MOLECULAR STUDIES ON GENES AND PROTEINS INVOLVED IN BIOMINERALIZATION AND DEVELOPMENT OF THE SEA URCHIN *Paracentrotus lividus*

Von der Fakultät für Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

vorgelegt von

Konstantinos M. Karakostis

aus Athen, Griechenland

Hauptberichter: **Prof. Dr.rer.nat. Franz Brümmer** (Universität Stuttgart, DE)

Mitberichter: **Dr. Valeria Matranga** (Consiglio Nazionale delle Ricerche, IT)

Mitberichter: **Dr. Frédéric Marin** (Université de Bourgogne, FR)

Tag der mündlichen Prüfung: 15.01.2014

Biologisches Institut der Universität

Abteilung Zoologie

2014

Originality Statement

‘I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at any educational institution, except due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have collaborated, is acknowledged is the thesis. I also declare that the intellectual content of this thesis is the product of my work except to the extent that assistance from others in the project’s design and conception or in style, presentation and linguistic expression is acknowledged’.

Table of Contents

Chapter I: General Introduction

Abstract and Scope of the study

This study is focused on the Mediterranean sea urchin species, *Paracentrotus lividus* (*P. lividus*). The major part of the study covered the characterization of genes and proteins involved in biomineralization of *P. lividus* embryo and adult. The aim was to establish *P. lividus* as a working model for biomineralization studies. The primary goal was to develop a set of molecular tools involving plasmids, labeling probes, recombinant proteins and antibodies which subsequently were used in developmental and biomineralization studies. Candidate genes were selected based on previous findings in other species and on *in silico* analysis and comparison to the invertebrates and the sea urchin. The selected biomolecules involved acidic proteins (*p16, p19),* lectins *(advillin, sm30α, sm50, galectin-8),* a signaling protein (*tetraspanin)* and an enzyme *(carbonic anhydrase).* All the above groups of encoded proteins are known to participate in biomineralization in various species and therefore they were selected for the study of their role in the formation of the sea urchin skeleton.

The protein-coding mRNA sequences *p16, p19, advillin, tetraspanin, sm30α, sm50, carbonic anhydrase* and *galectin-8* were identified and cloned, following molecular biology techniques. The putative protein domains were analyzed and a phylogenetic comparison among the sea urchin species of *P. lividus*, *Strongylocentrotus purpuratus* (*S. purpuratus*) and *Lytechinus variegatus (L. variegatus)* revealed their evolution relation. Furthermore, the localization of the temporal and spatial expression of each transcript throughout the embryo development was characterized by comparative qPCR and whole mount *in situ* hybridization (WMISH). The acquired experimental data, revealed the expression profile of each one of these genes in the developing embryo and the insights on the role of each gene in development and skeletogenesis were discussed. Additionally, a functional characterization of recombinant *carbonic anhydrase* and *galectin-8* proteins, cloned and expressed *in vitro* in E. *coli* and purified by affinity chromatography, gave insights of the role of each protein in development. Specific polyclonal antibodies were prepared and used to identify by Western Blot and ELISA, Pl-CA and Pl-Galectin-8-like proteins, in both the embryo and in the occluded matrix proteins of adult *P. lividus,* showing the importance of the two proteins in development. Functional assays, involving the *in vitro* esterase activity of Pl-CA and the lactose-specific hemagglutination activity of the lectin *Pl-GALECTIN-8,* confirmed that both recombinant proteins were active. Additionally, the biological role of *Pl-GALECTIN-8* in cell adhesion was tested *in vivo*, in a human cell adhesion assay.

Furthermore, aiming towards a complete characterization of the biomineralization proteins both in the embryo and in adult, the proteome of the occluded matrix proteins from adult *P. lividus* mineralized tests, was examined. Proteins were extracted from purified calcareous tests and identified by Liquid Chromatography coupled with Mass Spectrometry (LC-MS/

MS). For the first time, the effect of the sea urchin occluded matrix protein on the formation of calcite is studied by an *in vitro* crystallization assay.

Zusammenfassung

Diese Studie basiert auf der Charakterisierung von Genen und Proteinen bei der Biomineralisation des Steinseeigels des Mittelmeers, *Paracentrotus lividus* (*P. lividus*). Das Ziel war es, den P. lividus als Arbeitsmodell für Studien zur Biomineralisation zu etablieren, indem eine Reihe von molekularen Tools entwickelt wurden, die Plasmide, die Markierung von Proben, rekombinante Proteine und Antikörper einbeziehen und die anschließend bei Entwicklungs- und Biomineralisationsstudien verwendet werden. Kandidatengene, die an der Skeletogenese teilnehmen, wurden basierend auf früheren Forschungsergebnissen bei anderen Arten, ausgewählt. Die ausgewählten Biomoleküle beinhalten saure Proteine (p16, p19), Lektine (Advillin, sm30α, sm50, Galectin-8), ein Signalprotein (Tetraspanin) und ein Enzym (Carboanhydrasen). Die cDNA-Sequenzen wurden identifiziert und nachgebildet, basierend auf Techniken der Molekularbiologie. Die mutmaßlichen Proteindomänen wurden analysiert und ein phylogenetischer Vergleich zwischen den Seeigeln zeigte ihre Konservierung. Darüber hinaus wurden die mRNA Expressionsprofile während der Entwicklung des Embryos durch qPCR und "whole mount *in situ* Hybridisierung (WMISH)" charakterisiert, Erkenntnisse bringend über die Rolle jedes Gens während der Entwicklung und Skeletogenese. Zusätzlich hat eine zweckmäßige Charakterisierung von rekombinanten Carboanhydrasen und Galectin-8-Proteinen , die in E. *coli* hergestellt wurden, Erkenntnisse erbracht bezüglich ihrer Aufgaben während der Entwicklung. Antikörper wurden hergestellt und verwendet, um Pl-CA und Pl-Galectin-8-ähnliche Proteine zu identifizieren, im Embryo sowie im Erwachsenen. Funktionale Proben, die die *in vitro* Esterase Aktivität der Pl-CA und die Aktivität der Lactose spezifischen Hämagglutination des Lektin Pl-Galectin-8 beinhalten, haben bestätigt, dass beide rekombinanten Proteine aktiv waren. Zusätzlich wurde in einer Zelladhäsions Probe die biologische Rolle von Pl-Galectin-8 in der Zelladhäsion getestet, welche Aufschluss über dessen Rolle bei der Zelladhäsion gab. Weiterhin wurde, mit dem Ziel in Richtung einer vollständigen Charakterisierung der Proteine bei der Biomineralisation sowohl im Embryo als auch im Erwachsenen, das Proteom der verschlossenen Matrixproteine aus erwachsener *P. lividus* mineralisierter Probe untersucht. Proteine wurden von kalkhaltigen erwachsenen Teilen gereinigt, durch Flüssigchromatographie mit Massenspektrometrie-Kopplung (LC-MS/MS) identifiziert und ihre Wirkung auf die Bildung der Biomineralisation wurde durch eine *in vitro* Kristallisation Probe untersucht.

Outline of the thesis

This thesis is structured as a series of stand-alone manuscripts as most chapters are in the process of being prepared for submissions to peer-reviewed journals (after some minors modifications like adding names of authors etc). Therefore, there may be some

slight repetition between chapters. Chapter II involves a description of the experimental procedures followed in common, in the majority of the experiments. Chapters III to VI examine each topic separately involving an additional introduction to each topic, a description of the experimental procedures followed; and results' sections. Chapter VII is a general discussion chapter including comments and remarks of the chapters III to VI. Last, lists of published or under preparation manuscripts derived from this work, the most relevant references and abbreviations are provided.

Biomineralization

Biomineralization refers to the biological processes employed by living organisms to form minerals as a result of regulated biological processes. A biomineral represents a complex material which incorporates both mineral and organic components, exhibiting advantageous properties compared to its inorganically formed counterpart. Indeed, compared to abiotic minerals, biominerals possess additional physical and chemical characteristics which offer increased flexibility and durability. They vary in morphology, shape and size, as well as in element composition. The structure of a biomineral involves a mosaic of crystalline domains separated by an occluded proteinaceous material, forming a framework (Wilt 1999). This structure exhibits single crystal diffraction properties as shown by X-ray diffraction studies (Simkiss, 1986).

According to the degree of biological control over the precipitated mineral, biomineralization processes can be categorized into two groups: The "biologically-induced" (Lowenstam 1981) and the "biologically-controlled" mineralization (Mann, 1983).

In biologically-induced mineralization, cell surfaces may act as causative nucleation agents which lead to crystal growth. Mineral growth is indirectly affected, but not controlled, by the biological system. The adopted mineral form is favored by metabolic processes which define the chemical conditions of the micro-environment (i.e. the pH, pCO_2, concentration of secreted products) (Frankel and Bazylinski, 2003). As the environmental conditions play a potential role in the formation of the biologically-induced minerals, these biominerals exhibit heterogeneity in elemental composition, in water content and in particle size, resulting in various external morphologies.

On the other hand, in biologically-controlled mineralization, cellular activities direct all the stages for the precipitation of the mineral: from the determination of the initial deposition site, to nucleation, crystal growth and formation of the final crystal form. These biominerals acquire reproducible, species-specific, genetically determined structures and properties. Controlled biomineralization requires an isolated micro-environment which serves as the mineralization site. This micro-environment can be extra-cellular, inter-cellular or intra-cellular, with respect to the cells which control the process.

In general, biologically-induced mineralization is found in bacteria and lichens, whereas biologically-controlled mineralization is found in foraminifera, cephalopod statoliths, mollusks shells, bryozoan exoskeletons, scleractinian corals, echinoderms, human bones and teeth.

In nature, almost half of the biomineral types are calcium-bearing minerals (Lowenstam and Weiner 1989). Calcium content in living organisms at concentrations varying from 10 nM to 10 µM, is tightly regulated as it plays a key role in various metabolic processes. The most abundant calcium-bearing biominerals precipitate, acquiring one of the eight known calcium carbonate polymorphs. These include six crystalline forms: calcite, magnesium-calcite, aragonite, vaterite, monohydrocalcite, protodolomite, and one amorphous calcium carbonate (ACC) form which is also found hydrated (Addadi *et al.*, 2003).

Most biomineralization models invoke the involvement of membrane ion transporters (channels and pumps) in the delivery of Ca^{2+} and other ions to the calcification site (Simkiss and Wilbur 1989). This establishes a physico-chemical equilibrium with the micro-environment fulfilling an important parameter for the formation of a biomineral, which is the regulation of the so called “isotopic composition” at the mineralization site (Weber and Raup 1966). Additionally, the formation of the mineral requires a saturated medium. Super-saturation can be reached by the presence of additives such as magnesium in concentrations similar to those found in the sea water (Raz *et al.,* 2000), which prevent the rapid deposition of crystalline phases. Another inhibitory factor of uncontrolled crystallization is the presence of proteins which, serving as substrates, influence the solubility of the mineral phase and stabilize different polymorphs, either ACC (Aizenberg *et al.*, 1996) or crystalline calcium carbonate forms.

In conclusion, the shape of the crystal depends on both inorganic and organic factors: p*I*, temperature, micro-environment, solubility of the mineral phase, concentration of occluded macromolecules and ions. Through this medium, calcium carbonate crystal growth undergoes through a series of phases and morphologies to form the final calcite structure. In fact, the effect of whole extracts of proteinaceous matrices on the i*n vitro* precipitation of $CaCO_3$ was shown to selectively induce the precipitation of particular polymorphs (Marie *et al.*, 2012). Some authors have pointed out that the soluble or insoluble matrix extracted from nacre (Mollusk), controls the formation of calcium carbonate crystal polymorphs towards aragonite or calcite (Falini *et al.*, 1996). Others, used chitin as substrate and Mg^{2+} as additive to induce formation of aragonite double-layered composite film (Kato 2000). It becomes apparent that molecules can act as templates that favor the nucleation (precipitation) of the inorganic material. The surface chemistry of each template molecule and the arrangement of the amino acids and functional groups, guide the orientated nucleation of the cognate crystal face.

Biomineralization in echinoderms

Echinoderms appeared during the early Cambrian period, at least 520 million years ago (Bottjer *et al.*, 2006), forming a phylum of marine invertebrates, including Echinoids (sea urchins and sand dollars), Asteroids (sea stars), Ophiuroids (brittle stars), Crinoids (feather stars) and Holothuroids (sea cucumbers). Biomineral formation facilitates the construction of a calcium-bearing endoskeleton, offering mechanical support and protection to these organisms. Although bio-mineralized structures probably evolved independently in

echinoderms and in vertebrates, certain embryological and molecular programs already established in ancestral deuterostomes may have been exploited in similar ways in these two groups to carry out bio-mineralization (Sodergreen *et al.*, 2006; Matranga *et al.*, 2011).

Echinoderms employ a genetically regulated process to form their calcareous skeleton by the precipitation of magnesian calcite, a form of calcium carbonate that contains minor amounts of magnesium carbonate in a ratio given by the formula: $(Mg_xCa_{1-x})CO_3$. At high Mg/Ca ratio, nucleation of amorphous calcium carbonate or aragonite is favored over calcite (Mann, 2001; Raz *et al.*, 2000). The magnesium carbonate content varies from 2.5% to 39% depending on the classes and species. In general, the hardness of the biomineral is directly proportional to the magnesium concentration. Low magnesium biominerals are found in the softer Ophiuroids and high magnesium biominerals are found in the harder Asteroids (Dubois and Chen, 1989).

Structurally, the skeleton of all adult echinoderms is made of microscopic bony plates, also called ossicles, deposited as a three-dimensional meshwork called stereom, made of magnesian calcite and a network of interconnected holes filled with living tissue. Different types of stereom are indicative of the type of living tissue that penetrates the plates (as for example cells and tube feet). The skeleton is covered by an epidermis and contains a network of internal water-filled canals or encloses a coelomic cavity dipped in coelomic fluid. The structure of the biomineral reflects complexity. Skeletal plates may remain simple or fuse to form composite plates. They may also form tubercles, granules and fixed or movable spines. Among the classes, the most diverse skeleton organizations are evident in Echinoids and Holothuroids. In the first case, organisms are completely surrounded by a calcified test embedded in the mesodermal stroma tissue, with only a thin layer of tissue. On the contrary, Holothuroids possess an endoskeleton reduced to microscopic ossicles dispersed throughout the dermis, with a highly muscularized body wall (Smith *et al.*, 2010). In the last 40 years, advanced high resolution imaging technologies, such as scanning electron microscopy (SEM) and transmission electron microscopy (TEM), have provided researchers with high quality images of the fine structure and composition of the stereom giving information on the echinoderm calcification.

The sea urchin species and *Paracentrotus lividus* as a model system

Among echinoderms, the sea urchin embryo has been known for its versatility and suitability as a model since the end of the nineteenth century when embryologists performed the earliest studies on the basic mechanisms of embryo development, facilitated by the optical transparency of the embryo, along with its simplicity in shape and cellular organization (Hoerstadius 1939). Since that time, the sea urchin has been used to a great extent for studies in several research fields ranging from basic developmental biology to eco-toxicology and biotechnology.

P. lividus embryos have an extensive endoskeleton distributed as a calcified crystal, offering support and protection (Mann 2002). The skeleton is composed of magnesium calcite and occluded spicule matrix proteins (Decker and Lennarz, 1988; Wilt, 2002; Wilt *et*

al., 2003). In addition, the formation of its larval endoskeleton encourages its use in detailed studies on the mechanism underlying biomineralization.

In *P. lividus* as in all the sea urchin, it has become clear that biomineralization is a genetically controlled process involving various steps of biological control (Matranga *et al.*, 2011). The formation and remodeling of carbonate crystals involves over-saturation of calcium and bicarbonate ions, created by active transport of these ions from the sea water (Simkiss and Wilbur, 1989). The biomineral formation begins in early embryonic developmental stages (Fig.1) and is mediated by a specialized population of cells, the primary mesenchyme cells (PMCs), descendants of the micromeres. The PMCs, expressing specific genes under a complex signaling control of transcription growth factors, mediate the deposition of the biomineral into the blastocoel cavity of the developing embryo (Ettensohn 2009). Analytically, as gastrulation initiates by the ingression of the cells forming the archenteron into the blastocoel, PMCs migrate into the blastocoel arranging a characteristic ring-like pattern, which consists of two ventrolateral clusters of cells linked by cellular chains on the oral (ventral) and aboral (dorsal) surfaces of the blastocoel wall. The PMCs' patterning is accompanied by the fusion of filopodia-forming cells, forming cables which link the PMCs within a syncytial network (Okazaki, 1965; Hodor and Ettensohn, 1998). Within these filopodial cables, the PMCs of each ventrolateral cluster secrete a tri-radiate skeletal rudiment which later elongates in a stereotypical fashion forming the branched, bilaterally symmetrical spicules of the pluteus larva. Local cues from overlying ectodermic cells play an important role in regulating gene expression, skeletal rod growth and rod branching within the PMC syncytium (Ettensohn and Malinda, 1993; Guss and Ettensohn, 1997; Zito *et al.*, 2003) (See Fig.2).

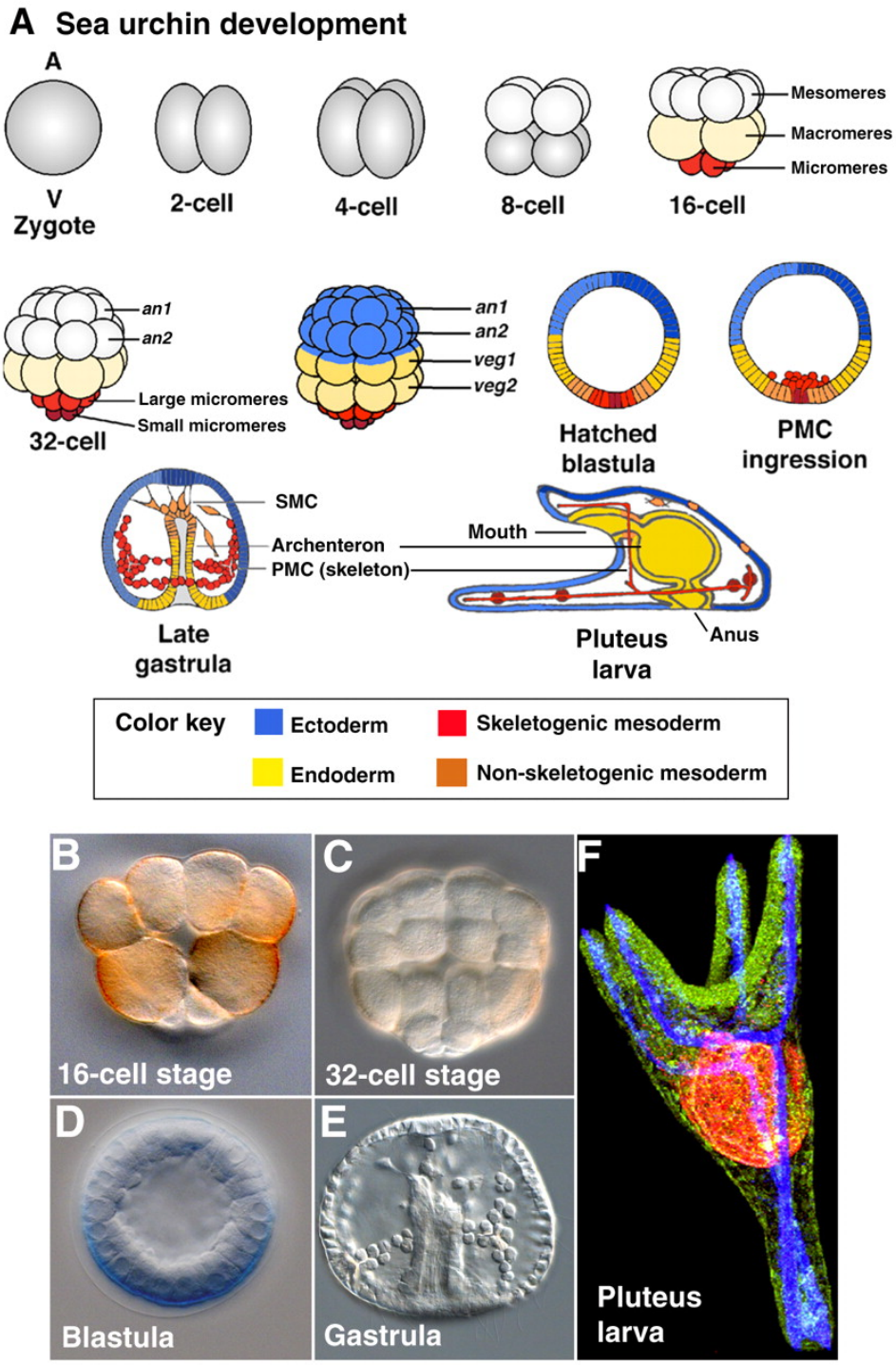


Figure 1. Early sea urchin development (Figure by McClay DR, Development, 2011;138:2639-2648). (A) Sequence of sea urchin development from the zygote to the pluteus larva stage. At the 16-cell stage there are four micromeres (red) at the vegetal (V) pole, four central macromeres (light yellow) and eight mesomeres (grey) at the animal (A) pole. From the hatched blastula stage onwards, the embryo is shown as a mid-sagittal section. The colors indicate when the cells begin to be specified toward ectoderm (blue), mesoderm (red) and endomesoderm (yellow). Later, the ectoderm becomes subdivided (as indicated by different shades of blue), and the mesoderm (orange) separates from endoderm (dark yellow). (B-E) Selected stages of *P. lividus* development: (B) 16-cell stage; (C) 32-cell stage; (D) blastula stage; and (E) mid-gastrula stage, showing the gut

invaginating and the skeletogenic cells forming a ring of cells around the gut and beginning to synthesize the skeleton. (F) Pluteus larva stained to show the gut (red), the skeleton (blue) and the ectoderm (green). an, animal; veg, vegetal; PMC, primary mesenchyme cells (skeletogenic cells); SMC, secondary mesenchyme cells (non-skeletogenic mesoderm).

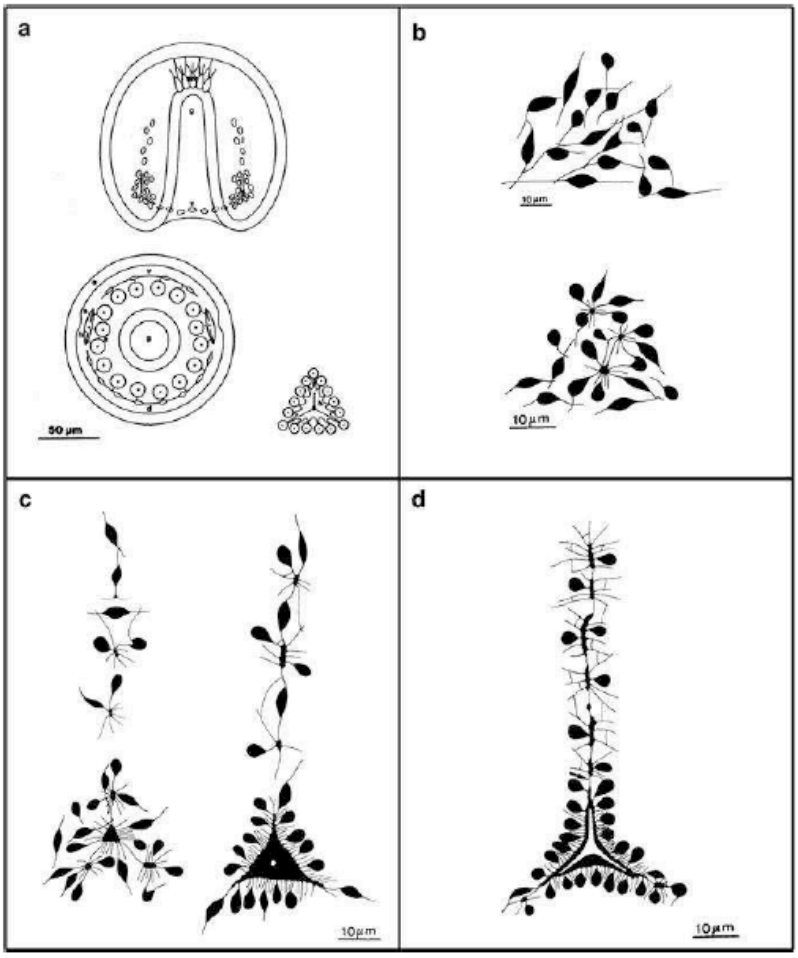


Figure 2. Schematic representation of the localization and arrangement of PMCs in the sea urchin embryo. Figure by Dubois and Chen, 1989. (a) Gastrula frontal and vegetal views showing the PMCs clusters, positioned at opposite sides of the archenteron and one of the clusters enclosing the spicule tri-radiate rudiment; (b) formation of PMCs aggregates and initial pseudopodial fusion; (c) alignment of PMCs and enlargement of the fused cytoplasmic space. The white dot represents the initial biomineral deposition site; (d) development of spicule rudiment inside the cell cluster.

The larval skeleton displays a considerable morphological diversity among sea urchin species, including variations in the number, shape and size of the rods, highlighting the importance of the diversity of larval skeleton morphology, from an evolutionary point of view (Zito and Matranga, 2009). In *P. lividus* embryo, the PMC chains give rise to different rods. As shown in Fig.3 , the ventral chain gives rise to the ventral transverse rod, the longitudinal chain will form the anterolateral rod and the dorsal chain gives rise to the body and postoral rods.

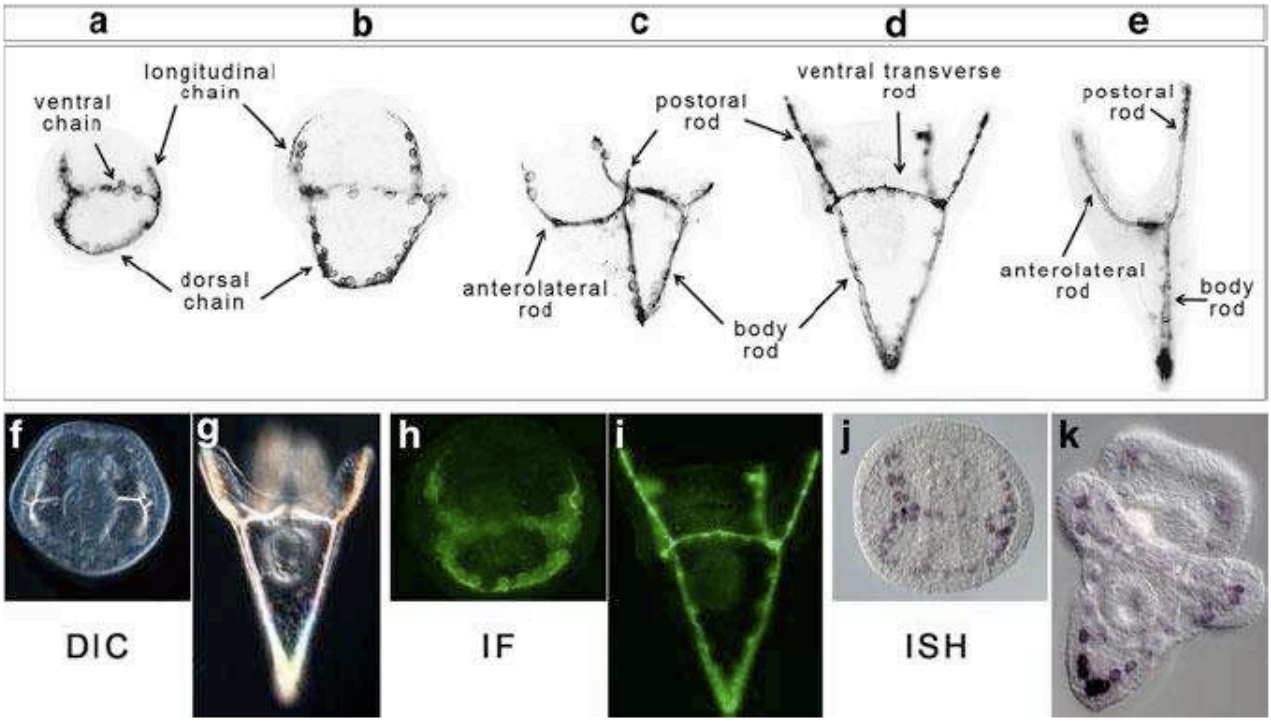


Figure 3. Development of *Paracentrotus lividus* skeleton. Figure by Dr. V. Matranga lab, IBIM, CNR. Schematic drawings of skeleton development observed at (a) late gastrula; (b) prism; (c) early pluteus (d) pluteus, ventral view; (e) pluteus, lateral view. Ventral chain, longitudinal chain and dorsal chain indicate set of PMCs which will give raise to the ventral transverse rod, anterolateral rod and body and postoral rod, respectively. Skeleton of embryos in late gastrula (f,h,j) and pluteus (g,i,k) as viewed by (f,g) Differential Interference Contrast (DIC) microscopy; (h,i) immunofluorescence (IF) with antibody against the PMC marker *MSP130*; and (j,k) *in situ* hybridization (ISH) with *msp130* RNA probe.

At the molecular level, several PMC-specific genes have been identified from an arrayed *S. purpuratus* PMCs' cDNA library (Zhu *et al.*, 2001; Illies *et al.*, 2002) as well as from the complete *S. purpuratus* Genome Project (Sea Urchin Genome Consortium, 2006). Furthermore, Livingston *et al.*, 2006 described a genome-wide search for biomineralization-related proteins. Sea urchins have proven to be particularly useful for the analysis of gene regulatory networks (GRNs) in early development. GRNs that underlie cell specification are presently understood in greater detail in the sea urchin compared to other metazoan embryos, although research is ongoing in several other experimental models as well (Koide *et al.*, 2005; Stathopoulos and Levine, 2005; Ge *et al.*, 2006; Satou *et al.*, 2008), Oliveri and Davidson, 2004, Ben-Tabou de-Leon and Davidson, 2007. The GRN concept offers an informative way to examine the ability of embryonic cells to switch developmental pathways (developmental plasticity), morphogenesis, and the evolution of developmental programs, including the regulation of skeletogenesis (Ettensohn *et al.*, 2009, Ettensohn, 2013). In the sea urchin (as in most metazoan embryos), maternal polarity entails early patterning (Brandhorst and Klein, 2002; Angerer and Angerer, 2003). Zygotic transcription initiates with the egg fertilization reaching a maximal rate during early

cleavage. By the 16-cell stage, various gene expression programs are deployed in the different tiers of blastomeres along the animal-vegetal (AV) axis. Many genes are expressed in each domain of the distinct territories of the late blastula (Fig.4).

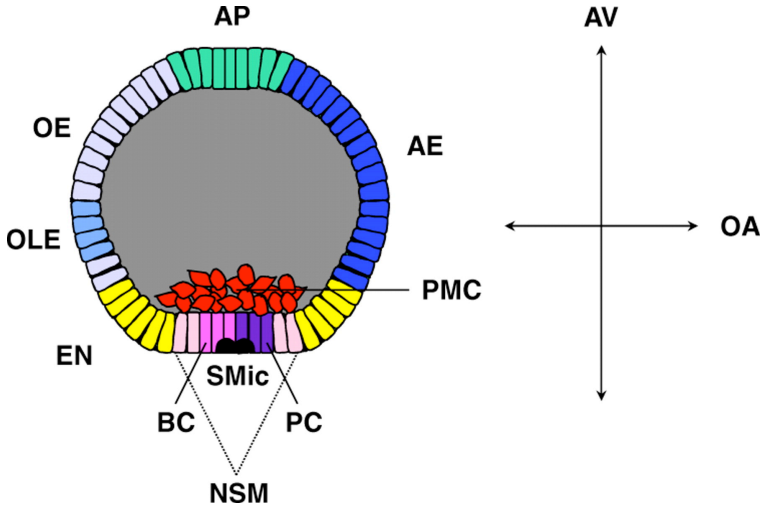


Figure 4. Territories of the late-blastula stage sea urchin embryo. Figure by Ettensohn *et al.*, 2009. The different cell territories of the embryo are shown in different colors and the central blastocoel cavity is shaded gray. (OA): oral-aboral axe, (AV): animal-vegetal, (AE): aboral ectoderm, (AP): apical plate, (BC): presumptive blastocoelar cells, (EN): endoderm, (NSM): non-skeletogenic mesoderm; (OE): oral ectoderm; (OLE): oral-lateral ectoderm; (PC): presumptive pigment cells; (PMC): primary mesenchyme cells; (SMic): small micromeres (Ettensohn *et al.*, 2009).

The micromere-PMC GRN consisting of more than 70 genes (a schematic view of the network is shown in (Fig.5), is activated by maternally derived components of the canonical Wnt signaling pathway, (reviewed by Ettensohn, 2006) and β-Catenin (Kitamura *et al.*, 2002; Oliveri *et al.*, 2002; Oliveri *et al.*, 2003; Nishimura *et al.*, 2004; Yamazaki *et al.*, 2005). It should be mentioned that although the β-catenin protein is present throughout the vegetal region of the embryo during early cleavage, it activates pmar1 only in the micromeres. It becomes obvious that local cues from different compartments, play an important role in the regulation of genes expressed in the PMCs, mediating skeletogenesis. Oliveri and co-workers (Oliveri *et al.*, 2008) expanded the micromere-PMC GRN by carrying out MO-mediated knockdowns of expression of various transcription factors (TFs), that are expressed selectively in this lineage.

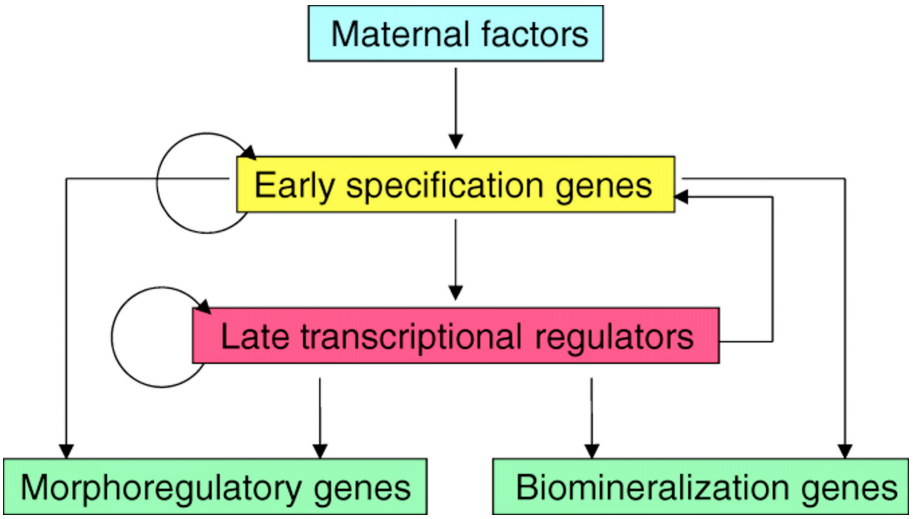


Figure 5. Schematic representation of the biomineralization-related regulatory network. Main layers of regulatory control within the PMC GRN. Regulatory interactions within and between levels are indicated by arrows (Ettensohn, 2009).

At the proteomic level, only recently Mann *et al.*, 2010, using Mass-Spectrometry analyzed the spicule matrix proteome of embryo *S. purpuratus* identifying 231 proteins some of which are also found in the adult mineralized parts (Mann *et al.*, 2008). Among the most abundant proteins, were various C-type lectins (*i.e.* SM30 and SM50), metalloproteases, acidic proteins and phosphoproteins (ie P16 and P19) and carbonic anhydrase (Alvares *et al.*, 2009). Most of the identified proteins were sea urchin-specific without apparent orthologues in other invertebrate, deuterostomes or vertebrates.

It should be mentioned that the increasing interest on the Mediterranean species *P. lividus*, initiated the *P. lividus* Genome Project which is currently in progress from Marine Genomic Europe Network (MGE).

Chapter II: General Experimental Procedures

This chapter describes the experimental procedures followed regularly and in common for the needs of this study. The additional experimental procedures which were applied for each subject, are described in the corresponding chapter.

Embryo culture, total RNA extraction and RT-PCR

Adult sea urchins (*P. lividus*) were collected from the North-Western coast of Sicily (Mediterranean Sea) and kept in aquaria with circulating seawater obtained from the collection site. Spawning was induced by intra-coelomic injection of 0.5 M KCl. After fertilization, embryos (4000 embryos/ml) were cultured in glass beakers in Millipore-filtered sea water containing antibiotics (50 µg/ml streptomycin sulfate and 30 µg/ml penicillin) with gentle stirring at 18°C. Unfertilized eggs and embryos were collected at different developmental stages by low-speed centrifugation and were either instantly frozen in liquid nitrogen before stored at -80 °C for subsequent RT-PCR experiments or fixed in 4% para-formaldehyde in seawater before stored at -20 °C for whole mount *in situ* hybridization (WMISH) experiments, as reported by Russo *et al*., 2010. Total RNA from embryos was isolated from various developmental stages using the Gene Elute Mammalian total RNA kit (Sigma). Residual DNA was digested by DNase I (Ambion) according to the manufacturer's instructions. Target cDNA was directly amplified from the total RNA in a one-step reaction, using the SuperScript One-Step RT-PCR kit (Invitrogen) following the manufacturer's instructions.

Identification of *P. lividus* full-length coding sequences (CDS) and cloning

cDNA sequences, previously annotated in the sea urchin *S. purpuratus* database, predicted to code for each protein, were screened by BLAST (Altschul *et al*., 1990) against EST databases of *P. lividus*, at NCBI (http://www.ncbi.nlm.nih.gov/) and at MPIMG (http://goblet.molgen.mpg.de/cgi-bin/webapps/paracentrotus.cgi). The *P. lividus* ESTs exhibiting the highest similarity were identified and used as templates for the amplification of each sequence by RT-PCR (One Step RT-PCR, Invitrogen), using purified total RNA extracted from the gastrula stage. Specific primers designed complementary to each homologous compiled EST sequence (purchased from MWG, Heidelberg, Germany). Each PCR amplification product was cloned in the *pGEM-T-Easy* Vector (Promega), following the manufacturer's instructions. Each PCR product was cloned in the *pGEM-T-Easy vector* (see relevant chapters). The obtained plasmids were sequenced by MWG (Heidelberg, Germany) and sequences were found to match with the *P. lividus* EST clones available in the databases. Each sequence was submitted in the EMBL Genebank database and is available on NCBI (see Results).

In the case of *Pl-ca, Pl-galectin-8-like* and *Pl-sm50,* as the obtained sequences revealed only partial CDSs, a 3'RACE kit (Invitrogen) was used to identify the 3'-terminal end of the sequence. For the identification of the full-length sequence of CA, 5' RACE (Invitrogen) was also applied for the identification of the N-terminal part of the CDS (see Chapter IV). Subsequently, new primers were designed for the amplification by RT-PCR of the full-length CDSs, which were cloned in the p*GEM-T-Easy* vector (Promega) and re-sequenced for validation.

All the used primers are outlined in the relevant chapters (for CA, see Chapter IV , for galectin-8, see Chapter V, for sm50, see Chapter III) with supplied information on the sequences and the cloning procedures.

Putative protein domain characterization and phylogenetic analysis

The deduced amino acidic sequences were analyzed *in silico* as described by Costa *et al.*, 2010, using the tools Compute pI/MW and Signal P3 server, at the server http://web.expasy.org (Gasteiger *et al.*, 2003). The functional domains were predicted by the Motif Scan Software (http://hits.isb-sib.ch/cgi-bin/PFSCAN; while putative signal peptide cleavage sites (Bendtsen *et al.*, 2004) and phosphorylation sites (Blom *et al.*, 1999) were predicted at the server of the Centre for Biological Sequence Analysis, BioCentrum-DTU Technical University of Denmark (http://www.cbs.dtu.dk/services/). The domain graph software (http://dog.biocuckoo.org) was used for the illustration of the domains of each protein.

Phylogenetic analysis of the amino acidic sequence was performed by ClustalW alignment (Thompson *et al.*, 1994) and by the Neighbor-Joining phylogenetic tree and the TreeTop (http://www.genebee.msu.su/services/phtree_reduced.html) as described by Costa *et al.*, 2010. BoxShade 3.21 was used for the graphical representations (http://www.ch.embnet.org/software/BOX_form.html). To spot the best local alignment among sequences we used BLAST analysis (Altschul *et al.*, 1990) and the LALIGN program (http://www.ch.embnet.org/software/LALIGN_form.html).

Comparative Real Time qPCR (ΔCCt)

Gene expression was measured by comparative q-PCR with the Comparative Threshold Cycle Method and SYBR Green chemistry (Livak and Schmittgen 2001) with the Applied Biosystems Step One Plus real time PCR cycler, following the manufacturer's instructions. cDNAs were synthesised using as templates total RNAs extracted from different stages (cleavage, blastula, gastrula, pluteus) and random hexamers, as described in the MultiScribe Reverse Transcriptase protocol (Applied Biosystems). Q-PCR primers were designed to amplify fragment of 80-100 bp, using the Primer Express software (v2.0.0; Applied Biosystems, Foster City, CA, USA). The primers are outlined in the corresponding chapters. The Pl-Z12-1 mRNA was used as an internal endogenous reference gene

(Costa *et al.*, 2012). q-PCR samples were performed in triplicates using the cDNA from blastula stage set as 1 (relative quantity value) as reference sample.

Whole-mount *in situ* hybridization (WMISH)

Whole-mount *in situ* hybridization was performed as described by Kiyomoto *et al.*, 2007. Digoxygenin (DIG) antisense probes were synthesized by asymmetric PCR using digoxygenin-labelled 2´-Deoxyuridine, 5'-Triphosphates (DIG-dUTP). Constructs built on the *pGEM-T-Easy* vector were used as templates for each amplification. Specific primers were designed from the amplification of each CDS. Alternatively, the T7 or Sp6 promoters, included on the backbone of the vector, were used for the asymmetric PCR. All probes were tested by Dot Blot on a Hybond filter prior to WMISH experiment to determine their specific activity.

All the pre-hybridization and hybridization steps were carried out in 96-well plates (Greiner Labortechnik, Longwood, FL, USA), using 30–40 embryos per well. The hybridization reactions were carried out in hybridization buffer containing 1% of each anti-sense DIG labelled probe, at 62°C for at least 20 h. After hybridization, specimens were extensively washed. DIG-labelled probes were incubated with an anti-DIG alkaline phosphatase-conjugated antibody (Roche) and immunodetection was carried out at room temperature with the chromogenic BCIP/NBT substrates. Stained embryos were mounted on glass slides and observed under a Zeiss Axioscope 2 plus microscope. Images were captured by digital camera. Negative control hybridization reactions with sense probes did not show any specific signal.

Production of recombinant *P. lividus* proteins in E. *coli*

Construction of recombinant plasmid for bacterial expression

The full length CDS of each gene, was amplified from the *pGEM-T-Easy* plasmids by PCR using suitable primers (see correpsonding chapters). Various expression vectors were used during the trial experiments of protein expression. Notably, *CA, galectin-8, P16, P19, advillin, tetraspanin, SM30a* and *SM50* were cloned in p*EXP-NT* (Invitrogen) and in p*TRC-CT* (Invitrogen). Additionally, CA was cloned in p*COLD-TF* (Takara) and in p*ET32b*$^{+}$ (Novagen); and galectin-8 in p*COLD-TF* (Table 1). p*EXP-NT*, p*TRC-CT* and p*ET32b*$^{+}$ contain the strong T7 promoter and a fusion His6 tag, while p*COLD-TF* contains the *csp*A (cold shock protein A) promoter, the N-terminal fused chaperone trigger factor of 48 kDa facilitating the expression of folded proteins and a fusion His6 tag. E. *coli* TOP10 (Invitrogen) was used as the host for gene cloning, and transformed cells were grown in Luria–Bertani (LB) medium with 100 µg/mL ampicillin selection. The cloned sequences were confirmed by direct sequencing at MWG (Heidelberg, Germany). Information on the primers used for each cloning are given individually in each respective chapter.

Table 1. Summary of the expression vectors used in the protein expression experiments. All vectors used express a fusion His6 tag.

	expression vectors	fusion tag
1	p*EXP5-NT*	N-terminal His6
2	p*TrcHis2-CT*	C-terminal His6
3	p*COLD-TF*	N-terminal His6 and Trigger factor chaperone
4	p*ET32b*$^{+}$	N' and C-terminal His6 and thioredoxin

Recombinant protein expression and purification

This project focused on the expression of a subset of the recombinant plasmids produced, namely two recombinant proteins, rPl-CA and rPl-Galectin-8-like, were produced. Trial expression experiments of Pl-P16, Pl-P19, Pl-SM30 and Pl-SM50 were also carried out (see Chapter III). A functional recombinant *Pl-CA* carrying an N-terminal and a C-terminal 6-His tags was expressed in p*ET32-Pl-CA*(BL.21(DE3)) while *Pl-galectin-8* was expressed as a fusion protein using the p*COLD-TF* (TAKARA) expression vector, in BL.21.A1 cells. The expression and purification procedures are explained in detail in Chapters IV (Pl-CA) and V (*Pl-galectin-8*).

