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

María D. Ganfornina*,†, 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
† 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 45×10^3 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 β-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); C8γ 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 (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% NaN3) 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 125I-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 (σ medium, Gibco-BRL), 1% antibiotic-antimycotic solution (Sigma). They were incubated, washed in PBS, and homogenized in 0.1 M Tris, pH 7.5, 150 mM NaCl, 0.1% Tween-20, 0.1 mM MnCl2, 0.1 mM CaCl2) and incubated for 1 hour with biotinilated 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 mM 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 cut, 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 dH2O at room temperature, chopped and dried in a vacuum centrifuge. The dried gel was soaked in 0.1 M NH4HCO3, 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 NH4HCO3 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.

**Molecular analysis of Lazarillo cDNA**

Degenerate oligonucleotides were designed from the peptide sequences to amplify different 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: MgCl2 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 radio labeled with [32P]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 × 106 plaques from a library made using nerve cords dissected from 55% embryos, and 1.3 × 106 plaques from an amplified library made using nerve cords dissected from 55% embryos, and 1.3 × 106 plaques from an amplified 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 compositions. 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, PIY, 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 mM guanidinium thiocyanate, 1% β-mercaptoethanol. The RNA was purified by CsCl centrifugation as described in Sambrook et al. (1989). Poly(A)+ mRNA was isolated from total RNA using the PolyATrac 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-
sitions, 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 MgSO4) 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 H2O) at room temperature, with a last incubation at 55°C for 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 color reaction was carried out as described in the paper. 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 MgSO4) 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 H2O) at room temperature, with a last incubation at 55°C for 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).

Lazarillo is a neuronal GPI-linked lipocalin

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 Mr of 45×10^3 is recognized by the mAb 10E6 (Fig. 2A, lane 2). In the absence of the mAb the secondary antibody recognizes a nonspecific band of 57×10^3 Mr (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 Mr was also detected by immunoblot analysis using adult membrane preparations (Fig. 2A, lane 4). The apparent Mr 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 µm; D, 25 µm.
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 45×10^3 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 57×10^3) is not extracted by PI-PLC. These results suggest that the 45×10^3 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 45×10^3 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 30-40×10^3 M_r is seen above a tight band of 28.2×10^3 M_r (Fig. 4A,B). Our interpretation is that this result could be due to

---

**Fig. 2.** Lazarillo is a protein with a M_r of 45×10^3. (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 45×10^3 (arrow) in the embryonic membrane preparation. (3) In the absence of the mAb the secondary antibody recognizes a nonspecific band of 57×10^3 M_r. (4) A single band of the same size and appearance than the embryonic 45×10^3 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 β-mercaptoethanol (50 mg/ml). NR: no reducing agent added. Molecular mass markers (M_r×10^{-3}) are shown on the left.

**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 45×10^3 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 μm; (B,D) 200 μm. Molecular mass markers in M_r×10^{-3} are shown on the left in E.
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) WYEYAK; (3) RPDSAASTEISWILLRSxSSxMTLERVEDELK; (4) xSVHFPNPSVGYxILSTDDxYSIV; 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 asparaginines 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 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)+ 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 in 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...
129 Lazarillo: a neuronal GPI-linked lipocalin

(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 (Coyne et al., 1993). The predicted mature protein after cleavage of both signal sequences would have a $M_r$ of approximately $20\times10^3$. 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. 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)$^+$ 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) Hydrophathy 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.
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...
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 (Icya). This clade is called the porphyrin-binding lipocalins because Bbp and Icya bind biliverdin IXγ (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+Icya 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 β-barrel structure composed of eight antiparallel β-strands forming two orthogonal sheets with a hydrophobic pocket in between. Two α-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 β-strands or α-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 β-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 Fasciclin 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 β-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 β-strands and α-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 (Urdea et al., 1989; McConathy and Alaupovic, 1976). The expression 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 (Urdea 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 (Urdea 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-
Lazarillo: a neuronal GPI-linked lipocalin

Carrying out genetic 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


Goodman, C. S. and Shatz, C. J. (1993). Developmental mechanisms that...
generate precise patterns of neuronal connectivity. Cell 72/Neuron 10 (Suppl.), 77-98.


(Accepted 30 September 1994)