensity of the signal indicates that expression in the mesectoderm is not the result of local stabilization of the maternal RNA but rather is the result of localized zygotic transcription. Though the significance of this labelling is unknown, KZ30 can now be added to the growing list of other genes (with no apparent functional relatedness) which also label the mesectoderm, e.g. single-minded (Thomas et al. 1988), Toll (Gerttula et al. 1988), Delta (Vassin et al. 1987), and c342 (Perkins and Perrimon, unpublished).

Second, KZ30 labels the presumptive epidermis and is polar along both the anteroposterior and dorsoventral axes. The early epidermal, as well as the later neural expression, places this gene among other developmentally important genes known to label subsets of cells within these two tissue types, e.g. fasciclin III (Patel et al. 1987); fushi tarazu, even-skipped, Ultrabithorax, and engrailed (Doe et al. 1985a,b; Doe & Scott, 1988). It will be interesting to determine how this tissue-specific expression is controlled and whether either transcript shows tissue specificity.

Third, expression is not observed in the CNS when neuroblasts are initially dividing or when the major axonal tracts are established (i.e. prior to 10 h of embryonic development) as opposed to the situation in vertebrates where GAP-43 is expressed in most, or likely all, neuronal cells during axonal growth. In part because its level dramatically increases during axonal elongation it has been proposed to be important for axonal growth. Clearly, this gene does not play a directly analogous role in *Drosophila* since it is not expressed in early developing neurones.

**KZ30 identifies subsets of cells in the CNS**

Following dorsal closure, KZ30 expression appears exclusively restricted to segmentally reiterated and bilaterally symmetrical cellular foci in the CNS. The topographic location of one of these foci in particular, that immediately dorsal to the longitudinal neuropile, strongly suggest that these labelling cells are not neurones but, in fact, glia. The glia dorsal to the longitudinal neuropile are believed to help establish the
orientation of this major axonal tract (Bastiani & Goodman, 1986; Jacobs & Goodman, unpublished). The remaining foci are located along the periphery of the ventral nerve cord and within the supra- and suboesophageal ganglia. These peripheral foci of expression can identify three possible cell types: a subset of glial cells, differentiated neurones, or nondividing neuroblasts, the latter of which are destined to contribute neurones to the adult nervous system during metamorphosis (Truman & Bate, 1988).

For helpful suggestions we thank especially Drs R. A. Raff at Indiana University, Robert Horvitz at MIT, J. Novotny at the MGH, and R. Klausner at the NIH. We thank Drs T. Kornberg, B. Meyer, K. Zinn and C. Goodman for cDNA libraries, Beth Noll for excellent technical assistance, and J. Jackson for manuscript preparation. This work was supported by Howard Hughes Medical Institute and an ACS grant (to N.P.). We thank R. Jacobs and C. Goodman for communicating results on glial cells prior to publication.

References


(Accepted 6 December 1988)