Development 122, 567-578 (1996)
Printed in Great Britain © The Company of Biologists Limited 1996
DEV8293

REGA-1 is a GPI-linked member of the immunoglobulin superfamily present on restricted regions of sheath cell processes in grasshopper

Elaine C. Seaver†, Ellen M. Carpenter* and Michael J. Bastiani

Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

*Present address: Department of Human Genetics, University of Utah School of Medicine, 5440 EIHG, Building 533, Salt Lake City, UT 84112, USA
†Author for correspondence

SUMMARY

REGA-1 is a glycoprotein localized to sheath cell processes in the developing CNS when NBs are producing progeny and neurons are maturing and extending processes. It is also present on a subset of muscles and on the lumenal surface of the ectoderm in the embryonic appendages when pioneer neurons are growing into the CNS. REGA-1 is associated with the extracellular side of the cell membrane by a glycosyl-phosphatidylinositol linkage. We have identified a cDNA clone encoding REGA-1 using a sequence from purified protein. Sequence analysis defines REGA-1 as a novel member of the immunoglobulin superfamily containing three immunoglobulin domains and one fibronectin type III repeat. Each Ig domain has distinct sequence characteristics that suggest discrete functions. REGA-1 is similar to other Ig superfamily members involved in cell adhesion events and neurite outgrowth.

Key words: glia, grasshopper, sheath cell, monoclonal antibody, immunoglobulin superfamily, GPI-linked, axon outgrowth, insect

INTRODUCTION

The generation of cellular diversity during neuronal development is an important prerequisite for establishing the complex cellular connections necessary for nervous system function. The stereotypy observed in the mature insect nervous system is derived from a defined pattern of neuronal precursor cells or neuroblasts (NBs) (Bate, 1976). The nervous system in insects arises from a specialized region of the ectoderm, the neurogenic region. Within the neurogenic region, NBs enlarge and delaminate to the interior of the ectoderm. Asymmetric divisions of NBs yield ganglion mother cells (GMCs), which later divide symmetrically to produce neurons or glia. In each hemisegment, 30 unique NBs produce distinct families of progeny. Cells of a neuroblast family are organized in a tightly packed column, generally perpendicular to the ectodermal surface, with adjacent family members having closely apposed cell membranes (Bate, 1976). The tightly juxtaposed family members are connected through gap junctions that become uncoupled at the onset of axonogenesis (Goodman and Spitzer, 1979; Raper and Goodman, 1982). This physical arrangement may function to maintain members of a single family in close physical contact and spatially segregate them from other families. Each NB family is physically surrounded by processes of a small group of sheath cells (Doe and Goodman, 1985a). Sheath cells are likely to play a role in maintaining the integrity of a family and are proposed to play a support role in neurogenesis (Doe and Goodman, 1985a), but it is not known if sheath cells are limited to providing only structural support. Sheath cells may also have an instructive role in the development of NB progeny. Can sheath cells influence the fate of progeny of a NB with positional information that they inherit? Is the presence of sheath cells necessary for regulating or promoting cell divisions of NBs and GMCs? Close physical contact between sheath cell processes and members of a NB family suggests that information may be transferred from sheath cells to NB family members via surface molecules. Identification of surface antigens present on sheath cells and determination of their function will aid in elucidating the role of sheath cells in neurogenesis.

The number of identified surface proteins in the developing nervous system that are members of the immunoglobulin (Ig) superfamily is impressive; they represent perhaps the largest category of identified cell surface proteins in the nervous system. The majority of these proteins are present on the surface of neurons and include NCAM, neuroglian, fasciclin II, fasciclin III, TAG-1/axonin-1, L1/NILE, contactin, MAG (reviewed by Bixby and Harris, 1991), and BIG-1 (Yoshihiro et al., 1994), unc-5 (Leung-Hagesteijn et al., 1992), lachesin (Karlstrom et al., 1993), amalgam (Seeger et al., 1988), OBCAM (Schofield et al., 1989), neurotrimin (Struyk et al., 1995) and Nr-CAM/Bravo (Grumet et al., 1991). Ig superfamily members are numerous and represent molecules found in a wide range of species. These proteins represent many functions including cell-cell adhesion, the immune response, and organization and regulation of muscle (Chothia, 1994), illustrating that presence of an Ig domain does not predict a single function.

We present a new member of the Ig superfamily called REGA-1, found exclusively on non-neuronal elements of the

nervous system. REGA-1 is expressed on processes of sheath cells in the developing CNS of grasshopper. In a previous paper, REGA-1 was characterized using the monoclonal antibody 7F7 and found to be present on the surface of a subset of glial cells in the developing CNS of the grasshopper embryo and adult (Carpenter and Bastiani, 1991). Initial characterization revealed that REGA-1 is a protein with a M_r of approximately 60×10^3. In this paper, we have characterized REGA-1 biochemically and shown it to be associated with the cell membrane via a glycosyl-phoshatidylinositol (GPI) linkage. We have also purified the protein and isolated the cDNA encoding REGA-1. Sequence analysis of clones encoding REGA-1 reveals that it is a member of the Ig superfamily, with sequence similarity to other family members present in the developing nervous system. REGA-1 is present on localized regions of sheath cell processes surrounding NB families at the time when NBs are actively dividing and progeny maturing into neurons.

MATERIALS AND METHODS

Grasshopper embryos were used from a colony maintained at the University of Utah and were staged according to Bentley et al. (1979).

Immunocytochemistry

Grasshopper embryos were labeled essentially as described in Seaver et al. (1991). The 7F7 polyclonal serum was generated against REGA-1 purified from embryonic lysates as described (Carpenter and Bastiani, 1991). Drosophila embryos were labeled using standard procedures (Ashburner, 1989). Staging was performed according to that outlined by Campos-Ortega and Hartenstein (1985).

Biochemical analysis

Analysis of the membrane attachment of REGA-1 was performed according to Chang et al. (1992). A 1:500 dilution of the 7F7 mAb was used to assay the presence of REGA-1 in embryos. Western analysis of proteins was performed using conditions outlined by Ganifornina et al. (1995), with a 1:1000 dilution of the 7F7 polyclonal sera.

The carbohydrate content of REGA-1 was analyzed as follows: membrane proteins of embryos were prepared according to Seaver et al. (1991), and proteins were deglycosylated using peptide-N-glycosidase F (PNGase F) from *Flavobacterium meningosepticum* (Boehringer Mannheim), using conditions described (Ganfornina et al., 1995).

Immunoaffinity purification and protein microsequencing

REGA-1 was purified from embryonic grasshopper lysates as described by Carpenter and Bastiani (1991). Purified protein was digested with sequencing grade trypsin (Boehringer Mannheim) according to methods in Karlstom et al. (1993), except that the protein was digested directly from a gel slice for 24 hours. Cleaved peptides were separated by reverse phase HPLC and sequenced by Edman degradation (Applied Biosystems 477A).

PCR amplification

A sense primer 4s (5′GAGAATTCGAC/TTTC/TGGIATITAT/CACITG) and an antisense primer TN (5′GAGAATTCCCIC-CIAAC/TTCA/GTTA/G/C/TGT) were designed from sequenced peptides and used to amplify via PCR a 350-base pair (bp) fragment from embryonic grasshopper cDNA using Taq Polymerase (Saiki et al., 1988). The following conditions for amplification were performed in a thermal cycler (Perkin Elmer Cetus): denaturation, 94°C for 10 seconds; annealing, 50°C for 30 seconds; and elongation, 72°C for 2 minutes for 40 cycles. The PCR product was reamplified with an internal primer PQ (5′GAGAATTCCCICAA/GCCICCIGGA/G/C/TGAA/GTAC/TGA), using the same conditions. The 350-bp PCR fragment (4STN) was subcloned into the pCR-II plasmid using the TA cloning system (Invitrogen) and sequenced as described below.

Library screening

Embryonic grasshopper libraries were screened to isolate clones encoding the entire ORF of REGA-1. The 4STN fragment generated by PCR was labeled with [^{32}P]dCTP by random priming (Pharmacia) for use as a probe. Initially, λgt11 and λgt10 libraries prepared by K. Zinn (Snow et al., 1988) were screened. To isolate full-length clones, a 45% embryonic cDNA library was constructed in λZAP. 5×10^5 pfu from an unamplified version of the library were screened and the resulting clones (1.5 kb, 4.5 kb, 4.5 kb, 5 kb) were further characterized by sequence analysis and found to contain the 5′ end of the REGA-1 ORF.

Northern analysis

Grasshopper mRNA was prepared from 45% embryos (see Fig. 2 legend) as described (Ganfornina et al., 1995). Radiolabeled 4STN was used as a probe.

DNA sequencing and analysis

Nucleotide sequences of REGA-1 cDNA clones were determined by the dideoxy chain-termination method of Sanger et al. (1977), using double-stranded plasmid DNA as template for T7 DNA polymerase (Sequenase; United States Biochemical). The sequence of the REGA-1 ORF was determined by sequencing both strands of 4stn-3 (3.1 kb) and 4stn-13 (1.5 kb). DNA and protein sequences were analyzed using the GCG Wisconsin sequence analysis package version 8 (Devereux et al., 1984). The REGA-1 sequence was compared with databases as a protein using BLAST (Altschul et al., 1990) and FASTA (Pearson and Lipman, 1988) search programs (ktup=2). The following databases were searched: PIR-Protein, GenBank, Swiss Protein and EMBL. From searches with individual domains of REGA-1, top-scoring candidates (high score for BLAST and Opt score from FASTA) were included in multiple sequence alignments using the PILEUP program of GCG. Individual domains of REGA-1 were used to search the databases with the ktup=1 setting, as suggested by Pearson (1991). Identification of the first Ig domain as a member of the I set of Ig domains was performed by comparing Ig domains of REGA-1 with the profile defined by Harpaz and Chothia (1994). The first Ig domain meets the criteria defined by the profile with no mismatches. The hydrophobicity plot of Fig. 3B was generated with the PEPTIDESTRUCTURE program of the GCG analysis package.

