

## Lazarillo, a new GPI-linked surface lipocalin, is restricted to a subset of neurons in the grasshopper embryo

María D. Ganfornina\*<sup>†</sup>, Diego Sánchez\* and Michael J. Bastiani

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

\*D. S. and M. D. G. contributed equally to this work

<sup>†</sup>Author for correspondence

### SUMMARY

**Lazarillo, a protein recognized by the monoclonal antibody 10E6, is expressed by a subset of neurons in the developing nervous system of the grasshopper. It is a glycoprotein of  $45 \times 10^3 M_r$  with internal disulfide bonds and linked to the extracellular side of the plasma membrane by a glycosyl-phosphatidylinositol moiety. Peptide sequences obtained from affinity purified adult protein were used to identify an embryonic cDNA clone, and in situ hybridizations confirmed that the distribution of the Lazarillo mRNA paralleled that of the monoclonal antibody labeling on**

**embryos. Sequence analysis defines Lazarillo as a member of the lipocalin family, extracellular carriers of small hydrophobic ligands, and most related to the porphyrin- and retinol-binding lipocalins. Lazarillo is the first example of a lipocalin anchored to the plasma membrane, highly glycosylated, and restricted to a subset of developing neurons.**

Key words: pathfinding, lipocalin, GPI, grasshopper, lipid-binding proteins

### INTRODUCTION

Embryonic neurons must extend growth cones along precise pathways, often over long distances, to reach and synapse with the appropriate target cell. The proper execution of this task by millions of neurons is essential to construct a functional nervous system. What cues are used by a neuron to direct its growth cone toward a distant target? Since the first observation of growth cones by Ramón y Cajal, a variety of chemical and physical cues have been invoked as guidance mechanisms (Bixby and Harris, 1991; Cypher and Letourneau, 1992; Tessier-Lavigne, 1992; Goodman and Shatz, 1993). Because neurons in the same embryonic environment often follow different pathways, we believe that the interaction of a growth cone with its environment requires a code of molecular signals and receptors to provide the necessary specificity. The 'labeled pathways hypothesis' (Goodman et al., 1982) proposes a mechanism for this specificity: axon fascicles in the embryonic nervous system are differentially labeled by surface molecules that are used for the guidance of growth cones. Many surface molecules restricted to subsets of axon fascicles have been found in both vertebrates and invertebrates. They belong to several families of proteins such as the immunoglobulin superfamily, the cadherins, some members of the integrin family, and others (reviewed by Bixby and Harris, 1991; Goodman and Shatz, 1993).

The lipocalins are a family of extracellular soluble proteins that transport small hydrophobic molecules. They have a well conserved structure: a calyx formed by a  $\beta$ -barrel with a hydrophobic ligand pocket. This family includes proteins with

diverse biological functions. For example, the serum retinol-binding proteins are transporters of vitamin A from the liver to various target tissues (Blomhoff et al., 1990); C8 $\gamma$  is one of the components of the complement cascade (Haefliger et al., 1987); the human brain prostaglandin D synthetase (PGDs) is a lipocalin with enzymatic activity (Nagata et al., 1991); the mouse oncogene 24p3 is proposed to regulate cell differentiation and mitogenesis (Hraba-Renevey et al., 1989). Within this large and diverse family there has been no indication that any are involved in the process of axonal pathfinding. However, the protein we describe in this work belongs to the lipocalin family and has a restricted expression pattern in the nervous system that makes it a candidate for a specific axonal receptor and/or guidance cue. In the following paper (Sánchez, Ganfornina and Bastiani, 1995) we indeed demonstrate that it is required for the navigation of identified commissural neurons. To the best of our knowledge, the localization, molecular characteristics and function of this new lipocalin are unique, both among the members of this protein family and among the proteins involved in axonal pathfinding during the development of the nervous system. For its putative role in axon guidance, we have named this protein Lazarillo after the main character of a sixteenth century Spanish novel, *Lazarillo de Tormes*, a crafty boy who guided a blind man.

### MATERIAL AND METHODS

Grasshopper (*Schistocerca americana*) embryos were obtained from a colony maintained at 31°C and 60% humidity at the University of

Utah. They were staged by percentage of embryonic development according to Bentley et al. (1979). The monoclonal antibody (mAb) 10E6 was generated by Carpenter and Bastiani (1990) against embryonic nervous tissue using a subtractive immunization method (Hockfield, 1987). Immunocytochemistry was carried out as described by Sánchez, Ganfornina and Bastiani. (1995).

#### Immunoblot analysis of embryonic and adult proteins

Embryonic and adult membrane proteins were prepared as described by Seaver et al. (1991), separated by electrophoresis (SDS-PAGE) and electrotransferred to nitrocellulose membranes. Membranes were washed with a Tris buffer (50 mM Tris, pH 7.6, 150 mM NaCl, 0.2% gelatin, 0.1% NaN<sub>3</sub>) and blocked for 2 hours with 2.5% BSA, 2.5% dry milk in the previous buffer. They were incubated with primary antibody (1:200 in blocking solution) for 2 hours, washed again, incubated with rabbit anti-mouse IgG (1:500 in blocking solution) for 2 hours, washed and incubated with <sup>125</sup>I-protein A solution (0.2 µCi/ml in blocking solution) for 1 hour. After washing, the membrane was air dried and exposed to film.

#### Biochemical characterization of Lazarillo protein

To analyze the association of Lazarillo with the membrane, high salt and/or basic pH extractions were carried out. Unsolubilized membrane proteins were subjected to a basic pH buffer (10 mM TEA, pH 11.3, 150 mM NaCl), a high salt buffer (10 mM TEA, pH 7.8, 500 mM NaCl) or a combination of both conditions for 30 minutes on ice. The mixture was diluted 10 times with the same buffer and centrifuged (100,000 g, 2 hours). Both supernatant and membrane associated proteins were analyzed by immunoblot.

Analysis of GPI-anchoring to the membrane was performed as described by Chang et al. (1992). 45% embryos were dissected in sterile RPMI medium containing 6 mg/ml glycine. They were transferred to culture medium (Cellgro) containing 50% Schneider's *Drosophila* Medium (Gibco-BRL), 49% Minimum Essential Medium (α medium, Gibco-BRL), 1% antibiotic-antimycotic solution (Sigma). They were incubated, without (control embryos) or with phospholipase C (PLC) 3 U/ml in medium, for 2 hours at 31°C. Two different enzymes were used, phosphatidylinositol-specific PLC (PI-PLC, from *Bacillus thuringiensis*, generously provided by Dr Martin Low, Columbia University) or phosphatidylcholine-specific PLC (PC-PLC, from *Bacillus cereus*, Boehringer-Mannheim). After the incubation period embryos were washed, fixed and labeled with antibodies. For the immunoblot analysis of proteins extracted by the treatment, the culture medium after the incubation was collected and concentrated with Centricon-10 (Amicon). Embryos were washed, then homogenized and a standard membrane protein preparation was followed as above.

To determine whether Lazarillo is a glycoprotein, the affinity purified protein was separated by SDS-PAGE and transferred to nitrocellulose membrane. Blots were preincubated for 30 minutes at room temperature in TTBS-Mn-Ca buffer (0.1 M Tris, pH 7.5, 150 mM NaCl, 0.1% Tween-20, 0.1 mM MnCl<sub>2</sub>, 0.1 mM CaCl<sub>2</sub>) and incubated for 1 hour with biotinylated concanavalin A (Vector) 10 µg/ml in TTBS-Mn-Ca solution. All subsequent steps were as described in Vectastain ABC kit (Vector). In addition, affinity purified protein was deglycosylated with peptide-N-glycosidase F (PNGase F) from *Flavobacterium meningosepticum* (Boehringer-Mannheim). The protein was denatured with 10 mM β-mercaptoethanol and boiling and then incubated with 3.3 U/ml of PNGase F at 37°C for 18 hours. Control protein was denatured in the same way, but incubated without enzyme. In parallel, pure protein was heat denatured and incubated with 0.15 mU/ml of neuraminidase (from *Arthrobacter ureafaciens*, Sigma).

#### Protein purification and microsequencing of Lazarillo protein

The protein was purified by affinity chromatography from detergent soluble fractions of embryonic or adult lysates, prepared as previously

described (Bastiani et al., 1987), using mAb 10E6 immobilized to Protein G Sepharose beads (Pharmacia). Lysate total protein concentration was measured with the micro-BCA assay (Pierce). Adult lysate was prepared as described for embryos but using whole adult grasshoppers whose gut, legs and wings were removed. Amino acid composition of Lazarillo was performed by transferring 2.7 µg of protein to PVDF membrane (Problott) followed by acid hydrolysis and amino acid analysis.

To obtain peptide sequence 13.5 µg of the protein Lazarillo were separated by SDS-PAGE. The gel piece containing the protein was rinsed in 250 mM Tris, pH 9.0, 250 mM EDTA and in dH<sub>2</sub>O at room temperature, chopped and dried in a vacuum centrifuge. The dried gel was soaked in 0.1 M NH<sub>4</sub>HCO<sub>3</sub>, pH 9.0 containing endoproteinase Lys-C (Lys-C from *Lysobacter enzymogenes*, Boehringer-Mannheim), and incubated at 37°C for 12 hours (enzyme to protein ratio, 1:20). Digestion products were eluted from the gel by extensive washing with NH<sub>4</sub>HCO<sub>3</sub> buffer. Peptides were then separated by reverse phase HPLC using a C-18 300 Å pore column. Fractions were subjected to microsequencing by automated Edman degradation.

Amino acid analysis, peptide separation and sequencing, as well as synthesis of custom oligonucleotide primers for PCR or DNA sequencing were carried out at the Protein/DNA Core Facility of the Utah Cancer Center under the direction of R.W. Schackmann.