Production of polyclonal antibodies in mice

Polyclonal antibodies (pAbs) were raised against the purified, recombinant carbonic anhydrase and the recombinant galectin-8 (*rPl-CA* and *rPl-Gal8*) proteins, in mice. About 1.5 mg of recombinant protein per injection was dissolved in phosphate-buffered saline (PBS). After three boosts, the serum was collected; the pAb against carbonic anhydrase was termed pAb-CA and against galectin-8, pAb-Gal8. The cross-reactivity was tested by ELISA against 10 µg/ml of each recombinant protein. The titer of the pAb-CA was 1:3000 and for the pAb-Gal8, 1:4000.

Immunoblotting (Western Blot Analysis)

Total cell lysates (30 µg) from different staged embryos were separated by electrophoresis on 10% SDS-PAGE gels and transferred to nitrocellulose membranes as described previously (Pinsino *et al.*, 2010). After blocking for 1 hour in BSA 1% in TBST, membranes were incubated overnight at 4°C, with the primary antibody in TBST (dilution 1-4,000). After extensive washing with TBST, membranes were incubated for 1 hour at RT with a 1:5,000 diluted in TBST of an alkaline phosphatase-conjugated anti-mouse IgG (SIGMA). Protein bands were visualized on Hyperfilm-ECL films using the ECL PLUS Western blotting Detection Reagents (Amersham).

Chapter III: Molecular studies on Biomineralization genes and proteins from *Paracentrotus lividus* embryo (P16, P19, ADVILLIN, TETRASPANIN, SM30α and SM50).

Abstract

The skeleton of the sea urchin is composed of magnesium calcite, a crystalline form of calcium carbonate containing small amounts of magnesium and occluded spicule matrix proteins (Decker and Lennarz, 1988; Wilt, 2002; Wilt *et al.*, 2003). Skeletogenesis begins in the early embryo development by the formation of a calcitic endoskeleton. Calcite is a stable crystalline form of calcium carbonate. In the sea urchin, bio-calcite involves small amounts of magnesium carbonate (5%) and organic material (0.1%), offering unique physical and chemical properties (Killian and Wilt, 1996; Wilt, 2002; Mann *et al.*, 2010). The primary mesenchyme cells (PMCs) are responsible for the deposition of the biomineral into the blastocoel cavity of the developing embryo. Several PMC-specific genes have been isolated and studied in various sea urchin species (Zhu *et al.*, 2001; Illies *et al.*, 2002, Livingston *et al.*, 2006).

In this chapter, the identification, isolation and characterization of five PMC-specific cDNA sequences: *p16, p19, advillin, sm30a, sm50* and of one cDNA of *tetraspanin* which is expressed in the ectoderm of the *P. lividus* embryo, are presented. The full-length coding sequences (CDSs) were identified by EST data mining and cloning. For the identification of the 3' end CDS of *Pl*-sm50, 3'RACE PCR was applied. Sequences were deposited in the EMBL database. Additionally, the tempo-spatial expression profile of each gene during the *P. lividus* early embryo development was monitored and correlated with skeletogenesis. Comparative studies with homologous sequences and expression patterns from *S. purpuratus* and *L. variegatus* (Illies *et al.*, 2002, Cheers and Ettensohn, 2005), revealed a high phylogenetic conservation of these sequences among the sea urchin species; with an interesting extent of diversity in the spatial expression profiles. Furthermore, the full-length CDSs were cloned in expression vectors for the preparation of recombinant proteins and specific antibodies.

This work provides with novel information on the characterization of six biomineralization protein-coding genes from *P. lividus.* A molecular tool-set including expression plasmids and labeling probes was constructed for the requirements of the study and for ongoing functional studies.

Introduction

The emerging interest for biomineral formation in biomedicine and materials science has attracted the attention of scientists towards the Echinodermata, the only (besides vertebrates) phylum among deuterostomes possessing an extensive biomineralized endoskeleton.

Paracentrotus lividus, being a well studied model for both Developmental and Ecotoxicological studies, offers various advantages as a system for biomineralization studies. Skeletogenesis in the sea urchin embryo involves regulated gene expression, mediated from the primary mesenchyme cells (PMCs), under a signaling control of transcription and growth factors synthesized by the neighboring ectodermic cells. A panel of PMC-specific genes has been identified from an arrayed PMCs' library from *S. purpuratus* (Zhu *et al.*, 2001; Illies *et al.*, 2002;) as well as from the complete *S. purpuratus* Genome Project (Sea Urchin Genome Consortium, 2006). A genome-wide search for 'biomineralization-related' proteins has been described (Livingston *et al.*, 2006) and, more recently, 231 proteins have been identified in the spicule matrix proteome extracted from mineralized parts of adult *S. purpuratus,* by Mass Spectrometry (MS) (Mann *et al.*, 2010). Some of them were previously identified in the spicule matrix of the embryo (Mann *et al.*, 2008). Studies in *S. purpuratus* and *L. variegatus* have indicated that *p16, p19, sm30a, sm50* and *advillin* are PMCs-specific genes which take part in the embryonic skeletogenesis (Illies *et al.*, 2002; Cheers and Ettensohn, 2005, Love *et al.*, 2007.)

It is surprising that, despite many developmental studies and the discovery of a high number of proteins reported to be involved in the sea urchin embryonic skeleton construction (Zhu *et al.*, 2001; Illies *et al.*, 2002; Sea Urchin Genome Consortium 2006; Livingston *et al.*, 2006; Mann *et al.*, 2010), the molecular/biochemical mechanisms associated with the biomineral formation and patterning remain relatively unknown. In this respect, of great importance are acidic proteins which seem to be effective crystal-modulators. Studies on crystal growth using acidic protein mixtures isolated from invertebrate biominerals suggest that these proteins are involved in the stabilization of the amorphous calcium carbonate, shaping the crystal orientation and morphology (Addadi *et al.*, 1989; Fu *et al.*, 2005). In accordance, former developmental studies on *S. purpuratus* and *L.variegatus* have indicated that *p16* and *p19* genes, coding for two small PMCs-specific acidic proteins, play a fundamental role in the sea urchin embryo skeleton growth and patterning (Illies *et al.*, 2002; Cheers and Ettensohn, 2005).

Biomineralization Proteins

P16 is a small acidic transmembrane protein which is localized at the plasma membrane in the sea urchin (Illies *et al.*, 2002) or secreted extra-cellularly (Veis, 2009) , with high Ca^{2+}-binding properties as indicated by direct binding studies of immobilized P16 from *L.variegatus* bound to ^{45}Ca, on nitrocellulose (Veis, 2009). Alvares *et al.*, 2009) have also found P16 proteins in the adult sea urchin tooth (UTMP16). The fundamental role of P16 in

skeletal rods' elongation has been demonstrated in *S. purpuratus* and *L. variegatus* by gene expression knockdown experiments (Cheers and Ettensohn, 2005). Nevertheless, since the protein is transmembrane, it has not yet been clarified whether it functions as a receptor receiving signals required for skeletogenesis, or it plays a more direct role in the biomineral deposition, like acidic proteins from other calcareous-forming systems (Takeshi Takeuchi *et al.*, 2008). Orthologues in other species were not identified.

P19 is a small acidic intracellular, possibly nuclear, protein with poor Ca^{2+}-binding properties, involved in biomineralization (Veis *et al.*, 2009). Lacking a signal sequence and transmembrane domains, it could probably be localized in the cytoplasm of the PMCs (Illies *et al.*, 2002). P19 has also been identified in the adult tooth of *L. variegatus* and in the proteome of *S. purpuratus* tooth tissue and was characterized as a probably intracellular phosphoprotein. The phosphorylation sites however, were not determined. Recently, P19 was found within the occluded matrix proteins of the adult sea urchin test and intact tooth (Mann *et al.*, 2010).

SM30 and **SM50** are two of the most well known C-type lectins involved in biomineralization, firstly discovered in the sea urchin embryo (Akasaka *et al.*,1994, George *et al.*, 2004, Killian *et al.*, 2010). For reviews on SM30: George *et al.*, 1991, Kitajima *et al* 1996, Wilt, 1999, Yamasu and Wilt, 1999, and for SM50: Sucov *et al.*, 1987, Sucov *et al.*, 1988; Benson *et al.*, 1987, Katoh-Fukui *et al.*, 1991, Makabe *et al.*, 1995 and Urry *et al.*, 2000. In the sea urchin *S. purpuratus*, there are six acidic glycoproteins of the **SM30** family, designated SpSM30A through SpSM30F (Livingston *et al.*, 2006). The SpSM30 proteins are found uniquely and abundant in embryonic and adult mineralized tissues of the sea urchin. SpSM30 proteins are occluded within the embryonic endoskeleton and adult mineralized tissues (Killian and Wilt,1996; Mann *et al.*, 2008; Urry *et al.*, 2000). SpSM30A, B, C, E and F are expressed exclusively in primary mesenchyme (PMC) cells and their descendants while SpSM30A mRNA expression is limited to the embryo. Conversely, SpSM30D mRNA is not expressed in the embryo, but is expressed in adult spines and teeth. SpSM30B, SpSM30C and SpSM30E are expressed in all mineralized adult tissues; while SpSM30F is expressed in adult tissues except the test. Relative levels of expression of the several family members in these different tissues vary widely. Authors conclude that SpSM30 proteins play a vital, but still unknown, role in biomineralization of these tissues during development (Killian *et al.*, 2010). Seto *et al.*, 2004, localized the expression of both SM30 and SM50 proteins in the adult mineralized parts using a gold particles-conjugated secondary antibody (Immunogold) and scanning and transmission electron microscopy (SEM and TEM). The **SM50** gene encodes a minor matrix protein of the sea urchin embryo spicule. SM50 protein is rich in Glu, Gln, Asp, Asn, Gly, Ser, and Ala amino acids, similar to the amino acid composition of other invertebrate skeletal organic matrix proteins (Weiner 1984). Western blotting analysis indicated that SM50 is the spicule matrix protein with the most alkaline isoelectric point (Killian *et al.*, 1995). A detailed regulatory functional domain analysis of a cis-regulatory region of this SM50 is given by Makabe *et al.*, 1995. C-type lectins are believed to guide the formation of the calcitic biomineral by acting as inhibitors of the phase transitions of the mineral.

Advillin was first isolated from mice and is localized at the dorsal root and trigeminal ganglia of the developing nervous system (Marks *et al.*, 1998). The villin / gelsolin family is composed of actin binding, capping, and severing proteins involved in cytoskeletal rearrangement. Advillin has a Ca^{2+}-regulated actin-binding function as demonstrated in filopodial retraction via actin severing and capping by Alan *et al.*, 2007). The expression profile in mice, suggests a retained role in the morphogenesis of adult structures within the rudiment before metamorphosis. Kim *et al.*, 2010, showed unique functions of advillin in morphogenesis of neural cells which form ganglia and outlined a role in ciliogenesis. Advillins contain a six tandem repeat structural domain (Kwiatkowski, 1999; McGough *et al.*, 2003). Our cDNA clone is identified as advillin rather than another member of the gelsolin family because it contains the distinguishing carboxy-terminus headpiece domain responsible for actin bundling found at the 3' end of the protein-coding region (Friederich *et al.*, 1999; Bartles, 2000; Vermeulen *et al.*, 2004). A putative orthologue to gelsolin has been described in the developing gut of sea urchins as a downstream target of *brachyury* (Rast *et al.*, 2002), pointing towards the concept that some gelsolin family members may coordinate the cytoskeletal actin rearrangement in distinct embryonic territories, via F-actin bundling and turnover in the filopodia of PMCs (Miller *et al.*, 1995). Advillin orthologues are also known to be involved in skeletogenesis of other sea urchin species (Love *et al.*, 2007). We focused on an advillin isoform specifically expressed in the PMCs of *P. lividus*.

Tetraspanins constitute a large super-family of membrane-spanning proteins involved in interactions regulating cell adhesion and mobility (Bronstein, 2000; Yanez-Mo *et al.*, 2000; Bouchiex and Rubinstein, 2001; Hemler, 2001, 2003; Tarrant *et al.*, 2003). They associate with other membrane proteins (Hemler 1998; Berditchevski 2001) to organize microdomains for signaling interactions and growth factor reception on the cell membrane surface (Hemler, 2003; Yunta and Lazo, 2003). In the sea urchin, tetraspanin is hypothesized to be involved in cell–cell interactions with the filopodia of the PMCs, either directly or indirectly through the recruitment of proteins (*e.g.* integrins) into microdomains that interact with the ECM via growth factors (Schubert, 1992; Hemler, 1998; Berditchevski ,2001; Baron *et al.*, 2003). These interactions could be recognized by PMCs through filopodial contacts (Miller *et al.*, 1995; Hodor *et al.*, 2000). We focused on a tetraspanin from *P. lividus*, orthologue of *Ht-tetraspanin (sup1)* which is expressed exclusively in ectoderm and the potential signaling with the PMCs. Apart from this, there are at least four other members of the Tetraspanin family expressed in the PMCs and the endoderm (Zhu *et al.*, 2001; Ransick *et al.*, 2002).

Skeletogenesis as a developmental process, involves the coordinated expression of genes. In the sea urchin embryo, Love *et al.*, 2007 concluded that the coordinated gene expression of advillin, tetraspanin and CA in *S. purpuratus*, evolved as part of the evolution of pluteus arms and is not required for larval or adult development (Fig.6).

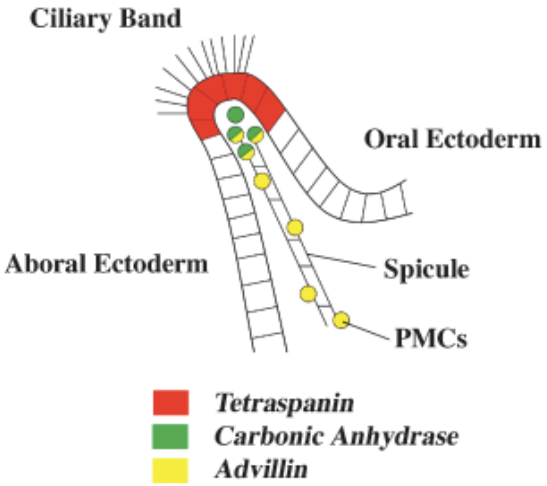


Figure 6. Diagram of the pluteus larval arm-tip as a distinct organ. Representation of a sectioned pluteus arm showing relevant ectoderm (aboral, ciliary band, and oral) and mesoderm (primary mesenchyme cell [PMCs]) domains with coordinated gene expression patterns. Tetraspanin (colored red) is expressed in the ectoderm of the arm-tip with carbonic anhydrase (colored green) concentrated in the PMCs immediately beneath. Advillin (colored yellow) is expressed in PMCs both within the arm-tip and along the forming spicule (Image from Love *et al.*, 2007).

Experimental Procedures

The experimental procedure for the embryo culture, total RNA extraction and RT-PCR is described in Chapter II.

Cloning of *Pl-p16, Pl-p19, Pl-advillin, Pl-tetraspanin, Pl-sm30α and Pl-sm50*

Each cDNA sequence, previously annotated in the sea urchin *S. purpuratus* (Table 2), was used for BLAST screening against various *P. lividus* EST databases including libraries acquired from blastula, gastrula and pluteus developmental stages (NCBI and MPIMG (http://goblet.molgen.mpg.de/cgi-bin/webapps/paracentrotus.cgi). Various *P. lividus* EST clones, similar to each cDNA were identified. The clones exhibiting the highest homology, are outlined in Table 2. The overlapping EST clones were merged and specific primers were designed complementary to the obtained sequences (Table 2) and purchased from MWG (Heidelberg, Germany). Each sequence was amplified by RT-PCR (One Step RT-PCR, Invitrogen), using purified total RNA extracted from the gastrula stage. PCR products were cloned in the p*GEM-T-Easy* Vector (Promega) and sequenced by MWG (Heidelberg, Germany). The obtained sequences of *Pl-p16, Pl-p19, Pl-advillin, Pl-tetraspanin, Pl-sm30α and Pl-sm50* were identical to *P. lividus* EST clones found in the database and were submitted cDNAs to Genebank (Table 2).

In the case of *Pl-sm50*, as the EST data mining study revealed only a partial 5' CDS coding sequence, 3'RACE PCR was applied using a 3'RACE kit (Invitrogen) for the

identification of the carboxylic end of the SM50 CDS. Total RNA was extracted from embryos at the gastrula stage and the first amplification was performed with the forward primer: 5'-GTCACTGAGATGCGCGCTTTC-3'. A nested amplification, using the internal forward primer: 5'-CAACCCAGGATTTGGTGGTCAG-3', amplified three cDNA fragments corresponding to three bands. The whole amplification reaction was cloned and the longest fragment was screened and selected by test PCR and sequenced at MWG (Heidelberg, Germany). The full-length coding sequence was re-amplified by RT-PCR with the primers: 5'-GTAACCATGAAGGGAGTTTTGCT-3', as forward and 5'-GTATGTTGCAGCATACGTGGTC-3', as reverse, giving an amplicon of 933 bp.

Table 2: A: EST clones of the biomineralization genes *Pl-p16, Pl-p19, Pl-advillin, Pl-tetraspanin, Pl-sm30α and Pl-sm50,* exhibiting high similarity with the annotated homologues from *S. purpuratus*. B: Primers designed on compiled *P. lividus* EST sequences.

	A	**B**		
	***P. lividus* EST clones of high homology with *S. purpuratus* annotated sequences (length)**	**Primer sequences for cDNA cloning on p*GEM-T-Easy***		
		forward 5'-3'	**reverse 5'-3'**	**PCR product length (bp)**
p16	MPMGp1173L0534Q (844 bp) [NCBI: NM_214646.1]	ATGAAGACTTTTGTTGC CCTCTTGTC	GAAGTCGTCTCAGCAC CTTTGG	633
p19	MPMGp1172J2299Q (799 bp) [NM_214647.1]	ATGACCAAGGAAGAGG CTGC	GTGGTCCCATTACCAT TTCACTG	515
advillin	NM_001114197	ATGTCCAAGGTTGACG CAGCATTC	TGGGATCGGAACACAA ACCCTCT	2685
tetraspanin	NP_001118229.1	ATGGGTGTGGAACTAG GAGGTTG	ACTGAACGATTCAAGG GGAAGAGGA	970
sm30a	Q26646.1	TCCCGAAAAGATGAGG AGTTTCGTGT	CGTTCAATGCACAATG GACGGACG	1034
sm50	NP_999775.1	GCGCGTACCCTACGGC ATGG	TGCCTGGTTGTTGCG GACCC	638

Domain characterisation and phylogenetic analysis

The putative aa sequences were characterized as described in Chapter II.

Comparative Real Time qPCR and Whole-mount *in situ* hybridization (WMISH)

Gene expression profiling was monitored throughout embryo development by comparative q-PCR and whole mount *in situ* hybridization, and as described in Chapter II. For the q-PCR, the primers used are outlined in Table 3.

Table 3: Primers used for the comparative qPCR of *Pl-p16, Pl-p19, Pl-advillin, Pl-tetraspanin, Pl-sm30α and Pl-sm50*

	qPCR primers		
cDNA	**forward 5'-3'**	**reverse 5'-3'**	**length (bp)**
Pl-p16	CGGGCAGCGATGACTCA	AAATGCCATACCGCTCTTCTGT	104
Pl-p19	CAGGCAGGAGACTAAGACAGAGACT	CTCCGCTCGCCTCTCCTT	86
Pl-advillin	CGCCAGCAGTCTCCATAAGTT	CATCATACATGCCCGTTCTTGA	80
Pl-tetraspanin	CCCTGCTACGTGTTGTGTGTTC	CCCATACAGGATGAGATGTTGACT	90
Pl-sm30α	GATCGACATGCTATCATCTGTGAGT	GGCGCAACTGTAGCACTGGTA	71
Pl-sm50	CCGTGAACGCACAAAATCC	GGGCCTGACGCTTCATGA	66

Whole-mount *in situ* hybridization was performed as previously described in Chapter II. Table 4 summarizes the primers of the asymmetric PCR (for *p16* and *p19*) and the restriction enzymes used for the linearization of *Pl-adv* and *Pl-tetraspanin* plasmids as well as the vector primer (either *T7* or *Sp6*); used for each amplification. Information on the insert direction of *pGEM-Pl-p16* and *pGEM-Pl-p19* plasmids are noted. The spatial localization of the expression of *sm30a* and *sm50* was not performed as these data are already well documented in the literature.

Table 4: Primers and plasmids used for the preparation of WMISH probes of *Pl-p16, Pl-p19, Pl-advillin* and *Pl-tetraspanin.* *The DNA probes p16 and p19 were prepared by asymmetric PCR; thus their length varies. The mentioned lengths correspond to the longest PCR product.

	WMISH probe synthesis				
cDNA	**forward primer 5'-3'**	**reverse primer 5'-3'**	**probe length (bp)**	**p*GEM-T-Easy* promoter antisense to the CDS**	**Restriction site for linearization**
Pl-p16	*AGCTTGTCTTCAA GCCATCTCGCC*	*GCAACTGCTACGA CGGCCCC*	548*	*Sp6*	*Sac*II
Pl-p19	*ATGACCAAGGAA GAGGCTGC*	*GTGGTCCCATTACC ATTTCACTG*	515*	*Sp6*	*Sac*II
Pl-advillin			2685	*T7*	*Sal*I
Pl-tetraspanin			970	*T7*	*Sal*I

Construction of expression plasmids

In view of the functional characterization of each gene, the full-length CDSs were cloned in expression vectors, suitable for the production of recombinant proteins in E. *coli*. The expression vectors p*EXP-NT* (Invitrogen) and p*TRC-CT* (Invitrogen) were used. Each cloning vector required different reverse primers but the same forward primer could be used, as outlined in Table 5. p*EXP-NT* vector generated a fusion protein introducing a tag consisted of six Histidines (His_6 tag) at the N-terminal of the recombinant protein; therefore the insert was amplified including a stop codon. On the contrary, p*TRC-CT* vector fuses a His_6 tag at the C-terminal of the recombinant protein; therefore the insert was lacking a stop codon. The full-length coding sequence of *Pl-advillin* was amplified by PCR with Taq polymerase (Invitrogen) from the *pGEM-T-Easy* constructs, while the full-length CDSs of *Pl-p16, Pl-p19, Pl-sm30a, Pl-sm50* and *Pl-tetraspanin,* were amplified *de novo* from total RNA extracted from embryos at the gastrula and pluteus stages by RT-PCR with the One Step Rt-PCR kit (Invitrogen). The generated inserts were ligated using the TA cloning strategy, into each expression vector and the cloned sequences were confirmed by direct sequencing from MWG (Heidelberg, Germany).

Table 5. Primers designed for the *in-frame* cloning of the full-length CDSs of *Pl-p16, Pl-p19, Pl-advillin, Pl-tetraspanin, Pl-sm30a* and *Pl-sm50* in the p*EXP-NT* and p*TRC-CT* expression vectors.

	Primers used for cloning in the expression vectors p*EXP-NT* and p*TRC-CT*		
cDNA	**Forward primer for cloning in expression vectors (5'-3')**	**Reverse primer for cloning in p*EXP-NT* vector (5'-3')**	**Reverse primer for cloning in p*TRC-CT* vector (5'-3')**
Pl-p16	ATGAAGACTTTTGTTGCCCTCTTGTC	GAAGTCGTCTCAGCACCTTTGG	CGCGTTTTGAAGCATCTGGGC
Pl-p19	ATGACCAAGGAAGAGGCTGC	GTGGTCCCATTACCATTTCACTG	CTGACCTTCGTCGGGTAC
Pl-advillin	ATGTCCAAGGTTGACGCAGCATTC	TGGGATCGGAACACAAACCCTCT	CTTCTTCAGGTTATCCTGTTTC CAC
Pl-tetraspanin	ATGGGTGTGGAACTAGGAGGTTG	ACTGAACGATTCAAGGGGAAGAGG A	GACAACATCTTCTCCCTTTGA GATGC
Pl-sm30α	ATGAGGAGTTTCGTGTGTGTTTTGG	CGTTCAATGCACAATGGACGGACG	CATGACGCCATAGTATCGGTT CATC
Pl-sm50	GTAACCATGAAGGGAGTTTTGCT	GTATGTTGCAGCATACGTGGTC	CCCGCTCCTGGTTGTTG

Results

Identification and cloning

Each CDS was amplified by RT-PCR from total RNA extracts from the gastrula primed by the oligos outlined in Table 2 (Fig. 7i-A). Each cDNA was cloned in the p*GEM-T-Easy* Vector. The ligation of each construct was confirmed by PCR before sequencing (Fig.7i-B).

The molecular sizes of the amplicants, were: for *p16*, 543 bp; for *p19*, 515 bp; for *sm30a*, 1034 bp; for *sm50*, 638 bp; for *advillin*, 2685 bp and for *tetraspanin*, 970 bp. The full-length CDSs of *p16, p19, sm30a, advillin* and *tetraspanin* were retrieved by EST data mining. The EST data mining revealed only a partial sequence of Pl-SM50. For the identification of the carboxylic end of the SM50 CDS, 3'RACE PCR was applied. Fig.7ii summarizes the 3'RACE procedure. By RT-PCR on total RNA extracts from gastrula and nested amplifications, we amplified three bands (Fig.7ii-B-C). The longest cDNA fragment was isolated by cloning in the p*GEM-T-Easy* vector. The plasmids from the positive colonies (Fig.7ii-B-D, lanes 2 and 3) were identified by PCR and sequenced. A reverse primer was designed on the basis of the sequence derived from the 3'RACE PCR for the amplification of the full-length *Pl-sm50* cDNA comprising 933 nucleotides, containing the entire coding region (1 - 891 bp) of *Pl-sm50*.

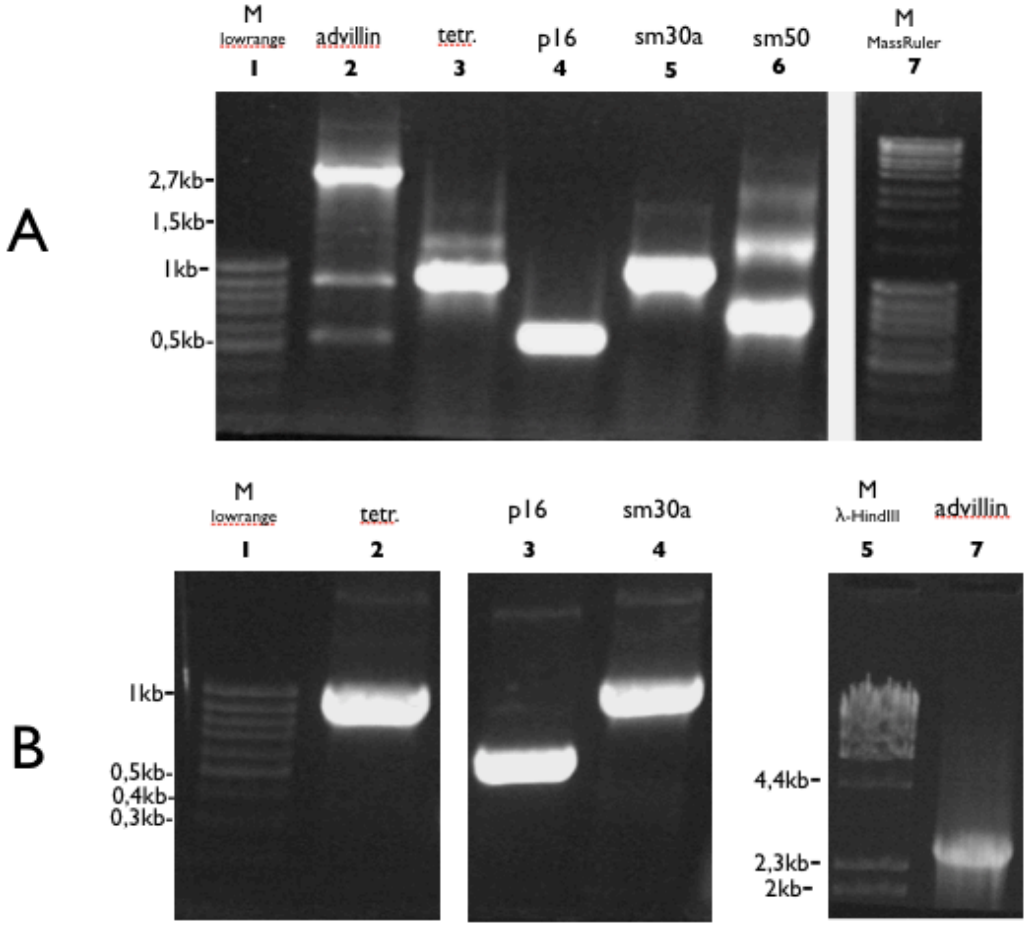


Figure 7i: Cloning of *Pl-advillin, Pl-tetraspanin, Pl-P16, Pl-sm30a* and *Pl-sm50*. A) Amplification of the EST CDS sequences, by RT-PCR from gastrula using the primers of Table 2. B) PCR amplification of each CDS templated by the previously cloned p*GEM-T-Easy* constructs.

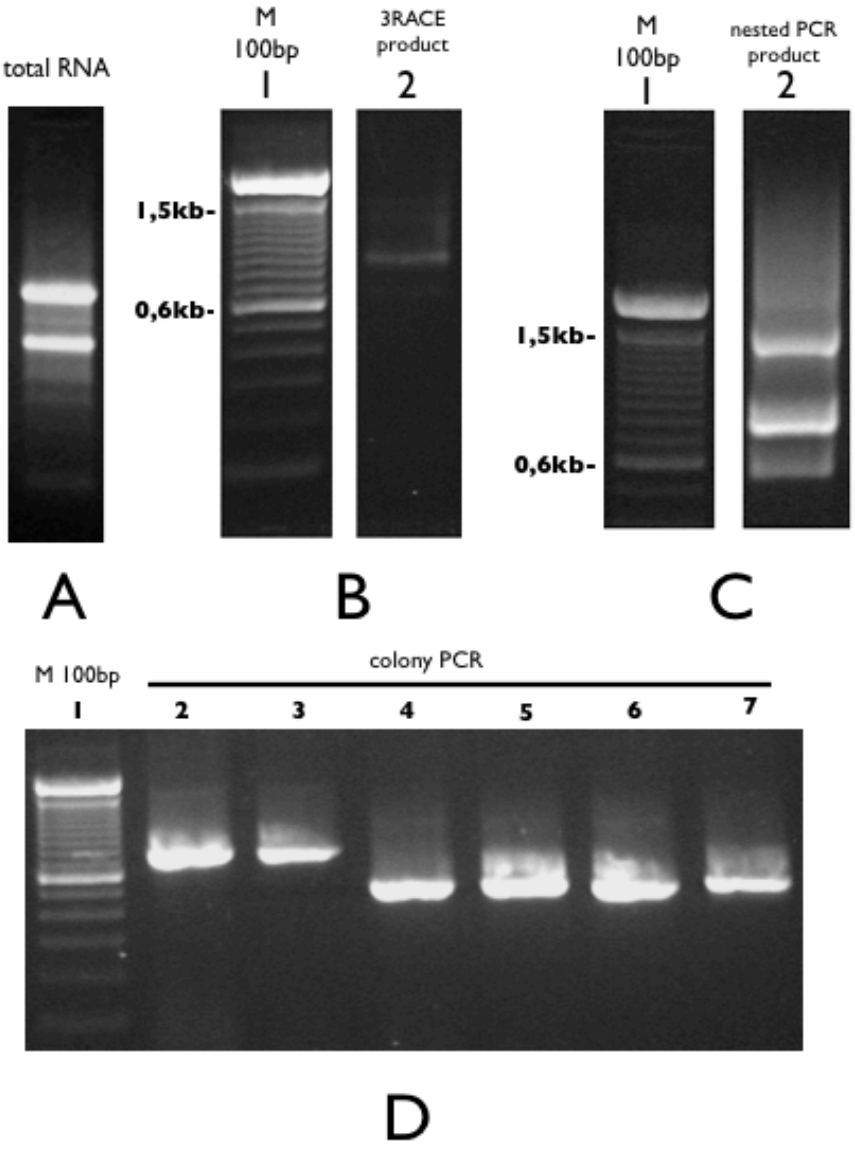


Figure 7ii: 3'RACE PCR for the identification of the carboxylic end of *PI-SM50*. A) Total RNA extract from gastrula stage; B) First amplification by 3'RACE; C) Amplification by nested PCR; D) Screening of positive colonies. Lanes 2 and 3 involve a longer positive sequence, while lanes 4-7 a shorter one.

Table 6: Identification and cloning of cDNAs from *Paracentrotus lividus*. Lane 1,2: Summary of sequence lengths of cDNAs retrieved from EST database and cloned in *P. lividus*. Lanes 3, 4, 5: Length and molecular characteristics of the putative proteins.

	Identification and cloning of cDNAs from *Paracentrotus lividus*					
cDNA	Identified sequence length (bp) in *P. lividus*	CDS length (bp) of database sequence in *P. lividus*	AA sequence length of Putative protein in *P. lividus*	MW (kDa) Putative protein in *P. lividus*	p*I* Putative protein in *P. lividus*	accession number of Deposited *P. lividus* sequences in EMBL
p16	633	543	181	17	3.69	FR693763
p19	515	495	165	18	4.84	FR693764
advillin	2685	2496	832	93	8.62	FR693766
tetraspanin	970	837	279	30	3.91	FR693765
sm30-a	1024	936	312	34	5.96	FR716470
sm50	933	891	297	31	9.10	

Domain homology and phylogenetic analysis

The deduced amino acid sequences were analysed *in silico.* The functional domains were predicted and the homology with homologous proteins from other species was documented by multiple alignments and a phylogenetic analysis. More details are given below.

Domain homology and phylogenetic analysis of P16

We amplified and cloned a p16 cDNA fragment of 633 bp which includes the complete CDS (543 bp) and a 3'UTR (90 bp). We deposited after sequencing the sequence in the EMBL database with the accession number FR693763.

The Pl-P16 deduced protein sequence is 181 aa long, with a predicted pI of 3.69 and an estimated MW of 17.1 kDa. The protein is rich in glycine (44/181 aa) and three other hydrophilic amino acids, including serine (15/181 aa), asparagine (12/181 aa) and threonine (11/181 aa). Four potential phosphorylation sites are predicted for serine (aa: 86,104,110 and 119) and two for threonine (aa: 95 and 114). In addition, the presence of 17 aspartic acid residues confer a total acidic nature to the protein. Pl-P16 contains a putative signal peptide (1-16), a transmembrane domain (aa: 148-168) and two 20 aa long repeats (AMGGAGAVGDTSTGGVDAMG), starting at positions 61and 82. The region between the signal sequence and the transmembrane domain contains three tandem copies of an imperfect GSDD(D/S) repeat (aa: 103-107, 109-113, 115-119). A potential N-glycosylation site is found at 108 aa position. The characterized domains are illustrated in Fig.8

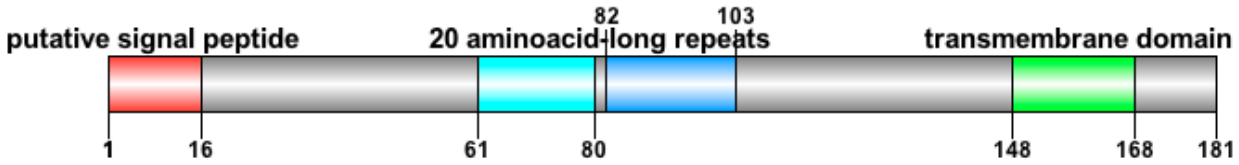


Figure 8. Illustration of the functional domains of *Pl-P16* protein

Results from the sequence alignment with *S. purpuratus* and *L. variegatus* orthologues showed very high conservation in the amino acidic composition, reaching 77% of similarity with *Sp-P16* and 71% with *Lv-P16*, despite the differences in the total length of the orthologues (*Sp-P16*, 172 aa while *Lv-Pp16*, 161 aa). A BLAST search for additional *Pl-P16* orthologues from other organisms, ranging from bacteria to mammals, showed some similarity to a few sequences reported in the databases (Table 7).

Table 7. Pl-P16 protein primary sequence similarities with proteins from distant phyla.

Species	***Uniprot Acc. N.***	***Length aa***	***Identity %***	***Similarity %***
Paracentrotus lividus	FR693763	181	100	100
Mycobacterium marinum	B2HG27	560	42	48
Crassostrea nippona[a]	B3ITC3	516	34	51
Sporisorium reilianum	E6ZYJ2	601	33	48
Mus musculus CD99L2	Q8BIF0	214	19.4	32

[a]predicted sequence

The phylogenetic analysis demonstrated longer distance relations between P16 from *P. lividus* and *L. variegatus* rather to *S. purpuratus*. Sequence analysis by BLAST indicated that *P16* is a protein found in sea urchin species. A more careful search, revealed protein sequences with low sequence similarity in various species. The *Mycobacterium marinum* (Accession number: B2HG27) showed 42% of identity and 48% of similarity. Lower similarities were identified on sequences from the oyster *Crassostrea nippona,* the fungus *Sporisorium reilianum* and *Mus musculus* (Bixel *et al.*, 2011). A multiple alignment P16 from different sea urchin species was prepared (Fig.9-A).

Figure 9-A. Multiple alignment of P16 from different sea urchin species. ClustalW alignment of the entire *Pl-p16* sequence with homologous proteins found in Echinidae. On the left, abbreviations of the species names and the total number of amino acids are presented. Protein GenBank numbers: Pl_181, FR693763; Sp_172, AM70485.1; Lv_161, AAY5933.1. Identical amino acids are shaded in black; conservative amino acids substitutions are shaded in gray; dashes correspond to gaps.

Domain homology and phylogenetic analysis of P19

We amplified and cloned the *Pl-p19* cDNA of 515 nucleotides, containing the entire coding region (1 - 495 bp) of *Pl-p19*. We deposited the *Pl-p19* sequence in the EMBL database with the accession number FR693764.

The deduced protein of Pl-p19, includes 165 aa with a predicted pI of 4,84 and a MW of 18.4 kDa. The protein is rich in glutamic acid residues (31/165 aa) which confers a total acidic charge; quite abundant are also four hydrophilic amino acids, namely threonine (14/165 aa), glycine (12/165 aa), glutamine (12/165 aa) and serine (11/165 aa). Five potential phosphorylation sites have been predicted for serine (aa: 51, 61, 77, 90 and 130) and nine for threonine (aa: 2, 11, 15, 19, 27, 62, 83 97 and 145). A potential N-glycosylation site is found at the 89 aa position. It is noteworthy that the Pl-p19 sequence has no Cysteine residues (*Sp-p19* has only one residue), thus excluding the possibility to form intrachain or interchain S-S bridges.

By database analysis at NCBI we identified another *P. lividus* EST sequence coding for a longer protein (179 aa), claiming for at least two Pl-P19 isoforms, in analogy with the two forms described for Sp-P19, including a short (166 aa) and a long (176 aa) form. The four proteins are highly conserved as their similarity is over 90%. As both the Pl-P19 and Sp-P19 shorter proteins differ from their longer forms because they lack a 14-10 aa stretch at their C-terminus parts, we suppose that they originate from an alternative splicing of their mRNA. Searching for Pl-P19 orthologues in other species we found a high similarity (91%) with the *L.variegatus* tooth matrix protein (UTMP19, Protein GeneBank ACU00092.1) that has been hypothesized to have different functions, including the mechanical strengthening of calcite plates due to the high magnesium content (Alvares *et al.*, 2009). A lower similarity (47%) has been found for the dentin matrix acidic phosphoprotein 1 (Dmp1, Accession Number: NP_058059; MacDougall *et al.*, 1998), a protein particularly expressed

in the mouse tooth odontoblast, ameloblast and cementoblast, but also in bone osteoblast (Qin *et al.*, 2007). Dmp1 is an extracellular matrix protein, a member of the small integrin binding ligand N-linked glycoprotein family. DMP1 protein, containing a large number of acidic domains, multiple phosphorylation sites, a functional arg-gly-asp cell attachment sequence, and a DNA binding domain, is critical for proper mineralization of bone. During osteoblast maturation the protein is exported, following phosphorylation to the extracellular matrix, where it regulates mineralized matrix formation. The similarity with Pl-P19 is restricted to the C-terminal region of DMP1, an acidic region which has been shown to be important for the biomineral formation and phosphate homeostasis regulation. A multiple alignment P19 from different sea urchin species was prepared (Fig.9-B). No particular functional domains were identified.

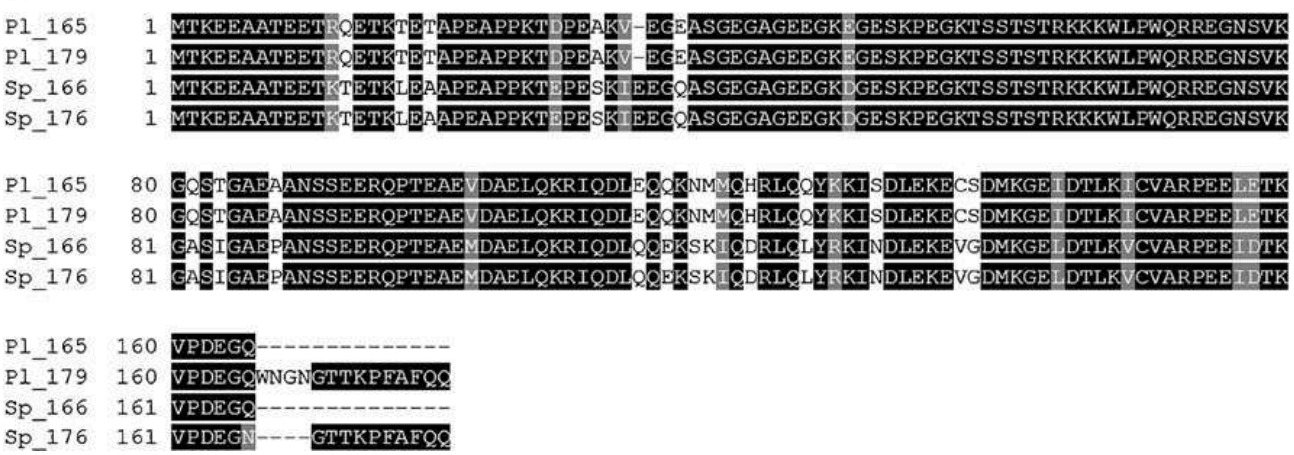


Figure 9-B. Multiple alignment of P19 from different sea urchin species. ClustalW alignment of the entire and *Pl-p19* sequence with homologous proteins found in Echinidae. On the left, abbreviations of the species names and the total number of amino acids are presented. Protein GenBank numbers: Pl_165, FR693764; Pl_179, AM196699; Sp_166, AAM70483.1; Sp_176, AAM70484.1. Identical amino acids are shaded in black; conservative amino acids substitutions are shaded in gray; dashes correspond to gaps.

Domain homology and phylogenetic analysis of Advillin

Based on EST data mining and using the primers outlined in Table 5, we amplified and cloned a cDNA sequence of 2685 bp, containing the full-length coding sequence of *Pl*-advillin, of 2496 bp. The sequence was deposited at the EMBL database with the accession number FR693766.

The protein contains a large extracellular domain (79 aa - 662 aa) and two cytoplasmic terminal regions. The deduced aa sequence, is composed of 831 residues with a predicted pI of 8.62 and a MW of 92.92 kDa. *In silico* analysis revealed the functional domains of the protein. Six gelsolin sub-domains were identified plus a carboxy-terminal villin headpiece domain (Fig.10). Various conserved acidic aa with a carboxy group in the side chain which potentially bind Ca^{2+}, were identified. Indeed, 31 aspartic acids (D) and 39 glutamic acids

(E) were identified. Additionally, 10 conserved cystein residues were identified throughout the structure. The topology of *Pl-advillin* was studied *in silico*. No signal peptide was identified (Fig.11).