In situ hybridization

In situ hybridization to whole-mount embryos was performed according to a protocol provided by J. Broadus (Broadus and Doe, 1995) and is outlined in Ganfornina et al. (1995). An RNA probe was synthesized from the clone 4stn-13 (see Fig. 1) containing most of the REGA-1 ORF.

RESULTS

REGA-1 is a glycoprotein anchored to the cell membrane via a glycosyl-phoshatidylinositol linkage

The 7F7 mAb recognizes a surface epitope of REGA-1, as demonstrated by labeling of live embryos (Carpenter and Bastiani, 1991). We investigated its carbohydrate content by treating membrane proteins with PNGase F, which cleaves N-linked carbohydrates. The relative mobility of REGA-1 is decreased by approximately 7×10^3 M_r in PNGase F-treated membrane proteins relative to untreated proteins (not shown),

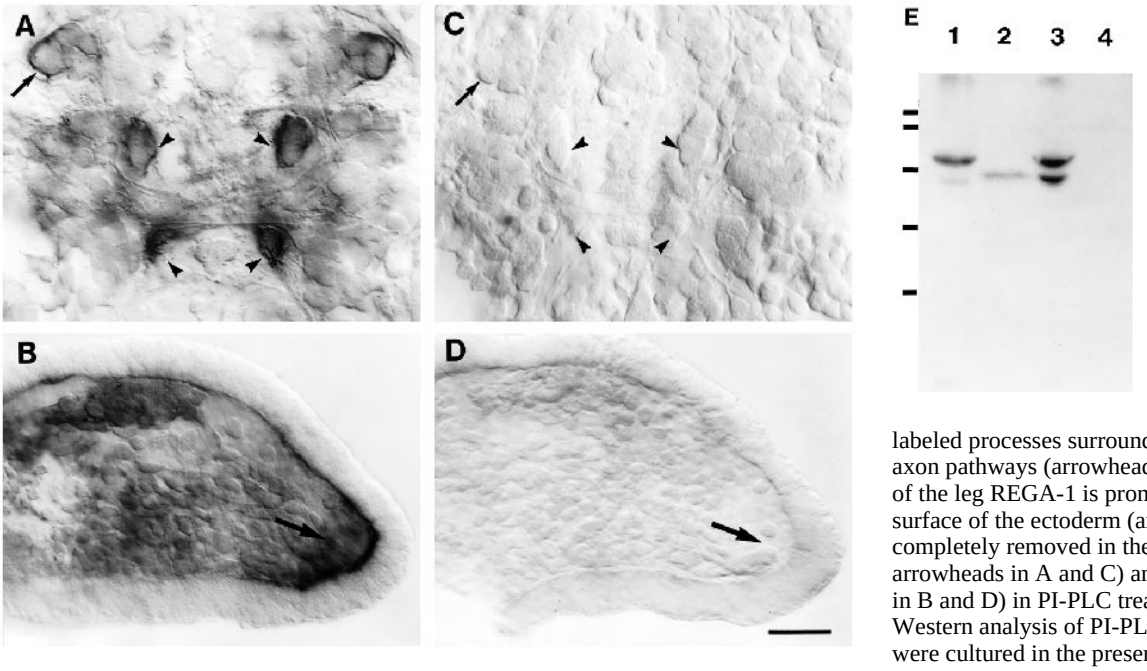

Fig. 1. REGA-1 is linked to the cell membrane via a glycosyl phosphatidylinositol anchor. Embryos treated with PI-PLC for 2 hours are exposed to antibodies that recognize REGA-1. (A) and (B) are untreated embryos and (C) and (D) exposed to PI-PLC. (A) In the CNS, labeled processes surround clusters of cells lining axon pathways (arrowheads). (B) In the distal tip of the leg REGA-1 is prominent on the lumenal surface of the ectoderm (arrow). Labeling is completely removed in the CNS (compare arrowheads in A and C) and leg (compare arrows in B and D) in PI-PLC treated embryos. (E) Western analysis of PI-PLC treatment. Embryos were cultured in the presence (lanes 1,2) or absence (lanes 3,4) of PI-PLC and then culture media and membrane fractions were collected. Lane 1, supernatent, control; Lane 2, membrane preparation, control; Lane 3, supernatent, PI-PLC treated; Lane 4, membrane preparation, PI-PLC treated. Molecular mass markers ($M_r \times 10^{-3}$) are: 116, 97, 66, 45, 31. Scale bar, 30 μm.

suggesting that REGA-1 is a glycosylated protein, containing N-linked carbohydrate.

We investigated the nature of its membrane association by exposing live embryos to phosphatidylinositol-specific phospholipase C (PI-PLC), which selectively cleaves GPI linkages. Enzymatic treatment of embryos with PI-PLC removes all 7F7 antibody labeling (Fig. 1C,D). In contrast, embryos cultured without enzyme exhibit characteristic labeling for REGA-1 in both the CNS (Fig. 1A) and leg (Fig. 1B, see below). Control embryos were labeled for fasciclin IV, a transmembrane protein expressed in the CNS and legs (Kolodkin et al., 1992). Fasciclin IV labeling is unaffected by treatment with PI-PLC (not shown). In contrast, fasciclin I, a known GPI-linked protein (Hortsch and Goodman, 1990), is removed by PI-PLC treatment (not shown). Immunoblot analysis demonstrates that the 60×10^3 M_r REGA-1 band is not detectable in membrane preparations of treated embryos (Fig. 1E, lane 4) but is associated with membranes of untreated embryos (lane 2). In treated embryos all of REGA-1 is released into the culture media (Fig. 1E, lane 3). These results demonstrate that REGA-1 is associated with the cell membrane via a GPI linkage throughout the embryo. A substantial portion of REGA-1 is present in the media of untreated embryos, suggesting the presence of a released form of REGA-1 (Fig. 1E, lane l). The released and membrane forms have slightly different mobilities through a polyacrylamide gel (compare lanes 1 and 2).

Isolation and sequence of grasshopper cDNAs encoding REGA-1

REGA-1 was purified from embryonic grasshopper lysates by affinity purification using the 7F7 mAb, as described previously (Carpenter and Bastiani, 1991). Tryptic fragments were generated from purified protein and several peptides sequenced. Degenerate primers representing peptide sequence were used to amplify by PCR a 350-bp fragment, 4STN, from embryonic grasshopper cDNA. Sequence analysis of 4STN revealed a continuous open reading frame (ORF) through the length of the fragment and a sequence corresponding to peptides internal to the primers. Northern analysis using 4STN as a probe reveals a single transcript of approximately 4.5-5 kb from embryonic mRNA at 45% of embryogenesis (Fig. 2), a stage when there is extensive REGA-1 labeling.

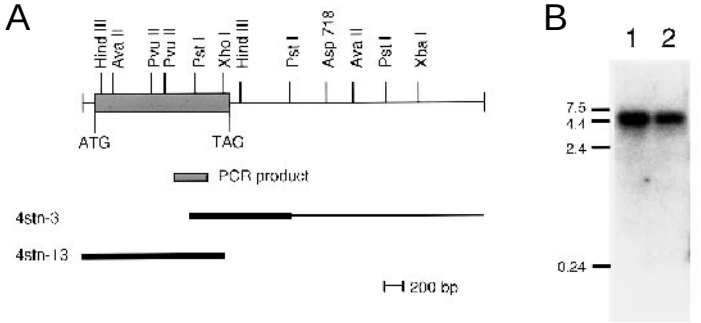

Fig. 2. (A) Schematic representation of the REGA-1 message. Below the diagram are clones used to derive sequence of the REGA-1 ORF, 4stn-3 and 4stn-13. The ORF is a shaded box. The 4STN probe used to screen libraries is a filled box labeled 'PCR product' positioned beneath the corresponding sequence in REGA-1. Sequenced regions of clones are in bold lines. (B) Northern analysis of REGA-1. 4STN was used to probe mRNA isolated from 45% embryos. A single transcript of approximately 5 kb is detected when 10 μg (lane 1) and 5 μg (lane 2) of poly(A)+ are loaded per lane.

4STN was used in an initial screen of 5×10^5 pfu from a λgt11 embryonic grasshopper cDNA library (Snow et al., 1988). A single positive with a 3.1-kb insert, 4stn-3, was found to contain a portion of the REGA-1 ORF (Fig. 2). An additional 5×10^6 pfu were screened from λgt11 and λgt10 libraries yielding 4stn-3 again, and a larger clone with a 3.5-kb insert containing an additional 200 bp of ORF sequence. To isolate the 5′ end of the REGA-1 ORF, a cDNA library was constructed in λZAP from 45% embryos. A screen of 5×10^5 pfu of the unamplified library yielded four positive clones containing inserts of 4.5 kb, 4.5 kb, 5 kb and 1.5 kb. The three larger clones are approximately the size of the transcript