#### Molecular analysis of Lazarillo cDNA

Degenerate oligonucleotides were designed from the peptide sequences to amplify DNA fragments from embryonic grasshopper cDNA using PCR with Taq DNA polymerase (Saiki et al., 1988). PCR was conducted in a thermal cycler (Perkin Elmer Cetus) and conditions were as follows: MgCl<sub>2</sub> concentration 2.5 mM; one cycle of 94°C for 2 minutes; 35 cycles of 94°C for 30 seconds, 48°C for 30 seconds, and 72°C for 1 minute; and a final cycle of 72°C for 5 minutes. The DNA fragments obtained were tested with a pair of internal primers that were designed to give a known size DNA fragment. A PCR product of approx. 400 bp, DL400, was cloned into pCR-II vector using the TA cloning system (Invitrogen) and sequenced (see below). The fragment DL400 was radiolabeled with [<sup>32</sup>P]dCTP using the random primer method (Prime-It II kit, Stratagene) and used as a probe.

Several embryonic cDNA grasshopper libraries were constructed and screened with the DL400 probe as described in the λZAP system (Stratagene). We screened 0.5×10<sup>6</sup> plaque forming units (p.f.u.) from an amplified library made using nerve cords dissected from 55% embryos, and 1.3×10<sup>6</sup> p.f.u. from an unamplified library made from embryos at 45% development.

Both strands of cDNA inserts were sequenced (Sanger et al., 1977) using Sequenase (v2.0, U.S. Biochemicals) and custom primers. dITP was employed to sequence through areas with GC compressions. DNA and protein sequences were analyzed with the GCG programs (Devereux et al., 1984) and the BLAST service (Altschul et al., 1990). The following data bases were screened: Swissprotein, PIR, GenBank/EMBL, and Brookhaven PDB. Sequence alignments were carried out with PILEUP GCG program using a symbol comparison table that takes into account structurally conservative amino acid substitutions (Risler et al., 1988).

#### Analysis of Lazarillo mRNA expression

Total RNA was prepared from grasshopper embryos at 45% development. 500 embryos were dissected in cold RPMI with 6 mg/ml of glycine, washed in PBS, and homogenized in 0.1 M Tris, pH 7.5, 4 M guanidinium thiocyanate, 1% β-mercaptoethanol. The RNA was purified by CsCl centrifugation as described in Sambrook et al. (1989). Poly(A)<sup>+</sup> mRNA was isolated from total RNA using the Poly-Atract mRNA isolation system (Promega) and analyzed by northern blot analysis with the DL400 probe. Briefly, 5 and 10 µg mRNA were separated by electrophoresis in a formaldehyde-agarose gel, transferred to a nylon membrane (Zeta-Probe, BioRad) under alkaline con-

ditions, and hybridized with the probe at 65°C in the same solution as for library screenings.

In situ hybridization was performed according to a protocol for grasshopper whole-mount embryos (J. Broadus, personal communication). Briefly, a digoxigenin-11-dUTP labeled RNA probe (Genius-4 kit, Boehringer-Mannheim) was synthesized using the entire Laz-5 clone as template. The resulting RNA was subjected to alkaline hydrolysis. Embryos were dissected and fixed in PEM-formaldehyde (37% formaldehyde 1:9 in 0.1 M Pipes, pH 6.9, 2 mM EGTA, 1 mM MgSO<sub>4</sub>) for 50 minutes. After washing, the solution was changed gradually to hybridization solution (50% deionized formamide, 4× SSC, 250 µg/ml yeast tRNA, 500 µg/ml boiled salmon sperm DNA, 50 µg/ml heparin, 0.1% Tween-20, 1× Denhardt's, 5% dextran sulfate in DEPC-treated H<sub>2</sub>O) at room temperature, with a last incubation at 55°C in hybridization solution for 1 hour. The labeled RNA probe was then added at 0.5 µg/ml and incubation proceeded at 55°C for 36–48 hours. Washes for 5 hours, with frequent changes, followed the hybridization step. To detect the labeled RNA, an alkaline phosphatase-conjugated anti-digoxigenin antibody was used and the color reaction was carried out as described in the Genius DNA labelling and detection kit (Boehringer-Mannheim).

## RESULTS

### The mAb 10E6 recognizes a surface molecule restricted to a subset of neurons and axon fascicles

The expression of the antigen recognized by mAb 10E6, the protein Lazarillo, is restricted to the surface of an identified subset of neurons. It is detected on cell bodies, axons, growth cones, and filopodia of live embryos suggesting that the mAb binds to an extracellular surface epitope. Fig. 1A shows a pair of labeled neurons that pioneer an anterior commissural pathway at 32% of embryonic development (the AcP cells). They direct their growth cones toward the midline of the embryo while extending many filopodia ahead (arrow) to contact the contralateral growth cone. Fig. 1B shows some of the fascicles labeled with the mAb 10E6 in a metathoracic ganglion at 40% of development. Unlabeled axon fascicles, not expressing Lazarillo, are indicated by open arrows. At this developmental stage pioneer neurons have established a scaffold of axon fascicles consisting of two commissures (ac, pc) connecting the hemiganglia, two longitudinal connectives (lc) linking adjacent ganglia, and a median fiber tract (MFT). The grasshopper central nervous system (CNS) is connected to the peripheral nervous system (PNS) by two main nerves; the segmental (SN) and intersegmental (ISN) nerves. Fig. 1C shows the ISN, where only the motoneurons (mn) and peripheral sensory neurons (sn) navigating along its posterior branch express Lazarillo. The anterior branch, pioneered by the U motoneurons, are never labeled by the mAb 10E6 (open arrow).

The fascicles expressing Lazarillo were identified and compared with those expressing other surface molecules restricted to subsets of fascicles in the grasshopper embryo. In Fig. 1D the distribution of Lazarillo is compared with that of Fasciclin I. Fig. 1E represents the 10E6 fascicle map compared to the fascicles expressing Fasciclin I, Fasciclin II, and Semaphorin I/Fasciclin IV (Bastiani et al., 1987; Harrelson and Goodman, 1988; Kolodkin et al., 1992). The longitudinal connectives at 40% of development are composed of three distinct fascicles differentially labeled by these surface glycoproteins. Lazarillo is currently the only one on the surface of axons in

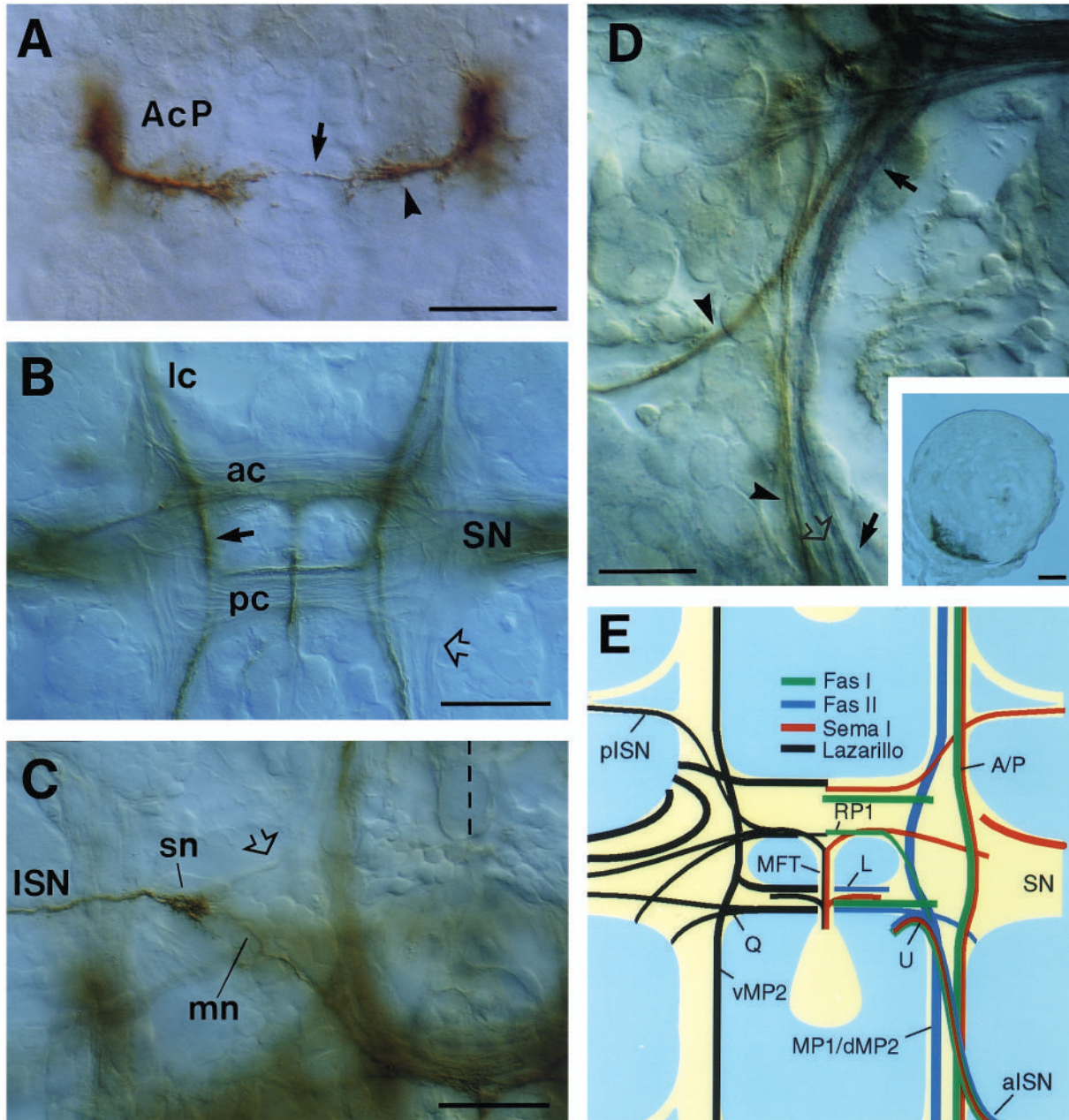
the vMP2 fascicle and some axons in this fascicle express Lazarillo through adulthood. A coronal section of an adult longitudinal connective labeled by mAb 10E6 is shown in the inset of Fig. 1D. Lazarillo shares with Semaphorin I its expression on the MFT, however, semaphorin I is restricted to the initial portion of the bifurcated axons while Lazarillo is located along the entire axonal length (Fig. 1E), reflecting a different spatial distribution. Two bundles in the posterior commissure are labeled by Fasciclin II and Lazarillo, but Fasciclin II expression is transient while that of Lazarillo remains, reflecting a different temporal regulation. The expression of the four molecules in distinct but overlapping subsets of pathways supports the hypothesis that a particular combination of molecules on the surface of a fascicle functions as the unique label used by neurons for their specific pathway decisions.