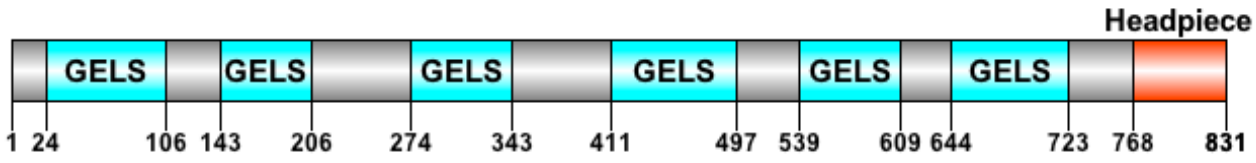


Figure 10. Illustration of protein domain structure of *Pl-advillin*. *Pl-advillin* involves six GELS (Gelsolin domains) and one headpiece domain.

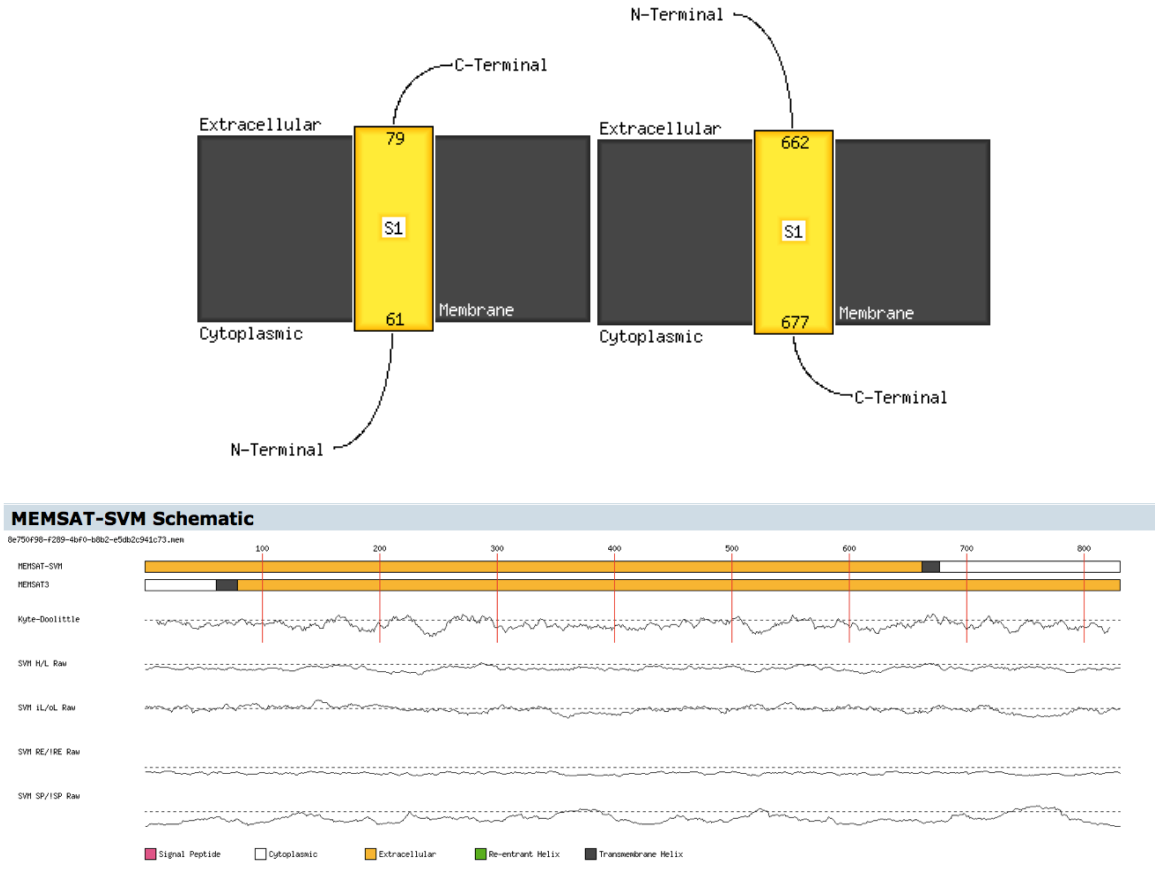
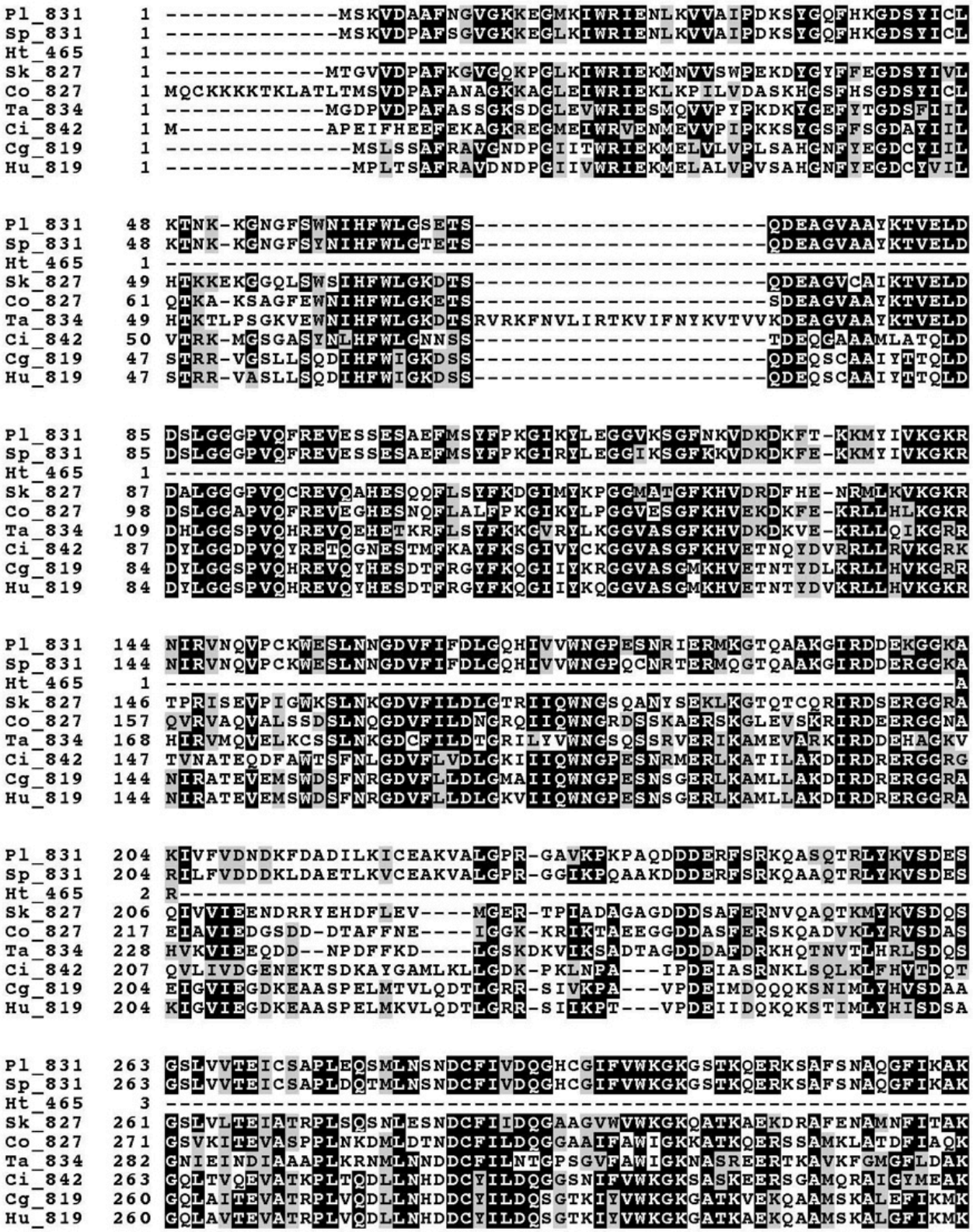
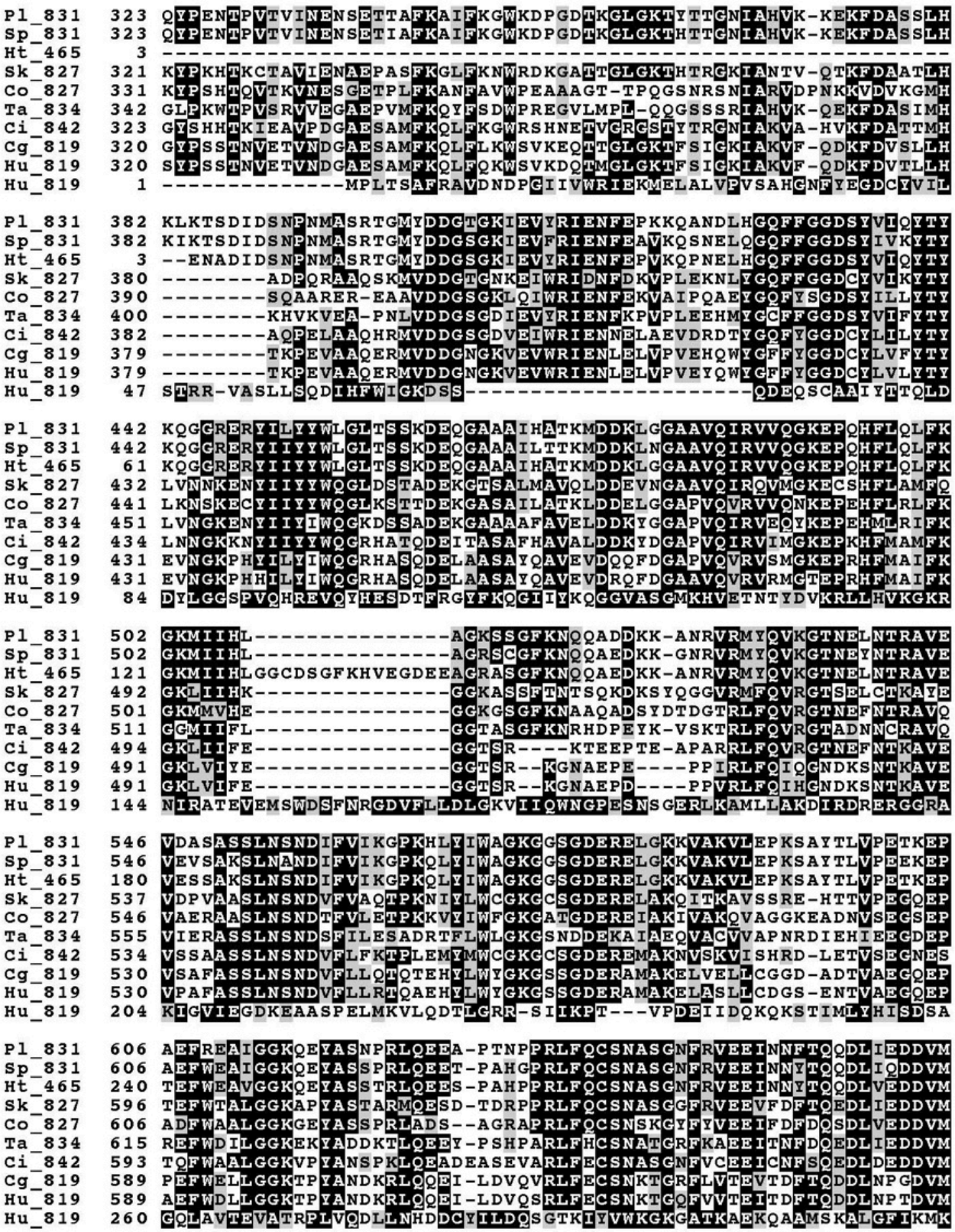


Figure 11. Illustration of the topology of *Pl*-advillin. The protein contains an extracellular domain (79 aa - 662 aa) and two cytoplasmic terminal regions. Analysis was performed using the method of McGuffin *et al.*, 2000 at the PSIPRED.

BLAST analysis identified two orthologues in other sea urchin species: an orthologue in *S. purpuratus* (Accession number: NP_001107669.1) of 831 aa in length, sharing a high sequence 90% identity and 96% similarity, and a partial orthologue from *H. tuberculata* (XP_002736967.1), of 465 aa in length, sharing 89% identity and 94% similarity. Additionally, the predicted advillin sequence from the hemichordate *Saccoglossus*

kowalevskii, belonging to the Hemichordata phylum (Accession number: XP_002736967.1), showed 53% identity and 71% similarity. *Pl-advillin* was also aligned with the 827 aa long villin protein from *Capsaspora owczarzaki* (EFW45931.1); a 834 aa long hypothetical protein from *Trichoplax adhaerens* (XP_002110432.1); a 842 aa long predicted advillin from *Ciona intestinalis* (XP_002125679.1); a predicted 819 aa long advillin from *Cricetulus griseus* (XP_003507565.1) and the 819 aa long advillin from human (NP_006567.3). The alignments are shown in (Fig.12). A phylogenetic tree was prepared demonstrating that advillin is well conserved in the sea urchins, especially compared to advillin from *S.purpuratus* (Fig.13); but also among various distant species such as *Ciona intestinalis* and human (Fig.13).





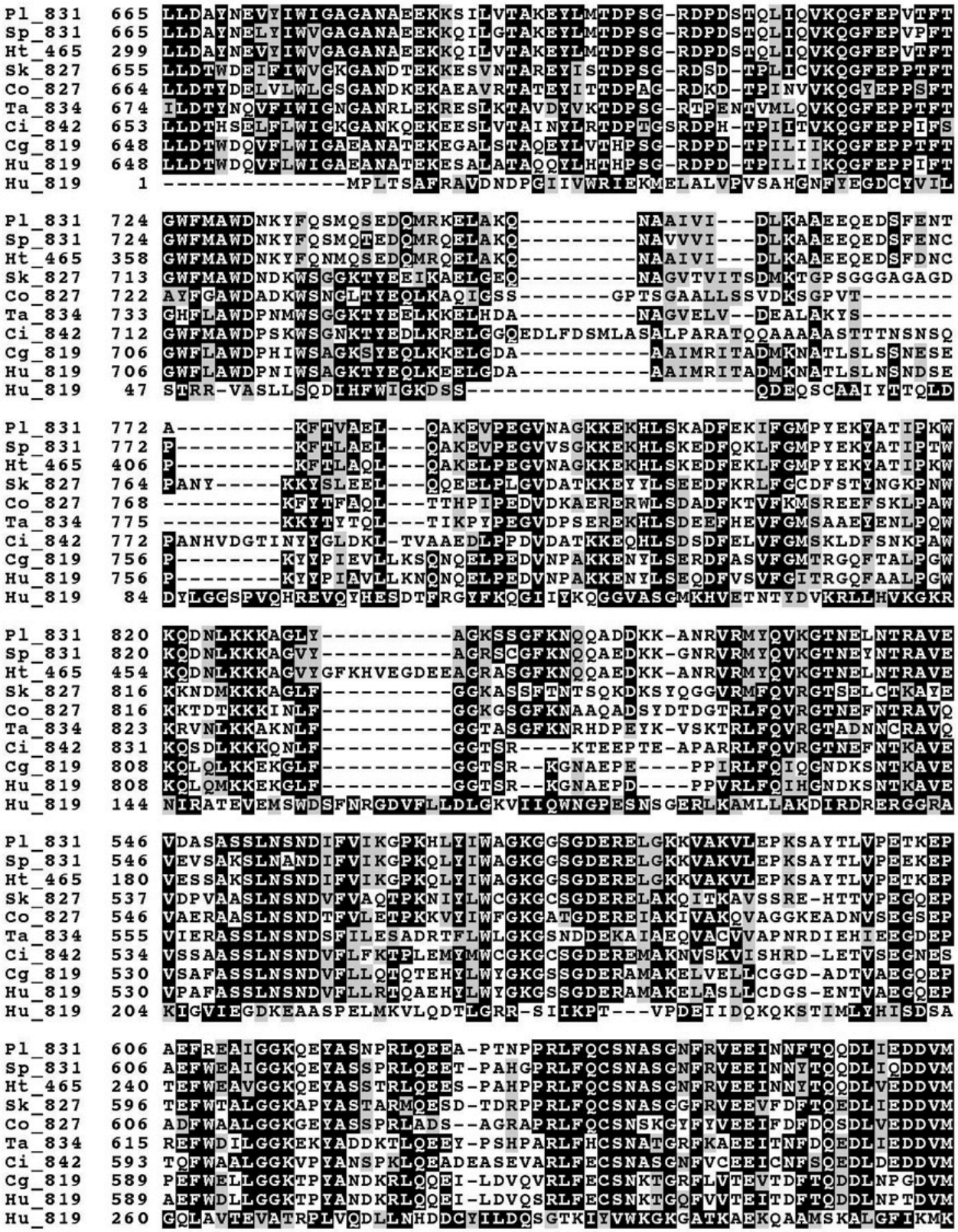


Figure 12. Multiple alignment of advillin from different species. ClustalW alignment of the entire CDS with homologous proteins. On the left, abbreviations of the species names and the total number of amino acids are presented. Identical amino acids are marked with asterisk; dashes correspond to gaps. Protein GenBank numbers: Pl_831, FR693766; Sp_831, NP_001107669.1; Ht_465, XP_002736967.1; Co_827, EFW45931.1; Ta_834, XP_002110432.1; Sk_827, XP_002736967.1; Cg_819, XP_003507565.1; Hu_819, NP_006567.3 ; Ci_842, XP_002125679.1.

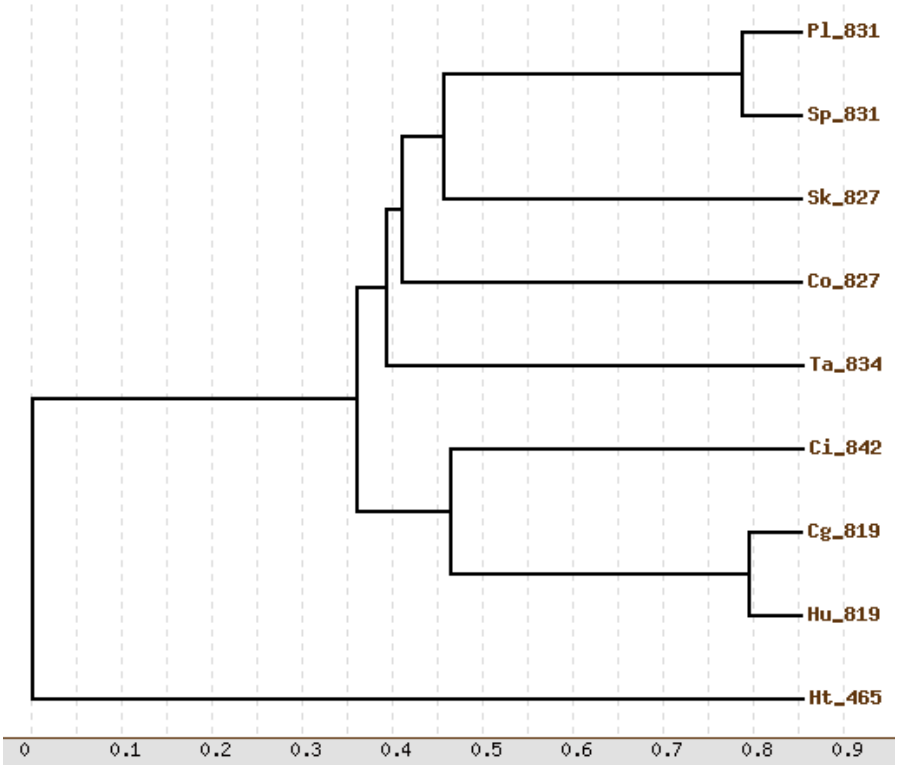


Figure 13. Phylogenetic tree of *advillins.* Advillin protein sequences from: *Heliocidaris tuberculata, Saccoglossus kowalevskii*, *Capsaspora owczarzaki*, *Trichoplax adhaerens* , *Ciona intestinalis*, *Cricetulus griseus* and Human. See text for reference sequences.

Domain homology and phylogenetic analysis of Tetraspanin

We amplified and cloned the *Pl-tetraspanin* cDNA of 970 nucleotides, containing the entire coding region (1 - 837 bp) of *Pl-tetraspanin*. We deposited the *Pl-tetraspanin* sequence in the EMBL database with the accession number: FR693765.

The deduced *Pl-tetraspanin* aa sequence is composed 278 residues with a predicted pI of 3,91 and a MW of 30.06 kDa. *In silico* analysis revealed the functional domains of the protein. Tetraspanin, involves two extracellular domains (38 aa - 57 aa and 112 aa - 242 aa), one small cytoplasmic (80 aa - 87 aa) and four transmembrane domains: 8 aa - 36 aa; 57 aa - 80 aa; 87 aa - 112 aa; 242 aa - 267 aa (Fig.14; Fig.15). A dimer interface region was identified from the 110 aa to 125 aa (Fig.14). A short repeat sequence at the N-terminal of the putative sequence (aa:19-44) was identified: FNLLFFLVGVALLAVGIYVIVQPYQL. No signal peptide was identified.

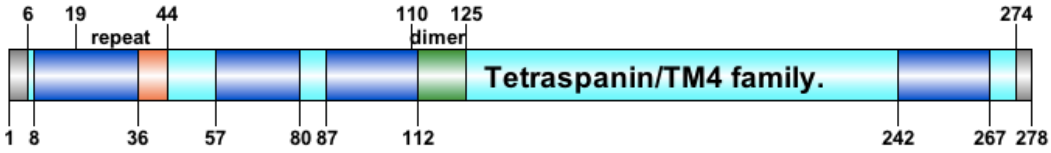


Figure 14. Illustration of protein domain structure of *Pl-tetraspanin*. The transmembrane regions are shown in blue. A short repeat of unknown function at the N-terminal (19 aa - 44 aa, red) and a dimer interface region (110 aa - 125 aa, green), are shown.

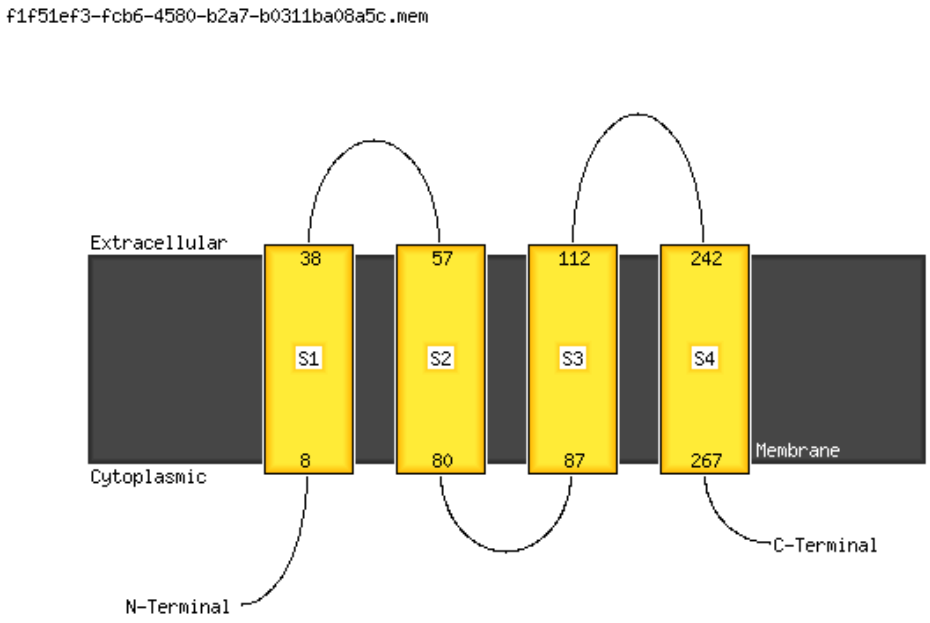
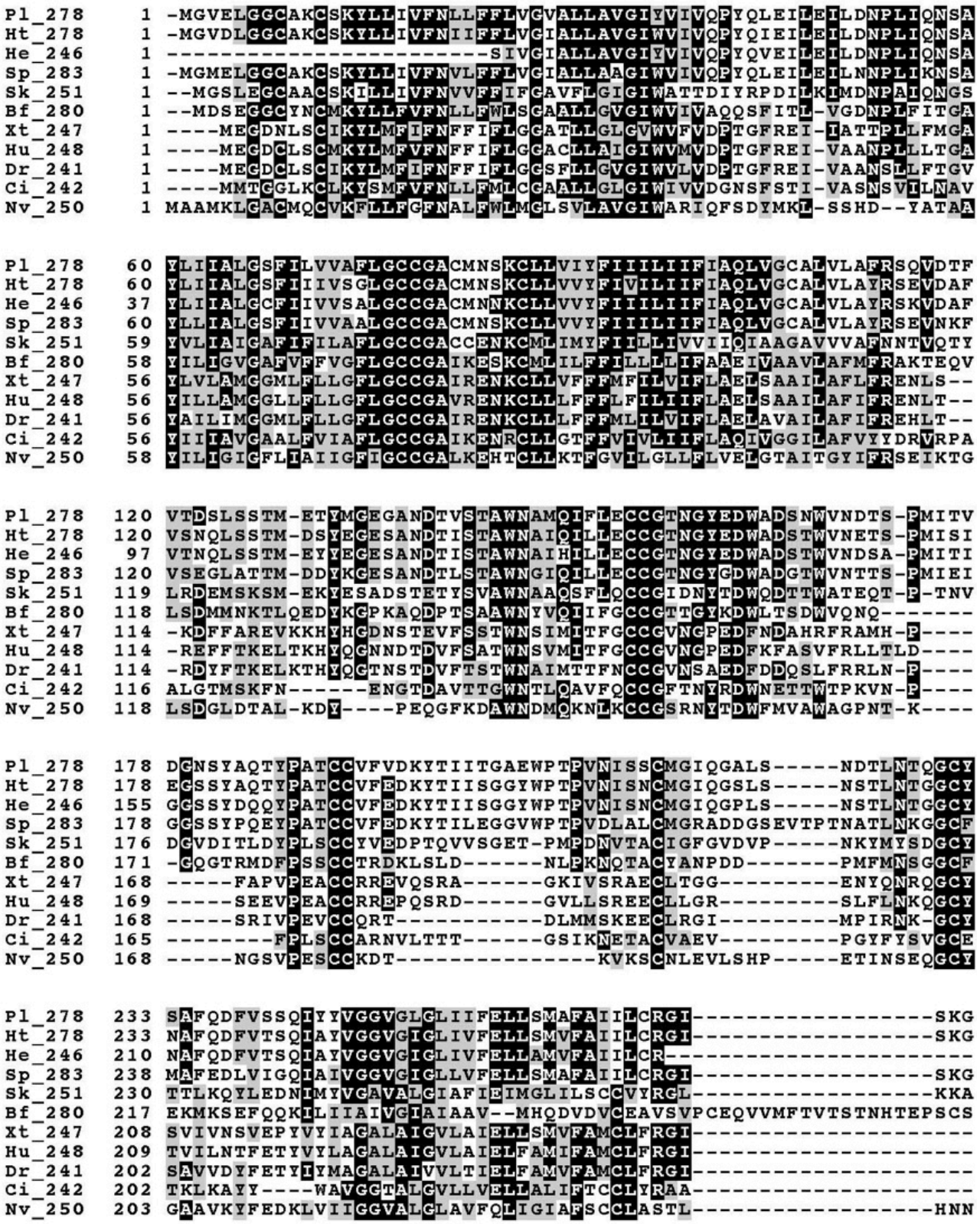


Figure 15. Illustration of the topology of *Pl-tetraspanin*. The putative protein involves two extracellular domains (38 aa - 57 aa and 112 aa - 242 aa), one small cytoplasmic (80 aa - 87 aa) and four transmembrane domains: 8 aa - 36 aa; 57 aa - 80 aa; 87 aa - 112 aa; 242 aa - 267 aa. Analysis was performed using the method of McGuffin *et al.*, 2000 at the PSIPRED.

BLAST analysis identified three homologues in the sea urchin species. *H. tuberculata* tetraspanin (Accession number: ABE27955) of 278 aa, shows an identity of 84% and a similarity of 95%. Tetraspanin from *S. purpuratus* (NP_001118229) of 283 aa length, shows an identity of 75% and a similarity of 86% with *P. lividus*. Last, tetraspanin from *H. erythrogramma* (ABE27956), of 246 aa, exhibits 82% of identity and 91% of similarity. *Pl-tetraspanin* exhibited high similarity with orthologues from other species, among others, belonging to the Chordata, Hemichordata and Cnidaria phyla and human. Indeed, the following sequences showed identities with *Pl-tetraspanin* varying from 32% to 72% and similarities varying from 49% to 84%. These sequences are: the predicted tetraspanin-like protein from *Saccoglossus kowalevskii* (XP_002734162.1) of 251 residues; the hypothetical protein from *Branchiostoma floridae* (XP_002608379.1); the predicted tetraspanin protein from *Ciona intestinalis* (XP_002123087.1); the tetraspanin 18 from *Xenopus (Silurana) tropicalis* (NP_001093673.1); the predicted protein from *Nematostella vectensis* (XP_001622937.1) and the 180 aa long tetraspanin from *Danio rerio* (NP_001002734.1). Additionally, the human tetraspanin-18 (NP_570139.3) exhibited 34% identity and 52% similarity.

The deduced sequence was used for multiple alignments (Fig.16). The phylogenetic analysis using the full-length *Pl-tetraspanin* protein sequences from various species belonging to evolutionarily divergent families, suggested that *Pl-tetraspanin* is a well conserved member of the tetraspanin family (Fig.17).


Pl_278 1 -MGVELGGCAKCSKYLLIVFNLLFFLVGVALLAVGIYVIVQPYQLEILEILDNPLIQNSA
Ht_278 1 -MGVDLGGCAKCSKYLLIVFNIIFFLVGIALLAVGIWVIVQPYQIEILEILDNPLIQNSA
He_246 1 ----------SIVGIALLAVGIYVIVQPYQVEILEILDNPLIQNSA
Sp_283 1 -MGMELGGCAKCSKYLLIVFNVLFFLVGIALLAAGIWVIVQPYQLEILEILNNPLIKNSA
Sk_251 1 --MGSLEGCAACSKILLIVFNVVFFIFGAVFLGIGIWATTDIYRPDILKIMDNPAIQNGS
Bf_280 1 --MDSEGGCYNCMKYLLFVFNLLFWLSGAALLGVGIWVIVAQQSFITL-VGDNPLFITGA
Xt_247 1 ----MEGDNLSCIKYLMFIFNFFIFLGGATLLGLGVWVFVDPTGFREI-IATTPLLFMGA
Hu_248 1 ----MEGDCLSCMKYLMFVFNFFIFLGGACLLAIGIWVMVDPTGFREI-VAANPLLLTGA
Dr_241 1 ----MEGDCLSCIKYLMFIFNFFIFLGGSFLLGVGIWVLVDPTGFREI-VAANSLLFTGV
Ci_242 1 ----MMTGGLKCLKYSMFVFNLLFMLCGAALLGLGIWIVVDGNSFSTI-VASNSVILNAV
Nv_250 1 MAAMKLGACMQCVKFLLFGFNALFWLMGLSVLAVGIWARIQFSDYMKL-SSHD--YATAA
Pl_278 60 YLIIALGSFILVVAFLGCCGACMNSKCLLVIYFIIILIIFIAQLVGCALVLAFRSQVDTF
Ht_278 60 YLIIALGSFIIIVSGLGCCGACMNSKCLLVVYFIVILIIFIAQLVGCALVLAYRSEVDAF
He_246 37 YLIIALGCFIIVVSALGCCGACMNNKCLLVVYFIIILIIFIAQLVGCALVLAYRSKVDAF
Sp_283 60 YLLIALGSFIIVVAALGCCGACMNSKCLLVVYFIIILIIFIAQLVGCALVLAYRSEVNKF
Sk_251 59 YVLIAIGAFIFILAFLGCCGACCENKCMLIMYFIILLIVVIIQIAAGAVVVAFNNTVQTY
Bf_280 58 YILIGVGAFVFFVGFLGCCGAIKESKCMLILFFILLLLIFAAEIVAAVLAFMFRAKTEQV
Xt_247 56 YLVLAMGGMLFLLGFLGCCGAIRENKCLLVFFFMFILVIFLAELSAAILAFLFRENLS--
Hu_248 56 YILLAMGGLLFLLGFLGCCGAVRENKCLLLFFFLFILIIFLAELSAAILAFIFRENLT--
Dr_241 56 YAILIMGGMLFLLGFLGCCGAIRENKCLLLFFFMLILVIFLAELAVAILAFIFREHLT--
Ci_242 56 YIIIAVGAALFVIAFLGCCGAIKENRCLLGTFFVIVLIIFLAQIVGGILAFVYYDRVRPA
Nv_250 58 YILIGIGFLIAIIGFIGCCGALKEHTCLLKTFGVILGLLFLVELGTAITGYIFRSEIKTG
Pl_278 120 VTDSLSSTM-ETYMGEGANDTVSTAWNAMQIFLECCGTNGYEDWADSNWVNDTS-PMITV
Ht_278 120 VSNQLSSTM-DSYEGESANDTISTAWNAIQILLECCGTNGYEDWADSTWVNETS-PMISI
He_246 97 VTNQLSSTM-EYYEGESANDTISTAWNAIHILLECCGTNGYEDWADSTWVNDSA-PMITI
Sp_283 120 VSEGLATTM-DDYKGESANDTLSTAWNGIQILLECCGTNGYGDWADGTWVNTTS-PMIEI
Sk_251 119 LRDEMSKSM-EKYESADSTETYSVAWNAAQSFLQCCGIDNYTDWQDTTWATEQT-P-TNV
Bf_280 118 LSDMMNKTLQEDYKGPKAQDPTSAAWNYVQIIFGCCGTTGYKDWLTSDWVQNQ-------
Xt_247 114 -KDFFAREVKKHYHGDNSTEVFSSTWNSIMITFGCCGVNGPEDFNDAHRFRAMH-P----
Hu_248 114 -REFFTKELTKHYQGNNDTDVFSATWNSVMITFGCCGVNGPEDFKFASVFRLLTLD----
Dr_241 114 -RDYFTKELKTHYQGTNSTDVFTSTWNAIMTTFNCCGVNSAEDFDDQSLFRRLN-P----
Ci_242 116 ALGTMSKFN------ENGTDAVTTGWNTLQAVFQCCGFTNYRDWNETTWTPKVN-P----
Nv_250 118 LSDGLDTAL-KDY----PEQGFKDAWNDMQKNLKCCGSRNYTDWFMVAWAGPNT-K----
Pl_278 178 DGNSYAQTYPATCCVFVDKYTIITGAEWPTPVNISSCMGIQGALS-----NDTLNTQGCY
Ht_278 178 EGSSYAQTYPATCCVFEDKYTIISGGYWPTPVNISNCMGIQGSLS-----NSTLNTGGCY
He_246 155 GGSSYDQQYPATCCVFEDKYTIISGGYWPTPVNISNCMGIQGPLS-----NSTLNTGGCY
Sp_283 178 GGSSYPQEYPATCCVFEDKYTILEGGVWPTPVDLALCMGRADDGSEVTPTNATLNKGGCF
Sk_251 176 DGVDITLDYPLSCCYVEDPTQVVSGET-PMPDNVTACIGFGVDVP-----NKYMYSDGCY
Bf_280 171 -GQGTRMDFPSSCCTRDKLSLD------NLPKNQTACYANPDD-------PMFMNSGGCF
Xt_247 168 -----FAPVPEACCRREVQSRA------GKIVSRAECLTGG---------ENYQNRQGCY
Hu_248 169 -----SEEVPEACCRREPQSRD------GVLLSREECLLGR---------SLFLNKQGCY
Dr_241 168 -----SRIVPEVCCQRT----------DLMMSKEECLRGI---------MPIRNK-GCY
Ci_242 165 --------FPLSCCARNVLTTT------GSIKNETACVAEV---------PGYFYSVGCE
Nv_250 168 -----NGSVPESCCKDT----------KVKSCNLEVLSHP-----ETINSEQGCY
Pl_278 233 SAFQDFVSSQIYYVGGVGLGLIIFELLSMAFAIILCRGI----------SKG
Ht_278 233 NAFQDFVTSQIAYVGGVGIGLIVFELLSMVFAIILCRGI----------SKG
He_246 210 NAFQDFVTSQIAYVGGVGIGLIVFELLAMVFAIILCR----------
Sp_283 238 MAFEDLVIGQIAIVGGVGIGLLVFELLSMAFAIILCRGI----------SKG
Sk_251 230 TTLKQYLEDNIMYVGAVALGIAFIEIMGLILSCCVYRGL----------KKA
Bf_280 217 EKMKSEFQQKILIIAIVGIAIAAV--MHQDVDVCEAVSVPCEQVVMFTVTSTNHTEPSCS
Xt_247 208 SVIVNSVEPYVYIAGALAIGVLAIELLSMVFAMCLFRGI----------
Hu_248 209 TVILNTFETYVYLAGALAIGVLAIELFAMIFAMCLFRGI----------
Dr_241 202 SAVVDYFETYIYMAGALAIVVLTIELFAMVFAMCLFRGI----------
Ci_242 202 TKLKAYY----WAVGGTALGVLLVELLALIFTCCLYRAA----------
Nv_250 203 GAAVKYFEDKLVIIGGVALGLAVFQLIGIAFSCCLASTL----------HNN

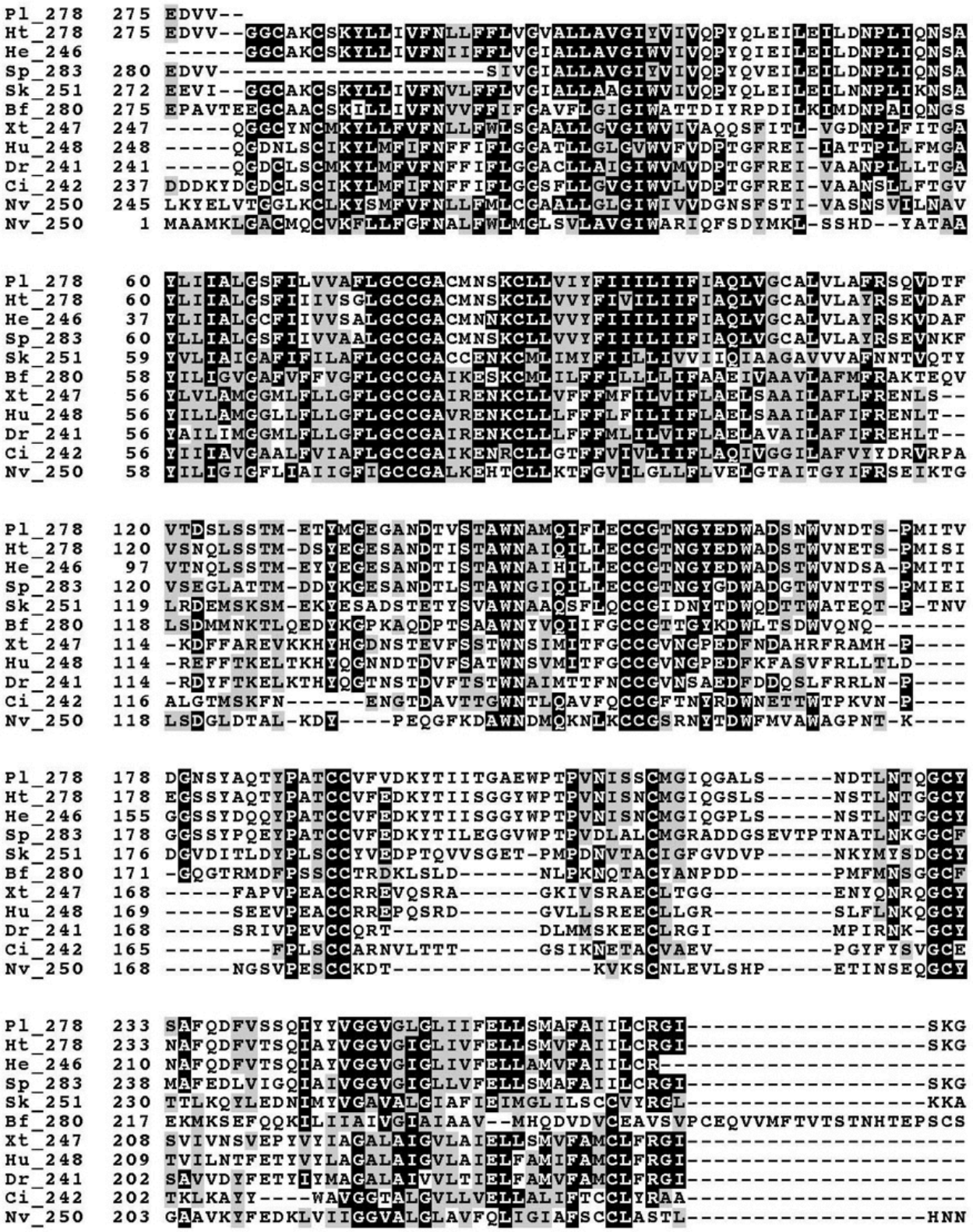


Figure 16. Multiple alignment of tetraspanin protein from different species. ClustalW alignment of the entire CDS with homologous proteins. On the left, abbreviations of the species names and the total number of amino acids are presented. Protein GenBank numbers: Pl_278, FR693765; Ht_278, ABE27955; Sp_283, NP_001118229; He_246, ABE27956; Sk_251, XP_002734162.1; Bf_280, XP_002608379.1; Ci_242, XP_002123087.1; Xt_247, NP_001093673.1; Nv_250, XP_001622937.1; Dr_241, NP_001002734.1 and Hu_248, NP_570139.3. Identical amino acids are marked with asterisk; dashes correspond to gaps.

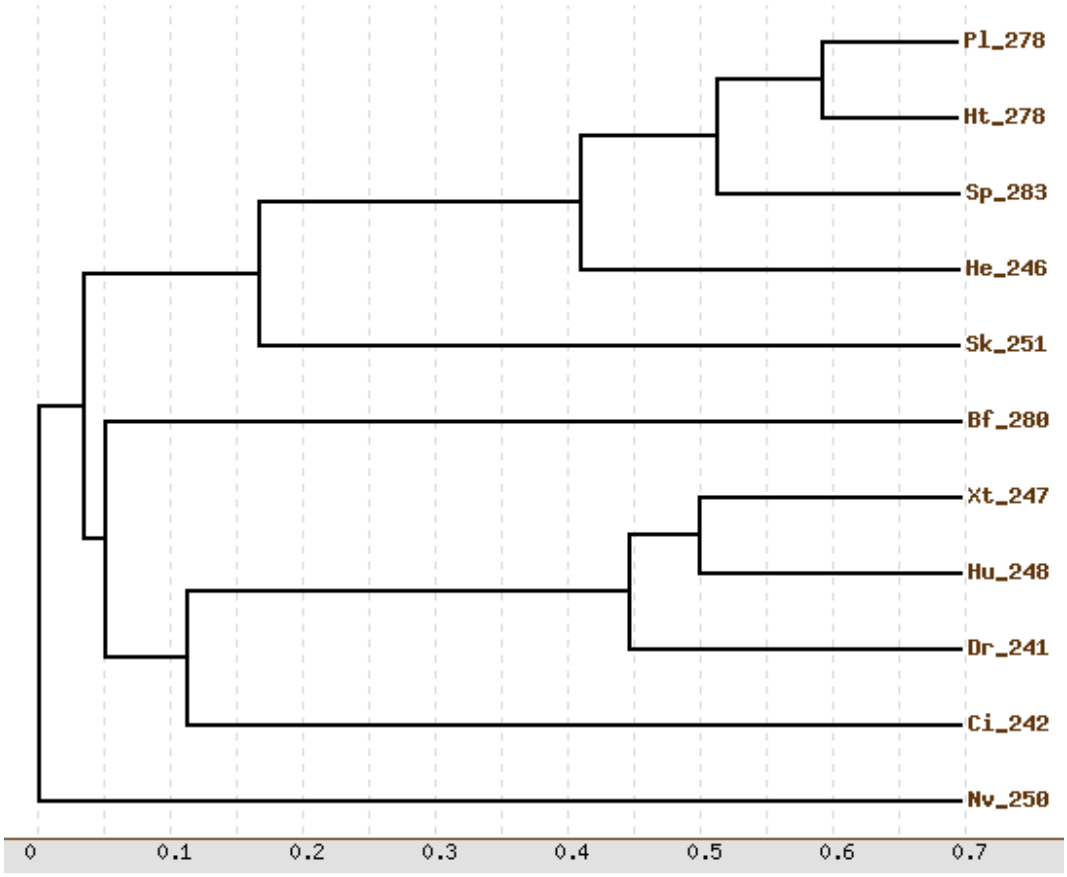


Figure 17. Phylogenetic tree of *Pl-tetraspanin.* See text for reference sequences.