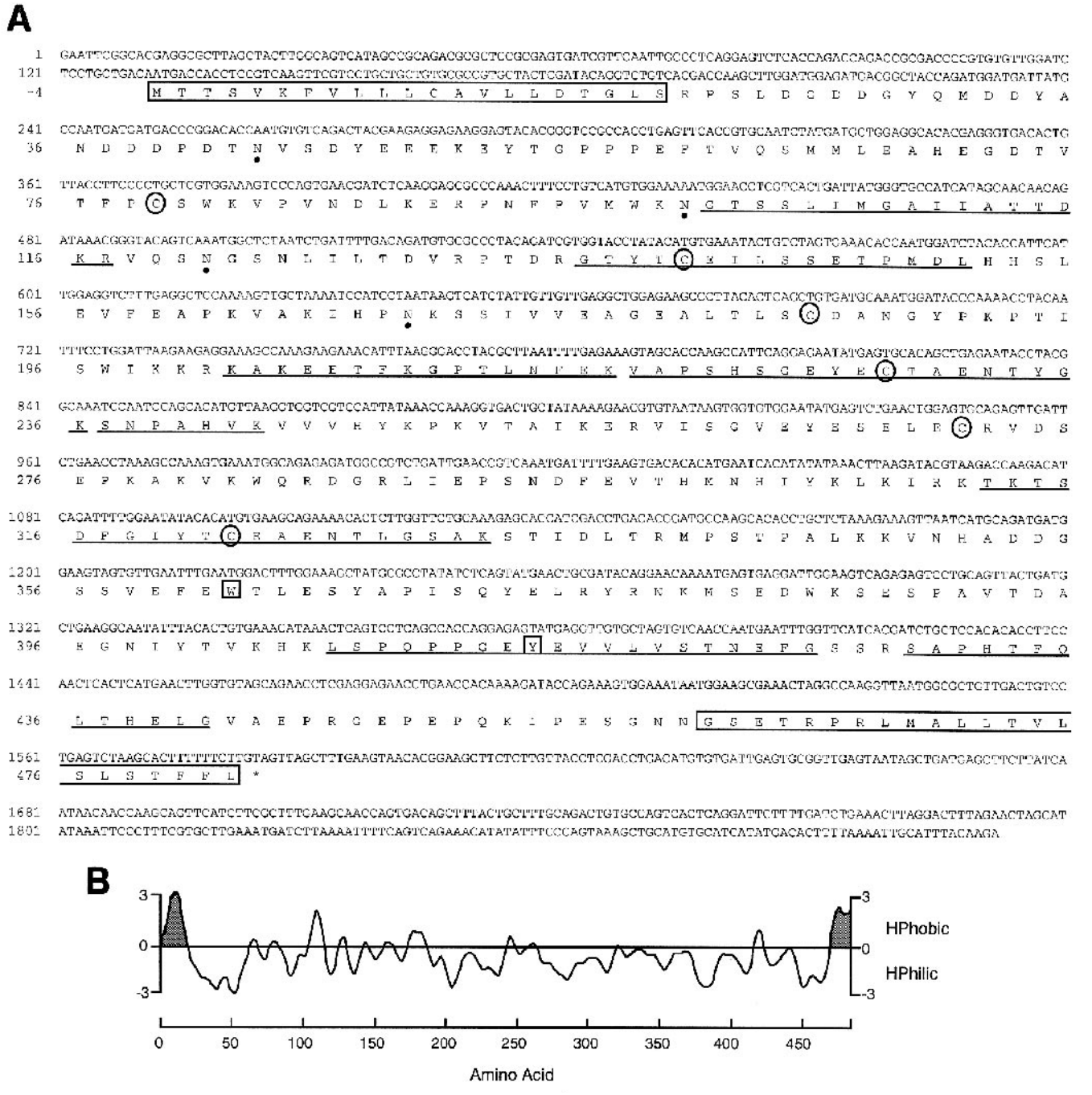

Fig. 3. The REGA-1 cDNA sequence predicts a glycosylated GPI-linked membrane protein. (A) cDNA sequence and predicted amino acid sequence of REGA-1. The putative signal peptide cleaved from the mature protein is boxed at the amino terminus. Boxed residues at the C terminus indicate peptide sequence removed during attachment of PI tail. Peptide sequences determined directly from purified protein are underlined. Conserved cysteines indicating the 3 Ig domains are circled and tyrosine and tryptophan aa characteristic of F3 repeats are highlighted in square boxes. Filled circles indicate potential N-linked glycosylation sites. (B) Hydropathy analysis (Kyte and Doolittle, 1982) of the predicted REGA-1 ORF indicates N-terminal and C-terminal hydrophobic stretches, characteristic of a signal sequence and addition of a GPI-tail, respectively. Both regions are shaded. Hydrophobic stretches appear as positive values. The EMBL accession no. for REGA-1 is X93601.

detected by northern analysis, and therefore probably represent full-length transcripts.

The REGA-1 ORF contains sequences coding for eight peptides (underlined in Fig. 3A) that correspond to tryptic fragments sequenced from purified REGA-1 protein. Sequence analysis of clones 4stn-3 and 4stn-13 reveals a single long ORF of 1455 nt (Fig. 3A), beginning with a methionine at nt 132. The methionine is the first in-frame methionine after a stop codon at position 60. In addition, nucleotides flanking the start site at positions −3 to −1 represent a conserved translation initiation sequence (Cavener and Ray, 1991). The stop codon TAG at nt 1584 is followed by approximately 3 kb of 3′ untranslated sequence. The ORF encodes a protein of 441 amino acids with a predicted mass of approximately 50×10^3 M_r. The difference between the predicted size and observed relative mobility of REGA-1 at 60×10^3 M_r may be accounted for by glycosylation, since there are four potential N-linked glycosylation sites (Fig. 3A). This is consistent with our biochemical findings that the relative mobility of REGA-1 is decreased by approximately 7×10^3 M_r when treated with PNGase F.

Hydropathy analysis (Fig. 3B) demonstrates that the initial methionine is immediately followed by 17 contiguous hydrophobic amino acids, forming a classic signal peptide sequence characteristic of proteins secreted or associated with the cell membrane (von Heijne, 1990). A putative peptidase cleavage site follows Ser21 (Fig. 3A). The carboxy terminus of the ORF also contains a hydrophobic stretch of 15 amino acids (Fig. 3B). A hydrophobic region immediately preceded by a short hydrophilic stretch at the carboxy terminus is characteristic of proteins linked to the cell surface via a GPI lipid anchor (Gerber et al., 1992). A putative cleavage site for addition of a GPI tail follows Asn460 (Fig. 3A).

Confirmation that the cDNA sequence is correct for REGA-1 was obtained from the following lines of evidence. Amino acid sequence representing peptides sequenced from affinity-purified REGA-1 is found in the predicted ORF of the isolated clones (underlined in Fig. 3A). In addition, fusion protein containing most of the REGA-1 ORF is recognized by the polyclonal sera generated against purified antigen (not shown). Conversely, antibodies generated against recombinant protein encoding the REGA-1 ORF recognize a band with a relative mobility of 60×10^3 M_r from embryonic grasshopper membrane preparations (not shown). Finally, in situ hybridization using a dixogenin-labeled antisense RNA probe generated from the 4stn-13 clone was used to localize the message in the embryo. The pattern of localization of the message and antibody is comparable in all regions of the embryo (Fig. 4 and see below). On the dorsal side of the CNS, 7F7-labeled processes surround cells along the anterior edge of the segmental nerve, the medial edge of the connective and the anterior commissure (Fig. 4A,E). In all of these positions REGA-1 mRNA is present (compare Fig. 4A,C). mRNA also co-localizes with immunoreactive processes surrounding cells along the lateral edge of the connective (arrow in Fig. 1A). No signal is seen with a sense probe control (not shown).

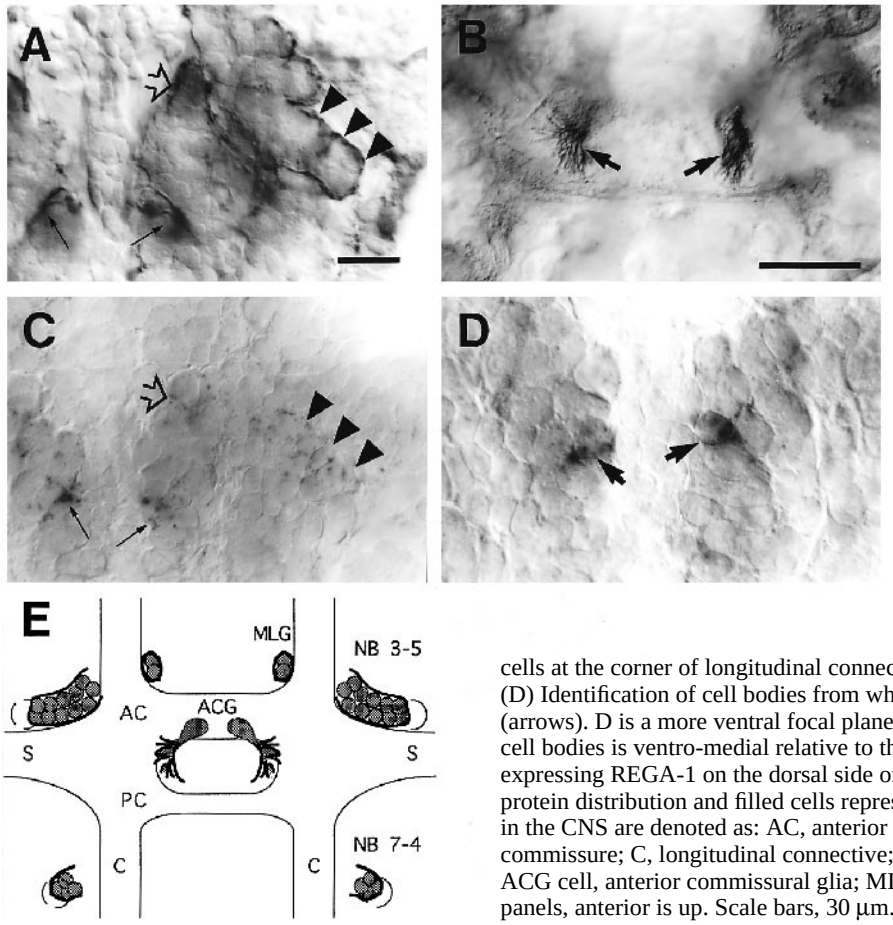

Fig. 4. Correspondence of mRNA and protein localization of REGA-1 in embryos at 34%. (A,B) Protein localization of REGA-1 by immunocytochemistry. (C,D) mRNA distribution by in situ hybridization. (A) Labeled processes surround a cluster of cells lining the anterior edge of the segmental nerve (arrowheads). Processes of MLG cells also label (open arrow). Arrows point to labeling of crescent processes (see B). (C) Similar view as A showing mRNA expression at position of the cresents (arrows), MLGs (open arrows) and cells anterior to the segmental nerve (arrowheads). A and C are a similar plane of focus. (B) REGA-1 is restricted to the processes of the ACG cells at the corner of longitudinal connective and anterior commissure. (D) Identification of cell bodies from which the crescents processes arise (arrows). D is a more ventral focal plane relative to C. The position of the cell bodies is ventro-medial relative to the crescents. (E) Diagram of cells expressing REGA-1 on the dorsal side of the CNS. Bold lines represent protein distribution and filled cells represent RNA expression. Major nerves in the CNS are denoted as: AC, anterior commissure; PC, posterior commissure; C, longitudinal connective; S, segmental nerve. Abbreviations: ACG cell, anterior commissural glia; MLG, medial longitudinal glia. In all panels, anterior is up. Scale bars, 30 μm.