Lazarillo is also expressed by a subset of neuroblasts in the CNS, every sensory neuron in the PNS, a group of neurons of the enteric nervous system (ENS) and by a few non-neuronal tissues. The developmental expression pattern is described in the following paper (Sánchez, Ganfornina and Bastiani, 1994).

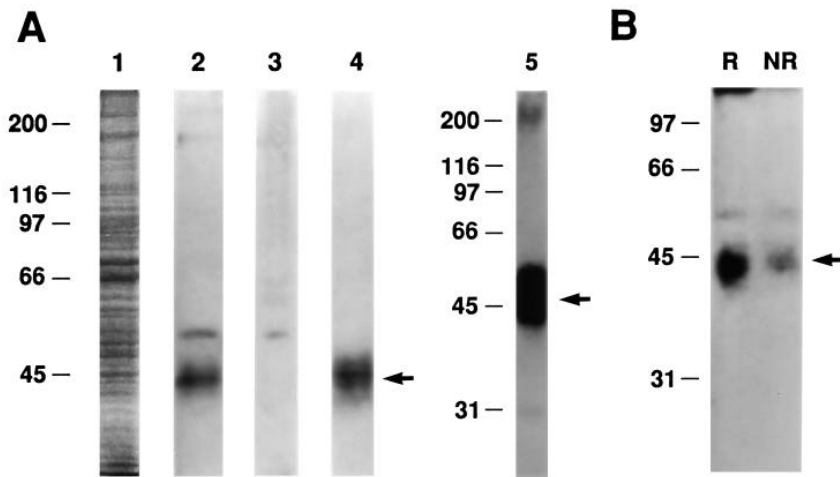
### Lazarillo is a glycoprotein anchored to the membrane by a glycosyl-phosphatidylinositol group

We identified Lazarillo by immunoblot analysis of embryonic membrane proteins separated by SDS-PAGE in the presence of reducing agents. A smeary band with an average apparent  $M_r$  of  $45 \times 10^3$  is recognized by the mAb 10E6 (Fig. 2A, lane 2). In the absence of the mAb the secondary antibody recognizes a nonspecific band of  $57 \times 10^3 M_r$  (Fig. 2A, lane 3) that also appears when other unrelated mAbs are used as the primary antibody. No signal was detectable in the soluble fraction of embryonic homogenates (not shown). A band of similar appearance and  $M_r$  was also detected by immunoblot analysis using adult membrane preparations (Fig. 2A, lane 4). The apparent  $M_r$  of Lazarillo does not change significantly under non-reducing conditions (Fig. 2B). Lazarillo was purified from embryonic lysates on a mAb 10E6 affinity column (Fig. 2A, lane 5) with a yield of approx. 1 ng/mg of total protein.

Several lines of evidence suggest that Lazarillo is associated with membranes, in addition to the 10E6 labeling obtained in non-permeabilized live embryos. The mAb recognizes the protein in the membrane but not in the soluble fraction of embryonic homogenates. Neither high salt, alkaline solutions, nor both combined, extract Lazarillo from the membrane fraction (not shown) suggesting that the protein is associated with the membrane by hydrophobic interactions. To determine whether Lazarillo is linked to the membrane by a GPI moiety, embryos were treated with PI-PLC prior to fixation and labeling with the mAb 10E6. Fig. 3 shows the results from control (A,B) and treated embryos (C,D). Following PI-PLC treatment, no signal was obtained with the mAb 10E6 in the CNS or in the PNS. As a control, we labeled embryos with mAb 3B11, which recognizes the GPI-anchored glycoprotein Fasciclin I (Hortsch and Goodman, 1990). As expected, the labeling disappeared after the enzymatic treatment. In contrast, when treated embryos were labeled with 8C6, a mAb against Fasciclin II, which has a transmembrane domain in grasshoppers (Harrelson and Goodman, 1988), the immunoreactivity did not change after PI-PLC treatment (not shown). To test the



**Fig. 1.** Lazarillo expression is restricted to a specific subset of neurons and axon fascicles in the embryonic nervous system of the grasshopper. (A-D) Presence of Lazarillo in the CNS visualized by immunocytochemistry with mAb 10E6. Whole embryos are viewed from the dorsal surface with DIC optics. Anterior is up. (A) The identified pioneer neurons of the anterior commissure (AcP cells) are labeled with mAb 10E6 in a subesophageal segment at 32% of embryonic development. Each neuron sends a growth cone (arrowhead) toward the midline where it will encounter its contralateral homologue. Filopodial extensions are darkly stained with this antibody. The arrow points to a long filopodium almost reaching the contralateral growth cone. (B) Thoracic ganglion at 40% of development. Only a subset of axon bundles is labeled. Compare the labeled medial longitudinal fascicle (solid arrow) with the unlabeled lateral longitudinal fascicles (open arrow). lc, longitudinal connectives; ac, anterior commissure; pc, posterior commissure; SN, segmental nerve. (C) The intersegmental nerve (ISN) exits between adjacent segmental ganglia. The growth cone of a sensory neuron (sn) finds the axon of a motoneuron (mn), both labeled by mAb 10E6, and enters the CNS by the posterior branch of the ISN. The anterior branch, or U fascicle, is not labeled by mAb 10E6 (open arrow). The midline is indicated by a dashed line. (D) Double labeling with mAb 3B11 (anti-Fasciclin I, brown) and mAb 10E6 (blue). Three distinct longitudinal fascicles are present at 38% of development that are differentially labeled with these surface molecules. Lazarillo is present only in the vMP2 fascicle (arrows) while Fasciclin I is in a lateral longitudinal fascicle (including the A/P fascicle) and the U fascicle (arrowheads). The unlabeled one (open arrow) is the MP1/dMP2 fascicle. Midline is on the right. The inset shows a transverse section of an adult longitudinal connective showing a ventral median tract labeled by mAb 10E6. Ventral is down and midline on the left. (E) Schematic diagram showing the complete subset of fascicles that expresses Lazarillo at 40% (on the left) compared to Fasciclin I, Fasciclin II and Semaphorin I expression (on the right). MFT, median fiber tract; pISN, posterior branch of ISN; aISN, anterior branch of the ISN (or U fascicle); SN, segmental nerve; A/P, MP1/dMP2, vMP2, Q, L and RP1 are fascicles named after the neurons that pioneer them. Scale bars: A-C and inset in D, 50  $\mu$ m; D, 25  $\mu$ m.

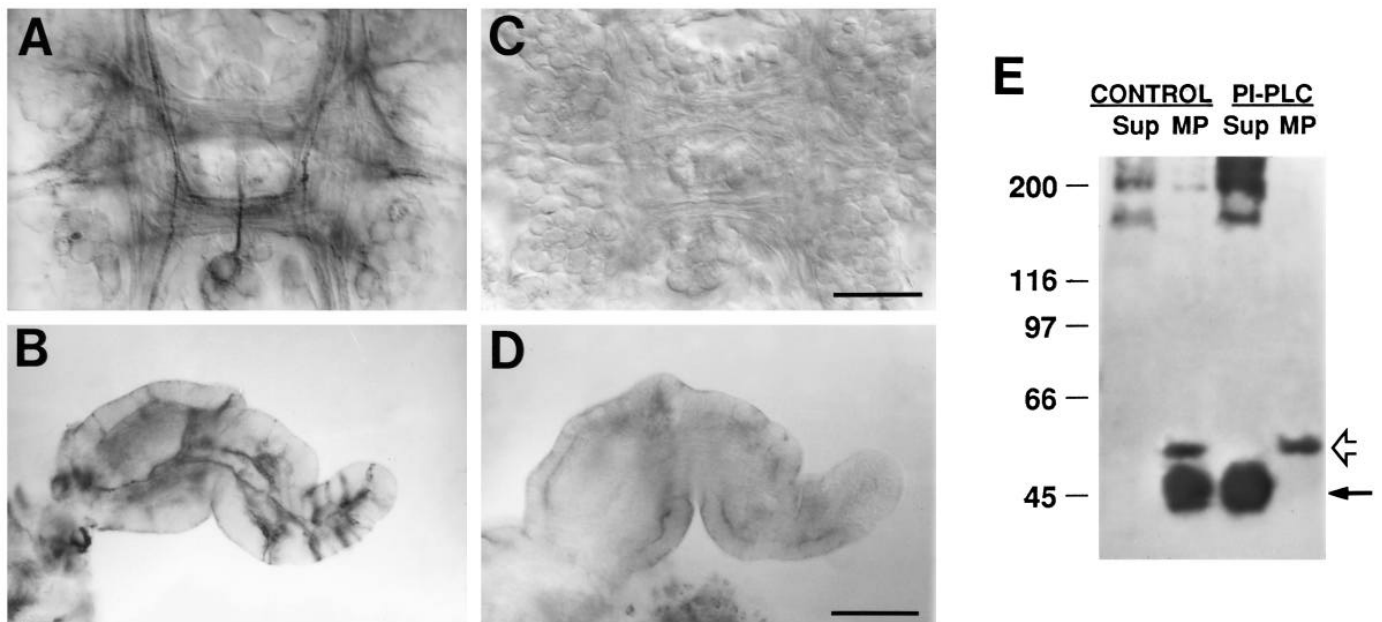


**Fig. 2.** Lazarillo is a protein with a  $M_r$  of  $45 \times 10^3$ . (A) Immunoblot analysis using mAb 10E6. (1) Coomassie blue-stained SDS-polyacrylamide gel showing total embryonic membrane proteins. (2) The mAb 10E6 specifically recognizes a protein migrating at an apparent  $M_r$  of  $45 \times 10^3$  (arrow) in the embryonic membrane preparation. (3) In the absence of the mAb the secondary antibody recognizes a nonspecific band of  $57 \times 10^3 M_r$ . (4) A single band of the same size and appearance than the embryonic  $45 \times 10^3 M_r$  protein is recognized with the mAb on total membrane proteins from adult grasshopper. (5) Lazarillo protein purified by affinity chromatography from embryonic lysates. (B) The redox state of the native protein