Domain homology and phylogenetic analysis of SM30α

We amplified and cloned the *Pl-sm30a* cDNA of 1024 nucleotides, containing the entire coding region (1 - 936 bp) of *Pl-sm30a*. We deposited the *Pl-sm30a* sequence in the EMBL database with the accession number: FR716470.

The deduced *Pl*-SM30a aa sequence comprises 311 residues with a predicted pI of 3,91 and a MW of 33521,87 Da. *In silico* analysis revealed the functional domains of the protein, including a glycine-rich region from the aa 25 to aa 57 (43 residues, 13.8 % of the whole protein); and a C-type lectin domain (aa 109 to aa 232) (Fig.18). It is noted that the sequence also includes 32 proline residues (10.3 %) and 27 alanines (8.7 %) spread throughout the sequence.

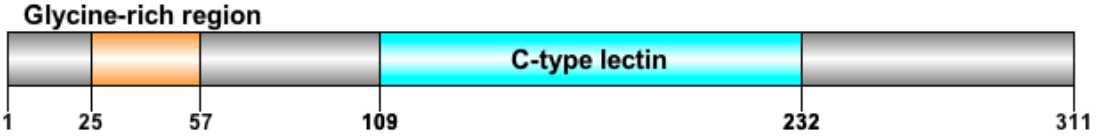


Figure 18. Illustration of protein domain structure of *Pl-SM30α*.

BLAST analysis identified seven homologues in *S. purpuratus* (Accession numbers: Q26646.1, NP_999766.1, XP_795963.1, XP_796721.1, XP_796701.1, XP_001177190.1, XP_001186324.1) showing similarities from 72% to 45% and identities from 64% to 32%. Additionally, Q25116.1 from *H.pulcherrimus*, exhibited 63% of identity and 70% of similarity. Sequences from other species, including *Danio rerio*, exhibiting 29% of identity and 45% of similarity; *Mus musculus*, exhibiting 33% of identity and 43% of similarity and the Human C-type lectin 19 member A precursor, exhibiting 28% of identity and 40% of similarity. Similarities with protein belonging to the Fras1 family related proteins, secretory phospholipase receptors and mannose receptors, were observed.

The phylogenetic analysis by the ClustalW software (Thompson *et al* 1994) using the full-length *Pl*-SM30a protein sequences from sea urchin species belonging to evolutionarily divergent families, suggested that *Pl*-SM30a is a member of the SM30 family (Fig.19). A phylogenetic tree was prepared (Fig.20) with the *S. purpuratus* SM30a sequence exhibiting the highest similarity (Accession number: Q26646.1).

```
Pl_311      1 MRSFVCVLVCVAAMAIHAQAQNWQGPGGGFPGQGGPGHGGGHGHGHGGQGHGGQGQGFPP
Sp_290      1 MRGFVYVLVCVLALASFSRAQ--------LPGGGGP-VLPG---------------GGPT
Hp_288      1 MRCFVYVLVCVVASVSYSRAQ--------LPGAGGPGVFPG---------------GVPT
Hu_186      1 MQRWT---LWAAAFLTLHSAQ----------------AFPQ---------------TDIS

Pl_311     61 IPPVNPDPTRQETCPKFWLEEGNSCYLFDSGAFLRQVAASAPVVVNNQDGLFQGAANAYC
Sp_290     37 IGPVNPDPTRTEVCAKFWVQEGNSCYLFDSGAFLRQVAASRPVVVNNENGLFQAAANMYC
Hp_288     38 IGPVIPDPTRTETCAKFWVQEGDSCYLFDSGAFLRQVAASRPVVVNNQDGLFQAAANMYC
Hu_186     27 ISPALPELPLPSLCPLFWMEFKGHCYRF--------------FPLNKT----WAEADLYC

Pl_311    121 GRM---YPGATLVTVNDLQENNFLYEWAVRMMVE-PQPVWIGLHVGPM-GQWQWFSGEPV
Sp_290     97 GQM---HPNASLVTVNSLAENNFLYEWAVRMMVE-PEPVWIGLHAGPM-GQWQWYSGEPV
Hp_288     98 GQM---HPNATLVTVNSLAENNFLYEWAVRMMVE-PEPVWIGLHVGPA-GLWQWYSGEPV
Hu_186     69 SEFSVGRKSAKLASIHSWEENVFVYDLVNSCVPGIPADVWTGLHDHRQEGQFEWTDGSSY

Pl_311    176 NFTNWEGMMAPMPEPGLGAVIFDADIRNQMFNNQVEITPQWVPEEAMNDRHAIICEYHPS
Sp_290    152 TYTNWERMTAPMAEPGLGAMIFDADIIAQMFNNQVEITPQWVPEQAINDRHALICEYHPS
Hp_288    153 TYTNWEGMVAPRAELGLGAMIFDADIIAQMFNNQVEITPQWVPEHGRNDRHSLICEYHPS
Hu_186    129 DYSYWDG---SQPDDGVHA------------------DPE--------EEDCVQIWYRPT

Pl_311    236 GMTAPTSATVAPTTAGVMPTGTTMAPASGGPVLMRNNPAPLQN---GGAFGGSRLFEVPR
Sp_290    212 GMTAAAAAATNAPTFPPMATAPPMAATTRGPVMFQNNPRNLVNSLTGGRFGGSLLHEIPR
Hp_288    213 GMT---AAATNAPTTPPMATAPPMAATTRSPVMFQNNPGNLVNRLTGGRFGGSLLHEIPR
Hu_186    160 SALRSWNDNTCSRKFPFVCKIPSLT-----------------------------------

Pl_311    293 RQRNRPSNYRMNRYYGVMP
Sp_290    272 RQRMRPSNYRKNPYFGIQP
Hp_288    270 RQRMRPSNYRKNPYFGIQP
Hu_186    185 ----------------IH
```

Figure 19. Multiple alignment of SM30a from different species. ClustalW alignment of the entire CDS with homologous proteins. On the left, abbreviations of the species names and the total number of amino acids are presented. Protein GenBank numbers: Pl_311, FR716470; Sp_290, Q26646.1; Hp_288, Q25116.1; Hu_186, NP_001243649.1. Identical amino acids are marked with asterisk; dashes correspond to gaps.

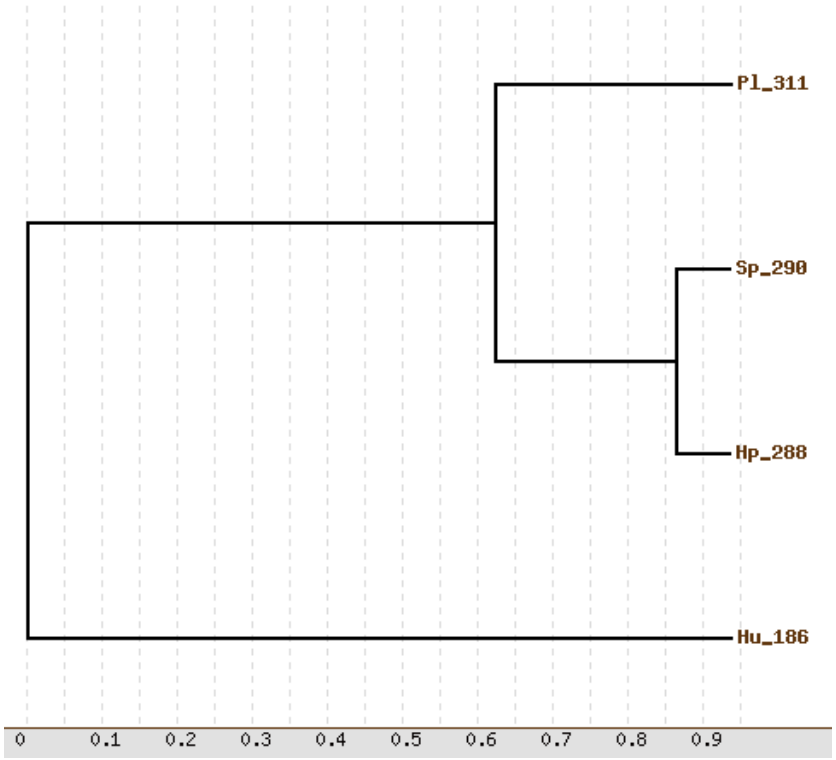


Figure 20. Phylogenetic tree of sea urchin *SM30a.* See text for sequences.

Domain homology and phylogenetic analysis of SM50

The deduced *Pl*-SM50 aa sequence comprises 297 residues with a predicted pI of 9,1 and a MW of 30.89 kDa. *In silico* analysis revealed the functional domains of the protein. Pl-SM50 involves an N-terminal C-type lectin domain (aa 29 to aa 128) and an extended C-terminal Proline-rich region from aa 129 to aa 270 (Fig.21). The proline-rich region includes several spread proline residues inter-dispersed in a fairly regular manner but also in short repeats. There are seven PGM, five QQPFM and five PQQ repeats. Furthermore, the region from aa 220 to aa 269 is rich in Q and G residues. In total, the sequence involves 49 glycine residues (16.6 %), 38 proline residues (12.8 %) and 29 alanine and glutamine residues (9.8 %).

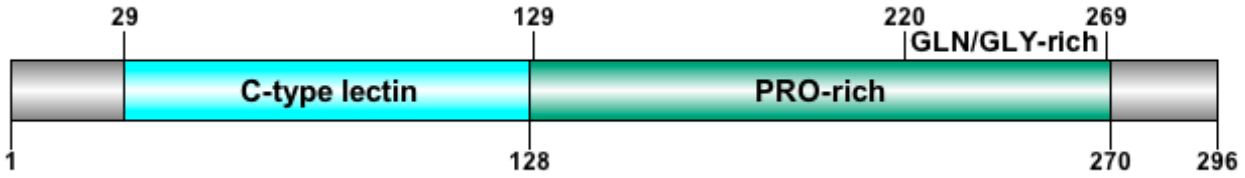


Figure 21. Illustration of protein domain structure of *Pl-SM50*. A conserved functional domain of C-type lectin is shown in light blue and a long *proline*-rich region is shown in green.

BLAST analysis identified homologues in three sea urchin species. The spicule matrix protein SM41 from *Hemicentrotus pulcherrimus* (Q26264.1) of 407 aa length showed the highest similarity (74%) and identity (83%). Two homologs were identified from *S. purpuratus*; NP_999775.1 of 445 aa length and NP_999803.1 of 289 aa length, exhibiting a similarity an identity of 69% and 60% a similarity of 78% and 70%, respectively. Finally, SM34 from *Lytechinus pictus* (Q05904.1), showed an identity of 65% and a similarity of 76%. The protein alignments are shown in Figure 21. A phylogenetic analysis using the ClustalW software (Thompson *et al* 1994) and the full-length Pl-SM50 protein sequences from sea urchin species, suggested that Pl-SM50 is a member of the SM50 family and Pl-SM50 is a conserved spicule matrix protein (Fig.22, 23).

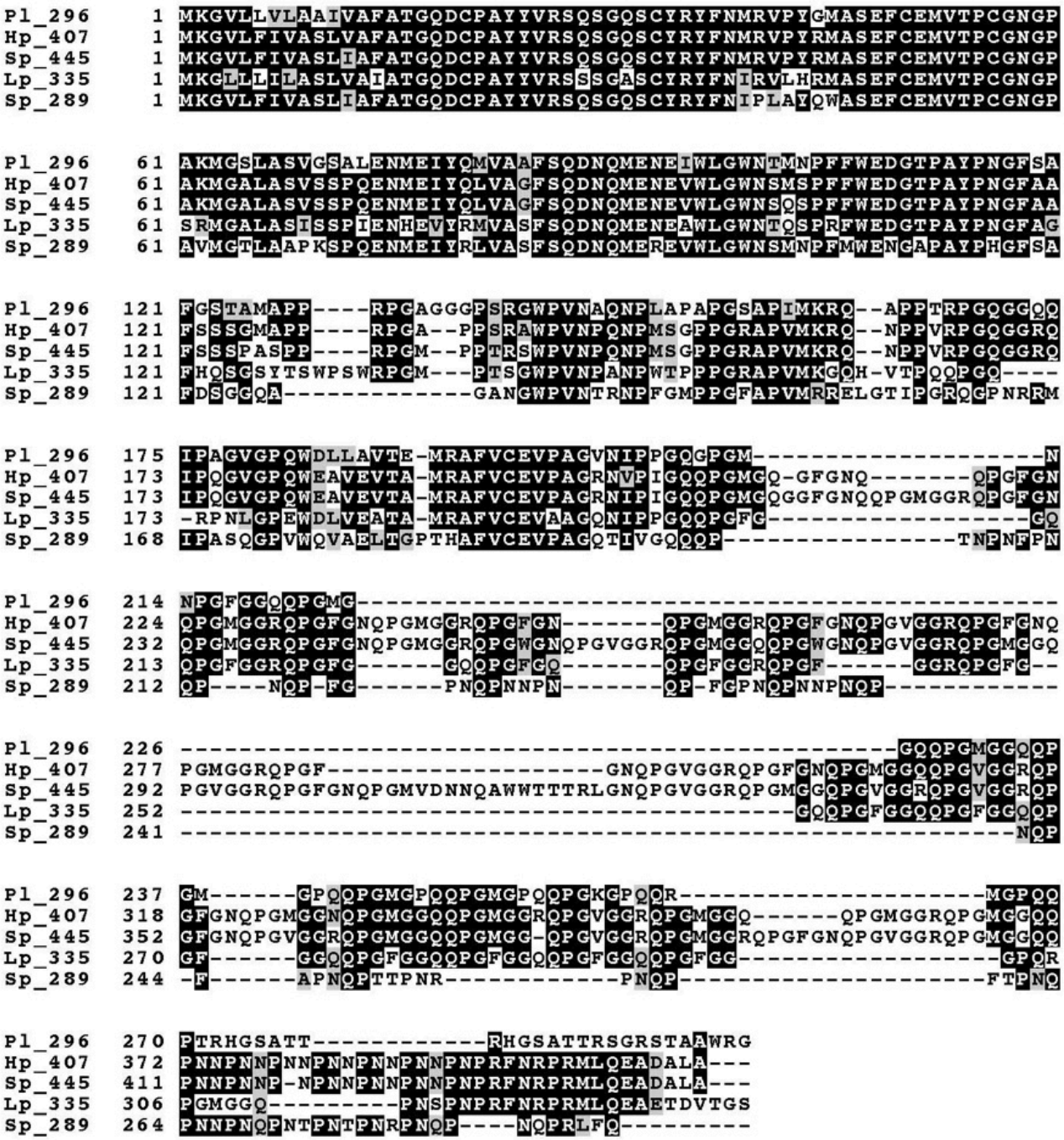


Figure 22. Multiple alignment of SM50 from different species. ClustalW alignment of the entire CDS with homologous proteins. On the left, abbreviations of the species names and the total number of amino acids are presented. Protein GenBank numbers: Pl_296; Hp_407, Q26264.1; Sp_445, NP_999775.1; Lp_335, Q05904.1 and Sp_289, NP_999803.1. Identical amino acids are marked with asterisk; dashes correspond to gaps.

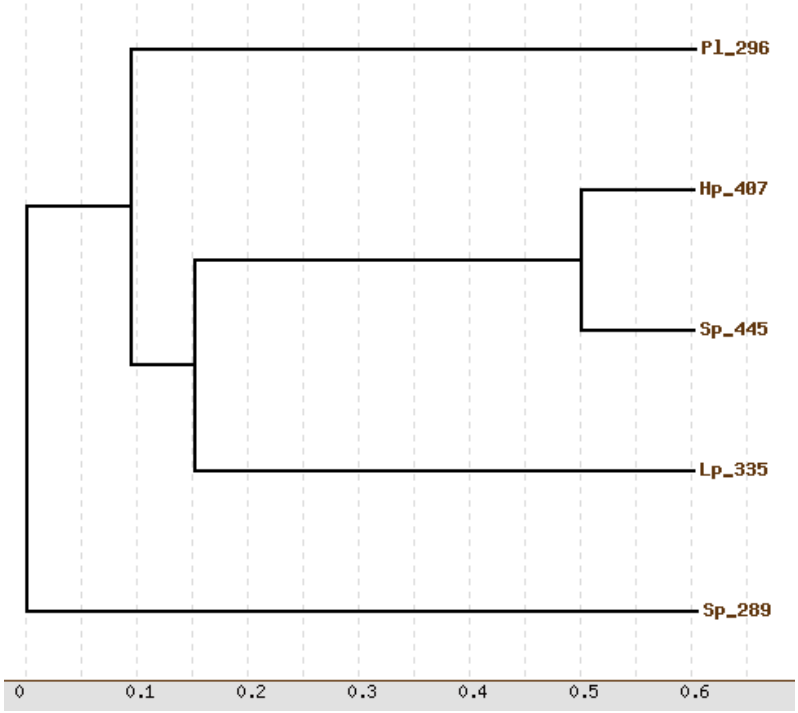


Figure 23. Phylogenetic tree of *Pl-SM50*

Tempo-spatial gene expression profiling of Pl-*p16, Pl-p19, Pl-advillin, Pl-tetraspanin, Pl-sm30a* and *Pl-sm50* during *P. lividus* embryo development

The expression profile of Pl-*p16,* Pl-*p19,* Pl-*advillin* and Pl-*tetraspanin* was monitored throughout development by comparative qPCR. Each primer set was tested by RT-PCR, prior to the qPCR. Each amplification yielded a unique product as viewed by 2% Agarose-TBE gel electrophoresis, confirming the specificity of the amplification (Fig.24). The maternal expression of each gene was investigated by RT-PCR from fertilized eggs (Fig. 25). Results indicated that none of these genes is expressed at detectable levels in the unfertilized egg or at the initial cleavage stages until the first 5 hours after fertilization; and their expression initiates at later stages.

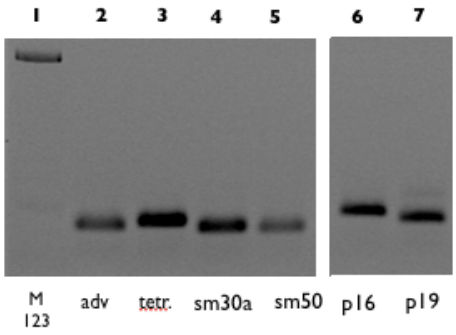


Figure 24. Test of the specificity of the qPCR primers by PCR and electrophoresis in gel (2% agarose in TBE). Each set of primers used for the comparative qPCR was primarily tested by RT-PCR. Total RNA extracted form gastrula was used as template.

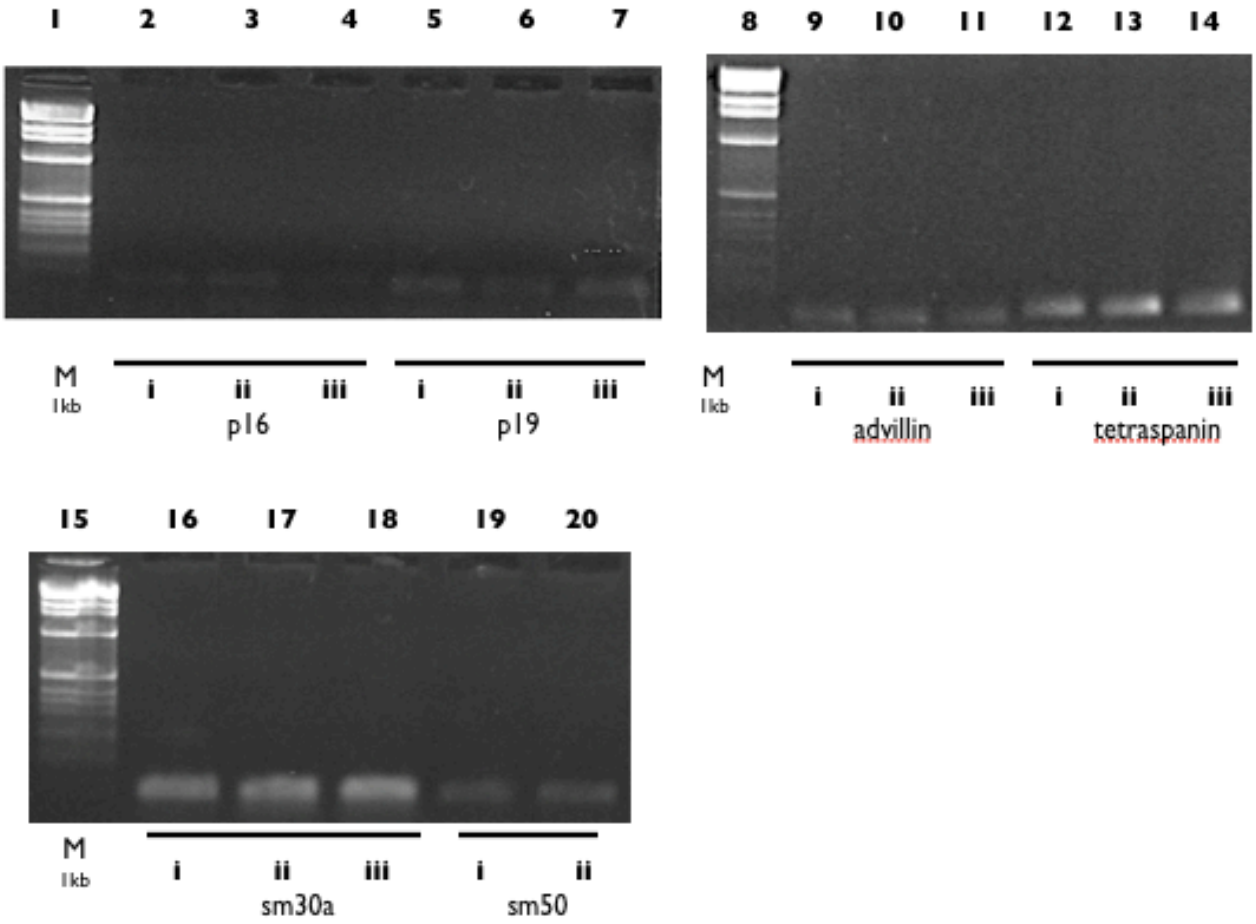


Figure 25. Testing of maternal expression of *p16, p19, sm30a, sm50, advillin* and *tetraspanin.* Amplification with the One step RT-PCR kit and the qPCR primers (Table 3) using as templates total RNAs from: (i) unfertilised eggs; (ii) two hours embryos (2-4 cell stage) and (iii) five hours embryos (morula stage). Maternal expression was not detected for any of the tested genes.

Additionally, the transcriptomic expression of each gene was localized by WMISH. For the WMISH, AS DIG-labelled probes were amplified from each *pGEM-T-Easy* plasmid by asymmetric PCR with the *T7* promoter, after linearization by digestion with *Sal* I. As an example, Fig.26 shows the linearization of *pGEM-T-Easy-Pl-advillin* and *pGEM-T-Easy-Pl-tetraspanin.* Each probe had a concentration of 0.5 µg/µl.

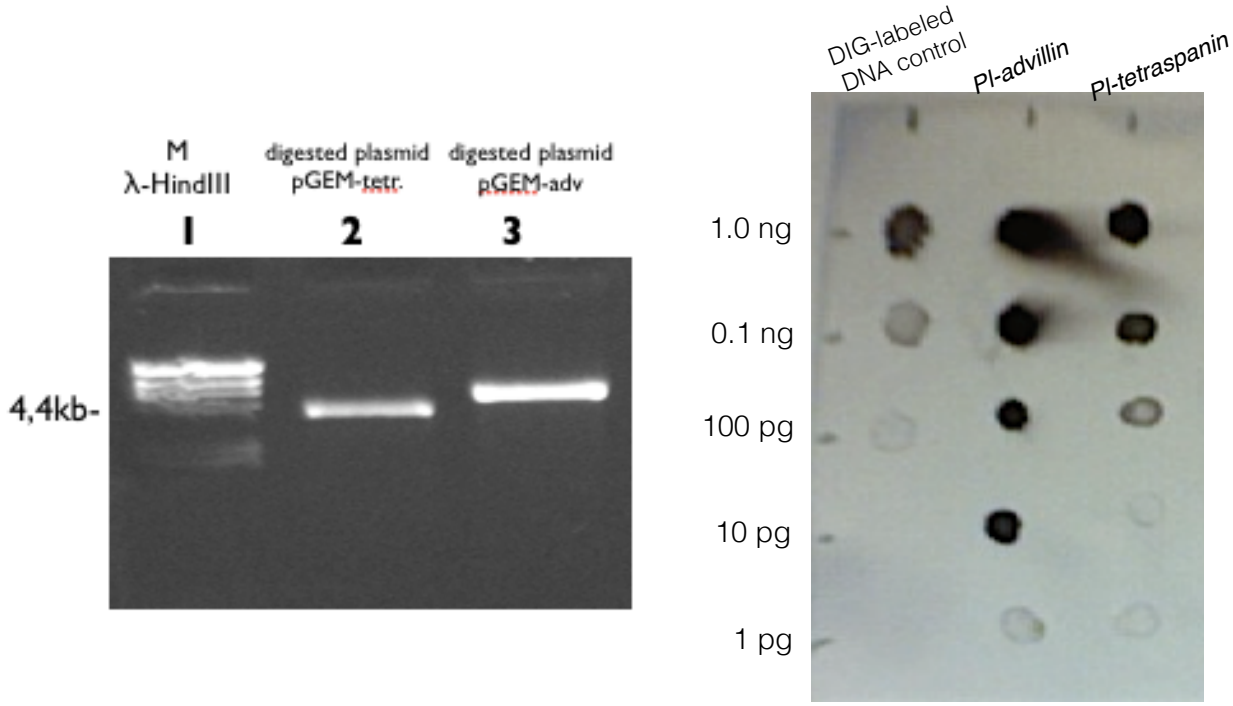


Figure 26. Preparation of WMISH probes: Left: Linearization of plasmids. Digestion for the linearization of *pGEM-T-Easy-Pl-advillin* and *pGEM-T-Easy-Pl-tetraspanin* by digestion with *Sal*-I. Linearized plasmids were then used for the preparation of the probes. The activity of the DIG probes was measured by Dot Blot (Right).

Temporal gene expression profiling of *Pl-p16 and Pl-p19*

The temporal expression of the two mRNAs has been analyzed by comparative real-time qPCR (ΔCCt Q-PCR). The two transcripts levels were measured throughout development, from cleavage to the pluteus stage (Fig.27). Compared to the cleavage stages, the expression of both *Pl-p16* and *Pl-p19* increased at the mesenchyme blastula stage and then sharply peaked at the gastrula stage to slightly decrease at the pluteus stage. The up-regulation in the embryo at the gastrula stage is similarly observed for both mRNAs. (Fig.27).

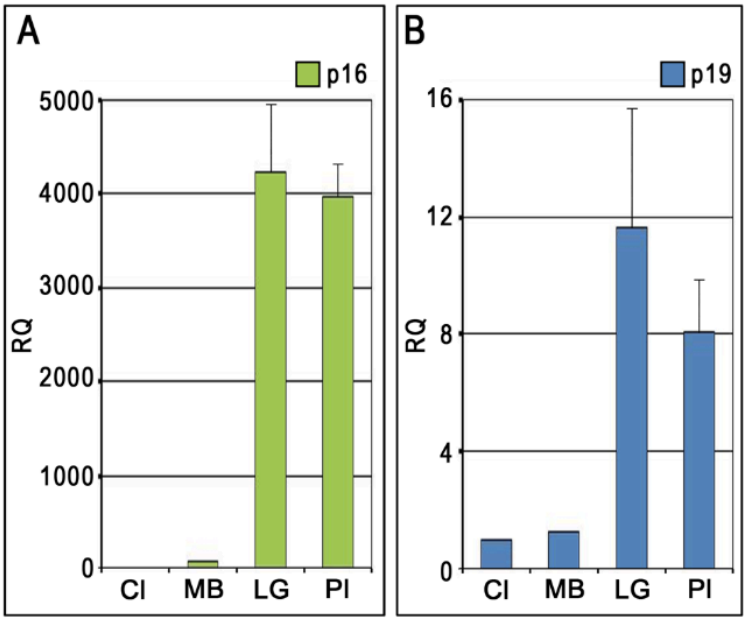


Figure 27. Temporal expression of *Pl-p16* and *Pl-p19*. Transcript levels from cleavage to pluteus were measured by comparative real-time Q-PCR: Cl cleavage stage, MB mesenchyme blastula, LG late gastrula, Pl pluteus. Each point represents the fold change (RQ value) calculated from the mean of three replicate reactions. The value of the cleavage stage was used as reference and assumed as 1. The endogenous gene *Pl-Z12-1* was used for normalization (Costa and Karakostis *et al.*, Dev Genes Evol., 2012)

Spatial expression patterns of *Pl-p16* and *Pl-p19*

To investigate the spatial expression patterns of *Pl-p16 and Pl-p19* during development, WMISH experiments were carried out on embryos collected at stages from swimming blastula to pluteus. Both *Pl-p16 and Pl-p19* were exclusively expressed in the PMCs throughout embryo development, but with slightly different expression patterns. Specifically, *Pl-p16* transcripts were initially fairly detectable in the presumptive PMCs localized at the center of the vegetal plate in swimming blastulae, prior to PMCs ingression (Fig.28-A). Then, *Pl-p16* transcripts were localised in all the PMCs at their ingression (Fig. 28-B) and during early gastrulation (Fig.28-C,D). At the late gastrula stage, expression was detectable mainly in the PMCs of the dorsal chain (Fig.28-E, arrow), ventrolateral clusters (Fig.28-E, asterisk) and longitudinal chains (Fig.28-E, arrow-head). PMCs of the ventral chain were not labelled. A further spatial down-regulation of *Pl-p16* mRNA expression was observed from the prism stage onward (Fig.28-F-I), restricted to a subset of the PMCs associated with sites of active spicule growth, namely the tips of the postoral rods (Fig.28-G, arrows), the anterolateral rods (Fig.28-G, arrowhead) and the scheitel (Fig.28-G, asterisk). At the pluteus stage only two or three stained cells per site of active spicule growth are detectable (Fig.28H,I). These observations are consistent with the expression profiles described for *Sp-p16* (Illies *et al.*, 2002) and *Lv-p16* (Cheers and Ettensohn, 2005).

Pl-p19 mRNA was initially detected in the presumptive PMCs at the vegetal pole of the early blastula embryos (Fig.28-J). Following ingression (Fig.28-K), during gastrulation (Fig. 28-L-O) and throughout later embryogenesis (Fig.28-P-T), all PMCs showed a strong signal with no apparent evidence of a spatially-regulated expression of *Pl-p19* mRNA within the PMC syncytium. Interestingly, we observed unforeseen peculiar distribution of *Pl-p19* transcripts in some PMCs at the late gastrula (Fig.28-N,O) and late pluteus stages (Fig.28-S), never observed by WMISH with any of the probes for skeleton matrix proteins in all sea urchin species investigated so far. In fact, in addition to the usual labeling of the perinuclear cytoplasm, the signal indicates the presence of *Pl-p19* mRNA at the filopodial cytoplasm of the PMC syncytium, both at the ventrolateral clusters (Fig.28-N,O, arrows) and scheitel (Fig.28-S, arrow). The localization of *Pl-p19* mRNA in both the cell bodies and filopodial cytoplasm described here suggests an accumulation of mRNA molecules where there is major request of specific protein synthesis for spicule development.

In conclusion, *Pl-p16* and *Pl-p19* mRNA, here described for the first time in *P. lividus*, might be regulated differently since they show dissimilar patterns of expression within the PMC syncytium.

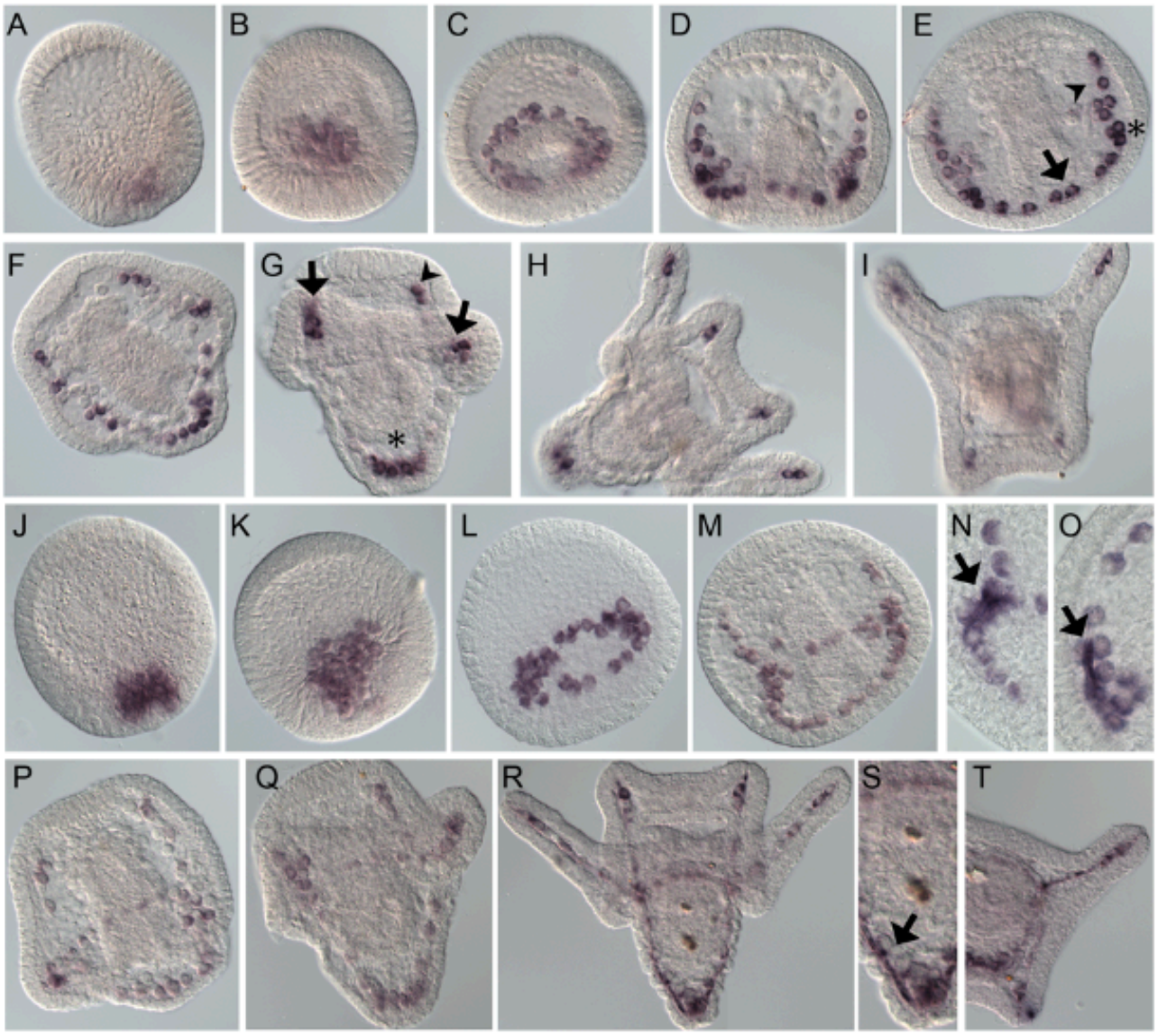


Figure 28. Spatial expression of *Pl-p16* and *Pl-p19*. Whole mount *in situ* hybridization showing *Pl-p16* (A-I) and *Pl-p19* (J-T) expression during *P. lividus* skeletogenic embryonic stages. A, J) early blastula; B, K) mesenchyme blastula; C, L) early gastrula; D) middle gastrula; E, M) late gastrula; F, P) prism; G, Q) early pluteus; H, R) late pluteus; I) apical view of late pluteus (*Pl-p16* probe); N-O) magnifications of late gastrula ventrolateral clusters (Pl-p19 probe), pointed by arrows; S) magnification of the apex of late pluteus (the scheitel) (*Pl-p19* probe), pointed by arrow; T) apical view of late pluteus (*Pl-p19* probe).The asterisk signs the ventrolateral clusters and the arrowheads sign the marked PMCs and territories as explained in the text, above (Costa, Karakostis *et al.*, Dev Genes Evol., 2012)

Advillin

Temporal gene expression profiling of *Pl-advillin*

The temporal expression profile, as revealed by comparative real-time qPCR (ΔCCt Q-PCR), showed that the expression of *Pl-advillin* is detectable from the cleavage stage and is up-regulated by 3 fold at the blastula and gastrula stages, to rapidly increase reaching a peak of 9.5 fold increase, at the pluteus stage (Fig.29).

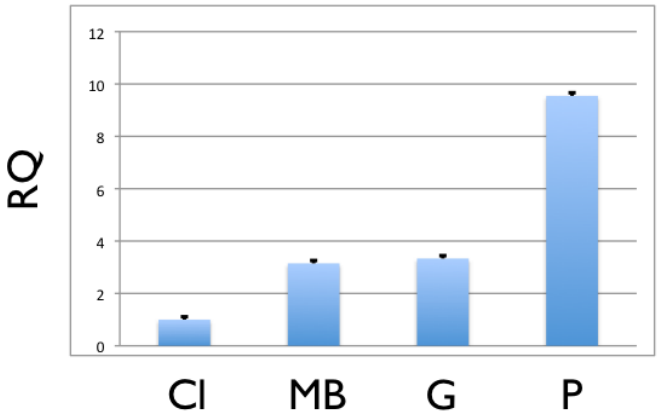


Figure 29. Temporal expression of *Pl-advillin*. Transcript levels from cleavage to pluteus stage were measured by comparative real-time q-PCR: Cl, cleavage stage; MB, mesenchyme blastula; G, gastrula; P, pluteus; Each point represents the fold change (RQ value) calculated from the mean of three replicate reactions. The value of the cleavage stage was used as reference and assumed as 1. The endogenous gene *Pl-Z12-1* was used for normalization.

Spatial expression patterns of *Pl-advillin*

To investigate the spatial expression patterns of *Pl-advillin* during developmental, whole-WMISH experiments were carried out on embryos collected from the mesenchyme blastula to the pluteus stage. *Pl-advillin* was exclusively expressed in the PMCs throughout embryo development. Expression was detectable in the presumptive PMCs along the vegetal plate, prior to PMCs ingression in embryos at the swimming blastula stage (Fig.30-A). Transcripts were localized in all the PMCs, at their ingression, during early gastrulation (Fig.30-B,C). After the archenteron elongation, until the late gastrula stage, expression was detectable at the PMCs of the ventral chain (Fig.30-D,E arrow) and the ventrolateral clusters (Fig.30-D,E asterisks). From the prism stage, *Pl-advillin* mRNA expression was mainly expressed in a subset of the PMCs which form the postoral (Fig.30-F, arrow) and longitudinal chain (Fig.30-G), while in pluteus expression was localized in the PMCs associated with sites of active spicule growth, namely the tips of the postoral rods (Fig.30-

H-K, arrows), the anterolateral rods (Fig.30-J,K, arrowhead) and on the scheitel (Fig.30-I,J asterisk). The ventral transverse rod was also labelled (Fig.30-J, double arrow).

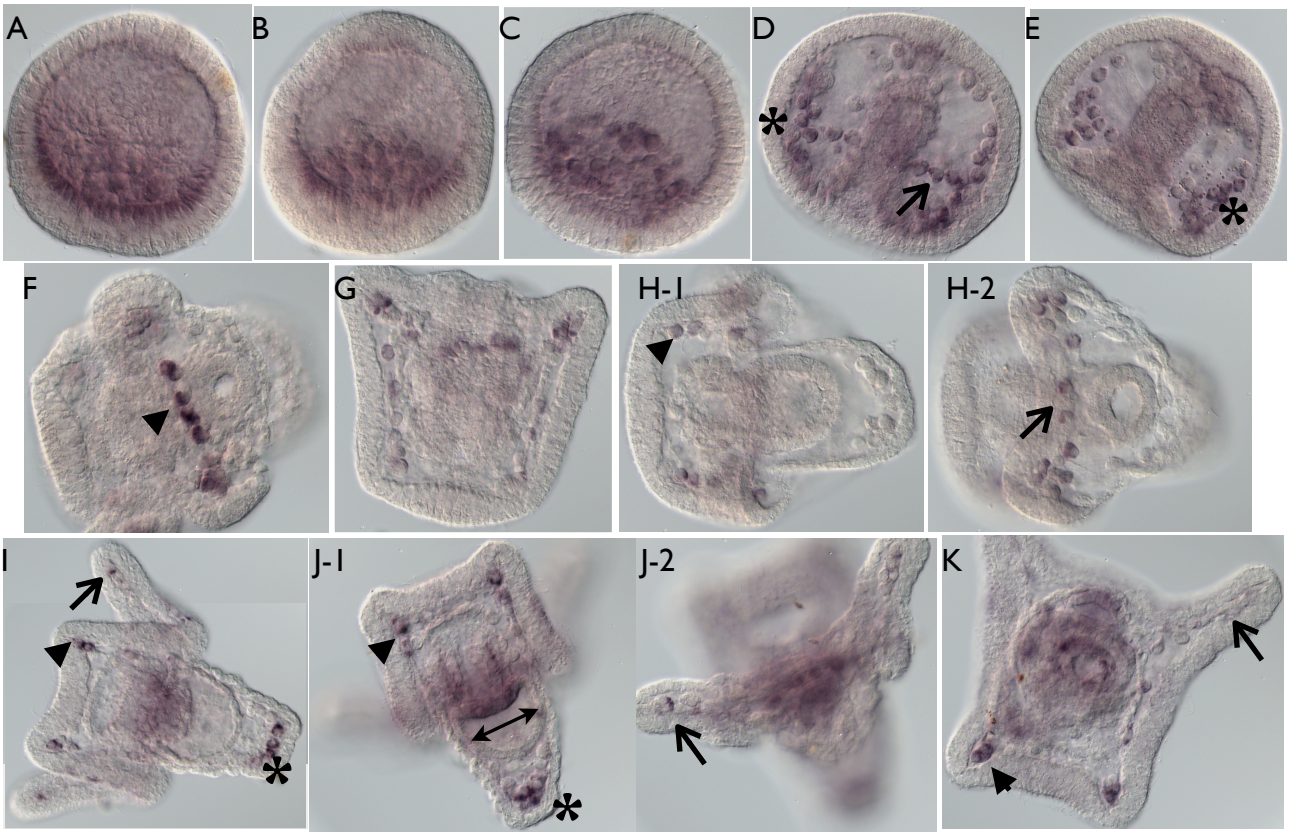


Figure 30. Spatial expression of *Pl-advillin.* Whole mount *in situ* hybridization showing the localization of the transcripts during *P. lividus* skeletogenic embryonic stages. A: swimming blastula; B: mesenchyme blastula; C: early gastrula; D: mid gastrula (Asterisk: ventrolateral cluster; Arrow: ventral chain); E: late gastrula (Asterisk: ventrolateral cluster); F,G: prisms (Arrowhead: postoral chain); H-K: pluteus (Arrows: postoral rods; Arrowheads: Anterolateral rods; Asterisk: the scheitel; Double-head arrow: ventral transverse rod). Each letter corresponds to an individual embryo and each number focuses different territories of the embryo.

Tetraspanin

Temporal gene expression profiling of *Pl-tetraspanin*

The temporal expression profile, as revealed by comparative real-time qPCR (ΔCCt Q-PCR), showed that the expression of *Pl-tetraspanin* is detectable and maintained in the cleavage and blastula stages, slightly increasing by two fold in gastrula and three fold in pluteus (Fig.31).

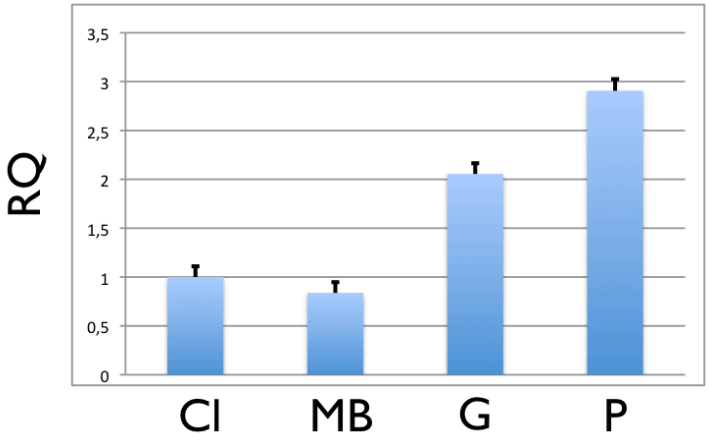


Figure 31. Temporal expression of *Pl-tetraspanin*. Transcript levels from cleavage to pluteus stage were measured by comparative real-time q-PCR: Cl, cleavage stage; MB, mesenchyme blastula; G, gastrula; P, pluteus; Each point represents the fold change (RQ value) calculated from the mean of three replicate reactions. The value of the cleavage stage was used as reference and assumed as 1. The endogenous gene *Pl-12-1* was used for normalization.