Together these findings confirm that the isolated clones encode sequence for the REGA-1 antigen.

Identification of glial cell bodies at the dorsal surface of the CNS

Initial characterization of the embryonic labeling pattern with the 7F7 mAb did not allow all cells expressing REGA-1 to be clearly identifed. The protein is regionally localized to cellular processes and many cell bodies do not label. To identify cell bodies that express REGA-1, we performed in situ hybridization analysis. Labeling is prominent in the processes of the ACG cells, the crescents (Carpenter and Bastiani, 1991), positioned on the dorsal side of the CNS where anterior commissure and longitudinal connective meet (Fig. 4B,E). In situ hybridization reveals that the cell bodies of the ACG cells are located ventrally, rostrally and medially to the cresent processes and ventral to the anterior commissure (Fig. 4D). Labeled processes along the medial edge of the connective anterior to the commissures also show restricted protein expression (Fig. 4A, open arrow). In this case REGA-1 is produced by a small group of 2-3 cells. We have named these cells the medial longitudinal glial cells, or MLGs. The relationship of the MLGs with the longitudinal glial cells overlying the connectives in *Drosophila* (Jacobs and Goodman, 1989) is unknown. A summary of the cells expressing REGA-1 at the dorsal region of the CNS is presented in schematic form in Fig. 4E.

REGA-1 is expressed on restricted regions of sheath cells

REGA-1 is also present on the ventral side of the CNS. This region contains NBs and several other cell types including cap cells, sheath cells and ganglion mother cells (GMCs). These groups of cells have a unique spatial organization and, along with the more dorsally located neurons, constitute NB families (Fig. 5C). The family members form an organized column surrounded by processes of sheath cells extending from the ventral to the dorsal surface of the CNS. Individual NBs can be identified by their position within the ventral sheet (Doe and Goodman, 1985a). In a ventral view of the embryo, 7F7 labeling appears to surround the NB, but it is difficult to determine if REGA-1 is localized to the NB itself or slightly more dorsal or ventral. In the previous paper (Carpenter and Bastiani, 1991) it was suggested that labeling associated with NB families could be present on sheath cells. To clarify this issue, individual ganglia were dissected, turned sideways and observed from a frontal view (Fig. 5B). REGA-1 is present on cellular processes along the edge of several cell layers (Fig. 5B). Cell bodies do not label, indicating that REGA-1 is regionally localized within the cell. The morphology of the labeled processes indicates that REGA-1 is present on processes of sheath cells. The sheath cell body is located at the ventral surface of the neural ectoderm and its processes extend to the dorsal surface of the CNS surrounding NB families (Fig.

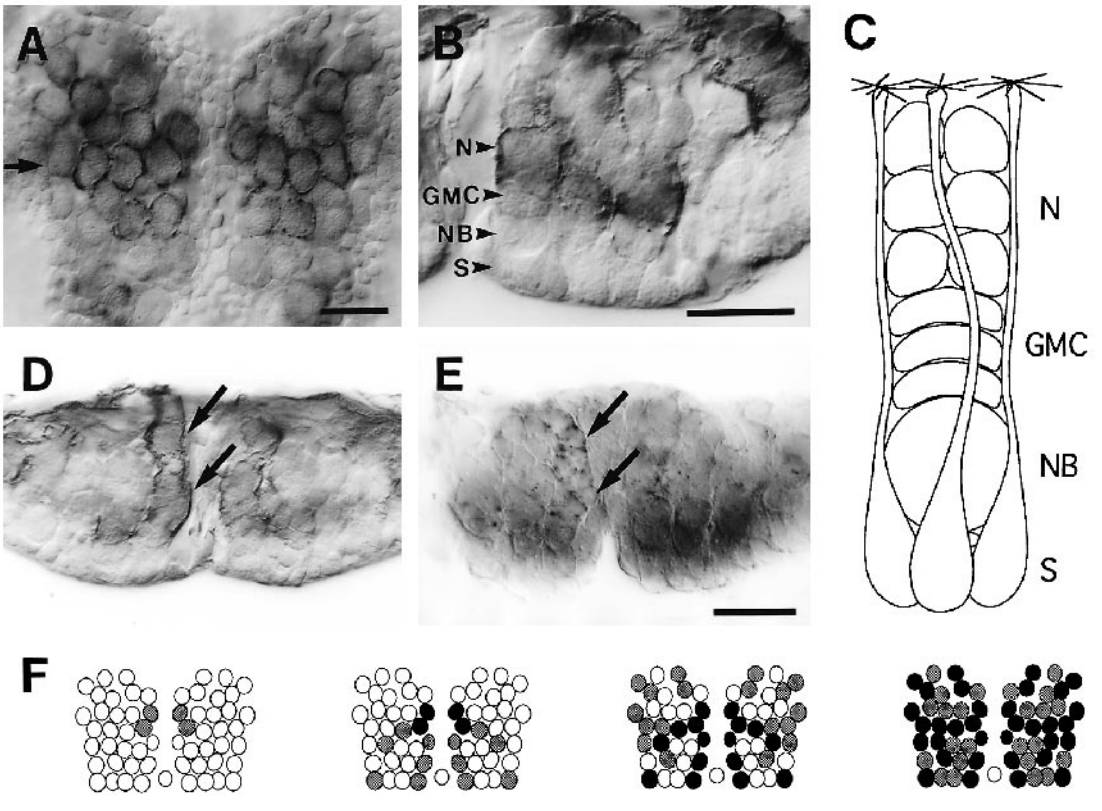

Fig. 5. REGA-1 is restricted to distinct regions of sheath-cell processes. (A,B,D,E) are views of single segments at 34% of embryogenesis. (A) Ventral view of a thoracic ganglion showing NB families surrounded by REGA-1 labeling. Arrow denotes position of NB row 4. (B-E) Frontal view of thoracic ganglion rotated relative to A. (B) Labeling is restricted to the region of sheath-cell processes that appose GMCs and neurons. Compare with other portions of sheath cells where NBs and sheath cell bodies are located. Families of NBs 4-1, 4-2, 4-3 and more laterally 3-5 are labeled. (C) Schematic of NB family depicting spatial arrangement of cells. (D) Sheath-cell processes of family 3-1 are extensively labeled from GMCs to dorsal surface of the CNS (arrows). (E) In situ hybridization with REGA-1 probe. Same view as D. The REGA-1 message is localized to GMC and neuron layers of NB families (arrows). (F) All lateral NB families label and labeling appears in a specific temporal sequence. Ventral view as in A. The first appearance of labeling of a NB family is in grey and families labeled from a previous stage are black. The order of appearance of labeling is as follows: 28-29%, NB 3-1 and 4-1; 30-31%, NB 4-2, 5-1, 5-4, 6-1, 7-1, 7-4; 32-33%, NB 1-1, 2-2, 2-4, 2-5, 3-5, 4-3, 4-4, 5-6 and 6-3; 34-35%, NB 1-2, 2-1, 2-3, 3-2, 3-3, 3-4, 5-2, 5-3, 6-2, 6-4, 7-2 and 7-3. The MNB does not label. Abbreviations: S, sheath cell; NB, neuroblast; GMC, ganglion mother cell; N, neuron. Scale bars, 40 μm.

5C). The processes of sheath cells span many cell layers, as does the REGA-1 labeling. No other described cell type in this region of the ganglion has these morphological characteristics. REGA-1 expression is restricted to regions of sheath cell processes adjacent to cell layers containing the GMCs and newly born neurons and is absent from the layer of NBs and sheath cell bodies. Most NB families are surrounded by labeling in the GMC layer and 1-2 cell diameters in the neuron layer. However, a few NB families are more extensively labeled. This is most striking for NB families 3-1, 3-5 and 7-4, where sheath cell processes are labeled from the most ventral GMC to the dorsal surface of the embryo (Fig. 5D). Several labeled processes visible at the dorsal surface of the CNS are processes of sheath cells. Arrowheads in Fig. 4A show labeled dorsal processes that emanate from the GMC layer of NB family 3-5. The progeny of NB 3-5 are oriented diagonally in the rostro-caudal plane at the dorsal surface. In addition, processes surrounding neurons along the lateral edge of the connective (arrow, Fig. 1A) are those of sheath cells associated with NB family 7-4. In situ hybridization reveals that the REGA-1 message is localized to GMC and neuron cell layers. The message distribution associated with NB family 3-1 and more lateral NB families mirrors the distribution of the protein (compare Fig. 5D,E). The REGA-1 message co-localizes to the same layers as the protein for the NB family 3-5 (compare arrowheads in Fig. 4A,C).