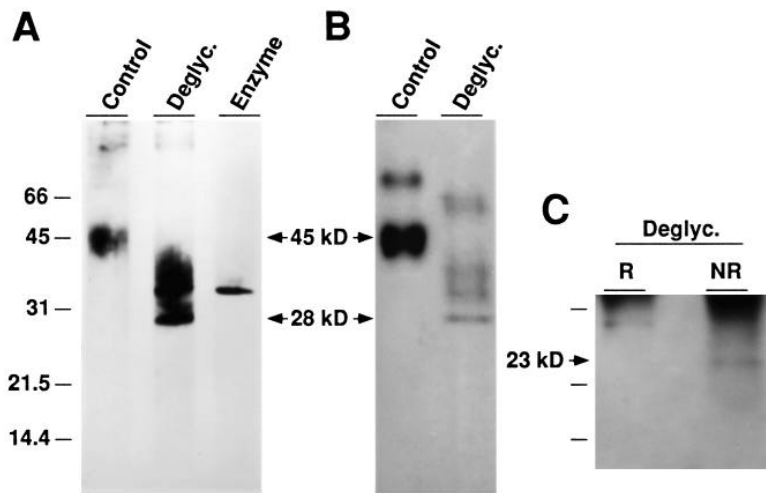
does not affect its electrophoretic mobility. Lane R: the electrophoresis sample buffer contains  $\beta$ -mercaptoethanol (50 mg/ml), NR: no reducing agent added. Molecular mass markers ( $M_r \times 10^{-3}$ ) are shown on the left.

specificity of the effects of PI-PLC, we treated embryos with PC-PLC. None of the labeling (10E6, 3B11 or 8C6) decreased after the treatment (not shown) suggesting that the conditions used selectively remove GPI-anchored proteins and do not cause nonspecific release or damage of membrane proteins. Accordingly, the  $45 \times 10^3 M_r$  band is present only in the membrane preparation from control embryos, but is entirely released to the culture supernatant after treatment with PI-PLC (Fig. 3E). The nonspecific band ( $M_r$ ,  $57 \times 10^3$ ) is not extracted by PI-PLC. These results suggest that the  $45 \times 10^3 M_r$  protein is linked to the membrane by a GPI tail and corresponds to the Lazarillo pattern observed in embryos.

The smeary appearance of the band seen under SDS-PAGE could be the result of incomplete SDS solubilization due to the presence of carbohydrates. The presence of oligosaccharides covalently linked to Lazarillo was assayed by incubating purified Lazarillo immobilized on nitrocellulose with Concanavalin A. The lectin does bind to the  $45 \times 10^3 M_r$  protein (not shown) suggesting the presence of D-glucose and D-mannose residues. As the latter are present in the core structure of N-linked oligosaccharides, we treated the protein with PNGase F, which cleaves N-linked carbohydrates. A diffuse band of  $30\text{--}40 \times 10^3 M_r$  is seen above a tight band of  $28.2 \times 10^3 M_r$  (Fig. 4A,B). Our interpretation is that this result could be due to



**Fig. 3.** A glycosyl-phosphatidylinositol group anchors Lazarillo to the extracellular side of the plasma membrane. (A,B) Metathoracic ganglion and leg from a control embryo at 45% of development, labeled with mAb 10E6. (C,D) The labeling disappears when embryos are cultured in the presence of a phosphatidylinositol-specific phospholipase C (PI-PLC) for 90 minutes. (E) Membrane proteins (MP) and soluble proteins in the incubation media (Sup) were analyzed by immunoblot in control and PI-PLC-treated embryos. The  $45 \times 10^3 M_r$  protein (solid arrow) is completely released from the membranes by the enzyme. The nonspecific band (open arrow) remains in the membrane fraction. Bars, (A,C) 50  $\mu\text{m}$ ; (B,D) 200  $\mu\text{m}$ . Molecular mass markers in  $M_r \times 10^{-3}$  are shown on the left in E.



the protein is deglycosylated due to the low efficiency of the PNGase F in these conditions (native protein not shown). The non-reduced core protein migrates as a  $23 \times 10^3 M_r$  band. Molecular mass markers ( $M_r \times 10^{-3}$ ) are shown on the left.

partial deglycosylation of the native protein and suggests that the  $28.2 \times 10^3 M_r$  band represents the fully deglycosylated protein. As revealed by immunoblot analysis (Fig. 4B) the mAb 10E6 is able to recognize the deglycosylated form of Lazarillo. This suggests that the 10E6 epitope is not on the N-linked oligosaccharides, but could be part of the core polypeptide. The protein was also treated with neuraminidase, which hydrolyzes the terminal bonds joining sialic acid to oligosaccharides, and no change in electrophoretic mobility of Lazarillo was observed (not shown).

Given the large influence of the carbohydrates on the electrophoretic mobility of Lazarillo, the effect of reducing agents shown in Fig. 2B is probably masked by the anomalous solubilization of the glycoprotein. To analyze the presence of intracatenary disulfide bonds we deglycosylated Lazarillo with PNGase F in the absence of  $\beta$ -mercaptoethanol (Fig. 4C). The deglycosylated protein shows a clear decrease in electrophoretic mobility when the reducing agent is present, consistent with the presence of internal disulfide bonds in Lazarillo core protein.

#### Generation of a Lazarillo DNA probe. Hybridization to a single species of embryonic mRNA

Lazarillo protein was purified from adult lysates by affinity to the immobilized mAb 10E6. Adult lysates were the source of choice for the protein purification due to their larger yield (15 ng/mg of total protein). Attempts to obtain N-terminal sequence were unsuccessful. An amino acid composition analysis was performed and we chose Lys-C to produce peptides of reasonable length. After HPLC separation, four peptides were sequenced: (1) NLQLDLNK; (2) WY EYAK; (3) RPDSAASTEISWILLRSR<sub>x</sub>SS<sub>x</sub>MTLERVEDELK; (4) xSVHFNPNSVGNy<sub>x</sub>LSTDYD<sub>x</sub>YSIV; where x stands for undetermined amino acids. The presence of a serine or a threonine two residues after some undetermined amino acids in peptides 3 and 4 suggested that the latter would be asparagines bound to oligosaccharides (N-linked).

Three pairs of degenerate primers (sense and antisense) were designed from peptide 2, and from the N- and C-terminal

**Fig. 4.** A large shift in relative mobility is observed in Lazarillo after deglycosylation and subsequent addition of reducing agents. (A) Silver stained gel showing the  $45 \times 10^3 M_r$  protein purified from adult grasshoppers by affinity chromatography (Control). Deglycosylation with PNGase F (Deglyc.) generates a smeary band ranging from  $30$ – $40 \times 10^3 M_r$  with a sharp lower boundary of  $28.2 \times 10^3 M_r$ . The enzyme band ( $M_r$ ,  $35.5 \times 10^3$ ) is shown in the third lane for comparison. (B) Analysis by immunoblot of the same samples. The mAb 10E6 recognizes the protein after removal of the N-linked carbohydrates (Deglyc.). In this case the lower boundary of the smear is resolved into a tight band of  $28.2 \times 10^3 M_r$  that we think represents the fully deglycosylated protein. The upper band is due to antibody leaking from the affinity column. (C) Lazarillo protein was deglycosylated in the absence of reducing agents and its electrophoretic mobility tested under non-reducing (NR) and reducing (R) conditions. Only a small proportion of