Spatial expression patterns of *Pl-tetraspanin*

To investigate the spatial expression patterns of *Pl-tetraspanin* during developmental, WMISH was carried out on embryos fixed at the stages from mesenchyme blastula to pluteus stage. Embryos at developmental stages prior to prism, did not show any evident labeling. Late gastrula: Fig.32-A. In fact, *Pl-tetraspanin* was exclusively expressed in the aboral ectoderm throughout embryo development from prism to pluteus (Fig.32-B-G, see arrow). Transcripts were exclusively localized in the ectoderm (Fig.32-B-D and E-2 arrow).

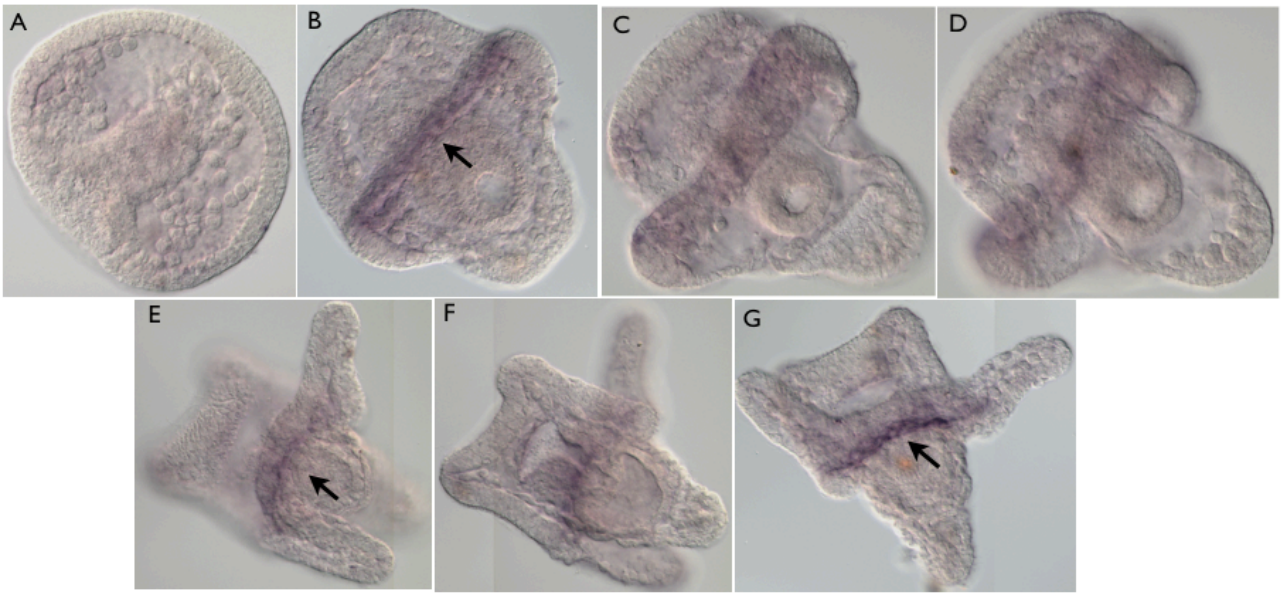


Figure 32. Spatial expression of *Pl-advillin.* Whole mount *in situ* hybridization showing *Pl-tetraspanin* expression during *P. lividus* skeletogenic embryonic stages. A: late gastrula; B: prism; C,D: early pluteus; E,F,G: pluteus. Arrows indicate the marked ectodermal expression. Images C and D and also E and F focus on different territories of the same embryos. Image G is a collage of two photos.

SM30α and SM50

Temporal gene expression profiling of *Pl-sm30α* and *Pl-sm50*

ΔCCt Q-PCR analysis of *Pl-sm30a* expression in early development revealed a well regulated expression pattern. Compared to cleavage stages, expression levels are increased by 4,9 fold at blastula and 12,6 fold in gastrula to reach a remarkably high level of 969 fold increase at pluteus (Fig.33). The temporal expression profile of *Pl-sm50,* showed that the copy number of *Pl-sm50* mRNA is increased by approximately 30 fold at the gastrula stage and by 100 fold at pluteus, compared to the blastula stage (Fig.33).

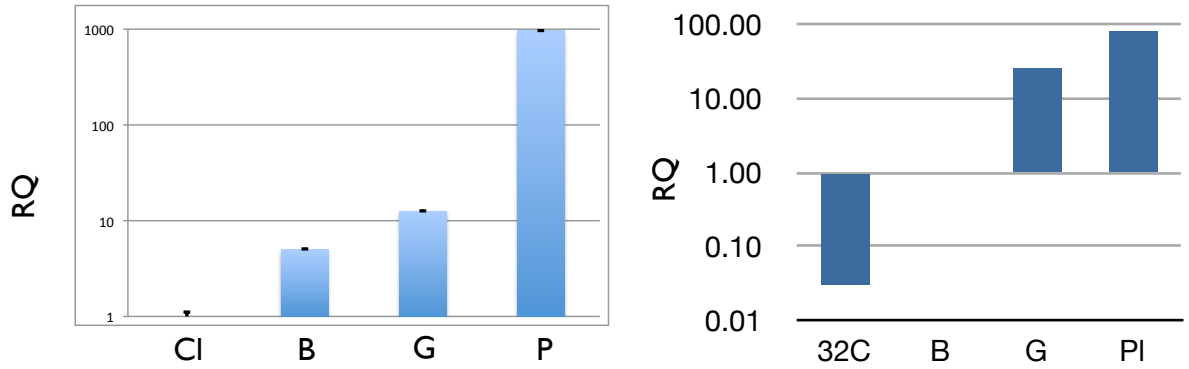


Figure 33. Temporal expression of *Pl-SM30α* and *Pl-SM50.* Transcript levels from cleavage to pluteus stage were measured by comparative real-time q-PCR. Left: sm30a; Right: sm50. Cl, cleavage stage; B, blastula; G, gastrula; P, pluteus; Each point represents the fold change (RQ value) calculated from the mean of three replicate reactions. In the qPCR measuring the expression of *Pl-sm30a*, the value of the cleavage stage was used as reference and assumed as 1, while for *Pl-sm50*, the value from the blastula stage was used as a reference equal to 1, for the better representation of the results. The endogenous gene *Pl-Z12-1* was used for normalization.

Spatial expression patterns of *Pl-sm30α* and *Pl-sm50*

The spatial localization of sm30α and sm50 transcripts is not shown as the results were identical to those found in the literature and to those previously carried out in the lab.

Construction of expression plasmids

For the production of recombinant proteins in E. *coli.*, in view of functional protein characterization, we cloned the full-length CDS of each gene in the expression vectors, p*EXP-NT* (Invitrogen) and p*TRC-CT* (Invitrogen). The full-length coding sequences were amplified by PCR from the *pGEM-T-Easy* plasmids with the primers outlined in the Table 5

and the generated inserts (Fig.34) were ligated in the expression vector p*EXP-NT* (Fig.34-C-H) and p*TRC-CT* (Fig.34-I).

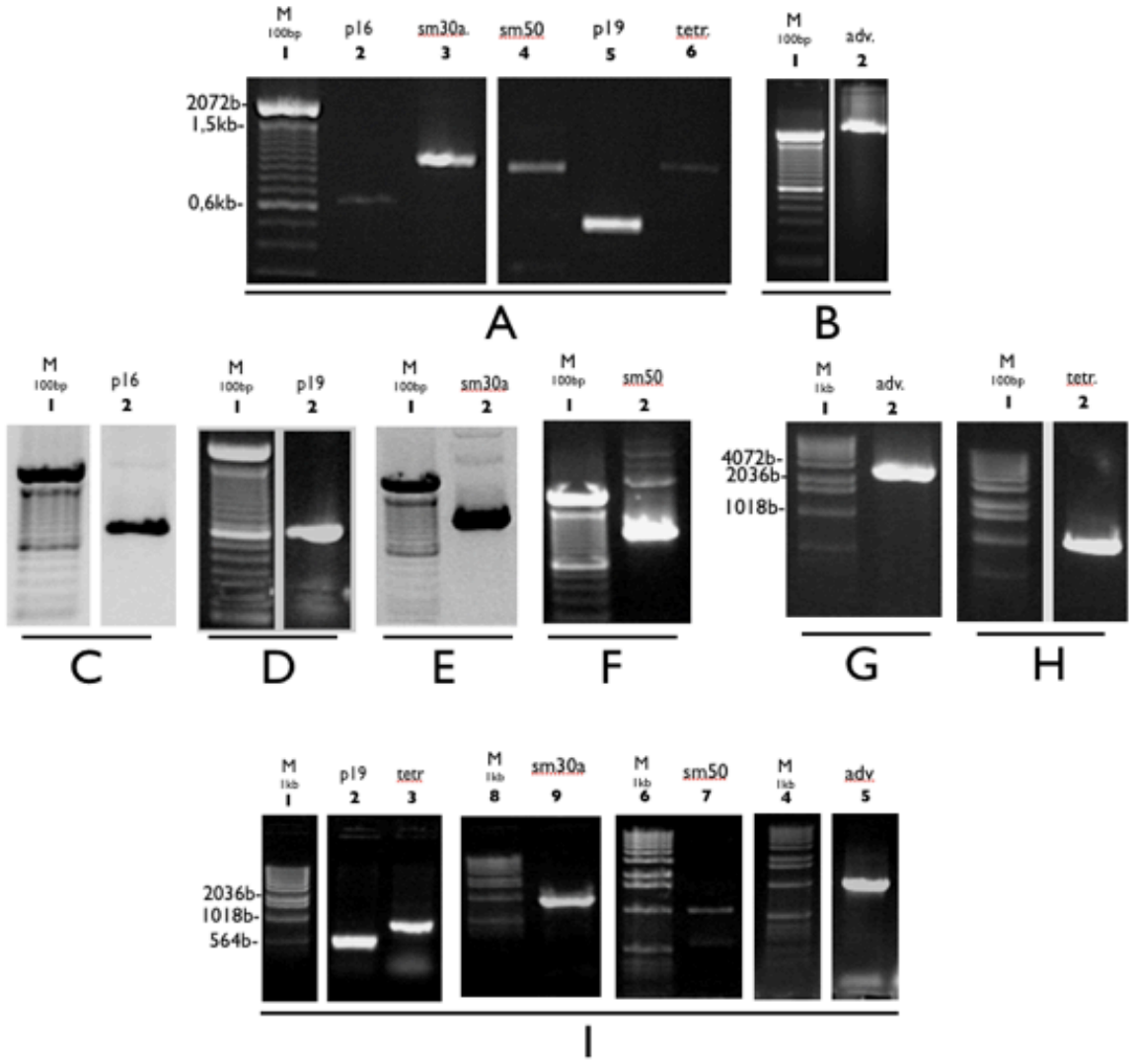


Figure 34. Representative images from the clonings of *p16, p19, sm30a, sm50, advillin* and *tetraspanin* CDSs in the p*EXP-NT* (A-H) and p*TRC-CT* (I) expression vectors. (A): PCR amplification of each insert from p*GEM-T-Easy;* (B) RT-PCR amplification of the *advillin* insert from total RNA extracted from gastrula; (C-H): PCR amplification of each insert from the cloned expression vector p*EXP-NT* with *T7* and insert-specific reverse primers; (I) PCR amplification of the insert from the cloned expression vector p*TRC-CT* with *T7* and insert-specific reverse primers.

Chapter IV: C*arbonic anhydrase* from *P. lividus,* is an active enzyme, involved in biomineralization and is expressed specifically in the Primary Mesenchyme Cells of the embryo and in the adult test.

Abstract

Carbonic anhydrases are zinc metalloenzymes which catalyze the reversible hydration of carbon dioxide to bicarbonate. There are five evolutionarily distinct CA families (alpha, beta, gamma, delta and epsilon) found in many species and in various forms without significant sequence similarity or common protein structure identity. In the sea urchin, carbonic anhydrase (CA) plays an important role in the formation of the calcitic skeleton, both during embryo spicule formation and in the adult mineralized tissue. Here, we report a PMC specific α-type carbonic anhydrase-coding mRNA sequence from the sea urchin *Paracentrotus lividus* (*P. lividus*) embryo, referred to as *Pl-CA*. The full-length CDS was identified by EST data mining and 3' and 5' RACE PCR. The deduced amino acidic sequence was used for the prediction of the functional domains and in phylogenetic analysis. The temporal and spatial expression profile was analyzed throughout embryo development by qPCR and whole mount *in situ* hybridization (WMISH). Differences in the localization of the expression were documented and compared to other sea urchin species. For a functional characterization, an enzymatically active recombinant *Pl*-CA was produced in E. *coli* and used for the raise of specific polyclonal antibodies in mouse, which identified by Western Blot, Pl-CA protein in embryos and in the organic matrix of adult *P. lividus* tests. This work provides with novel information and molecular tools for the characterization of an important biomineralization agent, the carbonic anhydrase.

Introduction

Carbonic anhydrase (CA) is a zinc containing metalloenzyme catalyzing the reversible hydration of CO_2. It has been purified from mammals, vertebrates, invertebrates, diatoms, cyanobacteria, algae, bacteria and archaea (Ilies *et al.*, 2004). Presently, CAs are divided into three main classes: α, β, γ and two secondary (δ and ε), which have no significant primary sequence similarity and are evolutionarily independent (Smith *et al.*, 2000; Tripp *et al.*, 2001; So *et al.*, 2004). They play important roles in a wide range of biological processes, such as photosynthesis, respiration; pH homeostasis, ion transport and inorganic carbon (Ci) transport (Pushkas *et al.*, 2000) and skeletogenesis (Mann, 2002).

Indeed, the formation of calcified crystals of phosphate or carbonate in vertebrates (Kakei *et al.*, 1996) and in invertebrates such as corals, shellfishes and sea urchins (Mann, 2002; Westbroek and Marin, 1998) serve in various functions such as supporting and protecting the body. Skeletogenesis involves the deposition of ions in the mineralization site, the nucleation, the growth/elongation and the patterning of the formed crystal. CAs from various species are evidenced to play important roles in all the stages of biomineral formation during skeletogenesis. The presence of CA in vertebrate hard tissues which bare minerals of calcium phosphate, is linked to the supply of carbonate ions which become incorporated into the initial mineralization site and initiate the nucleation (Kakei *et al.*, 1996). Similarly, CA is believed to be involved in the biomineralization process via the bicarbonate or carbonate supply in the sea urchin (Mann *et al.*, 2008; Livington *et al.*, 2006; Hofmann *et al.*, 2008), in Mollusca (Medakovic, 2000; Yu *et al.*, 2006); Marie *et al.*, 2008), in Cnidaria (Kingsley and Watabe, 1987, Ip *et al.*, 1991) and in the gland responsible for eggshell formation in birds (Gay and Mueller, 1973). In contrast, in marine sponges, the enzyme silicase belonging to the family of CAs (Müller *et al.*, 2007; Kirill *et al.*, 2011), provides silica depolymerization in a mechanism similar to zinc-dependent metalloenzymes hydrolyzing ethers (Schroeder *et al.*, 2007, Wang *et al.*, 2012)

In bones, calcium phosphate is resorbed by osteoclasts. Carbonate stimulates the carbonic anhydrase activity promoting the osteoclastic acidic secretion *in vitro* (Doi *et al.*, 1999; Leeuwenburgh *et al.*, 2001).

In the sea urchin, carbonic anhydrase is thought to play a central role in the formation of the endoskeleton of the embryo and in the biomineralization of the adult tissue. Livingstone *et al.*, 2006, reported 19 genes encoding CA from the *S. purpuratus* genome, three of which are present in the PMC EST library. The expression of one of them (Accession number: SPU_012518), was found prominent at the prism stage as well as in the adult spines (Zhu *et al.*, 2001). Studies on orthologs CAs in *H. tuberculata* and *H. erythrograma* demonstrated regulated expression of CA in the PMCs, at the growing ends of spicules (Love *et al.*, 2007). CA protein has been extracted and identified by Mass Spectrometry (MS) from the spicule matrix of *S. purpuratus* (Mann *et al.*, 2010) and from adult *S. purpuratus* test and spines (Mann, 2008). Nevertheless, the biological role of CA in the sea urchin biomineralization remains unclear; it has been evidenced though that, as in other biological systems, the formation and patterning of carbonate crystals is facilitated by the active transport of calcium and bicarbonate ions from the water, establishing an

oversaturated micro-environment at the calcification front (Simkiss and Wilbur, 1989; Hofmann *et al.*, 2008). Studies on the adult tooth concluded that CA facilitated the transportation of CO_2 from the external medium to the soft tissue and from the soft tissue to the calcareous parts (Chang-Po Chen *et al.*, 1987). Marshall, 1996 showed that CA is required for the normal skeletal development of the animal by contributing to the deposition of $CaCO_3$ in the sea urchin spicule (Mitsunaga *et al.*, 1986). Acetazolamide (Diamox), a potent inhibitor of CA, blocked the spicule formation or elongation in sea urchin embryos (Chow *et al.*, 1979), as well as the accumulation of bicarbonate in otoliths, resulting in the inhibition of otolith calcification (Tohse *et al.*, 2001). Recent studies on other non-mammalian species demonstrate the active role of proteins bearing carbonic anhydrase activity during the formation of calcified tissues, via the bicarbonate supply. In This is the case of nacrein, a protein which contains Gly–X–Asn repeats and two CA-active domains, isolated from the pearl oyster shell. The functional domains suggested that it has both calcium-binding and CA activities (Miyamoto *et al.*, 1996).

Here, we focused on the identification and characterization of a PMC specific α-type carbonic anhydrase-coding mRNA sequence from the sea urchin *Paracentrotus lividus* (*P. lividus*) embryo, referred as *Pl-CA*.

Experimental Procedures

The experimental procedure for the embryo culture, total RNA extraction and RT-PCR is described in Chapter II

Cloning of *Pl-CA*

A sequence predicted to code for a carbonic anhydrase protein in the sea urchin *S. purpuratus* database (accession number: XP_784328 (Love *et al.*, 2007), was used for BLAST screening against *P. lividus* EST databases available at NCBI and MPIMG (http://goblet.molgen.mpg.de/cgi-bin/webapps/paracentrotus.cgi). A single partial *P. lividus* similar EST clone (accession number: AM572836.1), demonstrating 81% of homology and 88% of similarity, was identified (see Results Section for description). Other clones, showing significant similarity, were the SP0AGASPL124YL19RM1 and the contig11217.

We followed different methods for the identification of the full-length CDS of CA in *P. lividus*. Our aim was to identify the CA isoform expressed in the PMCs. Various isoforms of CA were found to have similar cDNA sequences. To obtain the full-length CDS sequence of the PMC-specific CA, different strategies were followed. A total of fourteen primers were designed on the basis of the UTR and CDS of the carbonic anhydrase cDNA from *S. purpuratus*, and used in trial amplifications templated with total RNA extracted from *P. lividus* embryos at the gastrula and pluteus stages, by the One Step RT-PCR kit (Invitrogen) (Table 8). Fig.35-D-1 illustrates the primer sets 1-6, on the CA sequence from

S. purpuratus. Furthermore, two degenerated oligos were designed and purchased from MWG (Heidelberg, Germany). The primer sequences are outlined in Table 8.

Table 8. Primers Designed complementary to the *S. purputatus* CA sequence (XM_001177813.1) or to *P. lividus* ESTs (SP0AGASPL124YL19RM1 and the contig11217), used for amplification of PI-CA. The template EST used for the design of each primer set is noted at the Template column). The corresponding primer sets from #1 to #6, are illustrated on the CA coding sequence of *S. purpuratus* (Fig.35).

	primer set	Template	Forward primer (5'-3')	Reverse primer (5'-3')	PCR product length (bp)
1	LI1	CDS XM_001177813.1 (*Sp*)	GACGTTGTTGTACCAAGAATGTT TAG	TCCGTTGCCGTTCCACC AAC	388
2	LI2	CDS XM_001177813.1 (*Sp*)	GCACGGACCATTGGGACCGGA	TCCACTGCCACCTTGTCC ACCT	220
3	LI3	CDS XM_001177813.1 (*Sp*)	TGGGACCGGAAAACTGGCCGA	TGCCTCCCCATCCGTTAC CGT	324
4	LI4	CDS XM_001177813.1 (*Sp*)	GGTCGAGCGCTGGATCGCC	CTGTTGTGAACATTCTTT CGGGTCCGT	600
5	LI5	CDS XM_001177813.1 (*Sp*)	TCACTGGGGAACCACACCGGA	GCGGCTGTACCGGTCGG AAG	488
6	LO6	5' UTR XM_001177813.1 (*Sp*)	TCTCAGGCCCAAAACCTGTATT CCTCT	TAGCTCATGAAGAGCACA GCGGTGAA	160
7	LI7	*Pl* EST	-	AGCTCTCGGAACATGTCA	202
8	7I1	*Pl* EST CDS SP0AGASPL124YL19RM1 -	TGCCAGCTGACATGAGCTGCTT C	AGCTGAT CAGCCGGCAAACAAATG CCATTAGAG	362
9	7O1	*Pl* EST 5' UTR SP0AGASPL124YL19RM1 -	ACATCGGAATGAGTTAAAGCGC AAACA	GCTTAAGGGGAAGCAAG GCCGA	301
10	14I1	*Pl* EST CDS SP0AEGG64YC09RM1	GATGGTCGTGACACGTGGGC	TGCAGCTCGGCAGCATA GGT	336
11	14O1	*Pl* EST 5' UTR SP0AEGG64YC09RM1	CTCGACCGTGACGTTACACCGC	TGGGCGTTGGCCTTCTA CCAGA	150
12	CO1	*Pl* EST 3' UTR contig11217	CTCCCATTCTGATTGGTTCACG	GAGCCCGAGATGGCACA	470
13	DegCA1	CDS XM_001177813.1 (*Sp*)	GTCAG gacggtcnttbtggawgcgac	GGC -	-
14	DegCA2	CDS XM_001177813.1 (*Sp*)	gaWGGTCGANCGCTNGATC	-	-

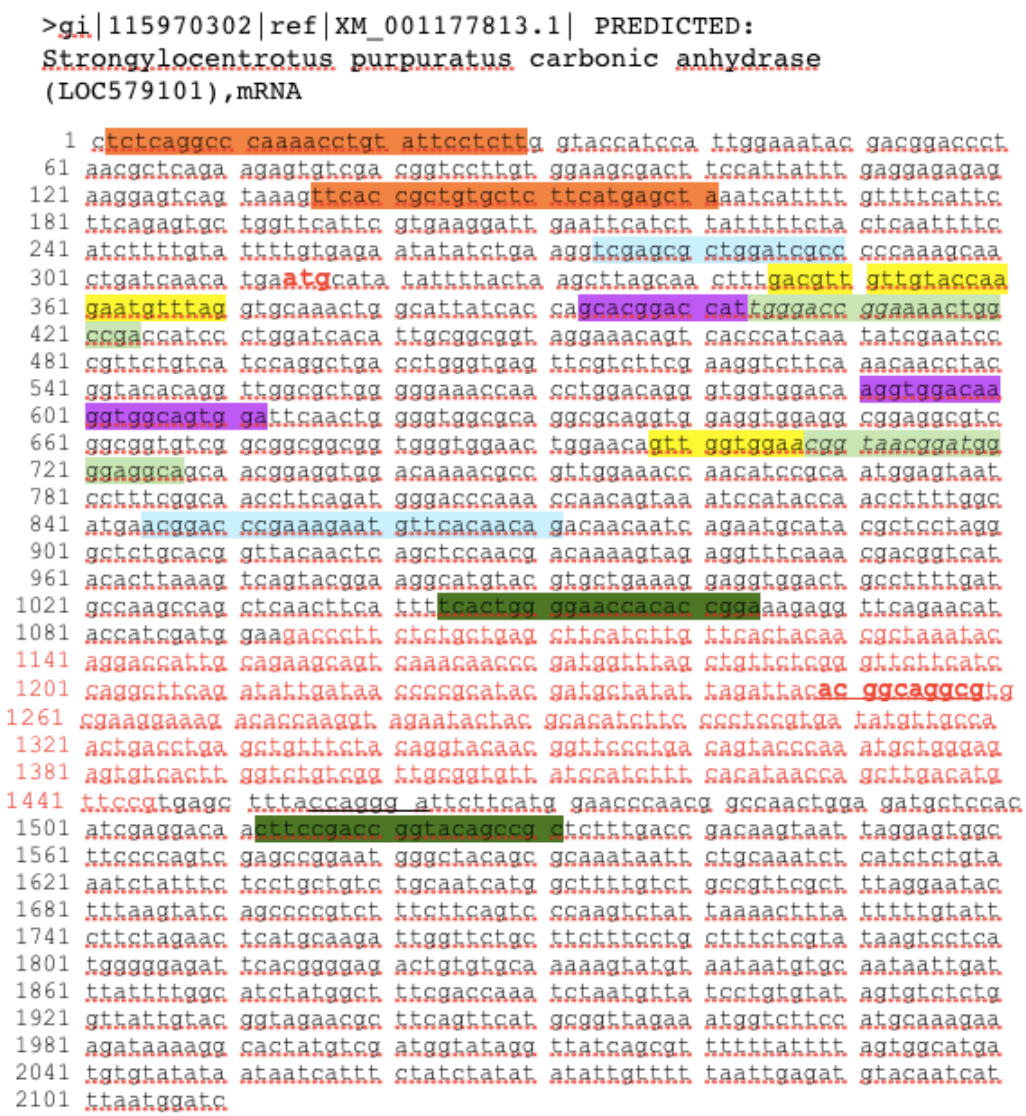


Figure 35. Illustration of the primers designed complementary to the *S. purputatus* CA sequence (XM_001177813.1), used for amplification of PI-CA. The corresponding primer sets from illustrated in colors, are summarized in Table 8.

Amplifications by RT-PCR (One Step RT-PCR, Invitrogen), using purified total RNA extracted from the gastrula and pluteus stages, gave PCR products, which were subsequently cloned in the p*GEM-T-Easy* Vector (Promega) and sequenced by MWG (Heidelberg, Germany). As the obtained sequence revealed only partial CDS, lacking both the 3' and the 5' terminals, 3' and 5' RACE primers were designed (Table 9) and 3' and 5' RACE was applied using a 3'RACE kit (Invitrogen) and a 5'RACE kit (Invitrogen) and total RNA extracted from embryos at the pluteus stage, as template. Primers were designed from the partial cDNA sequence, following the manufacturer's instructions. For the 3'RACE PCR, 5-'ACGCACATTTGCCCCTCCGT-3' was used for reverse transcription and 5-'TGGTCGGTCGGCTGCAATCC-3' for nested amplification as gene-specific primers (GSP). For the 5'RACE PCR, 5-'AGCTGATTATGTGAAAGATG-3' was used for reverse transcription and four primers were designed for nested amplifications. These were: 5-'GATTGCAGCCGACCGACCATG-3', 5-'CTCATGTCAGCTGGCAGCATATCAC-3',

5-'GCTAAACCATCGGGCTGTTTAAG-3' and 5-'GCATCGTAAGCAGTGTTATC-3' (Table 9). These primers, were coupled, respectively, to the primers with annealing site at the 3' and 5' extremity. The 5' and 3' end amplification products were then migrated on an agarose gel electrophoresis, purified and cloned in the p*GEM-T-Easy* vector (Promega) and the recombinant plasmids were then sequenced at MWG (Heidelberg, Germany).

Table 9. Summary of primers used in 3' and 5' RACE PCR for the identification of the CDS of *PI-CA* and for cloning of PI-CA in the expression vectors *pTRC-CT, pEXP-NT* and *pET32b*$^{+}$.

process	forward primer sequence	reverse primer sequence	amplification length (bp)
3'RACE reverse transcription	ACGCACATTTGCCCCTCC GT	-	(RACE)
3'RACE nested amplification	TGGTCGGTCGGCTGCAAT CC	-	(RACE)
5'RACE reverse transcription	-	AGCTGATTATGTGAAAG ATG	(RACE)
5'RACE nested amplification	-	GATTGCAGCCGACCGA CCATG	(RACE)
5'RACE nested amplification	-	CTCATGTCAGCTGGCA GCATATCAC	(RACE)
5'RACE nested amplification	-	GCTAAACCATCGGGCT GTTTAAG	(RACE)
5'RACE nested amplification	-	GCATCGTAAGCAGTGTT ATC	(RACE)
Second 5'RACE reverse transcription	-	ACGTACATGTTTTCTGTGC	(RACE)
Second 5'RACE reverse transcription	-	AGCCATGAAGACCTCTCG GAGAT	(RACE)
Second 5'RACE reverse transcription	-	GCCAAAAGGCATGTTTGG GTTGA	(RACE)
comparative qPCR	CCAAAATGCTGGGAAAGT GTAAC	TCGGAACATGTCAAGCT GATTATG	82
RT-PCR of partial CDS including the functional CA domain	GAACATCAGAACTGGGCT AAC	GACTTGCCAGAGTTGT TTGC	1027
cloning in p*EXP-5N*	ATGCCTTTTGGCCTACAT GG	-//-	955
cloning in p*Trc-CT*	-//-	TAGCAGCCGGCAAACA AATGC	858
cloning in p*Cold-TF*	GGACATATGCCTTTTGGC C	GCTCTAGACTATAGCAG CC	861
cloning in p*ET32b*$^{+}$	TTAGGATCCGATGCCTTTT GGCC	CATTACTCGAGTAGCAG CCGG	861

Finally, the derived sequences were merged and new primers were designed for the amplification of the CDS by RT-PCR. 5'-GAACATCAGAACTGGGCTAAC-3' was used as a forward and 5'-GACTTGCCAGAGTTGTTTGC-3' as reverse. A fragment of 1027 bp, involving a partial CDS, was cloned in the p*GEM-T-Easy* vector (Promega) and sequenced at MWG (Heidelberg, Germany). The plasmid was designated as p*GEM-T-Easy-Pl-CA.*

In silico analysis demonstrated that the compiled sequence derived from the EST data mining and the 3' and 5' RACE PCRs, involved the conserved CA functional domain but was partial, lacking the N'terminal. Further: the genome project of *P. lividus* revealed a new EST sequence involving the the missing N' end of the CDS of *Pl*-CA. This sequence involves 483 bp of additional 5' CDS and 307 bp of 5' UTR. Thus, a new set of primers was designed and purchased from MWG (Heidelberg, Germany): 5'-TTAGGATCCGCCAAAGCCCACAATCAACCCAAACATG-3' as forward and the 5'-CACTCGAGTAGCAGCCGGCAAACAAATGCC-3' as reverse and the full-length sequence was amplified by One Step RT-PCR (Invitrogen) from total RNA from embryos at pluteus stage. The amplification product was sequenced by MWG (Heidelberg, Germany). The full-length sequence was used for the characterization of the functional domains by *in silico* analysis and phylogenetic analysis.

Domain characterisation and phylogenetic analysis

The putative aa sequence was characterized as described in Chapter II.

Semi-quantitative characterization of the temporal expression profile of CA, by RT-PCR.

The temporal expression pattern of CA throughout the development of the embryo, was monitored by RT-PCR. Specific primers were designed to amplify a 620 bp fragment. 5'-GAATACTACGCACATTTGC-3' was used as a forward and 5'-GTTGTTCACGATGATTTGC-3' as a reverse. 100 ng of mRNA extracted from nine embryonic developmental stages were used as templates. Reverse transcription and PCR amplification were performed by the RT/Platinum Taq polymerase kit (Invitrogen), following the manufacturer's instructions. The gene S24 was used as a reference gene. PCR products were visualized in a 1% agarose gel.

Comparative Real Time qPCR and Whole-mount *in situ* hybridization (WMISH)

Gene expression profiling was monitored throughout embryo development by comparative q-PCR and whole mount *in situ* hybridization, as described in Chapter II. The primers used, were: 5'-CCAAAATGCTGGGAAAGTGTAAC-3', as forward and 5'-TCGGAACATGTCAAGCTGATTATG-3', as reverse, designed to amplify a fragment of 82bp. For the WMISH, the plasmid *pGEM-T-Easy-Pl-CA* was used as template, after

enzymatic digestion with *Sac*II (NEB) and *Sp6* primer was used for the production of an antisense full-length CDS DIG-labelled probe. Sense probes were used as negative controls.

Production of recombinant *PI-CA* plasmids in E. *coli*

Truncated CA CDSs of 955 bp, 858 bp, 861 bp and 861 bp, were respectively cloned in the expression vectors: p*EXP-NT* (Invitrogen), p*TRC-CT* (Invitrogen), p*COLD-TF* (TAKARA) and p*ET32b*$^{+}$ (Novagen) acquiring respectively the plasmids p*EXP-N-CA*, p*TRC-C-CA*, p*COLD-TF-CA* and p*ET*-CA. For each plasmid, different cloning procedure was followed. Table 9 summarizes the primers used for the cloning of each plasmid.

Briefly, p*GEM-T-Easy-PI-CA* was used as a template for all the amplifications of the truncated CA CDS. The p*EXP-NT* and p*TRC-CT* use the TA cloning strategy with the T7 promoter; thus, the insert was amplified with Taq Polymerase (Invitrogen) introducing an extra Adenine at the 3'end of the PCR product. This facilitated a direct ligation of the insert with the linearized vector carrying an extra Thimine at the 3'ends. p*EXP-NT* introduces a His_6 tag at the N' terminal of the recombinant protein, while p*TRC-CT* at the C' terminal. The p*Cold-TF* expression vector uses the *cspA* (cold shock protein A) promoter, introducing an N' terminal fused chaperone trigger factor of 48 kDa and an N-terminal hexahistidine (His_6) tag. The *Nde*I and *Xba*I sites were introduced in the primers and used for a directional cloning.

The pET32b$^{(+)}$ vector (Novagen) containing the strong T7 promoter, a fusion protein thioredoxin (Trx•Tag™) of 109 aa and an N' and a C'-terminal hexahistidine (His_6) tags, was used for high expression of the PI-CA. The coding region of PI-CA involving the CA functional domain, was amplified by PCR from the p*GEM-T-Easy-PI-CA* plasmid with the primers 5'-TTAGGATCCGATGCCTTTTGGCC-3' as forward and 5'-CATTACTCGAGTAGCAGCCGG-3') as reverse, fragment containing BamHI and XhoI restriction sites respectively and generating a 861 bp amplification product.

The plasmids were confirmed by direct sequencing at MWG (Heidelberg, Germany)

Protein expression and purification

During the trial experiments of protein expression, each one of the three plasmids p*EXP-CA*, p*TRC-CA* and p*COLD-TF-CA* were inserted in M15 and Bl.21(DE3) E. *coli*. cells following standard transformation procedure and variable expression conditions of induction, temperature and incubation duration, were tested. Expression was induced when reached a cell Optical Density (OD) ranging from 0.3 to 0.8 and expression length was tested from 1h to 16 h at 16 ^{o}C to 37 ^{o}C. Additionally, the S30 T7 High-Yield Protein Expression System (Promega) was used for cell-free expression with the p*EXP-CA* and p*TRC-CA* plasmids, following the manufacturer's instructions. However, the expression yield was found particularly low. Therefore, the recombinant p*ET-PI-CA* plasmid was used to express *PI-CA* protein in E. *coli* BL21(DE3) (Novagen), after induction with 1mM

isopropyl-b-D-thiogalactopyranoside (IPTG) (Sigma) at 28 ^{o}C for 8h, in presence of 0.1 mM $ZnSO_4$. The cells were harvested by centrifugation at 4,000 rpm at 4 ^{o}C for 10 min, resuspended in lysis buffer (50 mM Tris, 500 mM NaCl, 2 mM PMSF (phenylmethanesulfonylfluoride) and 5mM imidazole, pH 7.5), disrupted by sonication and centrifuged at 10,000g at 4 ^{o}C for 30 min. The supernatant was applied to a Ni-nitrilotriacetate column (Invitrogen), pre-equilibrated with lysis buffer. After washing of the column in: 50 mM Tris-HCl, 50 mM NaCl, 0.1 mM $ZnSO_4$, 6M Urea, 2 mM PMSF and 10 mM imidazole, pH 7.5), 1 ml fractions were eluted in: (50 mM Tris, 100 mM NaCl, 6M Urea, 2 mM PMSF, 300 mM imidazole, pH 7.5). The fractions containing the recombinant protein (checked by Bradford and 7,5% SDS–PAGE) were pooled and concentrated by centrifugation in 'Amicon Ultra' centrifugal filter units, 50 kDA cut-off (www.millipore.com). After dialysis against 50 mM Tris, 50 mM NaCl, pH 7.5, the protein concentration was determined using the Bradford assay (Bio-Rad) with bovine serum albumin (BSA)(Sigma) as a protein standard and the purity was analyzed on a 10% SDS polyacrylamide gel and refolding was facilitated by incubation at 4 ^{o}C in renaturation buffer: Tris-Cl 50mM, Glutathione reduced 9mM, Glutathione oxidised 1mM, Arginine 0,5M, NaCl 300mM and KCl 1mM, pH=8,5. Finally, the soluble protein was purified by dialysis at 4 ^{o}C against 50 mM Tris-SO_4 (pH 7.5) and 50 mM NaCl with two changes over the course of 24 h. The total protein concentration was determined using a Bradford assay (Bio-Rad) with bovine serum albumin (BSA; Sigma) as a protein standard and the purity was analysed on a 10% SDS polyacrylamide gel.

Raise of polyclonal antibodies in mice

Polyclonal antibodies (pAb) were prepared by Dr. B.Diehl-Seifert, (University of Mainz, Germany) against the purified, recombinant Pl-*CA-8* (*rPl-CA*) in mice. 1.5 µg of recombinant protein per injection was dissolved in phosphate-buffered saline (PBS). After three boosts the serum was collected; the pAb against *rPl-CA*, was termed pAb-CA. The titer of the pAb-CA was 1:4000. The cross-reactivity was tested against 10 µg/ml of recombinant protein by ELISA.

Western blot analysis

Total cell lysates from embryos at the gastrula and pluteus stages, were separated by electrophoresis in a 10% SDS-PAGE gel. The resolved proteins were transferred onto nitrocellulose membranes (Amersham) and immunodetection was carried out as described by (Pinsino *et al.*, 2010), at a dilution of 1-8000 of the anti-Pl-CA antibody.

Determination of esterase activity

The esterase activity of the refolded recombinant *Pl*-CA was measured spectrophotometrically using p-nitrophenyl acetate as a substrate according to the method described by Armstrong *et al.*, 1966 with modifications. The assay was carried out in a 96 well plate. 5 µg of recombinant *Pl-CA* protein were diluted in 15mM Tris-H_2SO_4 (pH 8.2) and freshly prepared 3 mM p-nitrophenyl acetate in acetone. The enzymatic activity was monitored through a Titertek Multiskan Plus (Bartolomey Labortechnik, Rheinbach; Germany), which allowed to measure the time-dependent increase in absorbance (405 nm for 30 min). A blank control was prepared using only the substrate in buffer. As positive control, carbonic anhydrase from bovine erythrocytes 5 mg/ml (Sigma) was used. In a separate series of experiments the effect on the reaction of acetazolamide (AZM) (0.1 mM) (Sigma), a specific CA inhibitor, was also tested, in order to obtain the net esterase activity from the value of total esterase (Carlsson *et al.*, 1998).

Results

Identification and cloning of Pl-CA

Various sets of primers designed on the homologous CA sequence of *S. purpuratus* (XM_001177813.1) (Table 8) were used during trial RT-PCR amplifications for the identification of Pl-CA. Fig.36 shows representative 1% agarose-TBE gels of the amplification products. An amplification product of 202 bp resulted from the amplification using the primer set LI7 (5'-CCGCAGGAGTTAAAAGAAAAGGCACA-3' as forward and 5'-AGCTCTCGGAACATGTCAAGCTGAT-3', as reverse), was found to overlap with known EST sequence coding for a partial CA CDS. The other sets of the amplification products were found to encode for different isoforms of CA or for different genes which share similarity with CA. Therefore, the amplification product of 202 bp was used for the identification of the full-length CDS of CA, by 3' and 5' RACE PCR.

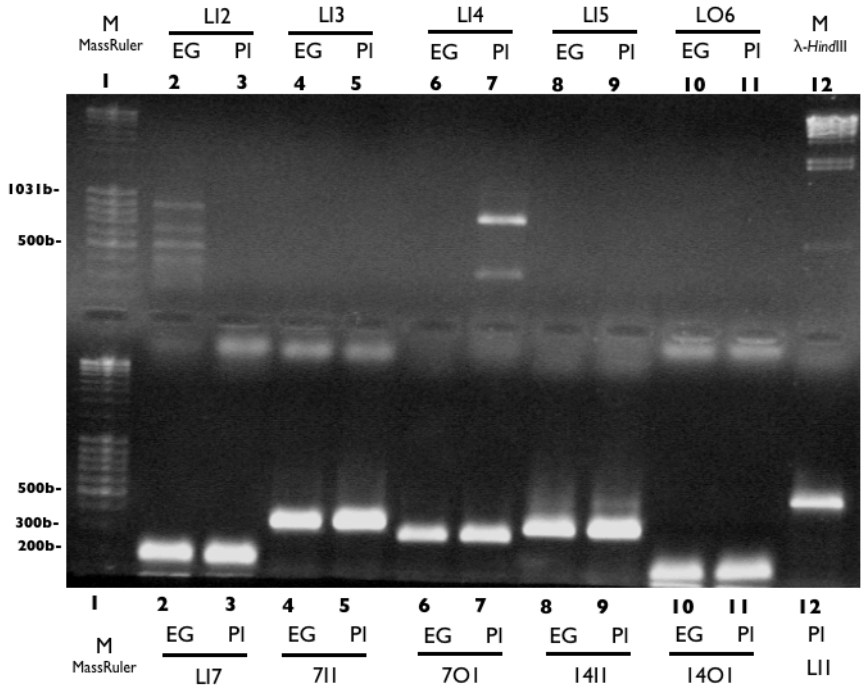


Figure 36. RT-PCR amplifications of cDNA fragments from *P. lividus* using primers designed on *S. purpuratus* orthologue, as viewed in a 1% agarose TBE gel.

Using a gene-specific primer designed on the basis of the sequence by NCBI accession number AM572836.1, a fragment of approximately 750 bp from total RNA cDNAs was amplified by the 3' RACE procedure. Subsequent sequencing analysis revealed that the fragment was 744 bp in length (Fig.37). On the other hand, a fragment of 478 bp was obtained after the 5' RACE procedure using gene-specific primers on the same cDNA sequence (Fig.38).

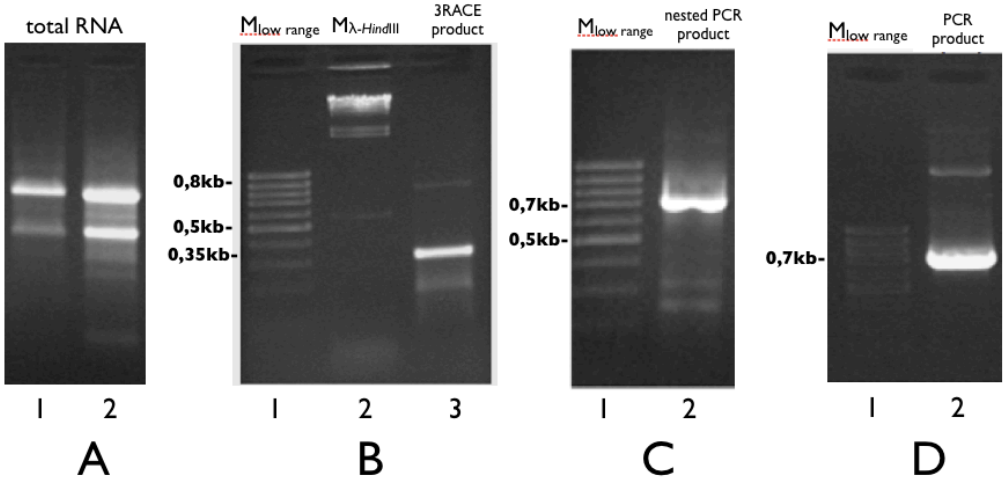


Figure 37: Summary of the 3'RACE cloning procedure of PI-CA. A: extraction and purification of total RNA; B: Amplification of PI-CA 3'RACE product by OneStep RT-PCR; C: Nested amplifications of PI-CA 3'RACE product by PCR; D: Amplification products by PCR templated with *PI-CA* 3'RACE product previously cloned in p*GEM-T-Easy* vector.