REGA-1 expression on sheath cells is temporally specific for individual NB families. When labeling first appears on sheath cell processes it has a discontinuous spotty appearance. Labeling intensifies with time, resulting in a continuous rind around NB families (Fig. 5A). The earliest labeling occurs at 28%-29% of embryogenesis, where REGA-1 is expressed on sheath cells surrounding NB 3-1 and 4-1 (Fig. 5F). In the next wave of REGA-1 expression from 29-30%, immunoreactivity spreads to other families, resulting in labeling of all NB families along the medial edge of the hemisegment with the exception of NB 2-1. At this stage labeling also appears around the families of NB 4-2, 5-4 and 7-4 (Fig. 5F). From 31-32% of embryogenesis, all of row 4 begin to express REGA-1 along with most lateral families and families of NB 6-3, 2-2 and 1-1. At this stage, labeling on row 4 is the first complete NB row to label. By 35-36% of embryogenesis, all lateral NB families express REGA-1. Labeling on sheath cell processes does not have equal intensity surrounding all families and is especially prominent on families of NB 2-3, 2-5, 3-1, 3-3, 3-5, 4-1, 4-2, 4-3 and 7-4. The families of NB 7-3 and 3-4 are weakly labeled. The median NB family never labels. Labeling fades gradually at 45% when other labeled components of the CNS lose their immunoreactivity.

REGA-1 is expressed in the appendages and transverse muscles

In addition to the complex expression pattern in the CNS, REGA-1 is present in all the appendages. The labeling pattern observed in legs, mouth parts, antennae and labrum is essentially the same and the time course of expression roughly corresponds with that of the legs. REGA-1 is present early in development of the leg on the lumenal side of the ectoderm (Fig. 6A). As the leg grows, labeling is most prominent at the distal tip of the leg (Fig. 6B). Intense labeling surrounds the Ti1 neurons as they begin to extend processes towards the CNS. By 35% of embryogenesis, additional elements are visible and two pairs of large irregular shaped cells on the inner surface of the ectoderm label (Fig. 6C, long arrows). The more proximal cell is aligned with the circumferential portion of the Ti1 process in the trochanter segment and the more distal cell is in the femur. There is also expression in mesodermal cells at the base of the leg (open arrow). This is clearly demonstrated by viewing the appendage from the base, as shown for a mouth part in Fig. 6E. The REGA-1 message is present in these mesodermal cells (Fig. 6E). Between 34% and 37% of embryogenesis, the ectodermal labeling resolves into three stripes (not shown) located in the femur, tibia and tarsus. Labeling on the ectoderm gradually fades and is gone by 43% of embryogenesis. The REGA-1 message co-localizes with the protein in the leg. In Fig. 6D the message is visible in mesodermal cells at the base of the leg (open arrow), in two cells in the medial portion of the leg (long arrows) and at the distal tip of the leg (arrowhead).

Because labeling on the lumenal surface of the ectoderm does not outline cell bodies, it was not initially clear which cells produce REGA-1 in this area. Localization of REGA-1 to the lumenal surface of the ectoderm in the appendages, together with observation of a released form of REGA-1, initially led us to postulate that REGA-1 may be released and incorporated into basal lamina. However, labeling in embryos treated with PI-PLC is absent, suggesting that REGA-1 is associated with cell membrane. In situ hybridization shows that REGA-1 is produced by ectodermal cells (Fig. 6D). Only ectodermal cells at the distal tip of the leg from a 34% embryo express the REGA-1 message corresponding to the regional restriction of the protein (compare arrowhead and short arrow in Fig. 6D).

REGA-1 is present in the transverse muscle bands overlying the CNS (Fig. 7). The transverse muscles arise from a mass of mesodermal cells in the lateral body wall that migrate medially and meet at the midline, organizing around the muscle pioneer cells (Ho et al., 1983). The mesodermal cells then differentiate into segmentally repeated muscle bands. REGA-1 is present in the developing muscle: mesodermal cells in the lateral body wall, medially migrating mesodermal cells and transverse muscle pioneer cells express REGA-1 (not shown). No labeling is observed in the nearby longitudinal muscles, which also arise from mesodermal cells in the body wall.

REGA-1 is a member of the immunoglobulin superfamily

To determine the domain composition of REGA-1 and its relationship with previously characterized molecules, we searched sequence databases with the REGA-1 ORF. Searches identify REGA-1 as a member of the Ig superfamily containing three Ig domains (Williams and Barclay, 1988) and one fibronectin type III repeat (F3) (Patthy, 1990) (Fig. 8C). All three Ig domains contain conserved cysteine residues characteristic of the Ig fold (Williams and Barclay, 1988) (circled residues in Fig. 3A), in addition to conserved amino acids (aa) surrounding the cysteines (Fig. 8A). Of the described subgroups of Ig domains, the second and third Ig domains have aa characteristics of the C2 type Ig domain (Williams and Barclay, 1988). The first Ig domain of REGA-1 is a member of the recently described ‘I set’ subgroup of Ig domain (Harpaz and Chothia, 1994). The percentage identity between the three possible pairs

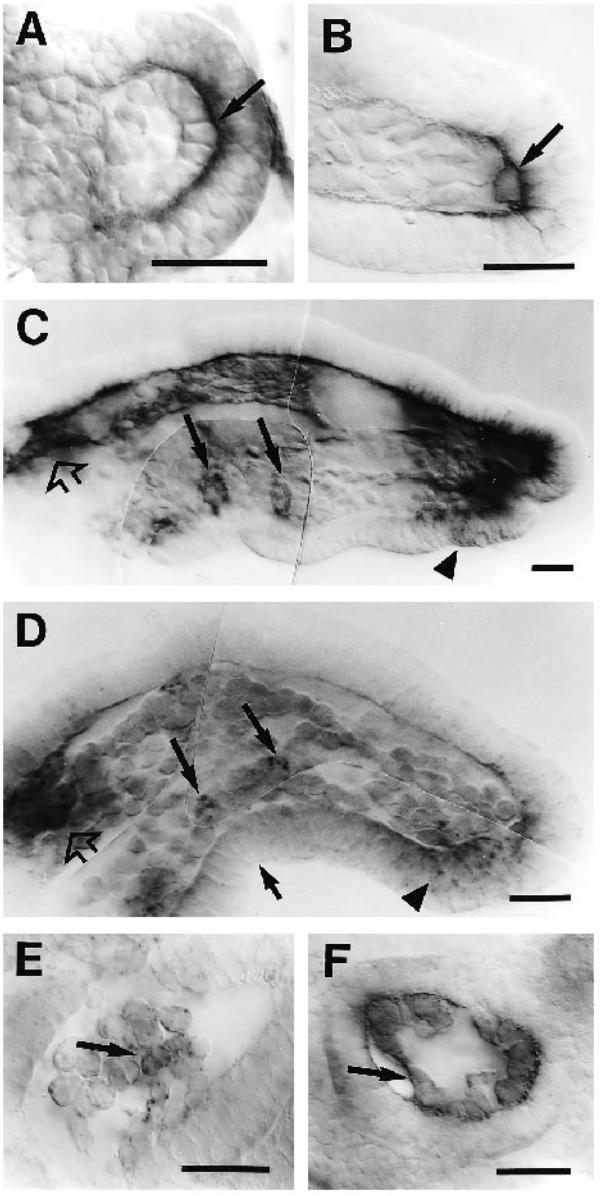

Fig. 6. REGA-1 is expressed in appendages. (A,B,C,F) show protein localization with the 7F7 mAb. (D,E) represent localization of REGA-1 message. (A) At 22% of embryogenesis REGA-1 localizes to the lumenal surface of the ectoderm in the leg (arrow). (B) At 31% of embryogenesis the most prominent labeling is at the distal tip of the leg. Arrow points to pioneer Ti1 neurons beginning grow along REGA-1-labeled surface. (C) At 35% of embryogenesis REGA-1 is expressed in several cell types in the leg. The lumenal surface of the ectoderm (arrowhead) is prominently labeled. Two cells inside the inner surface of the ectoderm with short processes oriented circumferentially express REGA-1 (arrows). Mesodermal cells label at the base of the leg (open arrow). (D) Localization of REGA-1 message in a leg from a 34% embryo. mRNA is present in ectodermal cells at the distal tip of the leg and is absent from proximal ectodermal cells (compare arrowhead with short arrow). Mesodermal cells at base of the leg express the REGA-1 message (open arrow). A pair of cells in the medial portion of the leg express REGA-1 message corresponding to cells labeled in C. (E) View into base of appendage of the first subesophaegeal segment in a 34% embryo. Mesodermal cells produce REGA-1 transcript (arrow). (F) Similar view and stage showing REGA-1 protein is localized to mesodermal cells (arrow). Scale bar, 30 μm.

of REGA-1 Ig domains is as follows: domain 1 versus domain 3, 18%; domain 1 versus domain 2, 25%; domain 2 versus domain 3, 33%. In addition, the size of each REGA-1 Ig domain is different: the distance between cysteines is 61 aa for Ig domain 1, 43 aa for Ig domain 2 and 51 aa for Ig domain 3. The F3 repeat of REGA-1 contains aas characteristic of this domain (Fig. 8B), including the highly conserved Trp and Tyr residues (boxed in Fig. 3A).

Many members of the Ig superfamily are present in the developing nervous system. When databases are searched with the REGA-1 ORF, the majority of molecules sharing amino acid identity with REGA-1 are neural cell adhesion molecules. Percentage identities with top-scoring candidates observed with REGA-1 over the length of the protein are: Nr-cam, 29%; amalgam, 28%; opcam, 25%; axonin-1, 25%; lachesin, 25%; NCAM, 24%; leukocyte antigen related protein, 23%; Cn3b, 22%; fasciclin II, 22%; vascular endothelial growth factor receptor, 22%; heparan sulfate proteoglycan, 22%; MAG, 22%; contactin, 21%; and neuroglian, 20%. Although some of the listed identities are statistically significant (Doolittle, 1987), none of the scores have a high percentage identity with REGA-1, indicating that we have identified a novel member of the Ig superfamily. Superfamily members often share sequence identities below significant values (Pearson and Lipman, 1988). Stretches of identity shared between REGA-1 and each molecule contain amino acids that define either an Ig domain or F3 repeat.