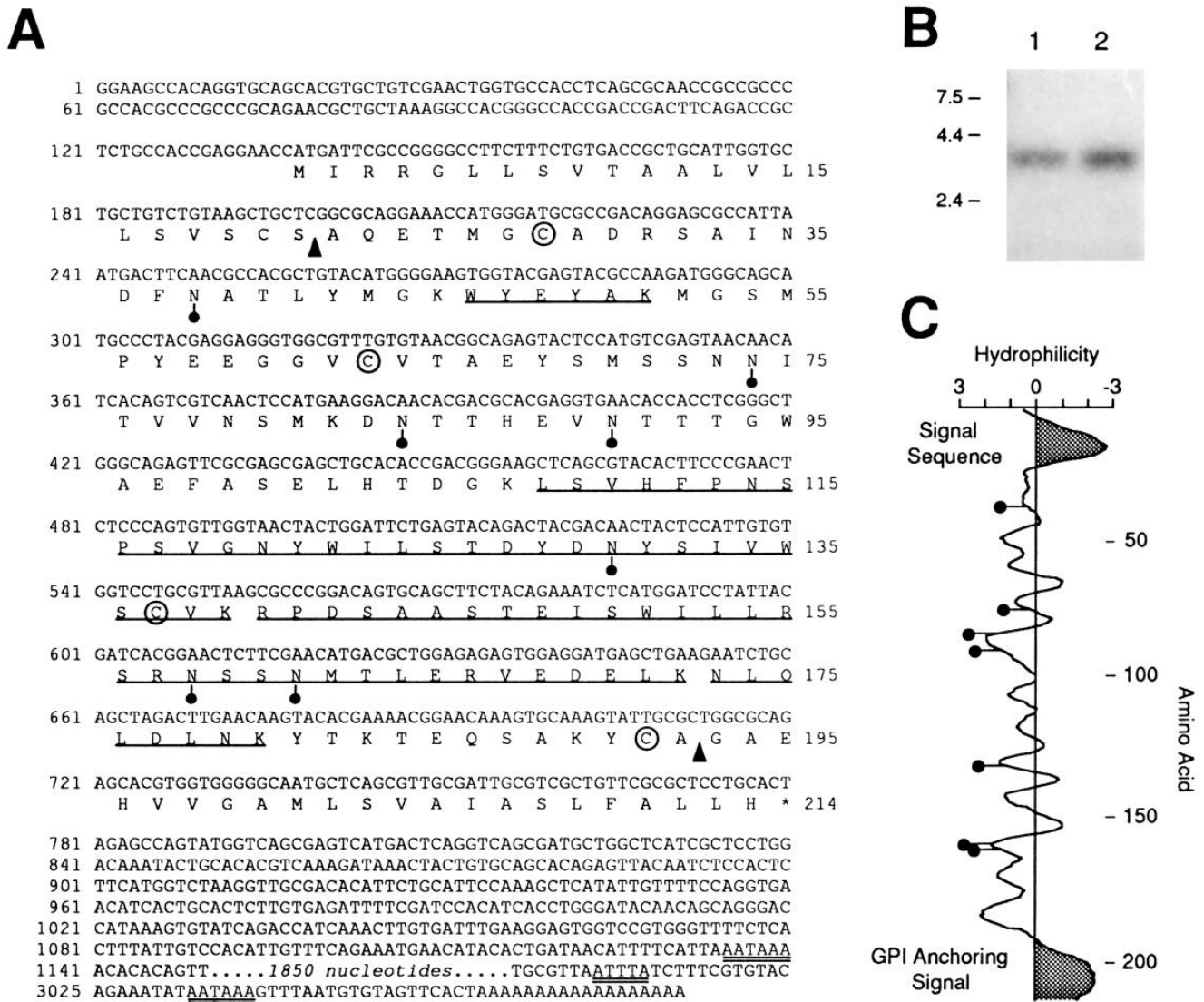
portions of peptide 3. A DNA fragment of approx. 400 bp (DL400) was amplified by PCR from embryonic grasshopper cDNA using the sense primer from peptide 2 and the antisense primer from the C-terminal end of peptide 3. An open reading frame (ORF) was found along the entire DNA fragment, which additionally contained peptide 4. The sequence confirmed that the mRNA coding for the peptides obtained from adult grasshoppers is present in embryos. The undetermined amino acids referred to as possible N-glycosylation sites were in fact asparagines.

The DL400 DNA was used as a probe for northern hybridization (Fig. 5B). A single band of approximately 3 kb is detected in a poly(A)<sup>+</sup> mRNA preparation from embryos at 45% of development, a stage that corresponds with protein expression as seen by immunocytochemistry with the mAb.

#### Isolation of Lazarillo cDNA clone. The predicted amino acid sequence has the characteristics of a GPI-linked membrane protein

We screened several  $\lambda$ ZAP embryonic grasshopper cDNA libraries with the DL400 probe and identified three positive clones. Two of them (Laz-9 and Laz-10) were identical clones from an amplified nerve cord specific library. They contain an insert of 2.8 kb and were truncated at the 5' end. The third clone (Laz-5) was from an unamplified library and contains an insert of 3075 bp that corresponds in size to the mRNA detected by northern hybridization. Both DNA strands of the Laz-5 insert were sequenced. Sequences corresponding to the four peptides are contained in this clone. The cDNA and the predicted amino acid sequence of Laz-5 are shown in Fig. 5A. It contains a 5' untranslated region (UTR) of 137 bp, an ORF of 642 bp, and a 3' UTR of 2296 bp. The sequence of the four peptides are underlined. The DL400 probe hybridizes to bp 273–653.

The first methionine codon encountered in frame in the Laz-5 clone is in an appropriate context to be the translation initiation site (Cavener and Ray, 1991). The length of the ORF agrees with the analysis of period three compositional constraint (TESTCODE GCG program) and the codon usage and compositional bias in the third position of each possible codon

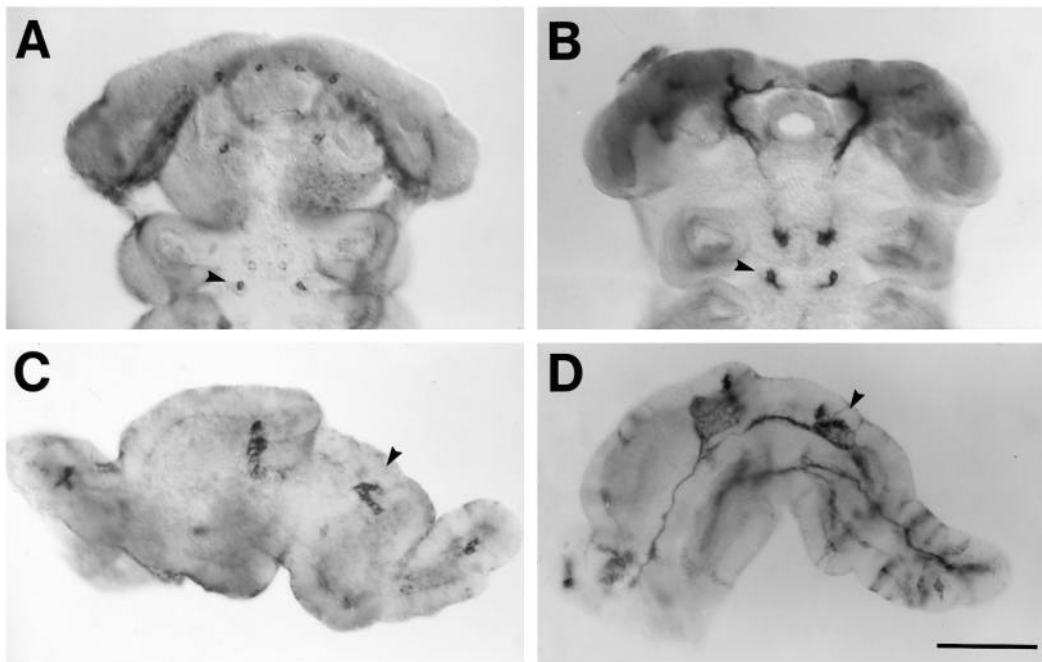


**Fig. 5.** The sequence of Lazarillo cDNA codes for a GPI-linked membrane glycoprotein and is represented by a single embryonic transcript. (A) Nucleotide and predicted amino acid sequence of Laz-5 cDNA clone. Nucleotide numbers are on the left and amino acid numbers are on the right (referred to the first methionine). Arrowheads point to the potential cleavage sites of the signal peptide and GPI tail. The potential N-linked glycosylation sites are noted by black dots. Cysteine residues are circled. The peptides obtained by Lys-C digestion are underlined. Some of the polyadenylation sites and mRNA instability sequences are shown double underlined. (B) Northern hybridization analysis. Poly(A)<sup>+</sup> mRNA (5 and 10 µg) from embryos at 45% of development was hybridized with DL400 probe after electrophoretic separation and transfer to a nylon membrane. A single transcript of approximately 3 kb is detected. Size markers in kb are on the left. (C) Hydropathy plot of the predicted protein sequence determined by the Kyte and Doolittle method (1982) using a window of 9 residues. The hydrophobic domains of the signal peptide and the GPI-anchoring signal are noted by the filled pattern. Black dots represent glycosylation sites.

(CODONPREFERENCE GCG program with a codon usage table made from *Schistocerca* and *Locusta* coding sequences found in the cited data bases).

Both the N- and C-terminal ends of the protein predicted by this ORF are hydrophobic (Fig. 5C). These fragments may represent, respectively, the signal peptide necessary for the translocation of the protein into the lumen of endoplasmic reticulum and the hydrophobic tail necessary for the attachment to a GPI group. The predicted cleavage site for Lazarillo is indicated in Fig. 5A before Ala22, though it could also be before Gln23. In both cases the rule '-3,-1' described by von Heijne (1990) is fulfilled. The 16 C-terminal amino acids of

Laz-5 are hydrophobic, and preceded by a hydrophilic spacer domain (Glu195 to Val198). Both domains define the potential GPI-anchoring signal with a cleavage site before or after Ala192 (Coyle et al., 1993). The predicted mature protein after cleavage of both signal sequences would have a  $M_r$  of approximately  $20 \times 10^3$ . The deduced protein has seven potential N-glycosylation sites (Fig. 5A and C), some of them indeed glycosylated as deduced from the experiments using PNGase F (Fig. 4). The four cysteine residues present, noted in Fig. 5A, can form internal disulfide bonds that would explain the different electrophoretic mobility produced by reducing agents (Fig. 4C). Four potential polyadenylation



**Fig. 6.** mRNA and protein co-localization in grasshopper embryos confirms the identity of the Laz-5 clone. (A,B) Anterior portion of whole-mount embryos at 32% showing the head and subesophageal segments. The mRNA (A) and the protein (B) colocalize, for example, in the AcP cells in the S2 segment (arrowheads). In the head the four labeled cells in A correspond to the ones extending axons along the commissure and the longitudinal connectives of the brain in B. (C,D) The sensory groups of cells in

the metathoracic legs of embryos at 40% of development (arrowheads) are shown as examples of the correlation between the Lazarillo mRNA (C) and the protein (D) in the PNS. Bar, 200  $\mu$ m.

signals and four message instability sequences are found in the 3' UTR (Fig. 5A).