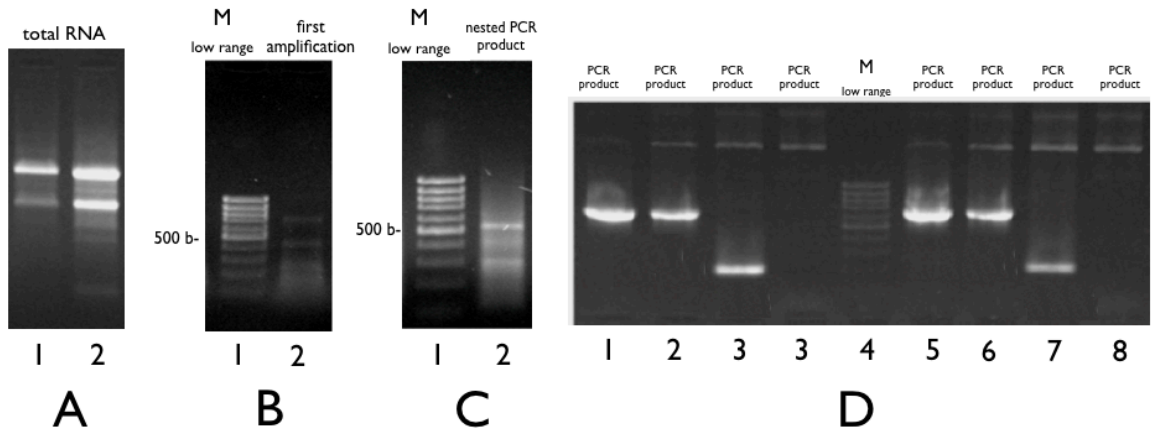


Figure 38: Summary of the 5‘RACE cloning procedure of Pl-CA. A: extraction and purification of total RNA; B: Amplification of Pl-CA 5‘RACE product by OneStep RT-PCR; C: Nested amplifications of Pl-CA 5’RACE product by PCR; D: Amplification products by PCR templated with *Pl-CA* 5’RACE product previously cloned in p*GEM-T-Easy* vector. Lanes 1,2,5 and 6 represent positive clones.

Both fragments resulted from the 3’ and 5’ RACE were assembled to give a merged cDNA sequence of 1429 bp in length with an ORF of 947 nucleotides (nt) and 446 nt of 3’-UTR including a putative polyadenylation signal (ATTAAA) which is 162 nucleotides upstream of the poly(A) tail starting at position 1389 bp. The derived CDS was amplified by RT-PCR, using the 5’-GAACATCAGAACTGGGCTAAC-3’ as forward and the 5’-GACTTGCCAGAGTTGTTTGC-3’ as reverse and cloned in the *pGEM-T-Easy* vector to produce the termed p*GEM-T-Easy-Pl-CA* involving a CDS of 1027 bp. The plasmid was tested by PCR (Fig.39) and sequenced by MWG (Hedelberg, Germany).

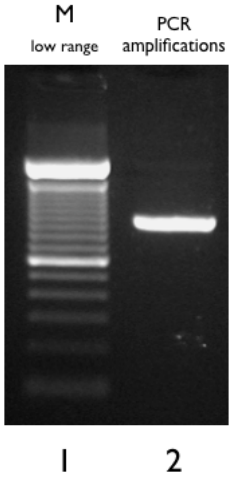


Figure 39. Amplification of 1027 bp CDS of *Pl-CA* by PCR from p*GEM-T-Easy-Pl-CA.*

It is mentioned that a second 5‘RACE PCR using the primers outlined in Table 9, was applied for the identification of the full-length CDS and the 5’ UTR of PL-CA. The resulted amplicon was a cDNA fragment of 600 bp which corresponded to a non CA coding sequence (Fig.40). Indeed, the sequencing revealed 89% similarity with the JNK-associated leucine-zipper protein from *S. purpuratus* (XM_001183252.1). It was concluded that the target sequence was of low complexity or shared by non-CA coding cDNAs.

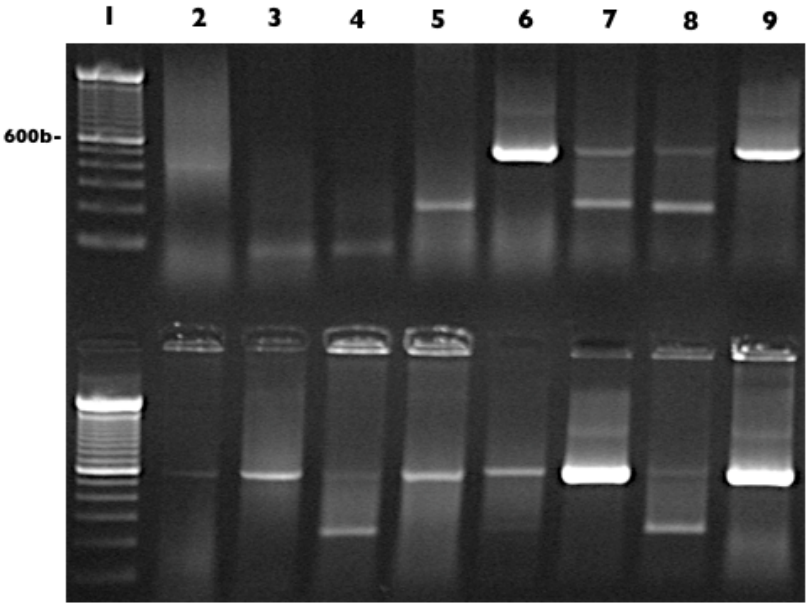


Figure 40. Resulted amplification products of an additional 5‘RACE. Samples represent the colony screening. Colonies in the upper lanes 6 and 9 and in the lower lanes 7 and 9 were identified by sequencing and found to be non-CA coding.

The partial *Pl-CA* compiled sequence lacking the 5’ end, involves the complete CA functional domain; therefore, it was used for the characterization of the tempo-spatial profile of Pl-Ca by qPCR and WMISH. It was also used for the preparation of a truncated version of a recombinant *Pl*-CA. On the other hand, for the characterisation of the functional domains by *in silico* analysis and for the phylogenetic analysis, the full-length sequence, comprising 1344 bp of CDS, was used.

Domain characterization of the full-length deduced amino acid Pl-CA sequence

The deduced protein is composed of 447 amino acids and has an estimated molecular mass of 48489.26 Da with an isoelectric point of 6.83. Protein domain analysis by the Motif Scan database revealed one alpha carbonic anhydrase isoform 2 domain from the 22th to the 406th amino acid and one glycine-rich region from the 78th to the 134th amino acid. The major residues of the protein were glycine (15%), leucine (8.9%) and asparagine (8.3%). The predicted secondary structure contains random coiled regions, 15 beta-sheet strands and 5 α-helix domains. Pl-CA has 20 putative phosphorylation sites as predicted by the

Centre for Biological Sequence Analysis server, BioCentrum-DTU Technical University of Denmark with high score which included 8 serine, 8 threonine and 4 tyrosine (Fig.44). A signal peptide from the 1st to the 21st amino acid residue (MNAYILLSLTTLTVLYQECLG) and a cleavage site between the 21st and 22nd amino acid, suggest that *Pl*-CA is a secreted protein (Fig.41, Fig.42 and Fig.43). After the cleavage, the predicted molecular weight of the protein is 46,15 kDa (427 aa). Finally, a hydrophobicity plot obtained by the Kyte and Doolittle calculation displays the overall hydrophilic character of the protein except at its C-terminal end (Fig.45). Altogether, these data indicate that Pl-CA, after the cleavage of the signal peptide, is an extracellular protein with the C-terminal of *Pl-CA*, embedded in the external side of the cell membrane.

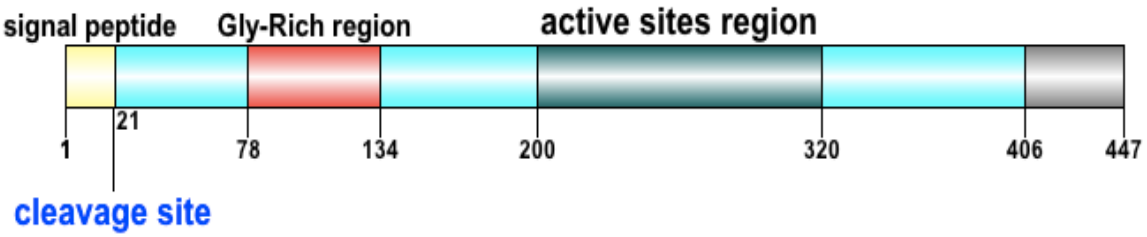


Figure 41. Illustration of protein domain structure of *Pl*-CA. In light blue, the CA-2 functional domain; in red, the Gly-Rich region and in yellow, the signal peptide. The cleavage site is marked (position 21). The active sites involving the Zn^{++} binding site are located within the 200th aa and 320th aa (green region). (Karakostis *et al.*, in preparation)

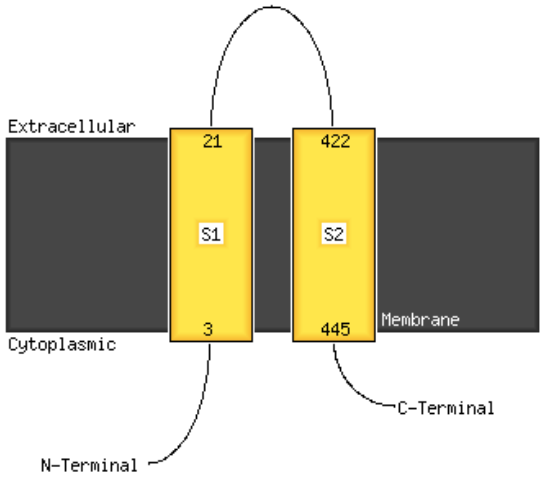


Figure 42. Illustration of the transmembrane topology of *Pl*-CA. Pl-CA proprotein, is predicted to be extracellular and attached to the membrane from the N' (21aa) and C' (23aa) terminals. The proprotein involves a signal peptide of 21 aa which is cleaved off. Analysis was performed using the method of McGuffin *et al.*, 2000 at the PSIPRED Protein Structure Prediction Server

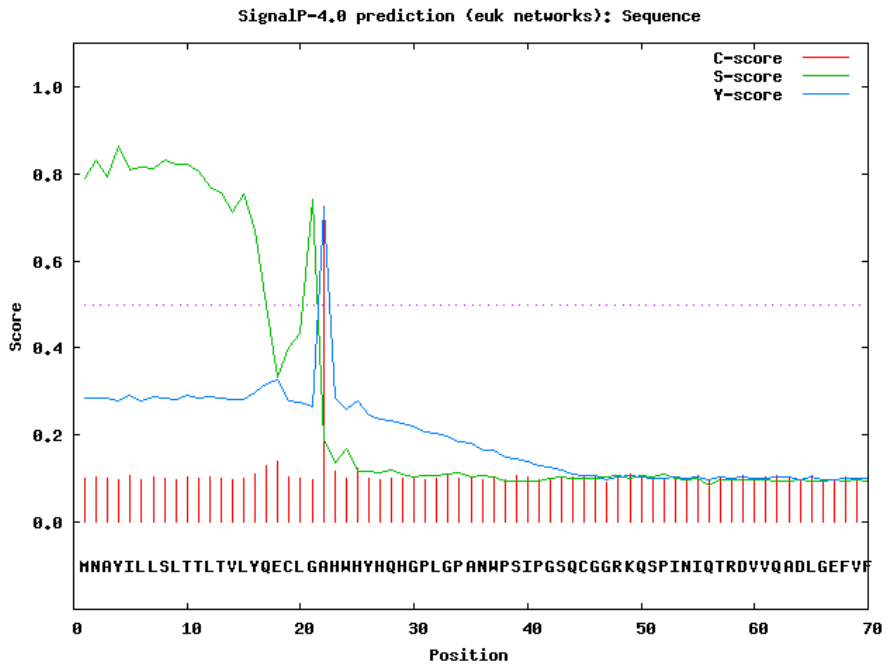


Figure 43. Prediction on the presence of a signal peptide and a cleavage site within the *Pl-CA* amino acid sequence. A signal peptide from the 1st to the 21st amino acid residue (MNAYILLSLTTLTVLYQECLG) and a cleavage site between the 21st and 22nd amino acid, suggest that *Pl*-CA is a secreted protein. (Prediction software: Bendtsen *et al.*, 2004).

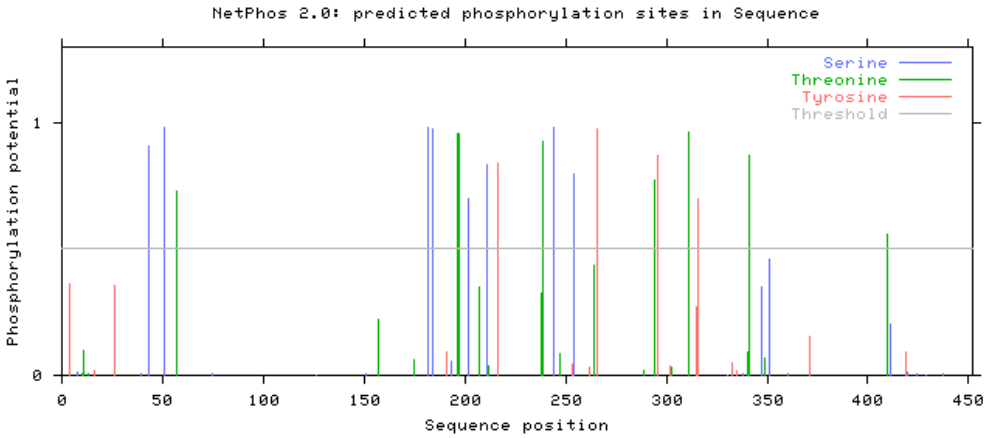


Figure 44. Prediction on phosphorylation sites for *Pl-CA*. (Blom *et al.*, 1999). 20 putative phosphorylation sites were predicted: eight serine residues, eight threonine residues and four tyrosine residues.

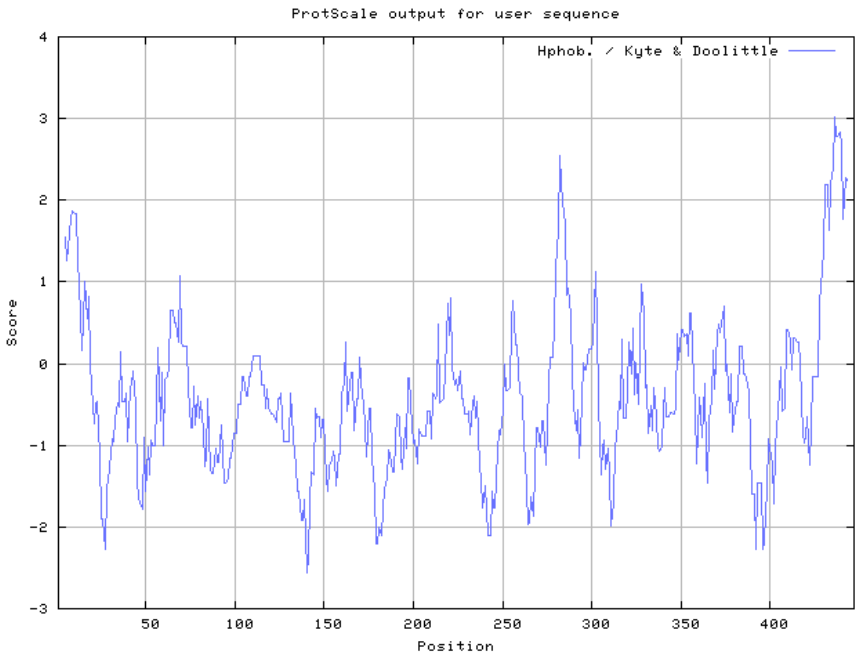
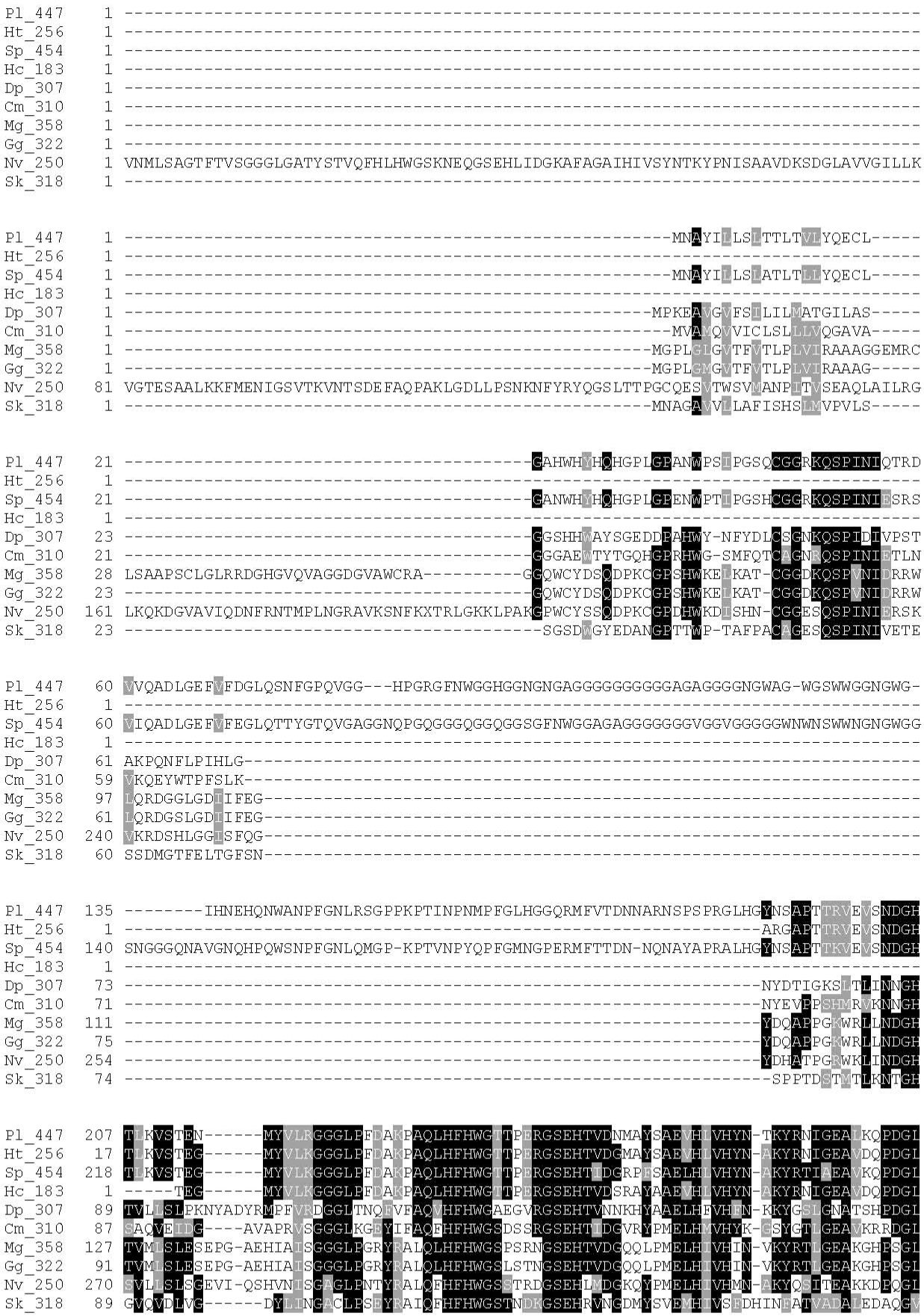


Figure 45. Hydrophobicity score along the linear sequence of the PI-CA protein (Prediction by Kyte and Doolittle, 1982.) The slope mainly distributed at the bottom of the diagram (below zero), shows that the molecule is hydrophilic except for three distinct regions with positive hydrophobic score (regions of aa: 5, 285 and 465).

Phylogenetic analysis of PI-CA

A phylogenetic analysis of *PI*-CA was performed. Homologous putative protein sequences of CA from other sea urchin species and from species belonging to evolutionarily divergent families, were identified by Blast screening. The homologue exhibiting the highest similarity was a predicted sequence from *S. purpuratus*, 454 aa long, with accession number: XP_784328, exhibiting 79% identity and 88% similarity. *PI*-CA, was also aligned with the partial CA from *H. tuberculata*, of 256 aa length, with accession number: ABE27961.1 and the partial CA from *H. erythrogramma*, of 183 aa long (with accession number: ABE27962). Homology was also found with CAs in species belonging in the phylum of Chordata, Hemichordata, Arthropoda and Cnidaria (Fig.46). Overall, the following CAs exhibited significant sequence similarities and were selected for the preparation of a phylogenetic tree (Fig.47): CA-XIV-like from *Saccoglossus kowalevsk*i, of 318 aa, with accession number: XP_002735859); *Daphnia pulex*, of 307 aa, with accession number: EFX81683.1; *Carcinus maenas*, of 310aa, with accession number: ABX71209.1; *CA-15*-like from *Meleagris gallopavo*, of 358aa, with accession number: XP_003211059; *CA-15*-like from *Gallus gallus*, of 322aa, with accession number: XP_415218; *Xenopus tropicalis*, of 310aa, with accession number AAI23070.1; *Nematostella vectensis*, of 250aa, with accession number: XP_001627923; and *CA-2* from Human, of 260aa, with accession number: NP_000058.1. These sequences exhibited identities with *PI*-CA ranging from 35% to 86% and similarities varying from 53% to 92% (Fig.46). Closer evolutionary relationships and higher conservation was found among sea

urchin species specially between *P. lividus* and *S. purpuratus*. The phylogenetic analysis demonstrated that *Pl*-CA shares higher similarity with Human *CA-2* than with CAs from Cnidaria or Chordata.



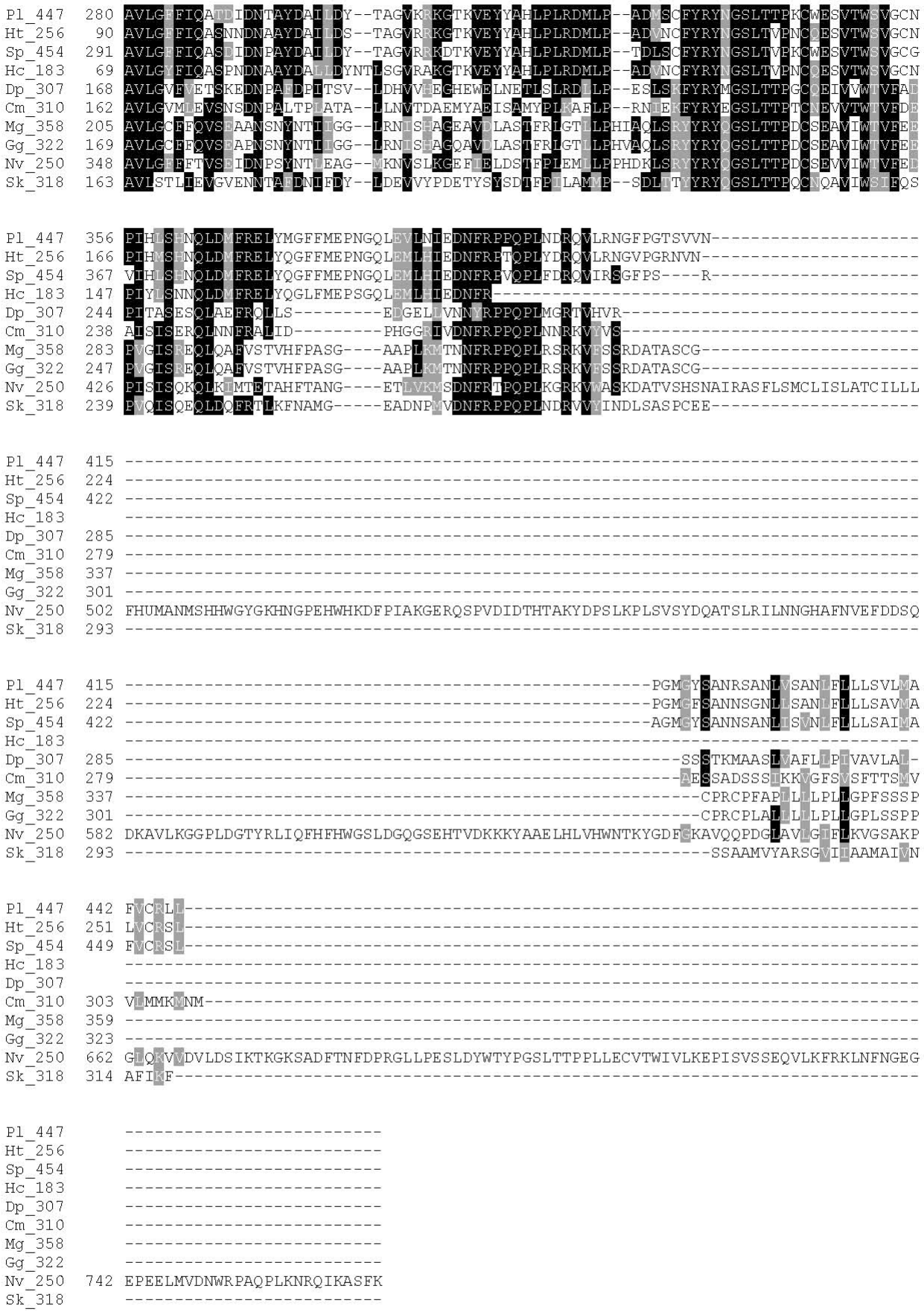


Figure 46. Multiple alignment of Pl-CA with Alignment was performed by ClustalOmega (http://www.ebi.ac.uk/Tools/msa/clustalo/). (Karakostis *et al.*, in preparation)

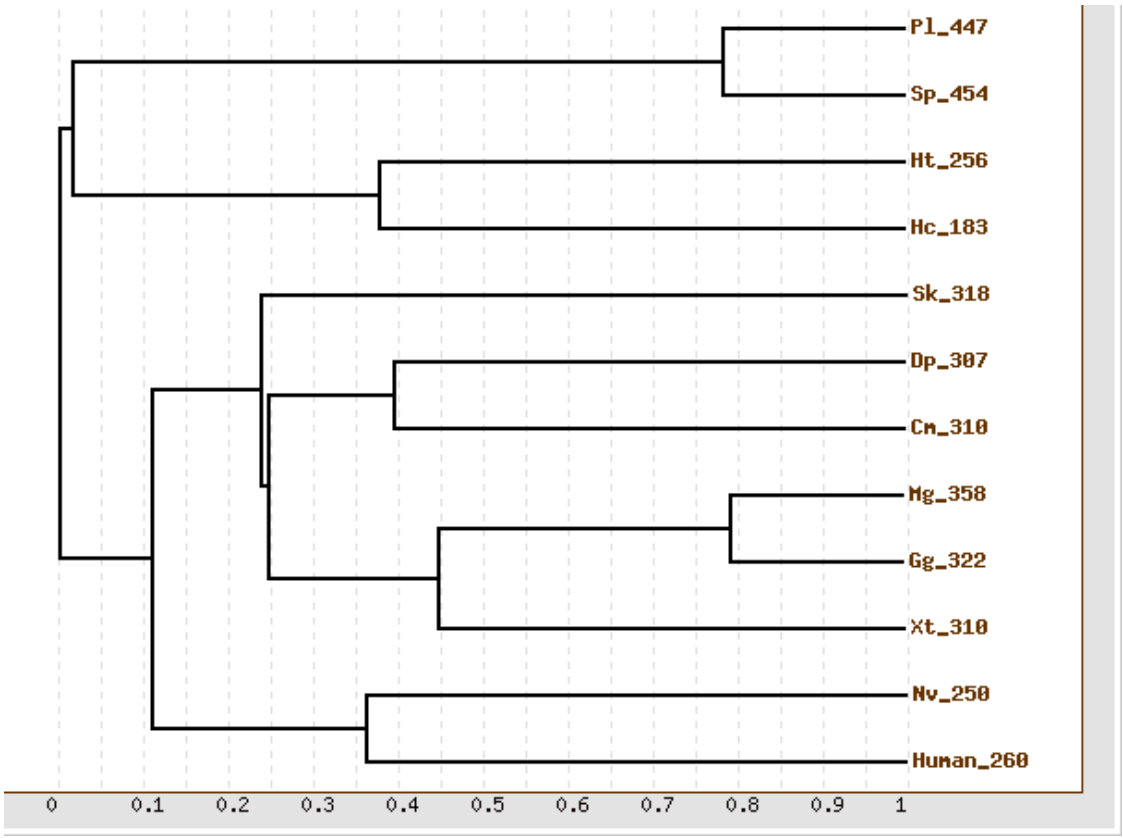


Figure 47. Phylogenetic tree of carbonic anhydrases.

Temporal gene expression profiling of Pl-CA

i. Semi-quantitive analysis by RT-PCR

The pattern of CA expression was analyzed by semi-quantitative RT-PCR analysis of gene expression. As shown in Fig.48, the expression initiates 15 h after fertilization, at mesenchyme blastula, after the formation of PMCs in the developing embryo. Expression increases at early pluteus as confirmed also by comparative qPCR.

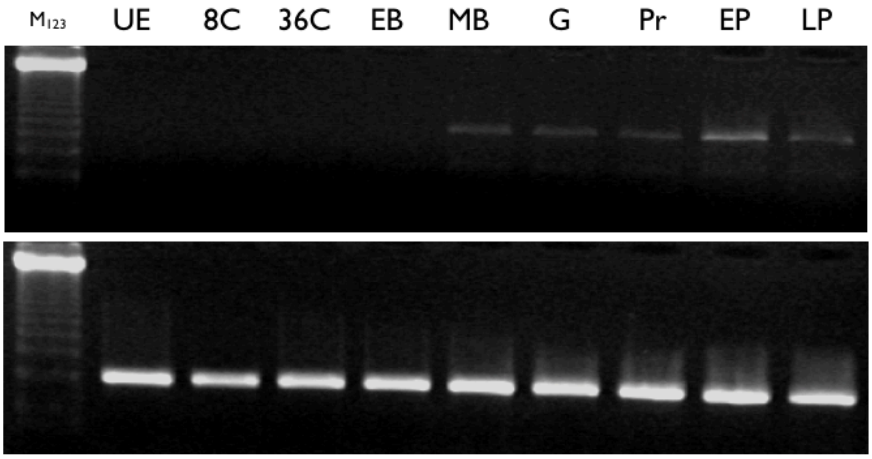


Figure 48. Analysis of CA expression by semi-quantitative RT-PCR throughout embryo development, by RT-PCR, visualized in 1% agarose-TBE gel. The gene S24 (lower image) was used as reference. UE: unfertilized egg; 8C: 8 cells stage; 32C: 32 cells stage; EB: early blastula; MB: mesenchyme blastula; G: gastrula; Pr: prism; EP: early pluteus; LP: late pluteus). The gene S24 (lower image) was used as internal reference control, its expression remained constant throughout the development.

ii. **Quantitative analysis by comparative qPCR**

The temporal expression profile of *Pl*-CA during embryo development was analyzed by comparative real-time qPCR (ΔCCt Q-PCR). Expression was at first detectable at the blastula stage, gradually increasing to reach a maximum at the pluteus stage, where *CA* was up-regulated by approximately 40 folds with respect to the cleavage stages (Fig.49).

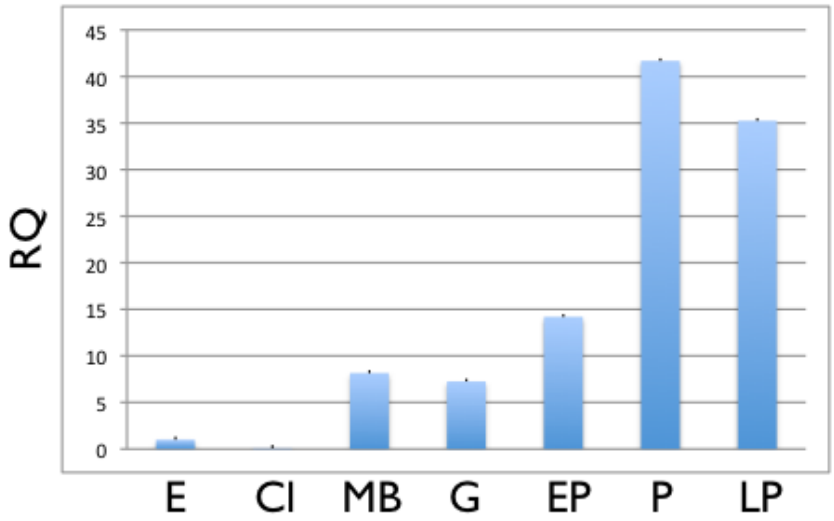


Figure 49. Temporal expression of *Pl*-CA during *P. lividus* embryo development. Comparative qPCR analysis of the *Pl*-CA transcription levels in sea urchin embryos at different developmental stages: E, egg; Cl, cleavage stage; MB, mesenchyme blastula; G, gastrula; EP, early pluteus; P, pluteus; LP, late pluteus. The *Pl*-Z12-1 mRNA was used as an internal endogenous reference gene; cDNA from the cleavage stage was used as the reference sample and was set to 1 in the graph. q-PCR experiments were performed at least three times (Karakostis *et al.*, in preparation).

Spatial gene expression profiling of Pl-CA

To investigate the spatial expression pattern of *Pl-CA* at early development, WMISH experiments were carried out on embryos collected from the swimming blastula to the pluteus stage. The antisense (AS) DIG-labelled RNA probe was amplified by asymmetric PCR from the *pGEM-T-Easy-3Pl-CA* plasmid after digestion with *Sac*II which results in linearization (Fig.50). The linear purified vector was of the expected calculated size of 4 kb and had a concentration of 60 ng/μl.

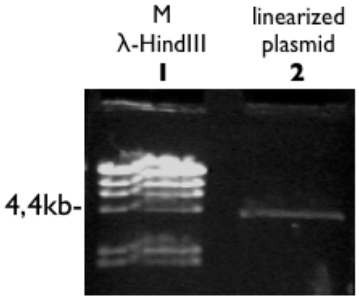


Figure 50. Preparation of WMISH probe. Linearization of *pGEM-T-Easy-Pl-CA* by digestion with *Sac*II for the preparation of the AS *Pl-CA* DIG RNA probe.

WMISH was carried out on embryos collected at different developmental stages, from early blastula to pluteus (Fig.51). *Pl-CA* transcripts were not detected at the blastula stage (Fig.51-A). Firstly detected at the mesenchyme blastula stage, they were exclusively localised within the primary mesenchyme cells (PMCs) throughout the embryo development from gastrula to pluteus. At the gastrula stage, expression was detectable exclusively in the PMCs of the ventrolateral clusters (Fig.51-B asterisks). At the prism stage, expression was restricted to a subset of the PMCs located mainly in the body rod (Fig.51-C,D,E arrows) but down-regulated in the dorsal and anterolateral chain. At pluteus, expression was localised in all PMCs, including the tips of the postoral rods (Fig.51-F,G arrows), the anterolateral rods (Fig.51-F,G arrowhead) and on the scheitel (Fig.51-F, asterisk).

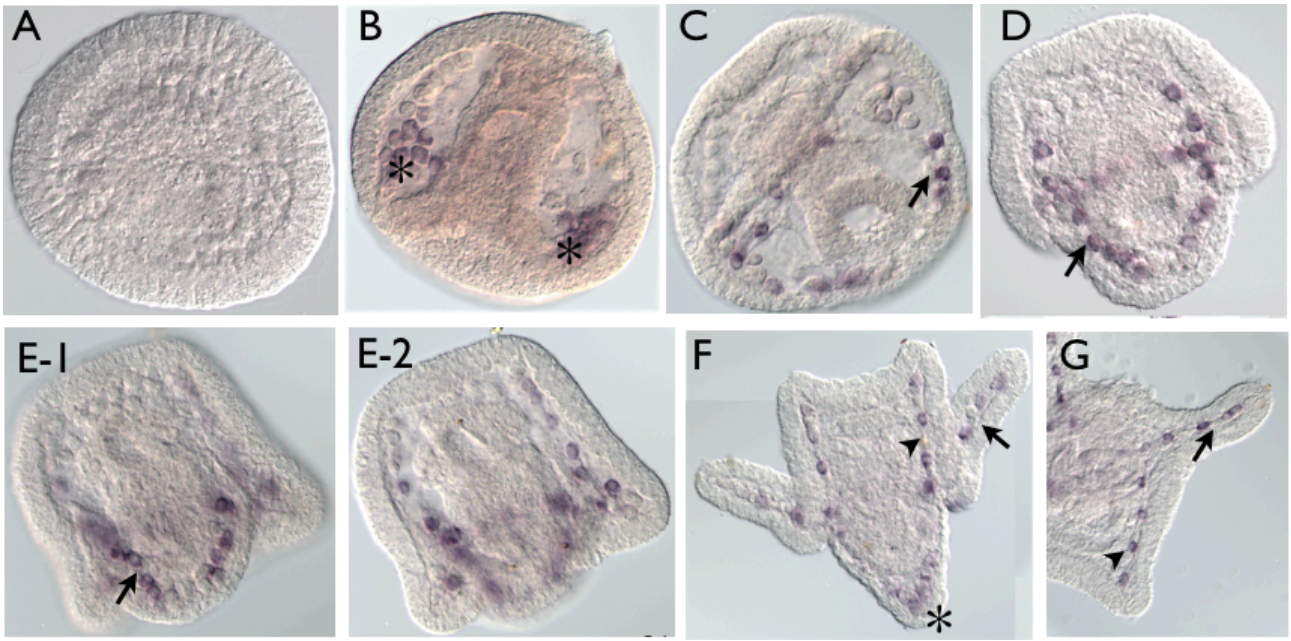


Figure 51. Spatial expression of *Pl*-CA at embryos at different developmental stages. Whole mount *in situ* hybridization using a DIG-antisense Pl-CA-8 RNA probe on embryos fixed at different developmental stages: A) blastula; B) gastrula; C - D): prism; E - G), pluteus. Arrows point to secondary mesenchyme cells; asterisks point to the ventrolateral clusters. The asterisk is B signs the ventrolateral clusters and in F, the scheitel. (Karakostis *et al.*, in preparation).

Preparation of recombinant PI-CA protein

In order to express the recombinant PI-CA protein, the subcloning into an expression vector was necessary. To this purpose truncated CDSs were amplified by PCR and cloned in the expression vectors p*EXP-NT*, p*Trc-CT* (Invitrogen), p*COLD-TF* (Takara) and p*ET32b*$^{+}$*.* Inserts were amplified using each set of CA primers (see Table 9) and ligated to the corresponding vectors. The recombinant plasmids p*EXP-PI-CA* and p*TRC-PI-CA* were prepared (Fig.52). The p*COLD-PI-CA* cloning required the digestion of the vector and the PCR product, before the ligation. Positive constructs were sequenced. The cloning process is briefly outlined in Fig.53.

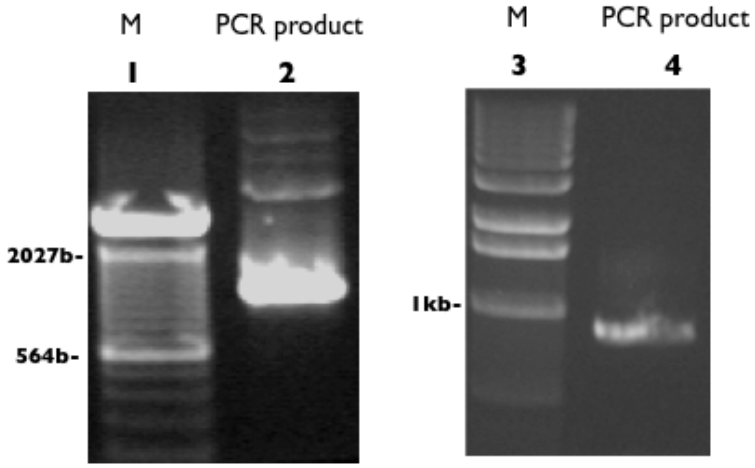


Figure 52. Cloning of p*EXP-PI-CA (lane 2)* and p*TRC-PI-CA* (lane 4). Test PCR products amplified by PCR, with T7 (on vector) and the respective gene-specific reverse p*EXP-PI-CA* and p*TRC-PI-CA* primers (on the insert).

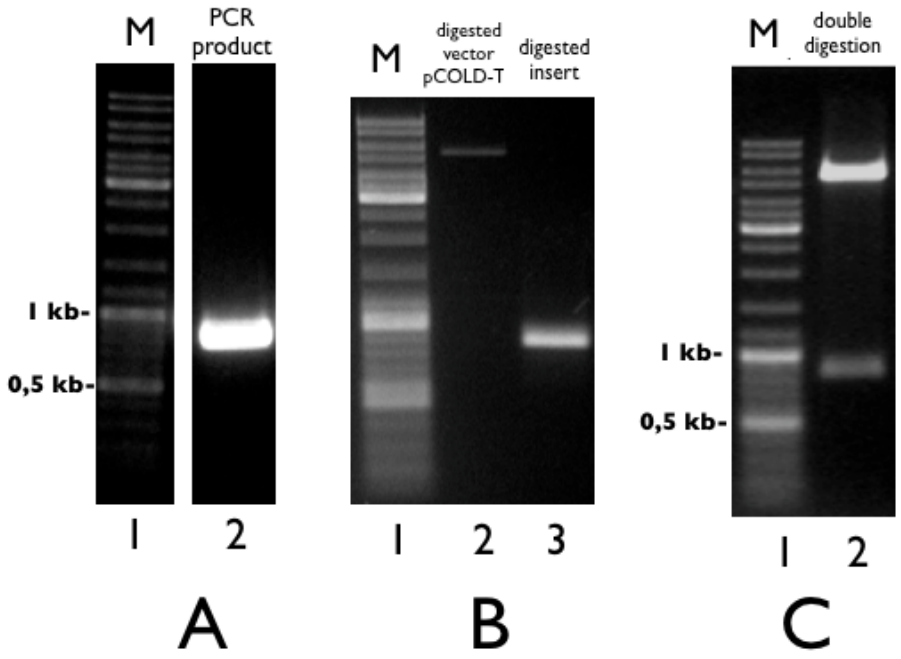


Figure 53. Cloning procedure of p*COLD-PI-CA*. A: PCR amplification from p*GEM-T-Easy-PI-CA*. B: Digestion of the vector (lane 2) and of the PCR amplification product (lane 3) with *Nde*I and *Xba*I. C: Test screening of positive colonies by double digestion (*Nde*I and *Xba*I) of the plasmid p*COLD-PI-CA*.

The expression plasmids p*EXP-N-PI-CA*, p*Trc-C-PI-CA* and p*COLD-TF-PI-CA* were used for expression of recombinant *PI*-CA protein, in Bl21(DE3) and M15 E. *coli* cells. The expression levels, are shown in the representative gel images of Fig.54. Additionally, the cell-free expression system, S30 T7 cell-free (Invitrogen), was used with p*EXP-N-PI-CA* and p*Trc-PI-CA*.

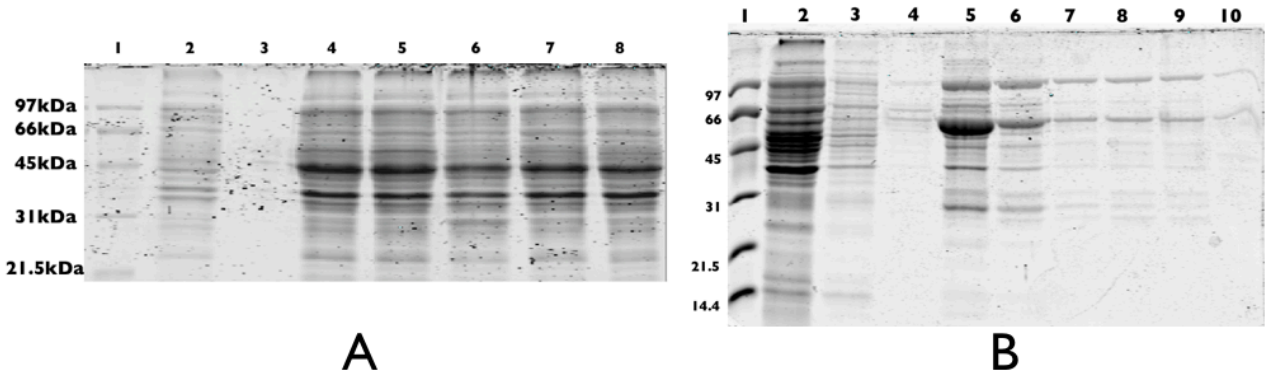


Figure 54. Recombinant *PI*-CA protein expression and purification using p*Trc-C-PI-CA* (35,7 kDa) expressed in E. *coli* Bl.21 (DE3). A) Monitoring of the expression (lane 1:M; lane 2: sample before induction; lanes 4-8: samples taken at hours 1, 2, 3, 4 and 5 after induction of expression. B) Purification by affinity chromatography (Ni-NTA) of the p*Trc-C-PI-CA.* (lane1: Molecular weight marker (M); lane 2: flow-through; lanes 3,4: washes; lanes 5-10: elution fractions with elution buffer).