In addition to searching databases with the entire ORF of REGA-1, we also searched with each of the four domains of REGA-1 individually. Each of the three Ig domains pulls up a distinct list of top scores (highest optimal score, see Materials and methods). Fig. 8A shows multiple alignments of each Ig domain with molecules containing the highest percentage identities for that domain. Domains most related by sequence to the first Ig domain of REGA-1, except Cn3b, all belong to the I set subgroup of Ig domains (see Materials and methods). The multiple alignment for the second Ig domain contains only molecules present in the nervous system, with the exception of perlecan (hspg2), a basement membrane heparan sulfate proteoglycan. The alignment of third Ig domain has several molecules represented that are expressed in muscle (5/8). Ig domains in other proteins often have a higher percentage

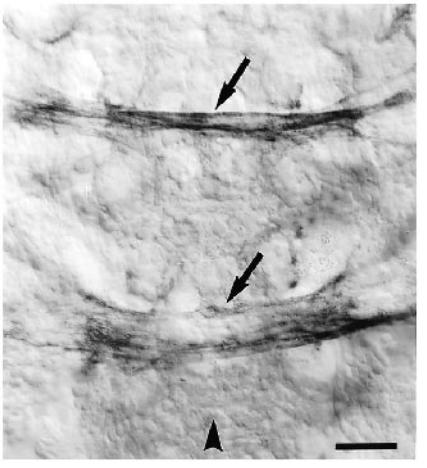

Fig. 7. REGA-1 is present in transverse muscle. View of subesophageal segments in a 45% embryo. The transverse muscle bands label with the 7F7 mAb (arrows). Arrowhead denotes the midline. Scale bar, 50 μm.

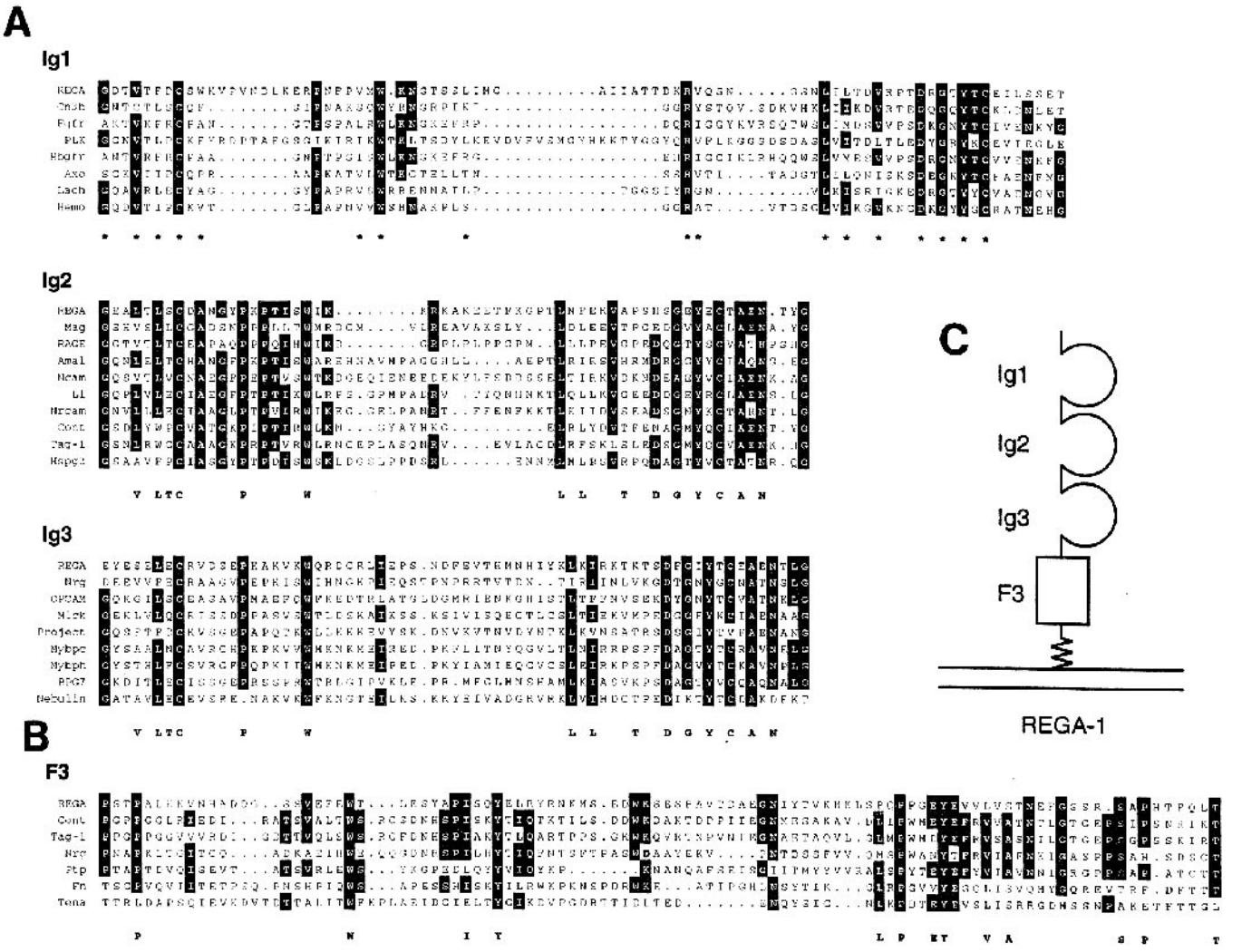

Fig. 8. Comparative analysis of REGA-1 domains. The three Ig domains and one F3 repeat of REGA-1 are aligned with similar domains from other molecules. Databases were searched with individual domains and alignments were made using the PILEUP program of GCG software (Devereux et al., 1984). (A) Alignments of Ig domains. Asterisks below first Ig domain alignment represent positions of conserved amino acids for the I set of Ig domain. Conserved residues of Ig C2-type are listed below the second and third alignments. (B) Alignment of the F3 repeat. Aas characteristic of F3 repeats are listed below alignment. Aas are highlighted when more than 50% of the group contain an identity at that position. (C) Structural features of REGA-1. Open circles denote Ig domains, box represents an F3 repeat and jagged line indicates GPI linkage to cell membrane. Abbreviations: Fgfr, fibroblast growth factor receptor ; CEK, tyrosine kinase receptor cek2; Hbgfr, heparin-binding growth factor receptor; ; axo, axonin-1; lach, lachesin; Hemo, hemolin; Mag, myelin-associated glycoprotein; RAGE, receptor for advanced glycosylation end products of proteins; amal, amalgam; Cont, contactin; hsp, heparan sulfate proteoglycan; HSPG2/perlecan; Nrg, neuroglian; Mlck, myosin light chain kinase; Project, projectin; Mybpc, myosin-binding protein C; Mybph, myosin-binding protein H; BPG7, mouse heparan sulfate proteoglycan BPG7; Ptp, protein-tyrosine-phosphatase; Fn, fibronectin; Tena, tenascin.

identity with one of the REGA-1 Ig domains than identities observed within REGA-1. For example, the following identities are observed between domains for the second Ig domain of REGA-1: MAG, 36%; amalgam, 40%; NCAM, 35%; L1, 39%; Nr-CAM, 42%; contactin, 38%; TAG-1, 35%; heparan sulfate proteoglycan, 37%. It is notable that all domains listed in the multiple alignment for the third Ig domain have only a 1-2 aa size difference (Fig. 8A). When databases are searched with the F3 repeat of REGA-1, one of the most closely related by aa identity is a F3 repeat from fibronectin, from which the F3 repeat was initially defined (Fig. 8B). With the exception of fibronectin, domains in the alignment with the F3 repeat are all from molecules found in the nervous system (Fig. 8B). Analysis of the REGA-1 ORF reveals that it is a member of the Ig superfamily and contains one F3 repreat and three Ig domains, each with unique sequence characteristics that associate each domain with distinct subgroups of domains found in a range of organisms.

DISCUSSION

REGA-1 is a member of the Ig superfamily that is expressed on the surface of sheath-cell processes surrounding NB families in the CNS. Among NB families, the extent and intensity of labeling on sheath-cell processes differs. For most

NB families, labeling within sheath-cell processes is very restricted. A few families express REGA-1 on processes from the GMC layer to the dorsal surface of the CNS. When labeling first appears on sheath-cell processes, all NBs have differentiated and are dividing (Doe and Goodman, 1985a; Shepherd and Bate, 1990). Sheath-cell processes surrounding NB family 7-4 begin to label at the birth of the first progeny, Q1 and Q2, one of the few NB families whose lineage is known (Raper et al., 1983). Appearance of labeling when the the first neurons are born is consistent with the spatial distribution of REGA-1 around the GMCs and newly born neurons. In situ analysis reveals that the REGA-1 message is localized to the same cellular layers as the protein. If REGA-1 is produced by sheath cells, the mRNA is localized within sheath cells. Alternatively, REGA-1 may be produced by GMCs and neurons and secreted locally to receptors on adjascent sheath-cell processes, consistent with the observation of a released form of REGA-1. The message of *amalgam*, an Ig superfamily member described in *Drosophila* (Seeger et al., 1988), is present in epidermal cells and the protein is localized to neurons (Mark Seeger, personal communication).