#### Lazarillo mRNA colocalizes with the mAb 10E6 antigen

The spatial localization of Lazarillo mRNA was studied by performing *in situ* hybridization in whole-mount grasshopper embryos. A digoxigenin-labeled RNA probe was generated using the entire Laz-5 cDNA. A comparison of the RNA hybridization pattern with the mAb 10E6 labeling in embryos of the same age is shown in Fig. 6. Lazarillo mRNA expression is detected in a group of neuronal cell bodies in the head (Fig. 6A). These same neurons, whose axons contribute to the primary commissure of the brain and to the fascicles connecting with the segmental ganglia, are labeled with the mAb (Fig. 6B). The AcP cells of the second subesophageal (S2) segment, specifically labeled by the mAb, also show hybridization with the Lazarillo RNA probe (arrowhead in Fig. 6A and B). Several clusters of cells expressing Lazarillo are seen in the metathoracic leg (Fig. 6C), which corresponds to the developing sensory organs recognized by the mAb (Fig. 6D). The distribution of Lazarillo mRNA matches that of the protein in all other areas of the CNS and PNS, and in other embryonic structures characteristically labeled by mAb 10E6, such as the ENS and subesophageal body (not shown). Hybridization is absent in embryos exposed to the sense RNA probe (not shown), indicating that the signal we observe with the anti-sense probe is due to specific hybridization with the endogenous mRNA. These experiments suggest that the antigen recognized by mAb 10E6, the surface glycoprotein Lazarillo, is encoded by the mRNA represented in the cDNA clone Laz-5.

#### Lazarillo belongs to the lipocalin family

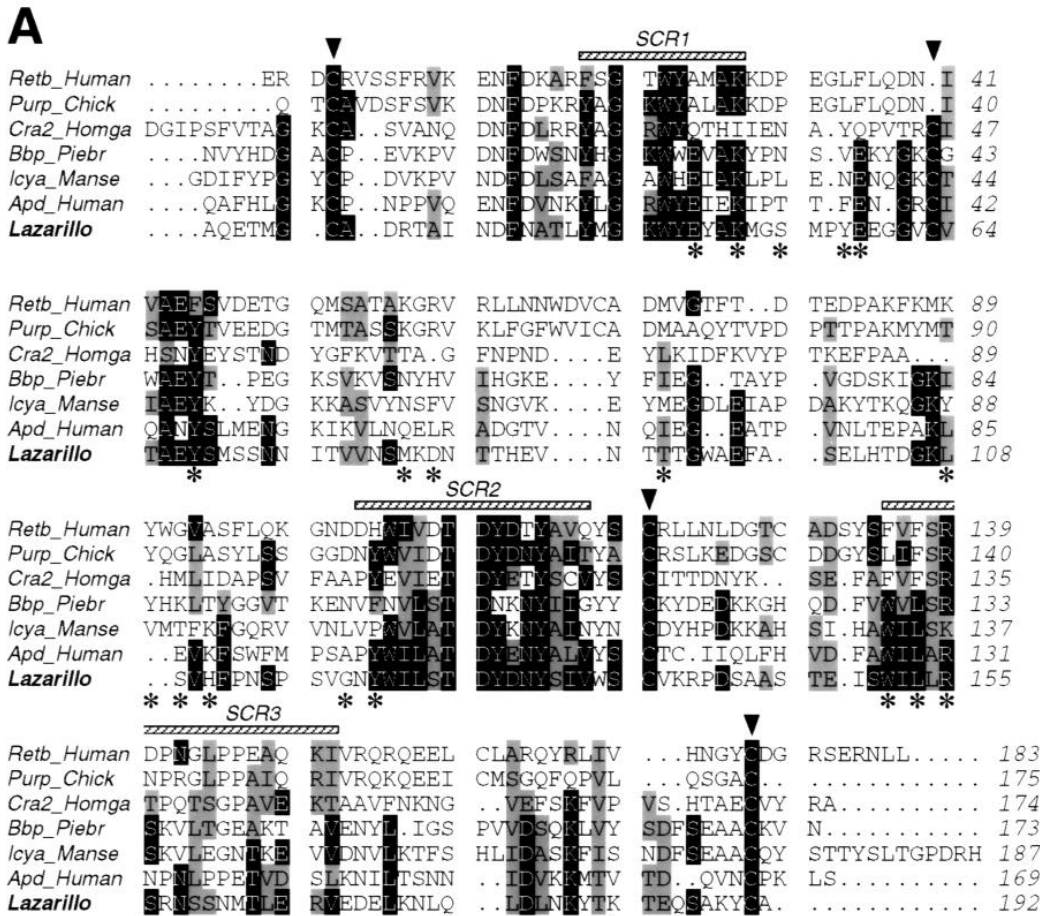
The deduced Lazarillo protein sequence was compared with

several data bases and the searches identified it as a member of the family of proteins called lipocalins; extracellular carriers of small hydrophobic ligands.

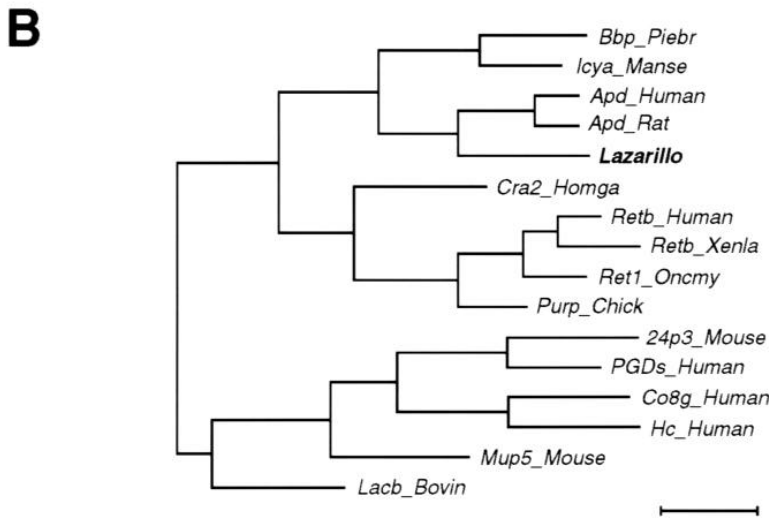
Lazarillo shows significant sequence similarity to most members of this family. It shares with its closest relatives a residue identity between 19-30%, values commonly found between family members (Flower et al., 1993), and a 72-85% similarity when taking into account structurally conservative amino acid substitutions (see Material and Methods). Lipocalins share a highly conserved tertiary structure despite the low identity values found in the family. Flower et al. (1993) have analyzed the structure and sequence relationships within the lipocalin family. By superimposing the known three-dimensional structure of four lipocalins they have defined three structurally conserved regions (SCR1-3) that correspond to the sequence motifs highlighted in Fig. 7A. They have proposed these SCRs as an essential requirement for belonging to the lipocalin family. Fig. 7A shows a multiple alignment of Lazarillo with the family members bearing the highest percentage identities. The three SCRs are largely conserved in Lazarillo, and other sequence stretches also show a considerable similarity.

The size and diversity of the lipocalin family is growing significantly; currently over 60 different proteins have been identified as family members. They can be further divided into several non-orthologous groups according to structural features that might reflect common functional properties, such as ligand binding specificities or interactions with other proteins (Cowan et al., 1990; Peitsch and Boguski, 1990). Because of this heterogeneous family organization we assessed whether Lazarillo belongs to a particular clade of lipocalins in order to form a hypothesis about its functional properties. A multiple sequence alignment with all lipocalins present in the data bases was performed. The dendrogram representing the clustering order





**Fig. 7.** Lazarillo belongs to the lipocalin family. (A) Multiple alignment of Lazarillo mature protein with several members of the lipocalin family. The alignment was done with a gap weight of 3 and a length weight of 0.5 (see Material and Methods). Black boxes highlight residue identity with Lazarillo with a threshold of 40%. Gray boxes highlight positions with structurally conserved amino acid substitutions with a threshold of 60%. The structurally conserved regions defining the lipocalin family are indicated (SCR1, 2 and 3). Cysteine residues conserved among Lazarillo and its closest relatives are indicated by black triangles. The residues indicated by asterisks are in close proximity to the hydrophobic ligand in Bbp, Icy, and Apd. (B) Phylogram of selected members of the lipocalin family derived from protein



sequence alignments conducted as in A. A heuristic search of the most parsimonious unrooted tree was carried out with the taxa shown. The total tree length is 1396, with a consistency index of 0.81 and a retention index of 0.58. Calibration bar equals a branch length of 50 residue changes. Bbp: bilin-binding protein from *Pieris brassicae*; Icy: Insecticyanin A from *Manduca sexta*; Apd: human and rat Apolipoprotein D; Cra2: A2 subunit of Crustacyanin from *Homarus gambarus*; Retb: serum retinol-binding protein from humans and *Xenopus laevis*; Ret1: serum retinol-binding protein from *Oncorhynchus mykiss*; Purp: chicken Purpurin; 24p3: mouse oncogene product; PGDs: human brain GSH independent prostaglandin D synthetase; Co8g:  $\gamma$  subunit of C8 component of complement cascade; Hc: human  $\alpha$ -2-microglobulin; Mup5: mouse major urinary protein 5; Lacb: bovine  $\beta$ -lactoglobulin.

employed in the alignment was used to choose the proteins most related to Lazarillo and also examples of more distantly related lipocalins. The program PAUP 3.0s (Swofford, 1991) was used to find an unrooted phylogenetic reconstruction of these proteins by a heuristic search approach. A single optimal tree was found (Fig. 7B) that clearly relates Lazarillo with the clade composed of Apolipoprotein D (Apd), Bilin-binding protein (Bbp) and Insecticyanin A (Icy). This clade is called the porphyrin-binding lipocalins because Bbp and Icy bind

biliverdin IX $\gamma$  (Holden et al., 1987; Huber et al., 1987) and Apd can bind bilirubin (Peitsch and Boguski, 1990). The internal phylogenetic organization of the two main branching lineages was further analyzed using an exhaustive search procedure, including in each case a member of the opposite group. This study confirmed the basic branching pattern within each group and that Lazarillo belongs to the Apd+Bbp+Icy clade. To evaluate the reliability of the inferred tree, a Bootstrap analysis was also conducted that revealed that Lazarillo is linked to a

node common to Apd+Bbp+Icya plus the retinol-binding proteins (Rbp) in 87% of  $10^3$  replicates (not shown).