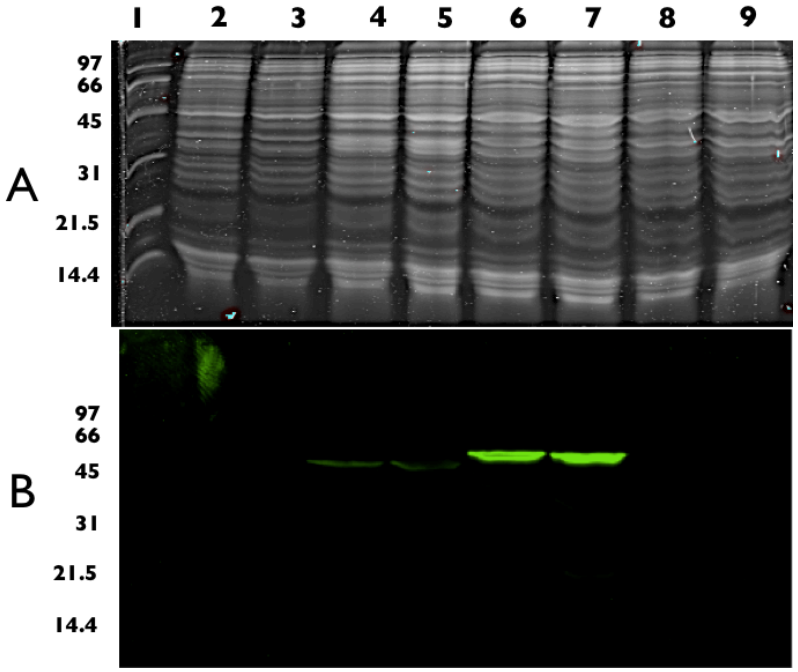


Figure 55. Recombinant *Pl*-CA protein expression and purification using p*EXP-Pl-CA*, p*Trc-Pl-CA* expressed by the S30 T7 cell-free expression kit. 12% SDS PAGE (A) and Western blot (B) with anti-His antibody. Lane1: M; lanes 2,3: control reaction without plasmid; lanes 4,5: reaction with the p*EXP-Pl-CA* plasmid (expected size: 34,4kDa), lanes 6,7: reaction with the p*Trc-Pl-CA* plasmid (expected size: 35,7kDa); lanes 8,9: positive control.

Unfortunately, the expression yield of the recombinant protein was particularly low, as viewed by SDS PAGE and Western Blotting with an anti-His antibody (Fig.55). Therefore, a different strategy was used, using the p*ET32b+* vector for the preparation of a recombinant Pl-CA. *Pl-CA* is an α-type CA. Among the five distinct CA families, a-type CAs are divided into four broad subgroups (i.e., cytosolic, mitochondrial, secreted, and membrane bound), which consist of several isoforms (Breton, 2001). Based on *in silico* analysis, we found that *Pl*-CA involves a signal sequence and an N-terminal Glycine-rich region. Thus, for acquiring cytoplasmic expression of high yield in E. *coli*, we used a truncated *Pl*-CA after fusion with the thioredoxin tag which facilitates the production of soluble proteins. Additional N' and C' terminal His_6 tags were linked in order to facilitate recombinant protein detection and purification. A partial CDS of 861 bp was amplified and cloned in the p*ET-32b+* expression vector (Fig.56). Protein expression using the p*ET32-Pl-CA* recombinant plasmid**,** produced a fusion protein comprising 460 aa, corresponding to a MW of 50,78 kDa. We observed a strong recombinant band of 50 kDa in the SDS PAGE analysis (Fig.57). The recombinant *Pl-CA* protein was mainly expressed as insoluble form (soluble in buffered 6M Urea) (Fig.57-A). Then, purified the tag recombinant protein from the 6M Urea-soluble fraction under denatured conditions using affinity chromatography (Fig.57-B). The recombinant protein was efficiently purified according to SDS–PAGE analysis (Fig.57-B).

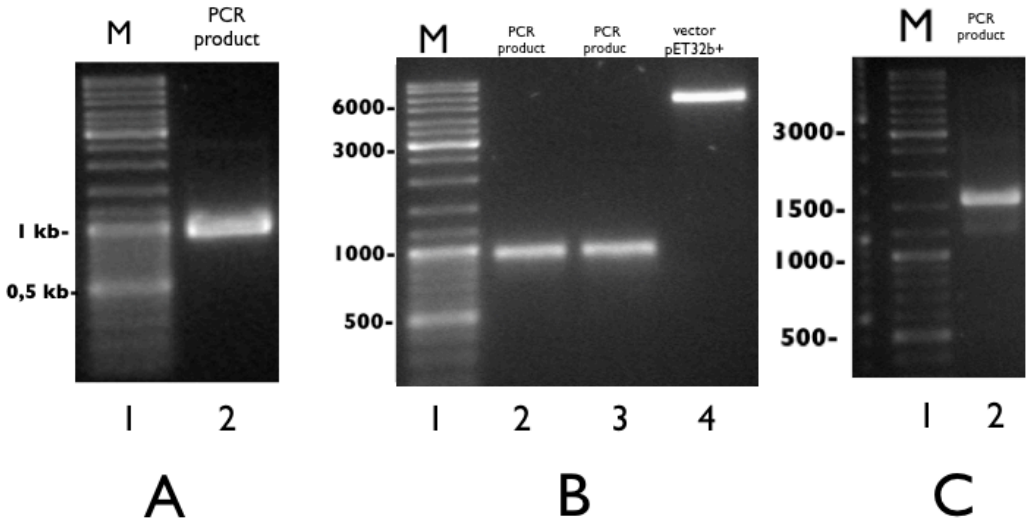


Figure 56. Clonings of p*ET32-PI-CA*. A: PCR amplification from p*GEM-T-Easy-3PI-CA*. B: Digestion of the insert and vector with *Xho*I and *BamH*I. C: Test PCR product of 1450 bp, amplified with T7 (on vector) and the gene-specific reverse p*ET32b-PI-CA* primer.

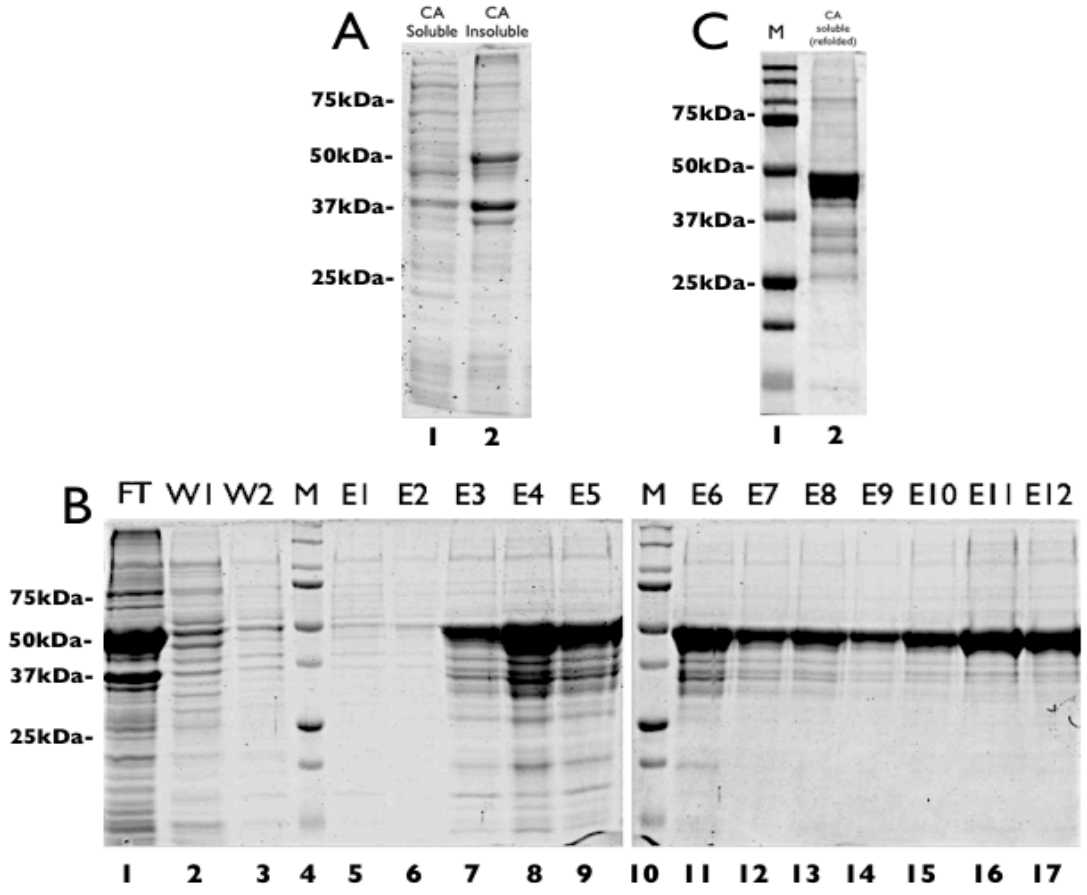


Figure 57. Protein expression using the p*ET32-PI-CA* recombinant plasmid. 12% SDS PAGE of PI-CA expression (A) and purification process of *PI*-CA (B). (M: Molecular mass marker, FT: Flow through, W: wash, E: Elution fractions of 1ml). Combined elution fractions after refolding and concentration (C).

Identification of Pl-CA protein in embryo and extracts from adult tests, by Western blot.

Polyclonal antibodies against the recombinant Pl-CA were prepared in mice and used in Western Blot analysis for the identification of Pl-CA in embryo lysates at the gastrula and pluteus stages (Fig.58) as well as in protein extracts from adult tests (See Chapter VI) (Fig.58).

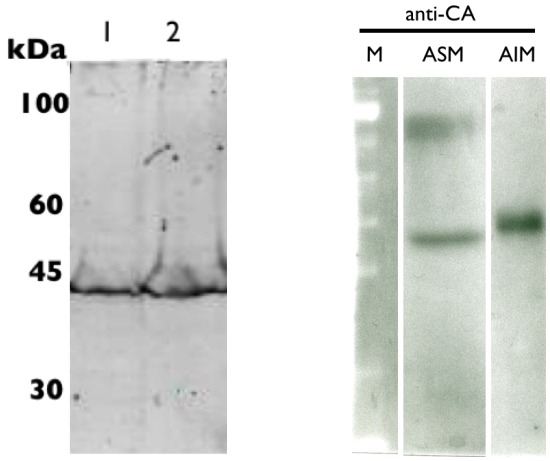


Figure 58. Western blot analysis of *Pl-CA* expression in protein extracts from *P. lividus* embryos(left) and adult tests (right). Left: Identification of *Pl-CA* on embryo lysates. Lane 1: protein extract from embryos at gastrula stage; Lane 2: protein extract from embryos at pluteus stage. Blots were incubated at 4 ^{o}C, overnight with the first anti-CA serum antibody (dilution: 1:1000) and at RT, for 2h with the secondary anti-mouse (dilution: 1:4000). Right: SDS-PAGE fractionation followed by Western Blot. Acid-soluble (ASM) and acid-insoluble (AIM) shell matrix proteins from *Paracentrotus lividus* adult tests cross-reacted with anti-CA antibody. Anti-CA antibody was used at a dilution of 1:500. Film was developed for 4 min. 20 μg of protein material were applied per well.

Functional characterization of recombinant Pl-CA

The activity of the recombinant *Pl-CA* was tested by its ability to catalyze the hydrolysis of p-nitrophenyl acetate. A protein concentration of 100 µg/ml accelerated the rate of the hydrolysis of p-nitrophenyl acetate. The activity was comparable to that of the commercial carbonic anhydrase used as control. The catalytic activity was reduced by the addition of acetazolamide, a specific inhibitor of the enzyme carbonic anhydrase (Fig.59), at levels similar to those of the Control AZM.

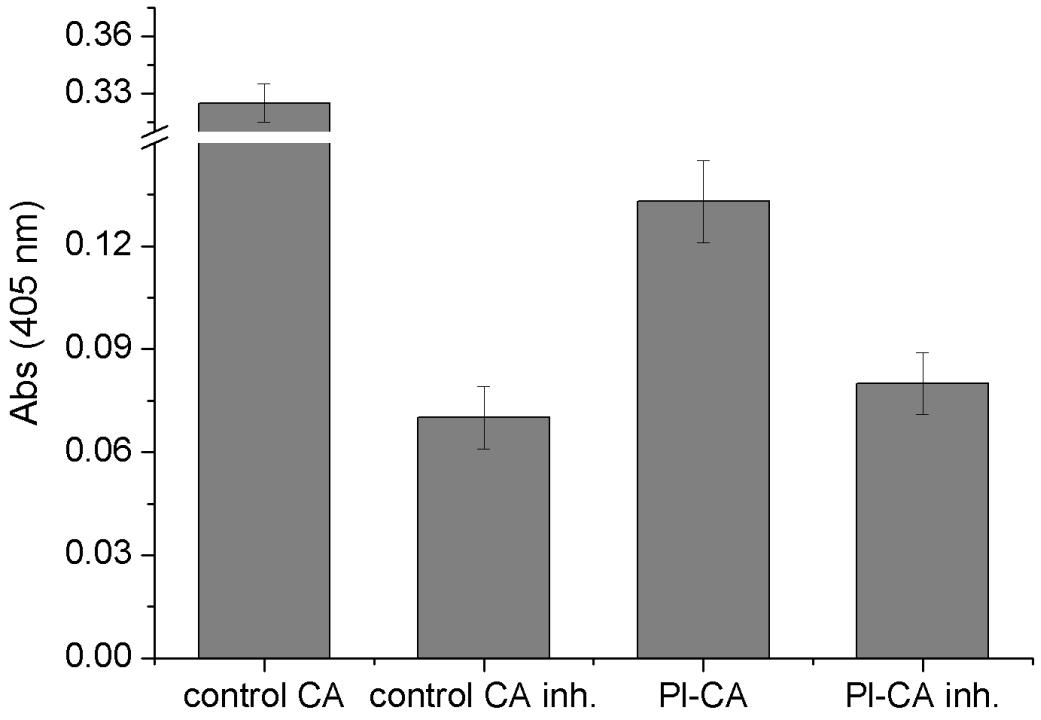


Figure 59. Hydrolysis of p-nitrophenyl acetate. The reaction was monitored reporting in the blue columns the increase in absorbance at 405 nm during 30 min (y-axis). Recombinant *Pl*-CA was added at a concentration of 100 µg/ml. As a positive control, carbonic anhydrase from bovine erythrocytes (Sigma) 20 µg/ml was used (Control CA). The CA activities of the *Pl-CA* and of the control CA were inhibited by the addition of acetazolamide (0.1 mM) (control CA inh. and Pl-CA inh.).

Chapter V: *Galectin-8* from *Paracentrotus lividus,* expressed in the archenteron and secondary mesenchyme cells of the embryo and in adult, is a lactose-specific galectin involved in cell adhesion.

Abstract

Galectins are carbohydrate-binding proteins that specifically bind beta-galactoside derivatives (Hirabayashi *et al.*, 2002; Pfeifer *et al.*, 21993; Liu *et al.*, 2011). The members of the galectin super-family interact with cell-surface glyco-conjugates and integrins (Nozomu Nishi *et al.*, 2003) and regulate diverse cellular events, including signaling, apoptosis as well as innate immune and inflammatory responses (Liu and Rabinovich 2005; Liu, 2005). Additionally, some galectins found in mammalian osteoblasts and osteocytes are involved in biomineralization (Tanikawa *et al.*, 2010). In sponges, the matrix guided formation of silicatein-mediated silica spicules is strongly increased when associated with a galectin (Schröder *et al.*, 2006).

This study was undertaken to identify and reveal the role of galectin-*8* in the sea urchin. A cDNA fragment of 1309 nt length was isolated and cloned by EST data mining, RT-PCR and 3' RACE. *In silico* analysis of the obtained coding sequence of 933 nt encoding a 34.7 kDa protein, revealed two tandem carbohydrate-recognition domains and homology comparison by BLAST, categorized *Pl-galectin*-8-like as a novel member of the Galectin-8 family. Furthermore, the putative protein 3D structure was modeled based on its high structural similarity with orthologous solved structures from Human Galectin-8. As studied by whole mount *in situ* hybridization and comparative qPCR, transcriptomic expression initiated at blastula and gradually increased reaching a 4,7 fold rate at pluteus. Expression was restricted to the gut of the developing embryo. A recombinant protein, produced in bacteria, agglutinated red blood cells (haemmagglutination assay), at a minimum concentration of 25 µM. This activity was inhibited by the addition of 5 mM lactose. The anti-adhesive biological activity of *Pl-galectin-8-like* was assayed by a cell adhesion assay in human cells. Last, the recombinant protein was used to the raise of polyclonal antibodies in mice which were used to identify the protein in protein matrices from the embryo and the adult.

Introduction

Lectins

Interactions between cells or cells with the extracellular matrix, are mediated by cell-surface carbohydrates and their binding proteins known as endogenous lectins (Liu *et al.*, 2002; Hughes, 2001). Lectins, initially described in plants (Sharon and Lis, 2004), are defined as proteins that preferentially bind carbohydrate complexes protruding from glycolipids and glycoproteins (Mody *et al.*, 1995; Gorelik *et al.*, 2001; Bies *et al.*, 2004; Minko, 2004). The term lectin has been generalized to encompass all non-immune carbohydrate-specific agglutinins regardless of blood-type specificity or source (Sharon and Lis, 2004). The carbohydrate-lectin interaction can be as specific as in the antigen-antibody or substrate-enzyme interaction (Minko, 2004). Lectins bind not only to oligosaccharides on cells but also to free-floating glycans including monosaccharides with relatively weaker dissociation constants, on the order of micromolar to millimolar range (Bouckaert *et al.*, 2005; Rabinovich *et al.*, 2007). They are also involved in cell homeostasis, which is regulated by the controlled transcription of numerous genes. Thus, multiple enzyme cascade and signaling pathways are influenced.

Lectins are perhaps the most widely studied molecules in glycobiology. Several studies have shown that certain lectins are expressed in a developmentally regulated fashion, either during embryonic stages (Colnot *et al.*, 1996; Lip *et al.*, 1999; Chiariotti *et al.*, 1999; Kaltner *et al.*, 2002) or postnatal development (Sanford *et al.*, 1993; Li *et al.*, 1992). The lectins family is rapidly expanding, presently counting 50 C-type lectins and at least 10 identified galectins (Gorelik *et al.*, 2001). In biomedicine, anti-adhesion therapy aims in treating microbial infections or diseases, by identifying specific carbohydrates analogs interfering with pathogen lectin–host carbohydrate interactions (Zopf and Roth, 1996; Karlsson, 1998; Kelly and Younson, 2000; Sharon and Ofek, 2000; Ofek *et al.*, 2003).

The Galectin family

Galectins represent a large ancient family of structurally-related, phylogenetically diverse lectins (Hughes, 1997; Ogawa *et al.*, 2004; Houzelstein *et al.*, 2004), present in all animal kingdoms, plants and fungi, with a carbohydrate binding specificity primarily against disacharides containing β-galactoside residues found in cellular glyco-conjugates and glycoproteins (Hirabayashi *et al.*, 2002). The first characterization of a D-galactoside-binding lectin in the sea urchins was from the specie *Anthocidaris crassispina* (Ozeki *et al.*, 1991; Ozeki *et al.*, 1995). The family name ‘Galectin’ replaced the previously used ‘S-type lectin’ designation (Drickamer *et al.*, 1988), indicating dependence on thiols (reducing conditions) for activity, a property of the first galectin studied, galectin-1 but not a property of many other galectins (Leffler *et al.*, 2004). The term galectin was introduced in 1994 (Barondes *et al.*, 1994) as follows: “Membership in the galectin family requires fulfillment of two criteria: affinity for β-galactosides and significant sequence similarity in the

carbohydrate-binding site, the relevant amino acid residues of which have been identified by X-ray crystallography (Lobsanov *et al.*, 1993). Therefore, proteins that are similar enough by overall sequence but lack proven β-galactoside binding activity or may lack some of the residues in the defining motif, are categorized as “galectin-like” (Leffler *et al.*, 2004).

Classification of Galectin types

The functional domain of galectins is one or two carbohydrate recognition domains (CRDs) within a single polypeptide chain. The galectin CRD is (with a few invertebrate exceptions) not associated with other types of well defined protein domains. Thus the galectin CRD acts mainly by itself or together with another galectin CRD. This is in striking contrast to many other types of protein domains, e.g. C-type lectin domains, which often occur together with other domain types in the same peptide chain. The mono CRD galectins can occur as monomers, dimers or higher order oligomers depending on specific case and conditions (concentration, presence of ligand). Based on their domain organization, galectin subfamilies are designated as proto-, chimera- and tandem repeat-types (Hirabayashi and Kasai, 1993) (Fig.60). The prototype is defined as similar to dimeric lectins with 14 kDa subunits including galectin-1 and all vertebrate mono-CRD galectins (except galectin-3) (Hirabayashi *et al.*, 2002). The chimera type defines galectins having a CRD and another type of domain, all of which are orthologues of galectin-3. In tandem repeat type, each of the two CRDs may have very diverse fine specificity.

Structurally, the galectin CRD is a beta-sheet sandwich of about 135 aa. Each sheet is slightly bent with 6 strands forming the concave side and 5 strands forming the convex side. The carbohydrate is bound in the groove formed in the concave which is long enough to hold about a linear tetra-saccharide.

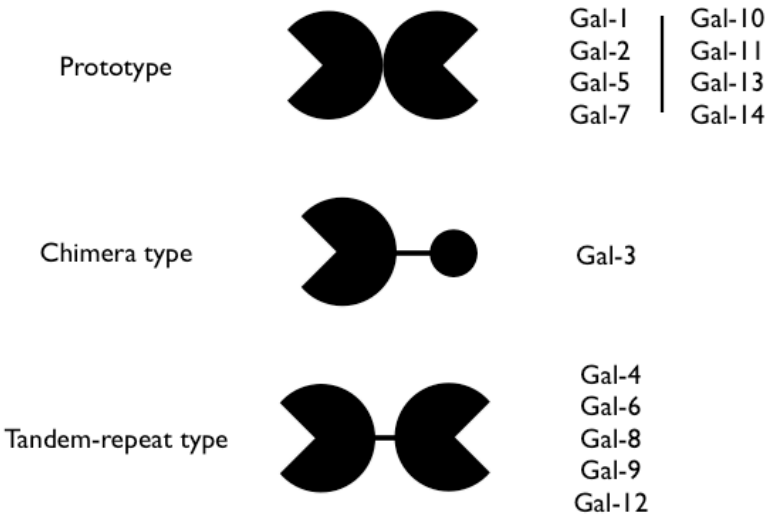


Figure 60. Classification of Galectin types.

Biological significance of Galectins

The relationships between the CRD fine specificity and the biological activity of the protein remain largely undefined (however see Amano *et al.*, 2003). Notwithstanding, the biological significance of galectins has started to unravel. Galectins were discovered in tissue extracts analyzed for their ability to agglutinate erythrocytes, based on the hypotheses that cell surface carbohydrates take part in cell adhesion (Barondes, 1997; Nowak TP *et al.*, 1976). They inhibit cell adhesion and induce apoptosis, by binding to integrins (Hadari, 2000); Various functional roles of galectins have been evidenced in cancer (Ghazarian *et al.*, 2011), in innate and adaptive immunity, inflammation and development, without offering though a unifying picture of their biological function. Instead galectins appear to be implicated in a particularly diverse, range of activities both inside and outside cells, in vertebrates, invertebrates and mammals. Indeed, additional studies identified galectins in mammals involved in cell cycle regulation (Yang *et al.,* 2001), apoptosis (Rabinovich *et al.*, 2002; Akahani *et al.*, 1997; Wada & Kanwar, 1997; Kuwabara *et al.*, 2002; Dettin *et al.*, 2003), and pre-mRNA splicing (Dagher *et al.*, 1997). In vertebrates, galectins participate in a variety of cellular functions including cell adhesion/ proliferation, development/morphogenesis, tumor cell metastasis and immune regulation/ innate immunity (Hughes, 2001; Vasta *et al.*, 2004; Zick *et al.*, 2004; Camby *et al.*, 2006). A galectin-3 null mutant mice showed a defect in bone development (Poirer *et al.*, 2002).

The Metazoan invertebrates possess multiple members of the galectin superfamily (Vasta *et al.*, 2004) as found in nematodes (Hirabayshi *et al.*, 1992; Greenhalgh *et al.*, 1999; Newlands *et al.*, 1999), arthropods (Pace *et al.*, 2002; Pace and Baum, 2004, Barat-Houari *et al.*, 2006; Huang *et al.*, 2007), tunicates (Parrinello *et al.*, 2007) and sponges (Pfeifer *et al.*, 1993; Stalz *et al.*, 2006). Galactose-binding lectins are also found in the Phylum Mollusca, (Suzuki and Mori, 1989; Mitra and Sarkar, 1998; Wilson *et al.*, 1992; Ozeki, 1998).

In the present study, a tandem-repeat type galectin, Pl-galectin-8, from the *P. lividus* sea urchin embryo, showing significant similarity with the human galectin 8, was identified and characterized for the first time in echinoderms. The expression profile in development was studied. A recombinant Pl-galectin-8, expressed in bacteria, was active in an hemagglutination activity assay revealing a lactose specificity and its role in cell adhesion was shown using human cells.

Experimental Procedures

The experimental procedure for the embryo culture, total RNA extraction and RT-PCR is described in Chapter II.

Cloning of *Pl-galectin-8*

A sequence predicted to code for a GALECTIN protein in the sea urchin *S. purpuratus* database (accession number XP_781871.1), was used for BLAST screening against *P. lividus* EST databases available at NCBI and MPIMG (http://goblet.molgen.mpg.de/cgi-bin/webapps/paracentrotus.cgi). Three overlapping *P. lividus* similar EST clones were identified (see Results Section). Specific primers were designed complementary to the obtained homologous merged EST sequences and purchased from MWG (Heidelberg, Germany) (Table 10, row 1). Amplification by RT-PCR (One Step RT-PCR, Invitrogen), using purified total RNA extracted from the gastrula stage, gave a 302 bp PCR product, which was subsequently cloned in the p*GEM-T-Easy* Vector (Promega) and sequenced by MWG (Heidelberg, Germany). As the obtained sequence revealed only a partial 5' CDS, 3'RACE primers were designed (Table 10, rows 2,3) and used a 3'RACE kit (Invitrogen) to identify the 3'-terminal end of the sequence. The latter amplification product was sequenced by MWG (Heidelberg, Germany). The compiled 1309 nt sequence was deposited at the EMBL databank, Acc Num: FR716469. New primers were designed for the amplification by RT-PCR of the full-length galectin-8 CDS (Table 10, row 5), which was cloned in the p*GEM-T-Easy* vector (Promega) and re-sequenced for validation. The obtained recombinant plasmid was referred to as p*GEM-T-Easy-Pl-Galectin-8*.

Domain characterisation and phylogenetic analysis

The putative aa sequence was characterized as described in Chapter II.

Modelling of *Pl*-GALECTIN-8

The Modeller software (http://salilab.org/modeller/) (Eswar *et al.*, 2008) was used for the construction of a 3D model predicting the structure of the N-terminal domain of *Pl*-GALECTIN-8, based on the structure of the human GALECTIN-8 N-terminal domain (pdb: 2yv8), retrieved from the protein data bank at http://www.rcsb.org/pdb and exhibiting a high localised sequence similarity (over 65%).

Comparative Real Time qPCR and Whole-mount *in situ* hybridization (WMISH)

Gene expression profiling was monitored throughout embryo development by comparative q-PCR and whole mount *in situ* hybridization, as described in Chapter II. For the q-PCR, the *Pl-galectin-8* primers 5‘-AAACATGGAGCTGGCAGCAT-3’ as forward and 5‘-TGAGTCTCCCTGGAGTCATTCC-3’ as reverse, were designed to amplify a fragment of 82bp (Table 10, row 4). For the WMISH, the plasmid *pGEM-T-Easy-Pl-galectin-8* was used as template, after enzymatic digestion with *Sac*II (NEB) and *Sp6* primer was used for the production of an antisense full-length CDS DIG-labelled probe.

Production of recombinant *Pl-Galectin-8* protein in E. *coli*

The full-length CDS was cloned in three different expression vectors for the optimal generation of an active recombinant protein.

The expression vectors: p*EXP-NT* (Invitrogen), p*TRC-CT* (Invitrogen) and p*COLD-TF* (TAKARA) were used, acquiring respectively the plasmids p*EXP-N-CA*, p*TRC-C-CA* and p*COLD-TF-CA*. For each plasmid, different cloning procedures and primers were used. Table 10 summarizes the primers used for the cloning of each plasmid. Briefly, p*GEM-T-Easy-Pl-Galectin-8* was used as a template for all the amplifications of the full-length CDS. The p*EXP-NT* and p*TRC-CT* involve a T7 promoter and a TA cloning strategy; thus, the insert was amplified by PCR with Taq Polymerase (Invitrogen) introducing an extra Adenine at the 3‘end of the PCR product. This facilitated a direct ligation to the linearized vector carrying an extra Thimine at the 3‘ends. p*EXP-NT* introduces an hexahistidine (6x His) at the N’ terminal of the recombinant protein, while p*TRC-CT* expresses a C’ terminal His_6 fusion tag.

Additionally, the full-length Pl-Galectin-8 CDS was amplified from the p*GEM-T-Easy-Pl-galectin-8* recombinant plasmid by PCR using the primers: 5’-GGACATATGGCATACCCATACC-3’forward and 5’-GCTCTAGATTACTGGAAGCGAATC-3’ reverse, containing NdeI and XbaI restriction sites (Table 10, row7). After digestion the amplified product (936 bp) was cloned in the pCOLD-TF (TAKARA) expression vector containing the cspA (cold shock protein A) promoter, the N-terminal fused chaperone trigger factor and an N-terminal hexahistidine (6x His) tag. The obtained recombinant plasmid was referred to as p*Cold-TF-gal-8*. The cloned sequence was confirmed by direct sequencing at MWG (Heidelberg, Germany). The constructed recombinant p*Cold-TF-gal-8* plasmid was used to express *Pl-Galectin-8* fusion protein in E. *coli* BL21(DE3) (Novagen), after induction with 1mM isopropyl-b-D-thiogalactopyranoside (IPTG) (Sigma) at 15 ^{o}C for 50 h. The cells were harvested by centrifugation at 4,000 rpm at 4 ^{o}C for 10 min, resuspended in lysis buffer (50 mM Tris, 500 mM NaCl, 2 mM PMSF and 5mM imidazole, pH 7.5), disrupted by sonication and centrifuged at 10,000g at 4 ^{o}C for 30 min. The supernatant was applied to a Ni-nitrilotriacetate column (Invitrogen), pre-equilibrated with lysis buffer. After washing of the column in: 50 mM Tris-HCl, 50 mM NaCl, 20 mM imidazole, pH 7.5), 1 ml fractions were eluted in: 50 mM Tris, 100 mM NaCl, 300 mM

imidazole, pH 7.5. The fractions containing the recombinant protein (checked by Bradford and 7,5% SDS–PAGE) were pooled and concentrated by centrifugation in Amicon Ultra centrifugal filter units, 50 kDA cut-off (www.millipore.com). After dialysis against 50 mM Tris, 50 mM NaCl, pH 7.5, the protein concentration was determined using the Bradford assay (Bio-Rad) with bovine serum albumin (BSA)(Sigma) as a protein standard and the purity was analyzed on a 7.5% SDS polyacrylamide gel.

Table 10: Primers designed for the identification and cloning of the full-length CDS of *Pl-galectin-8*.

	process	forward primer sequence	reverse primer sequence	amplification product length (bp)
1	RT-PCR of partial sequence	CGGCACTAGCACCAATCCC	GGTACCTGATGGAGTGGATGTGTACC	302
2	3‘RACE amplification	GAGCTGGCAGCATCGTCAAC	-	(RACE)
3	3‘RACE nested amplification	GTGGAAAGGTTCGGGCTAACC	-	679
4	comparative qPCR	AAACATGGAGCTGGCAGCAT	TGAGTCTCCCTGGAGTCATTCC	82
5	RT-PCR of full-length CDS and cloning in p*EXP-5N*	ATGGCATACCCATACCCACAAGC	GCACCCGAAAAATCATCCCTAC	1124
6	cloning in p*Trc-CT*	ATGGCATACCCATACCCACAAGC	CTGGAAGCGAATCTGGTGGATG	
7	cloning in pCOLD-TF	GGACATATGGCATACCCATACC	GCTCTAGATTACTGGAAGCGAATC	950

Raise of polyclonal antibodies in mice

Polyclonal antibodies (pAb) were prepared by Dr. B.Diehl-Seifert, (University of Mainz, Germany) against the purified, recombinant *GALECTIN-8* (*rPl-Gal8*) in mice. 1.5 µg of recombinant protein per injection was dissolved in phosphate-buffered saline (PBS). After three boosts the serum was collected; the pAb against *rPl-Gal8*, was termed pAb-Gal8. The titer of the pAb-Gal8 was 1:4000. The cross-reactivity was tested against 10 µg/ml of recombinant protein by ELISA and gave a titer of 1:4000.

Hemagglutination assay

The carbohydrates-binding activity of the recombinant *Pl-GALECTIN-8* was assayed by measuring its ability to agglutinate human erythrocytes (Tribulatti *et al.*, 2007). Briefly, 50µL of a 4% v/v suspension of PBS-washed fresh human red blood cells (RBCs) were mixed with 50 µL of serial dilutions of the recombinant Pl-GAL-8 in PBS, concentrations ranging from 0.01 µM to 1 µM, and were disposed in wells of U-shaped 96-wells microtiter plates. The test was performed in duplicates for each given concentration, plates were

incubated for 40 min at room temperature in the dark, and results were recorded by macro-photography. In control experiments, the recombinant Pl-GAL-8 protein was denatured at 100°C for 15 min prior to incubation with the RBCs, to show the requirement of the conformational folding of the protein for the maintenance of its biological activity. In addition, bovine serum albumin was used at different concentrations from 1 to 20 µM as negative control of a non specific protein. To exclude the possibility that the trigger factor linked to Pl-GAL-8 could contribute to the activity of the recombinant protein, control experiments were performed using the expressed and purified trigger factor protein alone. To test the sugar moiety specificity of the recombinant Pl-GAL-8, the inhibition of hemagglutination was monitored after different sugars: lactose, galactose, glucose, sucrose, maltose, mannose, (Sigma) were added to the wells at concentrations ranging from 1µM to 20 µM.

Cell adhesion assay

HepG2 cells, grown in complete medium containing 10% (v/v) Fetal Bovine Serum (FBS), were detached from tissue culture plates and resuspended (3x105 cells/ml) in the absence or presence of soluble recombinant human galectin-8 (h-Gal-8, R&D Systems) or recombinant *P. lividus* galectin-8 (*Pl-Gal-8*). The cells were seeded (100 µl per well) in 96-well tissue culture plates and incubated at 37°C. After 3 h, the cells were carefully washed three times with RPMI without FBS and three times with phosphate-buffered saline (PBS). Adherent cells were quantified according to Li *et al.*, 2010. Briefly, taking advantage of the reported autofluorescence induced by glutaraldehyde, cells were incubated with 100 µl of 0.5% glutaraldehyde in PBS for 5 min at room temperature. Then, cells were washed three times with PBS and the plate was scanned using an Odyssey Infrared Imaging System (LI-COR Biosciences, Lincoln, NE, USA) at 169 µm resolution (medium quality, 3.0 mm focus offset, intensity setting of 7) for 700-nm channel. Signal was quantified using Odyssey system application software. Mrs. Antonina Azzolina from Dr. M. Cervello's group (IBIM, CNR) is acknowledged and thanked for the HepG2 cells growth and manipulation. Dr. Francesca Zito is acknowledged and thanked for the analysis of the data.

Results

Identification, cloning and phylogenetic analysis

The complete galectin-8 cDNA was isolated and cloned from *P. lividus* sea urchin embryos. The compiled nucleotide sequence has a length of 1309 nt, including a 936 nt-long CDS and a 373 nt-long 3‘UTR (Acc. Num. FR716469).

The full length cDNA was obtained by amplifications of RT-PCR of a partial CDS (Fig.61) and 3‘RACE (Fig.62) using cDNA prepared from embryos at gastrula stage. The 3‘RACE amplification product included a fragment of 777 bp involving 353 bp of partial CDS

(partially overlapping with the previously known CDS from the EST data mining) and 424 bp of 3‘UTR. Gene specific primers were subsequently designed on the basis of the isolated overlapping fragments. The full-length amplification product was cloned into the *pGEM-T-Easy* vector (Promega) under the control of the T7 promoter and antisense to the Sp6 promoter. The plasmid was designated as *pGEM-T-Easy-Pl-galectin-8* (Fig.63).

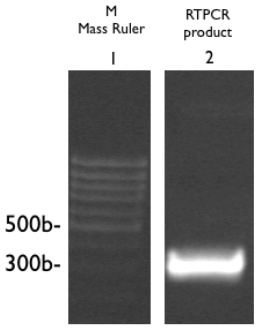


Figure 61. RT-PCR amplification of partial galectin-8 from total RNA extracted from *P. lividus* gastrula embryos in 1% agarose TBE gel.

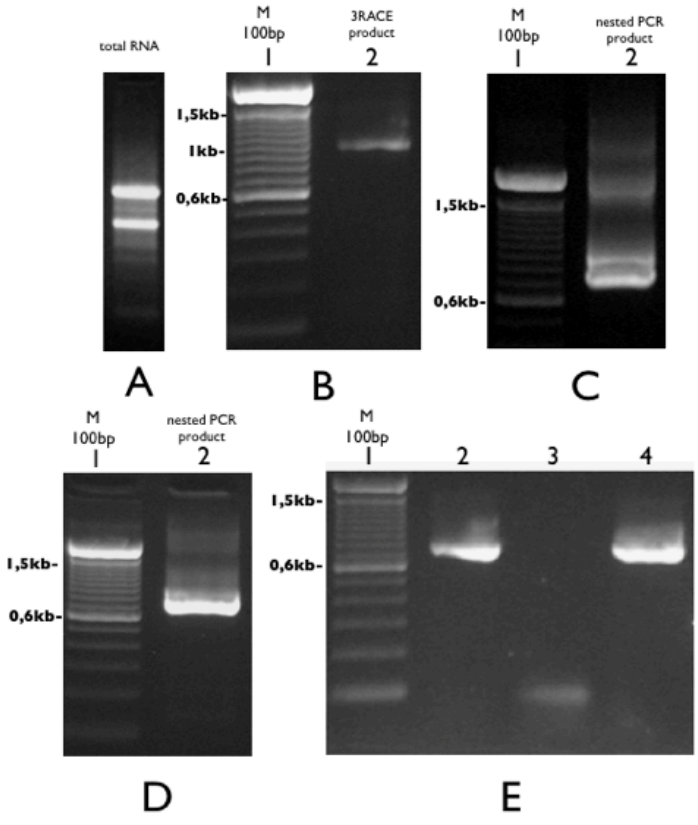


Figure 62: Summary of the 3‘RACE cloning procedure of *Pl-galectin-8*. A: extraction and purification of total RNA; B: Amplification of *Pl-galectin-8* 3‘RACE product by OneStep RT-PCR; C,D: Nested amplifications of *Pl-galectin-8* 3’RACE product by PCR; E: Amplification products by PCR templated with *Pl-galectin-8* 3’RACE product previously cloned in p*GEM-T-Easy* vector. Lanes 2 and 4 represent positive clones.

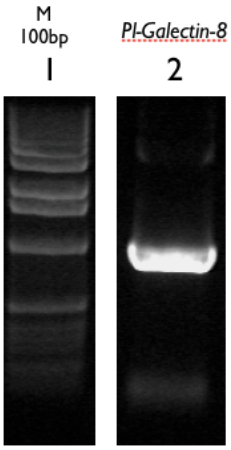


Figure 63. Amplification product of the full-length galectin-8 CDS, by PCR from p*GEM-T-Easy-Pl-galectin-8.*

The isolated cDNA showed the maximum identity (85%) with both *S. purpuratus* sequences: the predicted galectin-8-like, having a 2038 bp-long mRNA (Acc Num. XM_776778) identified at NCBI and the SPU_006306 sequence identified in the Sea Urchin Genome database (http://www.spbase.org/SpBase/search/) and annotated as Sp-Gal/lec3 (Cameron *et al.,* 2012).

The deduced Pl-GAL-8 amino-acidic sequence contains 311 residues, with a predicted pI of 9.39 and a MW of 34.73 kDa. The sequence contains two 135 aa-long and 134 aa-long carbohydrate-recognition (CRD) domains: the first spanning from V24 to Q158 and the second from Y178 to Q311 (Fig.64-A). The amino acidic sequence blasted at NCBI identified a number of uncharacterised galectin-8 proteins from many phyla and some proteins characterised as galectin-8 from mammalian. The highest identity (88%) and similarity (93%) was shown by the predicted *S. purpuratus* galectin-8 like (XP_781871). The other proteins showed decreasing percentage of identity: 54% (predicted; *B. floridae*: XP_002610290.1), 51% (predicted; *S. kovalevski*: XP_002731587.1), 46% (G. gallus: NP_001010843.1; Bath *et al*., 2011), 45% (*H.sapiens*: AAF19370.1; AAD45402.1;), 43% (*R. norvegicus*: NP_446314.2; Hadari *et al*., 2005), 42% (*M.musculus*: NP_061374.1; Tribulatti *et al*., 2012), 40% (Ostrea edulis: ADF80416.1; Morga *et al*., 2011). The *in silico* analysis predicted 10 putative phosphorylation sites, including six serine residues (S33, S98, S154, S170, S243, S252), three threonine residues (T84, T143, T187) and one tyrosine residue (Y164).

With the exception of the S33, only found in *P. lividus*, all the other putative phosphorylation sites were also found in *S. purpuratus.* The S243 putative site of phosphorylation was conserved in chicken and human. For what the tyrosines putative phosphorylation sites are concerned, T84 and S154 were found conserved in human, rat, chicken, and acorn worm, whereas T143 was found conserved in human, rat, chicken and amphioxus. One potential glycosylation motif was found at position 250-253: NASY (Fig. 64-B).

By the ClustalW software (Thompson *et al*., 1994) a phylogenetic analysis was performed comparing Pl-GAL-8 with homologous protein sequences, including *S. purpuratus*, and

sequences of different deuterostomes, from hemichordates to humans. Figure 64-B shows the alignment with sequences of *S. purpuratus* (XP_781871), *Saccoglossus kowalevskii* (XP_002731587), *Branchistoma floridae* (XP_002610290.1), *Ciona intestinalis* (XP_002126450.1), *Gallus gallus* (AAW51134), *Rattus norvegicus* (AAH72488), *Homo sapiens* (AAF19370.1). To note that, with the exclusion of *G. gallus*, *R. norvegicus* and *H. sapiens*, all the other sequences have been only predicted to code for galectin-8, the description of their cDNA identification being still awaited. As expected, sequence comparison showed the highest identity (88%) with the 278 aa-long *S. purpuratus* predicted Galectin-8. All other sequences showed medium-high identity that varied from 30% for *C. intestinalis* to 54% for *B. floridae.*

Based on the alignment, we generated a NJ phylogenetic tree (Fig.64-C) which showed that among chordates, that cluster separately from hemichordates, galectin-8 from cephalochordates is phylogenetically closer to echinoderms galectin-8 (*P. lividus* and *S. purpuratus*). Among vertebrates, birds and mammals cluster separately, the last one branch being divided in two parts, one for rodents and one for humans.