Sheath cells are postulated to play a support role for developing NB families, but the term 'support' is vague and may include many functions. The physical arrangement of sheath-cell processes around a NB family suggests that they may provide a physical barrier to keep family members isolated from surrounding cells or in close physical contact. Annulin is present throughout sheath cells (Bastiani et al., 1992) and has sequence similarity to transglutaminases (Singer et al., 1992), which have protein cross-linking activity. The molecular nature of annulin and its widespread distribution in the cell is consistent with it having a structural role. In contrast, REGA-1 is restricted to portions of sheath-cell processes, suggesting a more specialized function. Thus far, attention has focused on the role of NBs in neuronal determination. Molecular markers expressed in subsets of NBs have been characterized and the identity of a GMC is proposed to be specified through inheritance of a particular combination of transcription factors (reviewed in Doe and Technau, 1993). In addition, inductive events among equivalent cells influence cell fate in the neural ectoderm. Cell ablation studies in grasshopper have shown that as NBs enlarge they inhibit adjacent ectodermal cells from adapting a NB fate (Doe and Goodman, 1985b). The aCC and pCC neurons, the first progeny of NB1-1, are born equivalent and their unique identity is determined through cell interactions (Kuwada and Goodman, 1985). Little is known about the influence sheath cells have on neurogenesis. Sheath cells may prevent inappropriate induction events. Restricted localization of REGA-1 is consistent with sheath cells having a role in controlling timing of cell divisions of NBs and GMCs or providing positional information to NBs. Understanding how REGA-1 influences the development of NB progeny will advance the understanding of the range of sheath-cell functions.

A number of identified surface proteins in the nervous system are GPI-linked, although the functional significance of this membrane association is not clear. REGA-1 is restricted to the lumenal surface of epithelial cells in the leg and the GPI anchor may be important for this subcellular localization. Localization of particular proteins to the apical surface of epithelial cells is dependent upon the presence of a GPI anchor (reviewed in Ferguson, 1992). A role for GPI-linked proteins in axonogenesis has been demonstrated directly in the grasshopper leg (Chang et al., 1992). When GPI-linked proteins are enzymatically removed, the pioneer Ti1 neurons make pathfinding errors at multiple points as they navigate towards the CNS. Therefore, several aspects of the navigation of Ti1 neurons are due to the presence of GPI-linked proteins. Several GPI-linked proteins in the embryonic leg have been identified, including fasciclin I (Bastiani et al., 1987), lachesin (Karlstrom et al., 1993), lazarillo (Ganforninaet al., 1995) and REGA-1 (this manuscript). Growth cones of the Ti1 neurons extend along the surface of epithelial endfeet (Caudy and Bentley, 1986) and use bands of epithelial cells at segment boundaries and guidepost cells as guidance cues. The cues used for initial outgrowth of the Ti1 neurons are not currently known. The subcellular localization and membrane association of REGA-1 suggest it is localized to endfeet of epithelial cells at the distal tip of the leg. We propose that REGA-1 may act in a heterophilic manner to instruct growth in a proximal direction when the growth cones first extend, until contact is made with a guidepost cell.

REGA-1 may have multiple roles in the leg. The initial pattern of epithelial labeling resolves into three circumferential bands within the tibia, femur and tarsus segments by 37% of embryogenesis. Epithelial bands of fasciclin IV expression (Kolodkin et al., 1992) overlap with REGA-1 bands in the tibia and part of the femur, and there is no overlap with annulin expression at segment boundaries (Singer et al., 1992). Bands of REGA-1 labeling follow formation of visible segment boundaries and may provide positional information within the lumen of the leg. REGA-1 is also expressed by a pair of cells inside the lumenal surface of the ectoderm and their morphology suggests they may be glial cells. The origin and development of glial cells in the appendages is not well understood and REGA-1 may provide a molecular reagent for the study of these cells. REGA-1 is also present in mesodermal cells at the base of the limb and may function during their differentiation. The complex spatial and temporal pattern of REGA-1 expression in the leg suggests multiple functions at distinct developmental times.

REGA-1 is associated with the cell membrane via a GPI-linkage and a significant portion of it is released from the cells. The GPI-linked form and released form migrate with slightly different apparent molecular mass. Many GPI-linked proteins are present in both soluble and membrane-associated forms (reviewed in Ferguson, 1992) and there are several examples found in the nervous system, including fasciclin I (Hortsch and Goodman, 1990), TAG-1/axonin-1(Furley et al., 1990, Ruegg et al., 1989), BIG-1(Yoshihiro et al., 1994) and NCAM (He et al., 1987). The two forms of fasciclin I migrate with the same relative difference in mobility as observed for the two forms of REGA-1. The released form of these proteins may result from endogenous phospholipase activity. Cleavage by a GPI-specific phospholipase activity (GPI-PLD) in intact cell membranes was recently reported (Metz et al., 1994) and illustrates a potential mechanism of regulating molecules at the cell surface. The ability of a cell to remove REGA-1 from the membrane through GPI-PLD activity may serve as a fine temporal regulatory mechanism, or to increase the local concentration of a soluble form as suggested by Metz et al. (1994).

Sequence analysis of REGA-1 reveals that it is a novel

member of the Ig superfamily and is a mosaic protein containing three Ig domains and one F3 repeat. Both domains are found in a large number of molecules across species, the F3 repeat being especially common with over 300 variants described in about 70 proteins (Engel et al., 1994). In database searches, no other protein with the same domain organization as REGA-1 was identified. Many Ig superfamily members have counterparts with a similar domain organization. L1 and neuroglian both have six Ig domains and five F3 repeats, fasciclin II and NCAM have five Ig domains and two F3 repeats, TAG-1, BIG-1 and F11 all have six Ig domains and four F3 repeats and lachesin, OBCAM, neurotrimin and amalgam have three Ig domains. It has been suggested that proteins such as fasciclin II and NCAM (Grenningloh et al., 1990) and L1 and neuroglian (Bieber et al., 1989) may have arisen from a common ancestor.

Databases were also searched with individual domains to determine what this analysis might reveal about potential functions of REGA-1. Individual Ig domains of REGA-1 are more similar in aa sequence to domains in other Ig proteins than with each other. In addition, percentage identities comparing domains alone are often higher than percentage identities over the length of the compared proteins. Domains similar to the second Ig domain of REGA-1 are from proteins expressed in the developing nervous system and Ig domains similar to the third domain of REGA-1 are from proteins found in muscle. In both cases the list of proteins includes vertebrate and invertebrate examples, suggesting widespread conservation of amino acids. These subgroups include amino acids conserved beyond those characteristic of the C2 type Ig domain (Williams and Barclay, 1988) and may represent aas that confer specific function to these domains. Shared amino acids within subgroups of Ig domains reflect the evolutionary history of the domains and may represent domain-shuffling events. The unique character of each Ig domain raises the question of what constitutes a unit of function in REGA-1: can a single Ig domain act as a functional unit? Blechman et al. (1995) have demonstrated that the fourth Ig domain of the stem-cell-factor receptor is sufficient to produce ligand-induced dimer formation, which then causes a signal transduction event. Considering each REGA-1 Ig domain as an independent functional unit is consistent with the distribution of protein in the embryo. The second Ig domain may be an important part of the function of REGA-1 in sheath cells in the CNS. The third Ig domain may mediate function of REGA-1 in muscle, since REGA-1 is expressed in the transverse muscle bands that overly the CNS.

In summary, we have reported a new member of the Ig superfamily called REGA-1 that has a unique domain organization and is present in the developing grasshopper embryo. Its restriction to localized regions of sheath-cell processes surrounding NB families implies that it may have a role in influencing the development of NB families. The results reported here provide a basis for future in vivo studies of REGA-1 function in the grasshopper.

We wish to thank members of our laboratory for helpful discussions and critical reading of the manuscript; M. Low for the generous gift of PI-PLC; Julie Broadus for technical advice; and K. Zinn for technical advice and gift of λgt11 and λgt10 libraries. We also wish to thank R. Schackmann, who sequenced the REGA-1 peptides at the Protein/DNA facility at the Utah Cancer Center. This work was supported by NIH postdoctoral fellowship NS08404 to E. M. C., NIH genetics training grant (5T32GM07464-17) to E. C. S. and NIH NS25387 to M. J. B.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Ashburner, M. (1989). In *Drosophila A Laboratory Manual* pp. 214-216. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Bastiani, M. J., Harrelson, A. L., Snow, P. M. and Goodman, C. S. (1987). Expression of fasciclin I and II glycoproteins on subsets of axon pathways during neuronal development in the grasshopper. *Cell* **48**, 745-755.

Bastiani, M. J., de Couet, H. G., Quinn, J. M. A., Karlstrom, R. O., Kotrla, K., Goodman, C. S. and Ball, E. E. (1992). Position-specific expression of the annulin protein during grasshopper embryogenesis. *Devel. Biol.* **154**, 129-142.

Bate, C. M. (1976). Embryogenesis of an insect nervous system I. A map of the thoracic and abdominal neuroblasts in *Locusta migratoria*. *J. Embryol. Exp. Morphol.* **35**, 107-123.

Bentley, D., Keshishian, H., Shankland, M. and Toroian-Raymond, A. (1979). Quantitative staging of embryonic development of the grasshopper, *Schistocerca nitens*. *J. Embryol. Exp. Morphol.* **54**, 47-74.

Bieber, A. J., Snow, P. M., Hortsch, M., Patel, N. H., Jacobs, J. R., Traquina, Z. R., Schilling, J. and Goodman, C. S. (1989). *Drosophila* Neuroglian: A member of the immunoglobulin superfamily with extensive homology to vertebrate neural cell adhesion molecule L1. *Cell* **59**, 447-460.

Bixby, J. L. and Harris, W. A. (1991). Molecular mechanisms of axon growth and guidance. In *Ann. Rev. Cell Biol.* Annual Reviews Inc. **7,** 117-59.

Blechman, J. M., Lev, S., Barg, J., Eisenstein, M., Vaks, B., Vogel, Z., Givol, D. and Yarden, Y. (1995). The fourth immunoglobulin domain of the stem cell factor receptor couples ligand binding to signal transduction. *Cell* **80**, 103-113.

Broadus, J. and Dow, C. Q. (1995). Evolution of neuroblast identity: *seven-up* and *prospero* expression reveal homologous and divergent neuroblast fates in *Drosophila* and *Schistocerca. Development* **121**, 3989-3996.