Lipocalins share a highly conserved folding pattern: a  $\beta$ -barrel structure composed of eight antiparallel  $\beta$ -strands forming two orthogonal sheets with a hydrophobic pocket in between. Two  $\alpha$ -helices at the N- and C-terminal regions also contribute to the main scaffold of the protein. The tertiary structures of Bbp (Huber et al., 1987) and Icya (Holden et al., 1987) have been resolved and that of Apd was modeled from the Bbp and Icya coordinates (Peitsch and Boguski, 1990). This information makes it possible to predict some characteristics of Lazarillo secondary structure based on homology. The sequence alignment of Lazarillo, Apd, Bbp and Icya contains small gaps or insertions in Lazarillo that fall within loops between  $\beta$ -strands or  $\alpha$ -helices in the other lipocalins. The secondary structure of Lazarillo estimated with the statistical method of Chou and Fasman (1974) closely aligns with that of the related lipocalins whose tertiary structure is known (not shown). The four cysteines participating in the conserved pattern of alternating disulfide bonds in Bbp, Icya and Apd are identically aligned in Lazarillo (Fig. 7A). Bbp, Icya and Apd share amino acids that participate in four salt bridges, two of which are also present in Lazarillo. Disulfide bonds and salt bridges contribute to the position of certain  $\beta$ -strands that affect the size of the hydrophobic pocket. Furthermore, the characteristics of the binding pocket are conserved among the three proteins, including the location of hydrophobic patches and potential hydrogen bonds (Peitsch and Boguski, 1990). Lazarillo shares the six most conserved of the 18 residues known to be in close proximity to the ligand (noted with asterisks in Fig. 7A), and has conservative substitutions in seven other amino acids. We propose that Lazarillo contains the common structural core of the lipocalins, the SCRs, which are possibly involved in protein-protein interactions, and those features critical for the ligand binding specificity of the porphyrin-binding clade.

## DISCUSSION

The biochemical properties we have analyzed define Lazarillo as a highly glycosylated small protein, with internal disulfide bonds and linked to the extracellular surface of the plasma membrane by a GPI tail. The GPI linkage is a common feature of a growing number of proteins expressed during development and participating in a variety of biological processes in both vertebrates and invertebrates (Cross, 1990; Ferguson, 1992). Especially significant in this context are the experiments by Chang et al. (1992) that demonstrate that removal of GPI-linked proteins from the surface of cells in the developing limb bud of the grasshopper causes several pathfinding errors by the pioneer sensory axons. Thus, alone or in concert, GPI-linked proteins are certainly contributing to the signaling processes taking place between growth cones and their environment. A potential way to regulate cell-cell or cell-matrix interactions through GPI-linked proteins is to be able to release them from the membranes at appropriate times and locations. Although a soluble cleaved form of the GPI-linked molecule Fasciilin I has been proposed (Hortsch and Goodman, 1990), we do not have evidence for a released form of Lazarillo, at least at levels detectable by immunoblot analysis. Since our analysis of the

GPI linkage to the membrane was performed at 35-45% of development, we cannot rule out that a temporal regulation of release exists later in development or in adulthood.

Lazarillo protein sequence shows significant similarity to members of the lipocalin family, lipid-binding proteins that share a tertiary structure consisting of a  $\beta$ -barrel scaffold with a hydrophobic ligand binding pocket. The lipocalins are divided into several clades reflecting their diversity of functions and ligand specificities. We performed a phylogenetic analysis to assess which particular clade of proteins includes Lazarillo and found that it is associated with the porphyrin-binding lipocalins. Further comparisons revealed the conservation of important features: the main scaffold of  $\beta$ -strands and  $\alpha$ -helices is consistently predicted by homology, the pattern of cysteines forming disulfide bonds is conserved, as well as some of the amino acids participating in salt bridges and involved in ligand interaction. All these data unequivocally establish Lazarillo as a new member of the lipocalin family.

However, Lazarillo has several properties that are unique among lipocalins. Lazarillo is the most glycosylated of its clade and possibly of the lipocalin family. Only a few lipocalins are reported to be glycosylated, and none as extensively as Lazarillo (Urade et al., 1989; McConathy and Alaupovic, 1976). The presence of abundant glycosylation might be modulating the interactions of Lazarillo with other molecules in ways different from the rest of the lipocalins. The GPI tail links Lazarillo to the extracellular side of the plasma membrane and prevents it from being an extracellular carrier of hydrophobic ligands as is commonly the case for lipocalins. PGDs is the only other example of a lipocalin associated with membranes, the endoplasmic reticulum and nuclear membranes, but it is readily dissociated from them in the absence of detergents (Urade et al., 1985). Finally, the restricted tissue distribution of Lazarillo does not have a parallel in the lipocalin family. There are few lipocalins associated with the nervous system, and none of them is confined to a subset of developing neurons. PGDs is first expressed in developing neurons and then changes to oligodendrocytes in the adult animal (Urade et al., 1987). Apd is expressed in a variety of tissues, but accumulates at injured peripheral nerves where it is proposed to help remove toxic heme metabolites produced by hemorrhage (Boyles et al., 1989). Finally, Purpurin is a Rbp secreted by retinal photoreceptor cells that is involved in cell adhesion and transport of retinoids (Schubert et al., 1986; Berman et al., 1987).

Lipocalins are the extracellular elements of a trafficking system for small hydrophobic molecules that also includes the fatty acid-binding proteins (FABP), mainly located in the cytoplasm (Matarese et al., 1990), and nuclear receptors, such as those for retinoic acid that are involved in transcriptional regulation (Petkovich et al., 1987). Although different biological roles are attributed to these proteins, such as lipid transport and metabolism, evidence is accumulating that they function in signaling systems, regulating various developmental processes (Ross, 1993). The recently described brain lipid-binding protein (BLBP) is a neural-specific member of the FABPs whose expression is correlated with neuronal differentiation and is proposed to be required for the establishment of the radial glial fibers and the migration of cerebellar granule cells (Feng et al., 1994). In addition, cellular retinol and retinoic acid-binding proteins, CRBPs and CRABPs respec-

tively, could be involved in the guidance of spinal commissural neurons (Maden and Holder, 1991). Some members of the FABPs have been reported in invertebrates and proposed to play a role in nervous system morphogenesis (Muehleisen et al., 1993). However, the lipocalins described so far in arthropods are mainly circulating transporters of pigments related to camouflage, photoprotection and photoreception (Holden et al., 1987; Clarke et al., 1990; Huber et al., 1987).

The restricted expression, membrane localization, and the functional results described in the following paper (Sánchez, Ganfornina and Bastiani, 1995) suggest a role for the lipocalin Lazarillo in the signaling events necessary to direct the trajectory of growing axons in the grasshopper embryo. Based upon the information available on the molecular interactions described for the lipocalins, we suggest three testable hypotheses about how Lazarillo could function in a signaling system used for axon guidance. First, the ligand that binds to the hydrophobic pocket could act as a guidance cue by inducing Lazarillo to interact *in cis* with a transmembrane protein. The SCR of the lipocalins lie close to each other at the protein surface and are suggested to constitute the binding site for a lipocalin receptor (North, 1989; Flower et al., 1993). Likewise, lateral movement in the plane of the membrane and clustering of Lazarillo molecules could trigger the signal transduction as has been proposed for other GPI-linked proteins (Ferguson, 1992; Mayor et al., 1994). Second, Lazarillo could be an adhesion molecule. The lipocalin Purpurin has been shown to function as a cell adhesion molecule (Schubert et al., 1986). In this context, the observations that many lipocalins form dimers, tetramers or larger polymers (Drayna et al., 1986; Keen et al., 1991), suggest that Lazarillo could be involved in homophilic interactions, and therefore function in cell adhesion by binding monomers on opposite plasma membranes. Finally, Lazarillo might regulate the uptake and transport of a signaling ligand into the cytoplasm, and members of the FABP family would then transfer the ligand to nuclear receptors. Important for all these hypotheses is the observation that the formation of polymers as well as interaction with other proteins have been reported to be dependent on the ligand binding (Blomhoff et al., 1990; Keen et al., 1991).

To determine the role of Lazarillo in the development of the nervous system, it will be of great significance to find both the proteins with which Lazarillo can interact and any putative hydrophobic ligand. The phylogenetic analysis we have carried out links Lazarillo to the porphyrin-binding and retinol-binding lipocalins, suggesting that heme metabolites and retinoids are candidate ligands. In summary, Lazarillo, as a membrane bound lipocalin, adds a new element to the trafficking system of small hydrophobic molecules and a new family of proteins to the list of neuronal surface molecules involved in axon guidance. Its experimental analysis will help uncover new ways of coding the information required for accurate growth cone navigation, an essential process for assembling a functional nervous system.

It is our pleasure to thank the members of our laboratory and of the Biology Department for support and critical reading of the manuscript; M. Low for the generous gift of PI-PLC; C. Goodman for advice; J. Broadus for communication of unpublished protocols; G. Lark for advice on phylogenetic analysis; G. Gutiérrez for help on sequence analysis; M. Herrera for technical assistance; and E.

Carpenter for generating mAb 10E6. This work was supported by NIH grant to M. J. B. (NS25387). D. S. is the recipient of a NIH postdoctoral fellowship (1F05TW04686-01), and M. D. G. holds a fellowship from the Fulbright/MEC of Spain.

Accession numbers for Lazarillo sequence: GenBank U15656; EMBL Z38071.

## 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.
- 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.
- Bentley, D., Keshishian, H., Shankland, M. and Toroian-Raymond, A. (1979). Quantitative staging of embryonic development of the grasshopper, *Schistocerca nitens*. *J. Embryol. Exp. Morph.* **54**, 47-74.
- Berman, P., Gray, P., Chen, E., Keyser, K., Ehrlich, D., Karten, H., LaCorbiere, M., Esch, F. and Schubert, D. (1987). Sequence analysis, cellular localization, and expression of a neuroretina adhesion and cell survival molecule. *Cell* **51**, 135-142.
- Bixby, J. L. and Harris, W. A. (1991). Molecular mechanisms of axon growth and guidance. *Annu. Rev. Cell Biol.* **7**, 117-159.
- Blomhoff, R., Green, M. H., Berg, T. and Norum, K. R. (1990). Transport and storage of vitamin A. *Science* **250**, 399-404.
- Boyles, J. K., Kosik, L. M. and Wardell, M. R. (1989). Apolipoprotein D is synthesized by neural tissue and accumulates in the regenerating nerve. *Circulation* **80**, II-385, Abstr. 1528.
- Carpenter, E. M. and Bastiani, M. J. (1990). Expression and characterization of nervous system antigens in grasshopper embryos. *Soc. Neurosci. Abstr.* **16**, 303.
- Cavener, D. R. and Ray, S. C. (1991). Eukaryotic start and stop translation sites. *Nucl. Acids 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 glycosyl-phosphatidylinositol-anchored cell surface proteins. *Development* **114**, 507-519.
- Chou, P. Y. and Fasman, G. D. (1974). Prediction of protein conformation. *Biochem. J.* **13**, 222-245.
- Clarke, J. B., Eliopoulos, E. E., Findlay, J. B. and Zagalsky, P. F. (1990). Alternative ligands as probes for the carotenoid-binding site of lobster carapace crustacyanin. *Biochem. J.* **265**, 919-921.
- Coyne, K. E., Crisci, A. and Lublin, D. M. (1993). Construction of synthetic signals for glycosyl-phosphatidylinositol anchor attachment. *J. Biol. Chem.* **268**, 6689-6693.
- Cowan, S. W., Newcomer, M. E. and Jones, T. A. (1990). Crystallographic refinement of human serum retinol binding protein at 2 Å resolution. *Proteins Struct. Funct. Genet.* **8**, 44-61.
- Cross, G. A. M. (1990). Glycolipid anchoring of plasma membrane proteins. *Annu. Rev. Cell Biol.* **6**, 1-39.
- Cypher, C. and Letourneau, P. C. (1992). Growth cone motility. *Curr. Opin. Cell Biol.* **4**, 4-7.
- Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387-395.
- Drayna, D., Fielding, C., McLean, J., Baer, B., Castro, G., Chen, E., Comstock, L., Henzel, W., Kohr, W., Rhee, L., Wion, K. and Lawn, R. (1986). Cloning and expression of human apolipoprotein D cDNA. *J. Biol. Chem.* **261**, 16535-16539.
- Feng, L., Hatten, M. E. and Heintz, N. (1994). Brain lipid-binding protein (BLBP): A novel signaling system in the developing mammalian CNS. *Neuron* **12**, 895-908.
- Ferguson, M. A. J. (1992). Glycosyl-phosphatidylinositol membrane anchors: the tale of a tail. *Biochem. Soc. Trans.* **20**, 243-256.
- Flower, D. R., North, A. C. T. and Atkwood, T. K. (1993). Structure and sequence relationships in the lipocalins and related proteins. *Protein Sci.* **2**, 753-761.
- Goodman, C. S., Raper, J. A., Ho, R. K. and Chang, S. (1982). Pathfinding of neuronal growth cones in grasshopper embryos. In *Developmental Order: Its Origin and Regulation* (eds. S. Subtelny and P. B. Green), pp. 275-316. New York: Alan R. Liss.
- Goodman, C. S. and Shatz, C. J. (1993). Developmental mechanisms that

- generate precise patterns of neuronal connectivity. *Cell* **72/Neuron** **10** (Suppl.), 77-98.
- Haefliger, J. A., Jenne, D., Stanley, K. K. and Tschoopp, J.** (1987). Structural homology of human complement component C8 $\gamma$  and plasma protein HC: identity of the cysteine bond pattern. *Biochem. Biophys. Res. Commun.* **149**, 750-754.
- Harrelson, A. L. and Goodman, C. S.** (1988). Growth cone guidance in insects: Fasciclin II is a member of the immunoglobulin superfamily. *Science* **242**, 700-708.
- Hockfield, S.** (1987). Mab to a unique cerebellar neuron generated by immunosuppression and rapid immunization. *Science* **237**, 67-70.
- Holden, H. M., Rypniewski, W. R., Law, J. H. and Rayment, I.** (1987). The molecular structure of insecticyanin from the tobacco hornworm *Manduca sexta* at 2.6 Å resolution. *EMBO J.* **6**, 1565-1570.
- 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.
- Hraba-Renevey, S., Türlér, H., Kress, M., Salomon, C. and Weil, R.** (1989). SV-40-induced expression of mouse gene 24p3 involves a post-transcriptional mechanism. *Oncogene* **4**, 601-608.
- Huber, R., Schneider, M., Mayr, I., Müller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H. and Kayser, H.** (1987). Molecular structure of the bilin binding protein (BBP) from *Pieris brassicae* after refinement at 2.0 Å resolution. *J. Mol. Biol.* **198**, 499-513.
- Keen, J. N., Cáceres, I., Eliopoulos, E. E., Zagalsky, P. F. and Findlay, J. B. C.** (1991). Complete sequence and model for the A<sub>2</sub> subunit of the carotenoid pigment complex, crustacyanin. *Eur. J. Biochem.* **197**, 407-417.
- 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.
- Kyte, J. and Doolittle, R. F.** (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.
- Maden, M. and Holder, N.** (1991). The involvement of retinoic acid in the development of the vertebrate central nervous system. *Development Supplement* **2**, 87-94.
- Matarese, V., Stone, R. L., Waggoner, D. W. and Bernlohr, D. A.** (1990). Intracellular fatty-acid trafficking and their role of cytosolic lipid-binding proteins. *Prog. Lipid Res.* **28**, 245-272.
- Mayor, S., Rothberg, K.G. and Maxfield, F.R.** (1994). Sequestration of GPI-anchored proteins in caveolae triggered by cross-linking. *Science* **264**, 1948-1951.
- McConathy, W. J. and Alaupovic, P.** (1976). Studies on the isolation and partial characterization of apolipoprotein D and lipoprotein D of human plasma. *Biochemistry* **12**, 515-520.
- Muehleisen, D. P., Gray, R. S., Westbrook, A. L., Wang, Z. F., Chytil, F., Ong, D. E. and Bollenbacher, W. E.** (1993). Characterization of proteins immunologically similar to cellular retinoic acid and cellular retinol binding proteins (CRABP and CRBP) in the CNS of *Manduca sexta* and *Drosophila melanogaster*. *Soc. Neurosci. Abstr.* **19**, 1726.
- Nagata, A., Suzuki, Y., Igarashi, M., Eguchi, N., Toh, H., Urade, Y. and Hayaishi, O.** (1991). Human brain prostaglandin D synthase has been evolutionarily differentiated from lipophilic-ligand carrier proteins. *Proc. Nat. Acad. Sci. USA* **88**, 4020-4024.
- North, A. C. T.** (1989). Three-dimensional arrangement of conserved amino acid residues in a superfamily of specific ligand binding proteins. *Int. J. Biol. Macromol.* **11**, 56-58.
- Peitsch, M. C. and Boguski, M. S.** (1990). Is apolipoprotein D a mammalian bilin-binding protein?. *New Biol.* **2**, 197-206.
- Petkovich, M., Brand, N. J., Krust, A. and Chambon, P.** (1987). A human retinoic acid receptor which belongs to the family of nuclear receptors. *Nature* **330**, 444-450.
- Risler, J. L., Dlorne, M. O., Delacroix, H. and Henaut, A.** (1988). Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix. *J. Mol. Biol.* **204**, 1019-1029.
- Ross, A. C.** (1993). Cellular metabolism and activation of retinoids: roles of cellular retinoid-binding proteins. *FASEB J.* **7**, 317-327.
- 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.
- Sambrook, J., Fritsch, E. F. and Maniatis, T.** (1989). *Molecular cloning*, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Sánchez, D., Ganfornina, M. D. and Bastiani, M. J.** (1995). Spatial restriction of the lipocalin Lazarillo and its role in axonal pathfinding in the grasshopper embryo. *Development* **121**, 135-147.
- Sanger, F., Nicklen, S. and Coulson, A. R.** (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
- Schubert, D., LaCorbiere, M. and Esch, F.** (1986). A chick neural retina adhesion and survival molecule is a retinol-binding protein. *J. Cell Biol.* **102**, 2295-2301.
- Seaver, E., 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.
- Swofford, D. L.** (1991). *PAUP: Phylogenetic analysis using parsimony*, version 3.0s. Illinois Natural History Survey, Champaign.
- Tessier-Lavigne, M.** (1992). Axon guidance by molecular gradients. *Curr. Opin. Neurobiol.* **2**, 60-65.
- Urade, Y., Fujimoto, N. and Hayaishi, N.** (1985). Purification and characterization of rat brain prostaglandin D synthetase. *J. Biol. Chem.* **260**, 12410-12415.
- Urade, Y., Fujimoto, N., Kaneko, T., Konishi, A., Mizuno, N. and Hayaishi, O.** (1987). Postnatal changes in the localization of prostaglandin D synthetase from neurons to oligodendrocytes in the rat brain. *J. Biol. Chem.* **262**, 15132-15136.
- Urade, Y., Nagata, A., Suzuki, Y., Fujii, Y. and Hayaishi, O.** (1989). Primary structure of rat brain prostaglandin D synthetase deduced from cDNA sequence. *J. Biol. Chem.* **264**, 1041-1045.
- von Heijne, G.** (1990). The signal peptide. *J. Membrane Biol.* **115**, 195-201.

(Accepted 30 September 1994)