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

Carpenter, E. M. and Bastiani, M. J. (1991). Developmental expression of REGA-1, a regionally expressed glial antigen in the central nervous system of grasshopper embryos. *J. Neurosci.* **11**, 277-286.

Caudy, M. and Bentley, D. (1986). Pioneer growth cone morphologies reveal proximal increases in substrate affinity within leg segments of grasshopper embryos. *J. Neurosci.* **6**, 364-379.

Cavener, D. R. and Ray, S. C. (1991). Eukaryotic start and stop translation sites. *Nucl. Acid Res.* **19**, 3185-3192.

Chang, W. S., Serikawa, K., Allen, K. and Bentley, D. (1992). Disruption of pioneer growth cone guidance *in vivo* by removal of glycosylphosphatidylinositol-anchored cell surface proteins. *Development* **114**, 507-519.

Chothia, C. (1994). Protein families in the metazoan genome. *Development* **1994 Supplement**, 27-33.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nuc. Acid Res.* **12**, 387-395.

Doe, C. Q. and Goodman, C. S. (1985a). Early events in insect neurogenesis I. Development and segmental differences in the pattern of neuronal precursors. *Dev. Biol.* **111**, 193-205.

Doe, C. Q. and Goodman, C. S. (1985b). Early events in insect neurogenesis II. the role of cell interactions and cell lineage in the determination of neural precursor cells. *Devel. Biol.* **111**, 206-219.

Doe, C. Q. and Technau, G. M. (1993). Identification and cell lineage of individual neural precursors in the *Drosophila* CNS. *TINS* **16**, 510-514.

Doolittle, R. F. (1987). Of URFs and ORFS a primer on how to analyze derived amino acid sequences. Mill Valley, CA: University Science Books.

Engel, J., Efimov, V. P. and Maurer, P. (1994). Domain organizations of extracellular matrix proteins and their evolution. *Development* **1994/Supplement**, 35-42.

Ferguson, M. A. J. (1992). Glycosyl-phosphatidylinositol membrane anchors: the tale of a tail. *Biochem. Soc. Trans.* **20**, 243-256.

Furley, A. J., Morton, S. B., Manalo, D., Karagogeos, D., Dodd, J. and Jessell, T. M. (1990). The axonal glycoprotein TAG-1 is an immunoglobulin

superfamily member with neurite outgrowth-promoting activity. *Cell* **61**, 157-170.

Ganfornina, M. D., Sanchez, D. and Bastiani, M. J. (1995). Lazarillo, a new gpi-linked surface lipocalin, is restricted to a subset of neurons in the grasshopper embryo. *Development* **121**, 123-134.

Gerber, L. D., Kodukula, K. and Udenfriend, S. (1992). Phosphatidylinositol glycan (PI-G) anchored membrane proteins. *J. Biol. Chem* **267**, 12168-12173.

Goodman, C. S. and Spitzer, N. C. (1979). Embryonic development of idenified neurones: differentiation from neuroblast to neurone. *Nature* **280**, 208-214.

Grenningloh, G., Bieber, A. J., Rehm, E. J., Snow, P. M., Traquina, Z. R., Hortsch, M., Patel, N. H. and Goodman, C. S. (1990). Molecular genetics of neuronal recognition in *Drosophila*: evolution and function of immunoglobulin superfamily cell adhesion molecules. In *Cold Spring Harbor Symposia on Quantitative Biology*, Cold Spring Harbor Laboratory Press. **50,** 327-340.

Grumet, M., Mauro, V., Burgoon, M. P., Edelman, G. M. and Cunningham, B. A. (1991). Structure of a new nervous system glycoprotein, Nr-CAM, and its relationship to subgroups of neural cell adhesion molecules. *J. Cell Biol.* **113**, 1399-1412.

Harpaz, Y. and Chothia, C. (1994). Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a structural set which is close to that containing variable domains. *J. Mol. Biol.* **238**, 528-539.

He, H., Finne, J. and Goridis, C. (1987). Biosynthesis, membrane association, and release of N-CAM-120, a phosphatidylinositol-linked form of the neural cell adhesion molecule. *J. Cell Biol.* **105**, 2489-2500.

Ho, R. K., Ball, E. E. and Goodman, C. S. (1983). Muscle pioneers: large mesodermal cells that erect a scaffold for developing muscles and motoneurones in grasshopper embryos. *Nature* **301**, 66-69.

Hortsch, M. and Goodman, C. S. (1990). *Drosophila* fasciclin I, a neural cell adhesion molecule, has a phosphatidylinositol lipid membrane anchor that is developmentally regulated. *J. Biol. Chem.* **265**, 15104-15109.

Jacobs, J. R. and Goodman, C. S. (1989). Embryonic development of axon pathways in the *Drosophila* CNS. I. A glial scaffold appears before the first growth cones. *J. Neurosci.* **9**, 2402-2411.

Karlstrom, R. O., Wilder, L. P. and Bastiani, M. J. (1993). Lachesin: an immunoglobulin superfamily protein whose expression correlates with neurogenesis in grasshopper embryos. *Development* **118**, 509-522.

Kolodkin, A. L., Matthes, D. J., O'Connor, T. P., Patel, N. H., Admon, A., Bentley, D. and Goodman, C. S. (1992). Fasciclin IV: sequence, expression, and function during growth cone guidance in the grasshopper embryo. *Neuron* **9**, 831-845.

Kuwada, J. Y. and Goodman, C. S. (1985). Neuronal determination during embryonic development of the grasshopper nervous system. *Devel. Biol.* **110**, 114-126.

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

Leung-Hagesteijn, C., Spence, A. M., Stern, B. D., Zhou, Y., Su, M.-W. and Hedgecock, E. M. (1992). UNC-5, a transmembrane protein with immunoglobulin and thrombospondin type I domains, guides cell and pioneer axon migrations in *C. elegans*. *Cell* **71**, 289-299.

Metz, C. N., Brunner, G., Choi-Muira, N. H., Nguyen, H., Babrilove, J., Caras, I. W., Altszuler, N., Rifkin, D. B., Wilson, E. L. and Davitz, M. A. (1994). Release of GPI-anchored membrane proteins by a cell-associated GPI-specific phospholipase D. *EMBO J.* **13**, 1741-1751.

Patthy, L. (1990). Homology of a domain of the growth hormone/prolactin receptor family with type III modules of fibronectin. *Cell* **61**, 13-14.

Pearson, W. R. and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**, 2444-2448.

Pearson, W. R. (1991). Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics* **11**, 635-650.

Raper, J. A. and Goodman, C. S. (1982). Transient dye coupling between developing neurons reveals patterns of intracellular communication during embryogenesis. In *Cell Communication During Ocular Development* (ed. J. Sheffield and S. R. Hilfer), New York: Springer-Verlag.

Raper, J. A., Bastiani, M. and Goodman, C. S. (1983). Pathfinding by neuronal growth cones in grasshopper embryos. I. Divergent choices made by the growth cones of sibling neurons. *J. Neurosci.* **3**, 20-30.

Ruegg, M. A., Stoeckli, E. T., Lanz, R. B., Streit, P. and Sonderegger, P. (1989). A Homologue of the axonally secreted protein axonin-1 is an integral membrane protein of nerve fiber tracts involved in neurite fasciculation. *J. Cell Biol.* **109**, 2363-2378.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Ehrlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-494.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

Schofield, P. R., McFarland, K. C., Hayflick, J. S., Wilcox, J. N., Cho, T. M., Roy, S., Lee, N. M., Loh, H. H. and Seeburg, P. H. (1989). Molecular characterization of a new immunoglobulin superfamily protein with potential roles in opioid binding and cell contact. *EMBO J.* **8**, 489-495.

Seaver, E. C., Karlstrom, R. O. and Bastiani, M. J. (1991). The restricted spatial and temporal expression of a nervous-system-specific antigen involved in axon outgrowth during development of the grasshopper. *Development* **111**, 881-893.

Seeger, M. A., Haffley, L. and Kaufman, T. C. (1988). Characterization of amalgam: a member of the immunoglobulin superfamily from *Drosophila*. *Cell* **55**, 589-600.

Shepherd, D. and Bate, C. M. (1990). Spatial and temporal patterns of neurogenesis in the embryo of the locust (*Schistocerca gregaria*). *Development* **108**, 83-96.

Singer, M. A., Hortsch, M., Goodman, C. S. and Bentley, D. (1992). Annulin, a protein expressed at limb segment boundaries in the grasshopper embryo, is homologous to protein cross-linking transglutaminases. *Devel. Biol.* **154**, 143-159.

Snow, P. M., Zinn, K., Harrelson, A. L., McAllister, L., Schilling, J., Bastiani, M. J., Makk, G. and Goodman, C. S. (1988). Characterization and cloning of fasciclin I and fasciclin II glycoproteins in the grasshopper. *Proc. Natl. Acad. Sci. USA* **85**, 5291-5295.

Struyk, A. F., Canoll, P. D., Wolfgang, M. J., Rosen, C. L., D'Eustachio, P. and Salzer, J. L. (1995). Cloning of neurotrimin defines a new subfamily of differentially expressed neural cell adhesion molecules. *J. Neurosci.* **15**, 2141-2156.

von Heijne, G. (1990). The signal peptide. *J. Membr. Biol.* **115**, 195-201.

Williams, A. F. and Barclay, A. N. (1988). The immunoglobulin superfamily-domains for cell surface recognition. *Annual Review of Immunology* **6**, 381-405.

Yoshihiro, Y., Kawasaki, M., Tani, A., Tamada, A., Nagata, S., Kagamiyama, H. and Mori, K. (1994). BIG-1: A new TAG-1/F3-related member of the immunoglobulin superfamily with neurite outgrowth-promoting activity. *Neuron* **13**, 415-426.

(Accepted 16 November 1995)