By blasting the *Pl-galectin-8* sequence at the ESTs database (*P. lividus* embryonic stages libraries) we identified three partial cDNA sequences, named AM551138, AM211139, AM599276, having sizes of 866, 822, 807 nucleotides, containing 5'UTRs of 170bp, 90bp, 199bp, coding regions of 696 bp, 740 bp, 608 bp and having 98.7%, 90.5%, 92,7% identities with *Pl-galectin-8*. Local sequence similarities by comparison of Pl-GAL-8 with each of the deduced amino acid sequences (232 aa, 246 aa, 203 aa) showed high values of identity: 99.1%, 92.3% and 98.8%. From this *in silico* analysis it is hypothesized that they are *Pl-galectin-8* isoforms, possibly generated by alternative splicing.

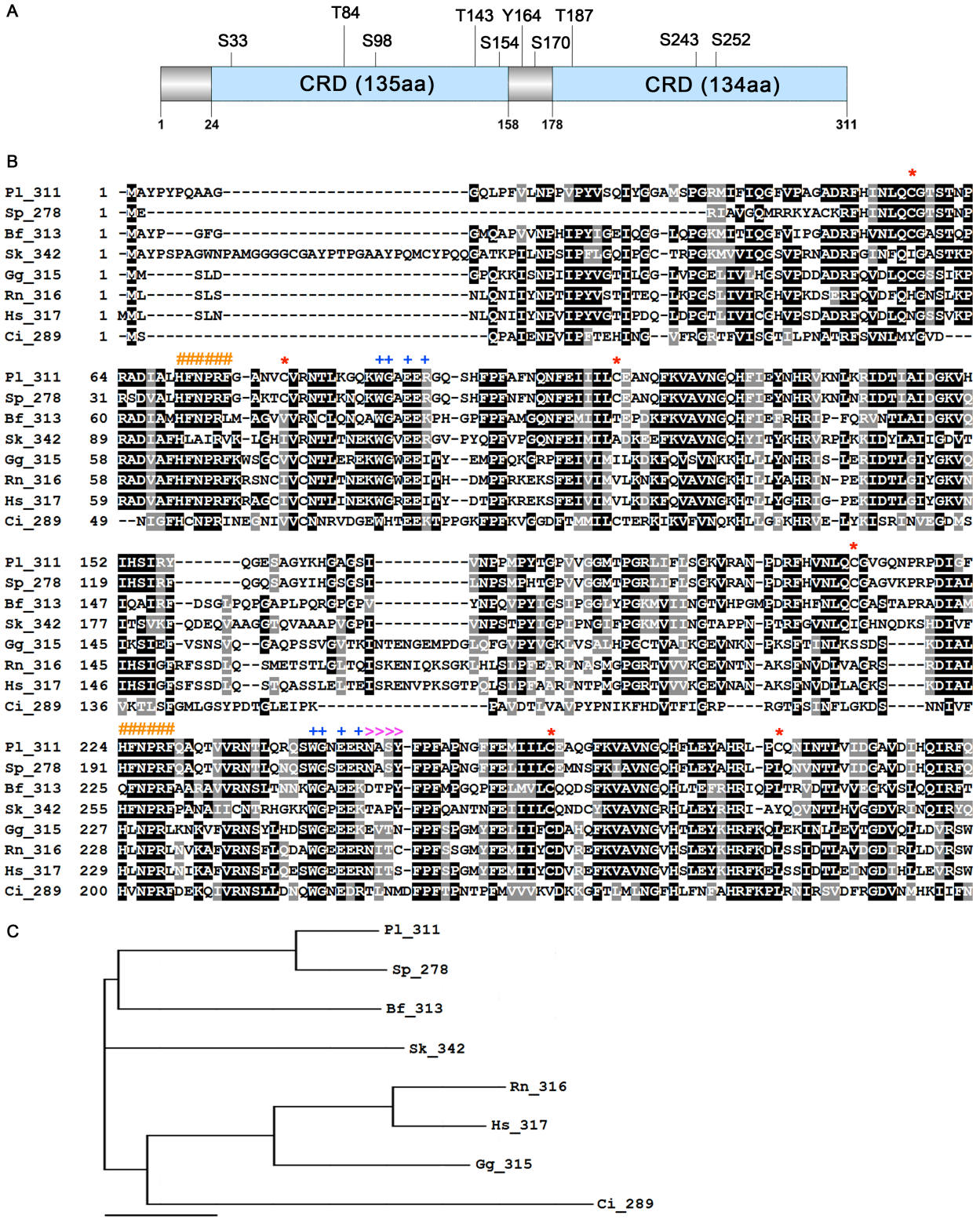


Figure 64. Sequence analysis of *Paracentrotus lividus* Galectin-8. A) Graphic representation of *Pl-Galectin-8* protein structure composed of a 23 aa-long N-terminal region and two tandem repeat CRD domains (24-158 and 178-311), joined by a linker region (159-177). Positions of putatively phosphorylated serine, threonine and tyrosine residues are indicated. B) Alignment of the Pl-Galectin-8 deduced sequence with

homologs from: Sp, *Strongylocentrotus purpuratus* (XP_781871); Bf, *Branchistoma floridae* (XP_002610290.1); Sk, *Saccoglossus kowalevskii* (XP_002731587); Gg, *Gallus gallus* (AAW51134); Rn, *Rattus norvegicus* (AAH72488); Hs, *Homo sapiens* (AAF19370.1); Ci, *Ciona intestinalis* (XP_002126450.1). Numbers after the species acronyms refer to the size in amino acids of the deduced proteins. Identical amino acids are shaded in black, conservative amino acids substitutions are shaded in grey. Broken lines indicate the gaps inserted for maximizing the match. C) Phylogenetic relationships among members of the galectin-8 family. Rooted phylogenetic tree was derived from (B). Branch lengths are proportional to evolutionary distance showing the divergence among different species. The scale bar indicates an evolutionary distance of 0.1 aa substitutions per position in the sequence (Karakostis *et al.*, in preparation).

Domain homology and modeling

The homology between the two *Pl-Galectin-8* tandem domains was studies by alignment of their amino acidic sequences using the LALIGN program (http://www.ch.embnet.org/software/LALIGN form.html). As expected for CDR domains, a high sequence identity (62.8%) and similarity (83.7%) was found, indicating an evolutionary-driven gene duplication typical of most tandem repeated domains (Vasta, 2012). To further characterize the Pl-GALECTIN-8 domains, the sequence of each of the two domains was Blasted at NCBI and found that the N-terminal domain has the highest identity (49.6 %) and similarity (75.6%) values with the N-terminal domain of the human GALECTIN-8. The *Pl-galectin-8* C-terminal domain showed lower values (identity 46%- similarity 63%). So far, fifteen 3D X-Ray structures of N- and C- terminal domains of the human galectin-8 have been described and deposited at the protein data bank (PDB) (http://www.rcsb.org/pdb), including their association to carbohydrates or oligosaccharides. Figure 65 shows the 3D structure model of Pl-galectin-8 N-terminal domain, obtained with the Modeller software (http://salilab.org/modeller/), based on the high primary sequence identity and structural similarity with the PDB structure of the N-terminal domain of human Galectin-8 (pdb file: 2yv8; Kishishita *et al.*, 2008). The model shows a bend-shaped structure, with two opposing sheets, each one containing antiparallel β-strands forming a β-sandwich arrangement (Fig.65, insert top left). One of the two β-sheets forms a concave region (cleft), described as containing the carbohydrates binding sites of the human galectin-8 N-terminal domain (Ideo *et al.*, 2011). By comparison with the human 3D structure (pdb: 3AP4), using the Calculate Structure Alignment 1.1 software developed at the RCSB Protein Data Bank, we identified the nine lactose-binding sites on the *Pl-gal-8* model. A part for H52, that in the human galectin-8 corresponds to Q47, all the other lactose-binding amino acids identified in *Pl-gal-8* are evolutionary conserved. Specifically, by computational analyses we identified the following amino acids: R50, H70, N72, R74, N83, and E93 shown to directly interact with lactose via hydrogen bonding (Ideo *et al.*, 2011); W90 forming hydrogen bonds with lactose and participating in van der Waals interactions

with the galactose ring (Ideo *et al* 2011); H52 and R64 forming a water-mediated hydrogen bond with lactose (Ideo *et al.*, 2011).

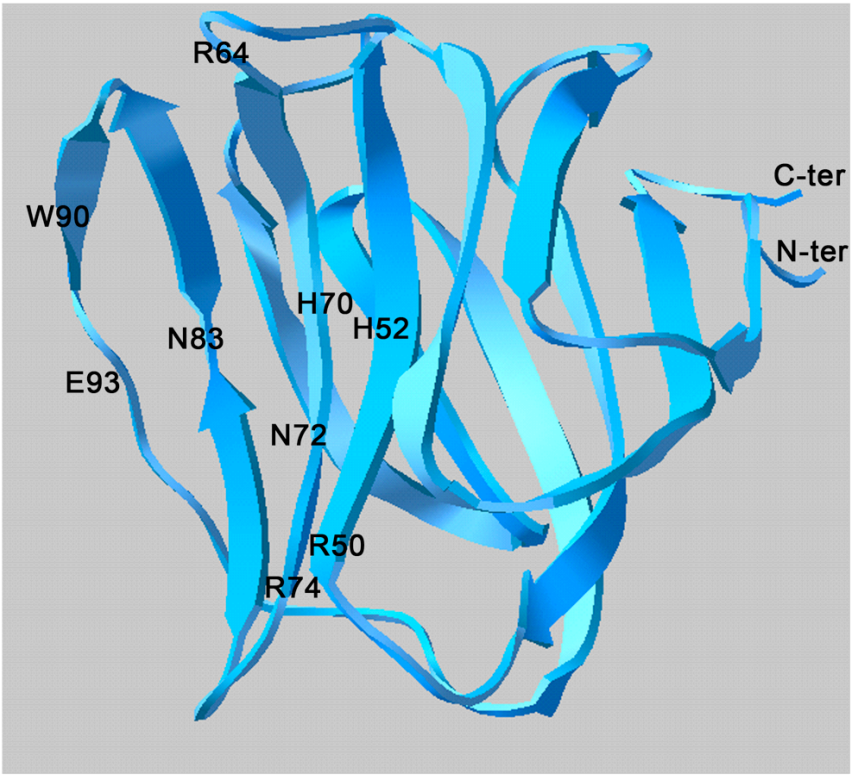


Figure 65. Three dimensional structure model of *Pl-galectin-8* N-terminal domain. Ribbon model of *Pl-galectin-8* N-terminal domain obtained by homology modelling using as template the N-terminal domain of the human Gal-8 (pdb file: 2yv8; Kishishita *et al.*, 2008). In the insert (top left), the domain is spatially oriented to show the CRD cleft on the left. The indicated amino acid residues point to specific carbohydrates binding sites shown by Ideo *et al.*, 2011 to interact with lactose via hydrogen bonding (R50, H70, N72, R74, N83, W90 and E93) or water-mediated bonds (H52 and R64) and with the galactose ring by van der Waals interactions (W90). N-ter, N-terminal; C-ter, C-terminal. (Karakostis *et al.*, in preparation).

Temporal and spatial gene expression profiling of *Pl-galectin-8*

The temporal expression profile of *Pl-galectin-8* during embryo development was analyzed by comparative real-time QPCR (ΔCCt QPCR). *Pl-galectin-8* mRNA was barely detectable at cleavage stage and gradually increased with time, at the blastula and gastrula stages, to reach a maximum at the pluteus stage, where it was highly up-regulated (Fig.66).

It was of some interest to determine if the *Pl-galectin-8* transcripts were restricted to specific territories during development. To this aim, embryos at different developmental stages were hybridized with a *Pl-galectin-8* antisense (AS) DIG RNA probe, containing the

complete open reading frame. The AS DIG-labelled RNA probe was amplified by asymmetric PCR with the *Sp6* promoter from the *pGEM-T-Easy-Pl-Galectin-8* plasmid after digestion by *Sac*II which resulted in linearization (Fig.67).

WMISH was carried out on embryos collected at different developmental stages, from early blastula to pluteus (Fig.68). We found that *Pl-galectin-8* transcripts were not detected at the blastula stage (Fig.68-A), firstly detected at the mesenchyme blastula stage, localized in the few cells which will give rise to the presumptive endoderm territory at the vegetal plate of the embryo (Fig.68-B). At gastrulation, when the archenteron starts to invaginate from the vegetal plate, transcripts were also localized at the tip of the archenteron where the so called secondary mesenchyme cells (SMC) originate and elongate to touch the animal pole, opposite to the vegetal plate of the embryo (Fig.68-C,D, see arrows). At the prism stage *Pl-galectin-8* mRNA was detected in the three differentiated parts of the gut (foregut, midgut and hindgut), with a higher signal at the level of the two sphincters, namely the constrictions separating the oesophagus from stomach and stomach from intestine (Fig.68-E). No other territories appear labelled at the latest developmental stages, such as the pluteus larva, where the *Pl-galectin-8* mRNA is highly restricted to intestine (Fig.68-F).

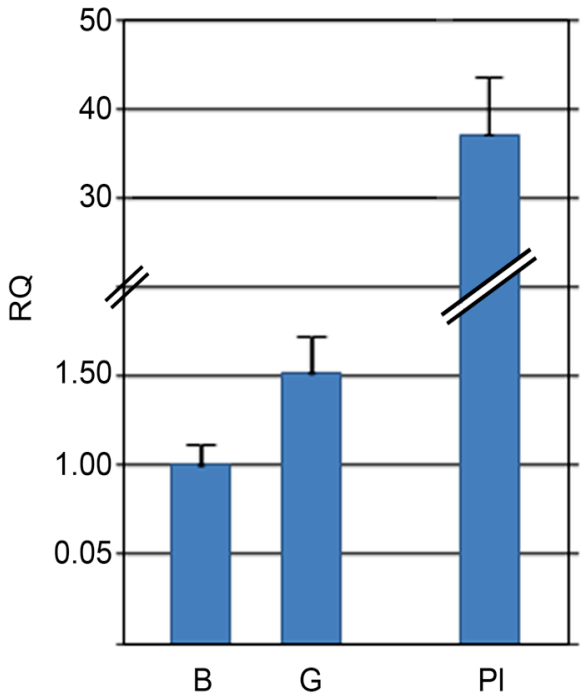


Figure 66. Temporal expression of *Pl*-galectin-8 during embryo development. Comparative QPCR analysis of the *Pl*-Galectin-8 transcription levels in sea urchin embryos at different developmental stages: B, blastula; G, gastrula; Pl, pluteus. The *Pl*-Z12-1 mRNA was used as an internal endogenous reference gene; cDNA from the cleavage stage was used as the reference sample and was set as 1. q-PCR experiments were performed at least three times (Karakostis *et al.*, in preparation).

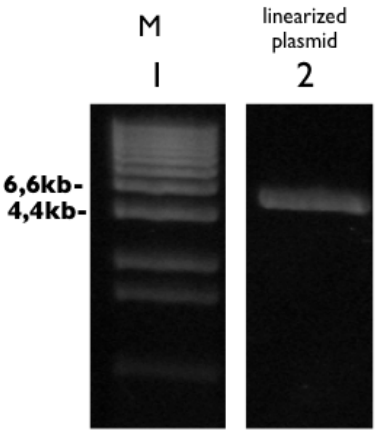


Figure 67. Preparation of WMISH probe. Linearization of *pGEM-T-Easy-Pl-Galectin-8* by digestion with *Sac*II for the preparation of the AS *Pl-galectin-8* DIG RNA probe.

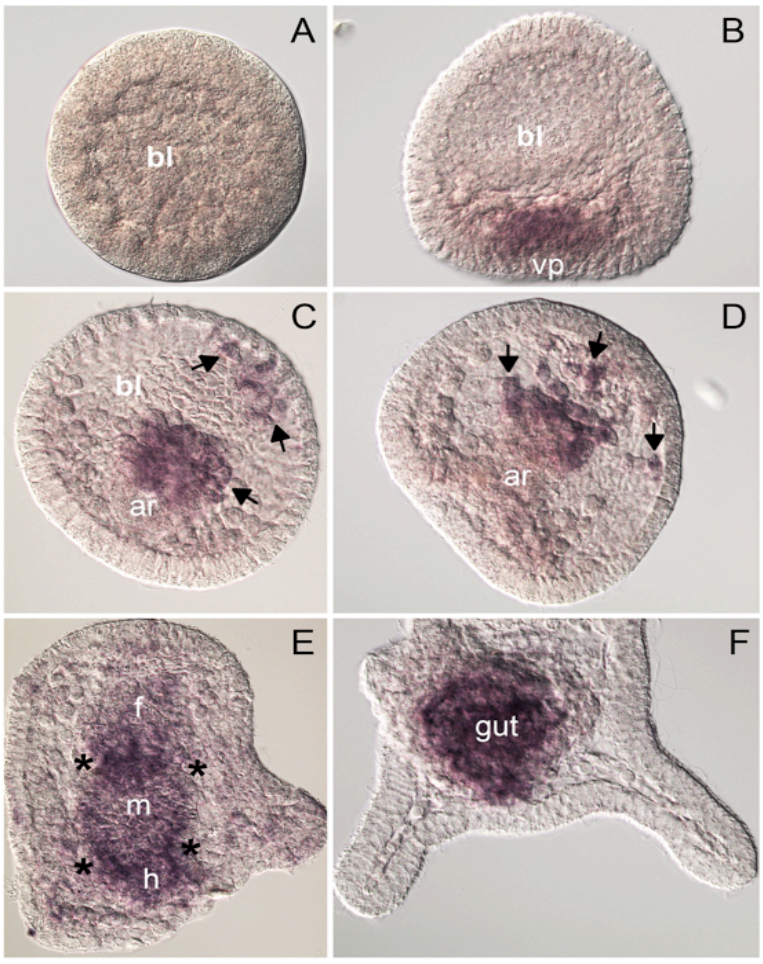


Figure 68. Spatial expression of *Pl-galectin-8* during development. Whole mount *in situ* hybridization using a DIG-antisense *Pl-galectin-8* RNA probe on embryos fixed at different developmental stages: A) blastula; B) mesenchyme blastula; C) middle gastrula; D) late gastrula; E, F) pluteus. Arrows point to secondary mesenchyme cells; asterisks to intestine constrictions where muscle cells are located (Karakostis *et al.*, in preparation).

Preparation of recombinant *Pl-GALECTIN-8* protein

Galectin-8 CDS was cloned in p*EXP-NT*, p*Trc-CT* (Invitrogen), p*COLD-TF* (Takara) and p*ET32b*$^{+}$*.* Inserts were amplified by PCR and ligated to the corresponding vectors. p*EXP-N-Pl-Galectin-8* (Fig.69-A), p*TRC-C-Pl-Galectin-8* (Fig.69-A) and p*COLD-TF-Pl-Galectin-8* were prepared (Fig.69-B). The p*COLD-TF-Pl-Galectin-8* cloning involved the digestion of the vector and the PCR product, before the ligation. After screening the plasmids extracted from the positive colonies, the correct construct was tested and sequenced. The cloning processes is briefly outlined in Fig.69.

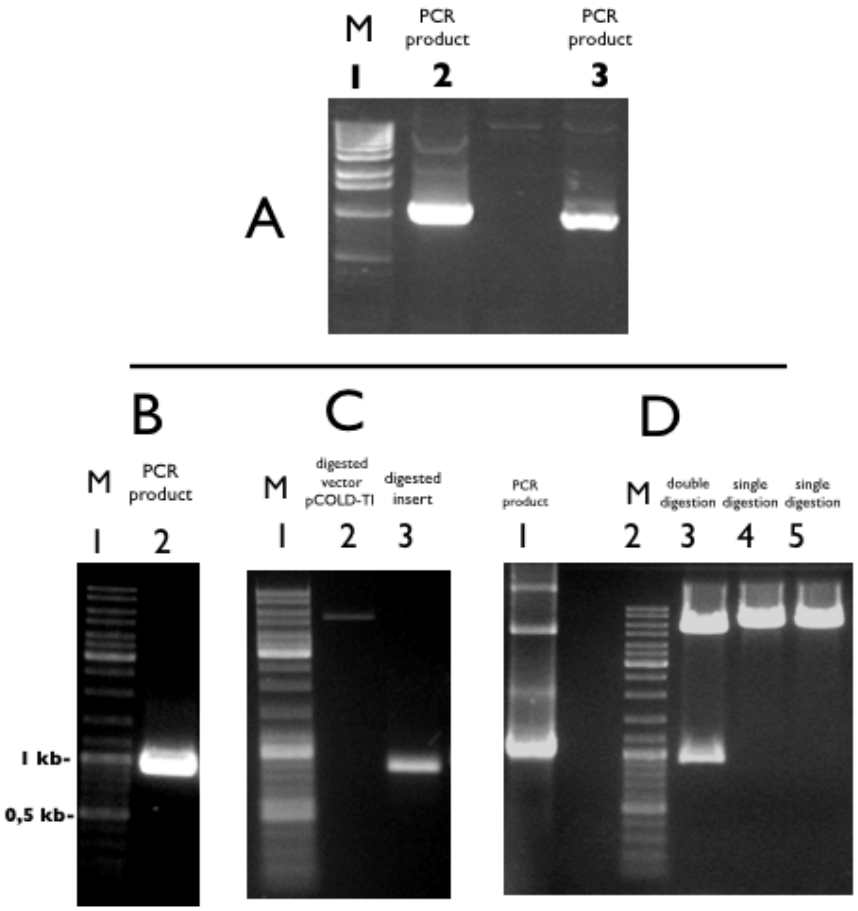


Figure 69. 1% Agarose-TBE gel. Cloning of full-length CDS of *Pl-Galectin-8* in expression vectors. A) Lane 1: Molecular ladder. Lane 2: Amplification by PCR of the CDS from the cloned p*EXP-N-Pl-Galectin-8;* Lane 3: Amplification by PCR of the CDS from the cloned p*TRC-C-Pl-Galectin-8.* B,C,D) Cloning process of p*COLD-TF-Pl-Galectin-8.* B: Amplification by PCR of the *Pl-gal-8* from p*GEM-T-Easy-Pl-gal-8*; C: digestion of the pCold vector and the the *Pl-gal-8* CDS with equivalent restriction enzymes; D: Screening of positive colony. Lane 1: Test PCR amplification of the full-length *Pl-gal-8* from the previously cloned p*COLD-TF-Pl-Galectin-8* plasmid. Lane 3,4,5: Test digestions of p*COLD-TF-Pl-Galectin-8* with *Nde*I and *Xba*I (Lane 4), *Nde*I (Lane 5) and *Xba*I (Lane 5).

The expression plasmids p*EXP-N-Pl-Galectin-8* and p*Trc-C-Pl-Galectin-8* were tested for expression of recombinant *Pl-Galectin-8* in Bl21(DE3) E. *coli* strain cells and yielded low amounts of recombinant protein. Thus, to produce large amounts of active Pl-GAL-8 recombinant protein to be used later in functional assays, the full-length CDS of *Pl-galectin-8* was used that was inserted into *pCold* vector containing the chaperone trigger factor (TF), generally used for the production of soluble and functioning proteins. The resulting construct contains also a 11aa N-terminal 6x His-tag, which is needed for protein detection and purification (Fig.70-A). The bacterial extracts containing the fusion protein were analysed by SDS-PAGE, as shown in Fig.70-B; a thick band was detected in the affinity chromatography eluted fractions, having an apparent protein size consistent with the calculated molecular weight of 89.0 kDa of the fusion protein. The *Pl-GAL-8* recombinant purified protein was obtained after elimination of the low molecular weight contaminants by ultracentrifugation (50 kDa cut-off) and dialysis of the pooled fractions (Fig.70, right lane). The dialysis was necessary to reduce salt concentration and remove compounds, such as imidazole, that could eventually interfere with the protein folding and activity.

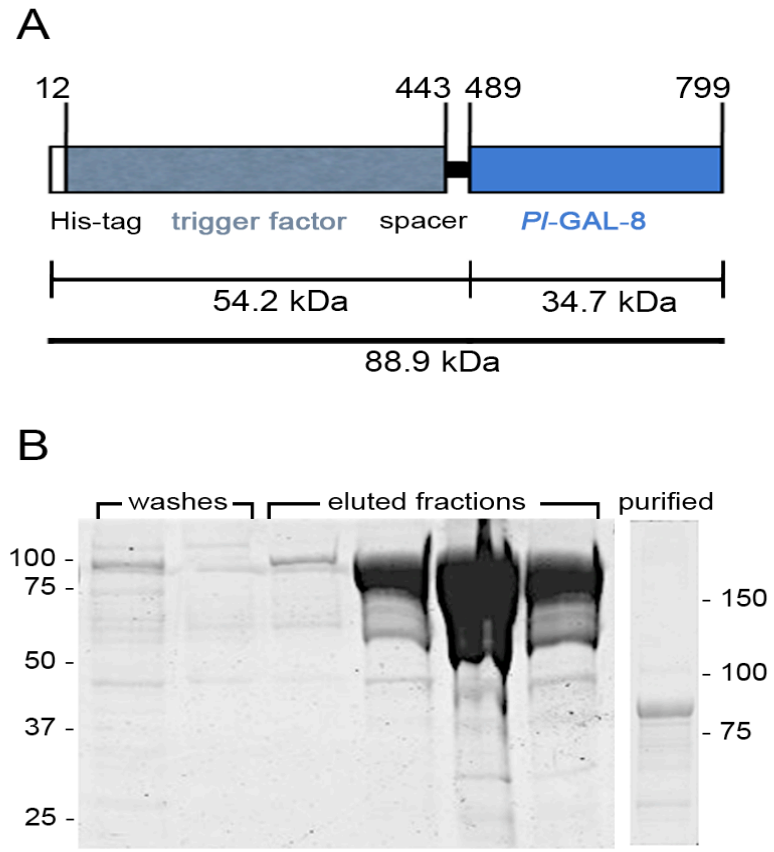


Figure 70. Recombinant *Pl*-GALECTIN protein purification. A) Schematic representation of the construct used for the expression of the recombinant *Pl*-GALECTIN fusion protein, including: a N-terminal His-tag, the trigger factor chaperon protein, a 45aa spacer and the *Pl*-galectin protein. Numbers above the scheme refer to amino acids of the construct. The calculated sizes of the protein segments are listed below, expressed in kDa. B) SDS-PAGE (7.5%) of washes and eluted fractions from the Ni-nitrilotriacetate column affinity chromatography of extracts containing the fusion protein from transformed bacteria by recombinant pCOLD-TF-*Pl*-galectin-8 plasmid. Molecular weight markers are indicated on the left. In the separated lane on the right, the purified fraction obtained after 50kDa cut-off, dialysis and concentration; molecular weight markers on the right. (Karakostis *et al.*, in preparation).

Pl-GAL-8 carbohydrate binding activity and specificity

As a first trial to establish the biological activity of the recombinant Pl-GAL-8 we tested its ability to induce agglutination of RBCs by serial dilutions in a microtiter well plate assay as described in the Materials and Methods section. We found that the minimum concentration able to induce agglutination was 0.25μM (Fig.71). Loss of Pl-GAL-8 activity was evident after inactivation of the recombinant protein at 100°C for 15min. BSA used as negative controls did not induce RBC agglutination, even at the highest dose tested (1 μM). The isolated and purified TF, obtained from the expression of the pCOLD vector, was also tested and shown to be inactive, thus demonstrating that it does not contribute to the

activity of the Pl-GAL-8 recombinant protein. The high similarity of *Pl-galectin-8* with other members of the family suggested it might have a unique sugar-binding specificity. To address this possibility we studied the ability of increasing concentrations of different mono- and di-sacharide sugars to interfere with the hemagglutinating activity of 0.25µM recombinant Pl-GAL-8. As shown in Figure 71, agglutination of RBC cells was efficiently inhibited by lactose at all doses tested (from 1 µM to 20 µM) and, to a certain extent, by galactose at 20 µM. No inhibition was observed with other sugars, like glucose, sucrose, maltose or mannose.

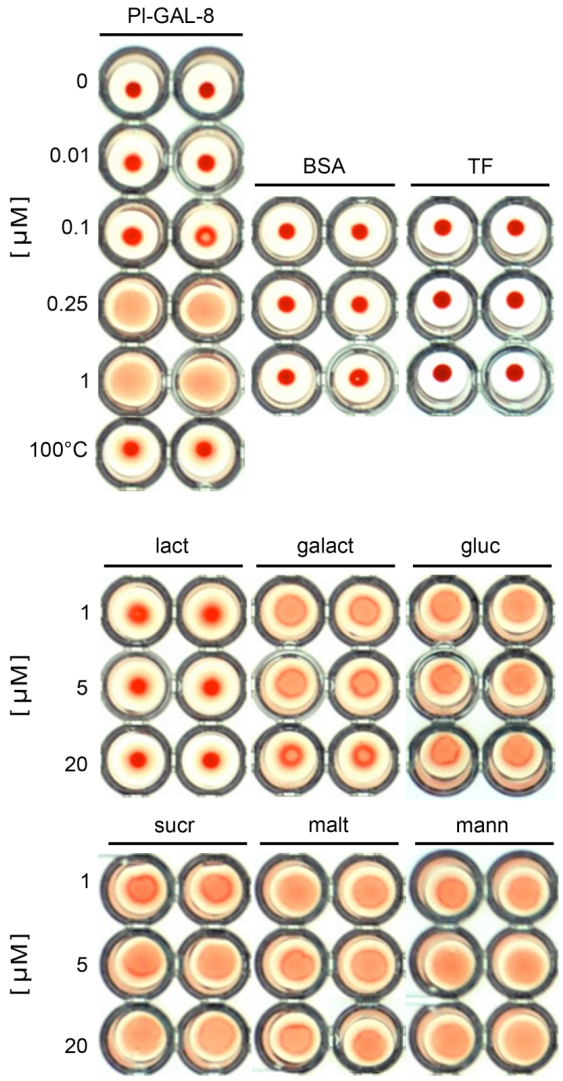


Figure 71. Carbohydrates-binding activity of recombinant *Pl*-GAL-8. Hemoagglutination assays on human red blood cells run in duplicates in U-shaped wells at the micromolar concentrations indicated on the left. Negative controls: 100°C, contains the highest concentration of recombinant protein denatured at 100°C for 5 min before testing; BSA, serum bovine albumin used as non specific protein; TF, trigger factor alone, to show that it does not contribute to the activity of the recombinant protein. Inhibition of hemoagglutination was tested in the presence of mono- and di-sacharides sugars at the micromolar concentrations indicated on the left: lact, lactose; galact, galactose; gluc, glucose; sucr, sucrose; malt, maltose; mann, mannose (Karakostis *et al.*, in preparation).

Anti-adhesive activity of the recombinant Pl-GAL-8

To further characterize the functional activity of the purified recombinant protein, given the sugar binding features, we tested its effects on the modulation of cell adhesion. For this purpose we tested the recombinant Pl-GAL-8 activity on the human hepatocellular carcinoma cell line (HepG2) by measuring cell adhesion when increasing concentrations of recombinant Pl-GAL-8 were mixed in the medium with suspended HepG2 cells before seeding them onto tissue culture plates. As shown in Figure 72, Pl-GAL-8 effectively inhibited the adhesion of cells in a dose-dependent manner, as assayed 3h after plating. About 50% of inhibition was observed at the concentration of 0.2 μM, while using the highest concentration (1.6 μM), only 10% of cells adhered to the plate (Fig.72). Recombinant galectin-8 from human, used as positive control, showed similar inhibitory effect, although with the highest concentration used (1.6 μM) about 40% of cells still remained attached to the substrate.

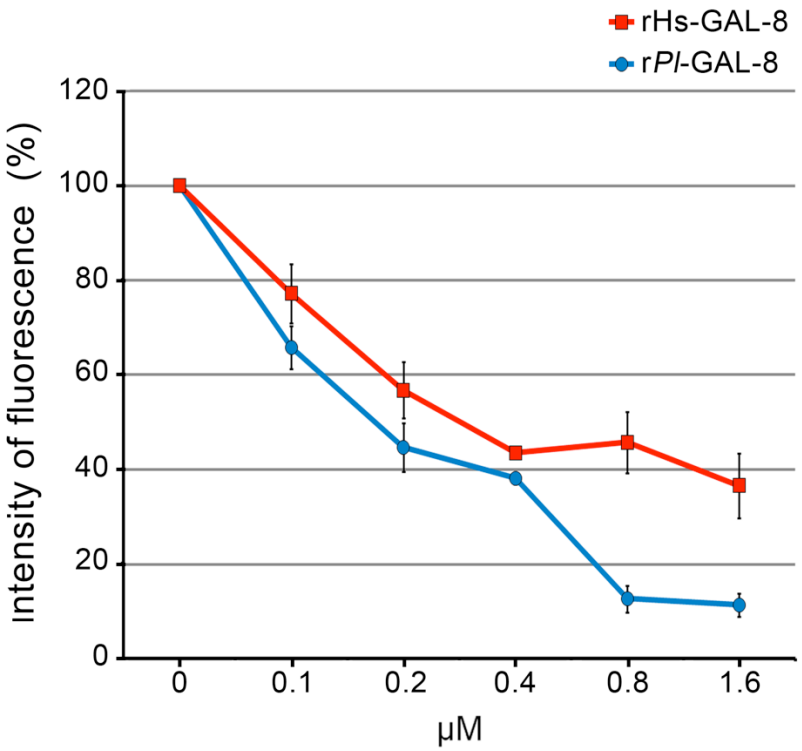


Figure 72. Inhibition of cell adhesion to culture plates by soluble Pl-galectin-8. HepG2 cells were resuspended in the absence or presence of soluble recombinant human galectin-8 (h-galectin) or recombinant *P. lividus* galectin-8 (Pl-galectin) at the concentrations shown. After seeding in 96-well tissue culture plates, cells were incubated at 37°C for 3 h. After several washings, cells were fixed with 0.5% glutaraldehyde in PBS for 5 min at room temperature. Then, the plate was scanned using an Odyssey Infrared Imaging System (LI-COR Biosciences) and the intensity of autofluorescence of adhered cells was quantified using an Odyssey system application software. Measurements of fluorescence intensities obtained in arbitrary units are expressed as % of adhered cells. Values are the mean of three replicates from a representative experiment ± SD. (Karakostis *et al.*, in preparation).

Chapter VI: Biochemical and Proteomic analysis of the organic matrix from the adult test of the sea urchin *Paracentrotus lividus*

Abstract

The formation of biomimetic materials of superior mechanical properties requires the study of naturally formed biominerals. Adult sea urchins involve several calcified structures including the test, the teeth and the spines. The sea urchin test is a remarkable example of a natural composite biomaterial, synthesized in ambient conditions. The mineral phase is composed of magnesium calcium carbonate in the form of calcite, while the occluded organic matrix regulates the crystal growth, facilitating the nucleation and growth of the crystal towards privileged directions. While, *Paracentrotus lividus* (*P. lividus*) is the most well studied Mediterranean sea urchin species, used as a model organism in developmental biology and ecotoxicology, the role of the occluded matrix proteins in the calcite deposition mechanism is not well documented and remains uncharacterised.

In the present study, we used liquid chromatography coupled with Mass Spectrometry (nanoLC-MS/MS) for the identification of matrix proteins extracted from adult *P. lividus* tests in acetic acid-soluble and insoluble form. Additionally, the presence of two biomineralization proteins (carbonic anhydrase and Galectin-8) was confirmed using previously developed specific polyclonal antibodies. Fourier transform infrared (FTIR) spectroscopy was applied for the identification of inorganic compounds of the acetic-acid-insoluble fraction. Finally, the presence of protein extracts during the precipitation of calcite in vitro resulted in variations in the shape and form of the crystal.

Introduction

Echinoderms, including Echinoids (sea urchins and sand dollars), Asteroids (sea stars), Ophiuroids (brittle stars), Crinoids (feather stars) and Holothuroids (sea cucumbers), is a phylum belonging to marine invertebrates. Echinoderms switched on a genetically regulated process to form a calcareous skeleton with variable chemical content of magnesium and remarkable variety of shapes and sizes. Among Echinoderms, sea urchins are the best studied and understood. Studies on the development of the embryo have revealed a complex gene regulatory network (GRN) employed soon after the fertilization of the embryo, which guides the development through cell cleavages, blastula, gastrula and pluteus stages, followed by the metamorphosis (Davidson, 2006). Skeletogenesis, as part of the development of the embryo, is regulated by a subset of the GRN (Ettensohn, 2009; Sharma and Ettensohn, 2010; Olivery *et al.*, 2008). The formation of the skeleton initiates during the early embryo development, 24h after the fertilization of the egg in a priviledged extracellular space. It is mediated by a set of matrix proteins specifically expressed by a cellular population devoted to skeleton formation, the primary mesenchyme cells (PMCs) and signaling pathways from overlying ectodermal cells.

Structurally, the echinoderms skeleton reflects complexity. As in all the sea urchin species, *P. lividus* exhibits an extensive endoskeleton composed of calcium carbonate containing magnesium carbonate and 0.1% w/w occluded matrix proteins (Wilt *et al.*, 2008). Adult sea urchins involve several calcified structures forming the stereom. The mineral is deposited as microscopic bony plates, called ossicles, which are deposited as a three-dimensional meshwork called the stereom.

Proteomic studies on *S. purpuratus* adult test and spines, have revealed a large number of proteins (110 proteins) within the organic matrix. The identified biomineralization proteins include C-type lectins, metalloproteases, small acidic proteins, collagens and enzymes such as carbonic anhydrase (Mann *et al.*, 2008). The extraction and purification procedure is important for the separation of the true components of the occluded matrix proteome from contaminating cellular remnants and extracellular proteins. The biochemical characterization of a subset of the identified proteins has indicated a variety of biological roles individually (Alvares *et al.*, 2009; Adomako-Ankomah *et al.*, 2012; Kitajima *et al.*, 1996). Wilt *et al.*, 2008 focused on the secretion of the mineral and the organic phase, introducing the importance of protein-mineral interactions in biomineralization.

In the present study, we used liquid chromatography coupled with Mass Spectrometry (LC-MS/MS) for the identification of matrix proteins extracted from adult *P. lividus* tests. Tests were initially thoroughly cleaned by treatment with sodium hypochloride and the occluded matrix was extracted by slow titration with acetic acid. The acetic acid soluble matrix (ASM) and insoluble matrix (AIM) fractions were concentrated and analysed separately. SDS PAGE revealed the protein content of each fraction, while the inorganic compounds of the AIM were identified by FTIR spectroscopy. Additionally, the presence of the carbonic anhydrase (CA) and galectin-8 was confirmed by specific polyclonal antibodies. Finally, the effect of the ASM on the precipitation of calcite, involving variations in the shape and form of the crystal, was assayed *in vitro.*

Experimental Procedures

Extraction of matrix proteins from mineralized parts (tests) of *Paracentrotus lividus*

Adult sea urchins (*P. lividus*) were collected from the North-Western coast of Sicily (Mediterranean Sea) and kept in aquaria, with circulating sea-water obtained from the collection site. Tests of adult *P. lividus* were mechanically separated from spines and teeth by brushing under running water. Superficial organic contaminants were removed from the surface of the test fragments with incubation for 24h in bleach (0.26% NaOCl) and were subsequently rinsed with Milli-Q water and dried at 25 ^{o}C. Samples were controlled by a scanning electron microscope (SEM) (Hitachi), before being crushed into approximately 1 mm^2 fragments and then into a fine powder (> 200 µM), in liquid nitrogen. All protein extractions were performed at 4 °C as described in B.Marie *et al.*, 2007. Briefly, the organic constituents were obtained by overnight decalcification of the biomineral (25 g) in cold acetic acid (5% v/v) at 4 ^{o}C which was added by an automated titrator (Titronic Universal, Schott, Mainz, Germany) at a flow rate of 20 µL every 1 s. The solution (final pH around 4.2) was centrifuged at 4000 g for 30 min. The resulting pellet, corresponding to the acetic acid-insoluble matrix (AIM), was thoroughly rinsed with Milli-Q water, freeze-dried and weighed. The supernatant, containing the acetic acid-soluble matrix (ASM) was collected, filtered through a mess of 5 µm mesh, concentrated by ultrafiltration through an Amicon ultra-filtration system on Millipore® membrane (YM10; 10 kDa cut-off), extensively dialyzed against Milli-Q water, freeze-dried and weighted.

Protein Matrix Analysis on SDS-PAGE

The extracted ASM was tested on monodiamentional mini SDS–PAGE under denaturing conditions (12% acrylamide gel) following standard procedure, involving heat denaturation in the Laemmli buffer (Laemmli, 1970). Briefly, each batch of ASM was diluted in Laemmli sample buffer containing mercaptoethanol. After heat denaturation (100°C for 3 min), aliquots were run on 12% polyacrylamide SDS gel. Similarly, the Laemmli-soluble AIM matrices were tested on the same gel. Staining with Coomassie Brilliant Blue (CBB) or with Silver Nitrate staining (Morrissey, 1981) revealed the protein bands.

Protein Matrix Analysis by ELISA testing

Antibodies against SM30 and SM50 were used to identify *Pl-SM30* and *Pl-SM50* at the ASM of *P. lividus* test extracts, by ELISA, following a standard procedure (Adams, 1977). Anti-SM30 and anti-SM50 antibodies produced from *S. purpuratus*, were kindly provided by Prof. Dr. F. Wilt, University of Berkeley, USA.

Protein Matrix Analysis by Immunoblotting (Western Blotting)

The ASM samples were run on two mono-dimensional gels and electro-transferred onto two separate PVDF2 Immobilon-P membranes (Millipore Corp.) with the Mini-Trans Blot module (Bio-Rad) for 90 min at 120 mA. Negative controls were performed with the pre-immune serum. The blots were blocked with 1% gelatin in Tris-buffered saline (TBS) for 30 min and then incubated overnight at 20 ^{o}C with a specific polyclonal antibody developed in mice, following standard Western Blotting procedure (Towbin *et al.*, 1979). After washing with TBS/Tween 20, a secondary goat anti-mouse antibody coupled to alcaline phosphatase (Sigma, A3687) was used for detection by incubation for 90 min, in a dilution of 1:3000 in TBS. After extensive washing, the membranes were incubated with luminase buffer for 2 min and then briefly exposed to X-Omat film (Eastman Kodak Co) for development. Blots were also stained with NBT/BCIP (Sigma, B5655).

***In vitro* interaction of the ASM with calcium carbonate**

In vitro precipitation of $CaCO_3$ was performed by slow diffusion of ammonium bicarbonate vapours into calcium carbonate chloride solution (Ramos-Silva *et al.*, 2012; Addadi, *et al.*, 1985.) (Fig.73). In detail, 200 µl of 10 mM $CaCl_2$ solutions containing serial dilutions of the ASM, were applied into a 16-well culture slide (BD Falcon, Becton Dickinson, Franklin Lakes, NJ, USA). Blank controls without any sample, were used. The samples were put in an hermetically sealed desiccator containing ammonium bicarbonate crystals. The system was air-sealed and *in vitro* precipitation of calcium carbonate occurred after incubation at 4 ^{o}C for 48h. Subsequently crystals were dried well by suction of the liquid without disturbing the crystals and by incubation at 50 ^{o}C for 2 hours. They were subsequently observed under a tabletop scanning electron microscope (Hitachi TM-1000).

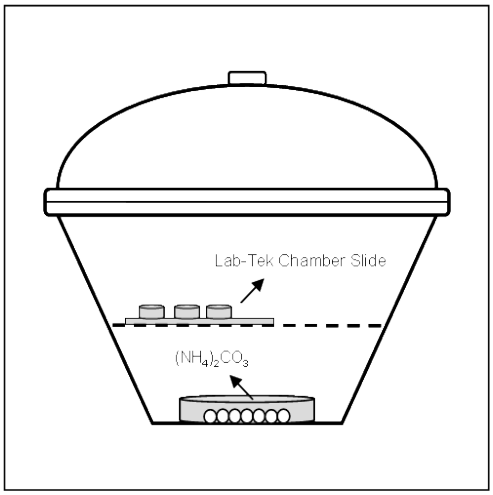


Figure 73. Scheme of desiccator. Ammonium carbonate was placed in the lower compartment. The released CO_2 vapors enter in the chamber slides containing the sample reactions.

Proteomic analysis of the ASM and the AIM; and data acquisition - nanoLC-MS/MS

The LC-MS/MS data were used for database searches using an in house version of the MASCOT search engine (Matrix Science, London, UK; version 2.1). EST sequences and nucleotide sequences derived from *P. lividus* EST libraries were downloaded (April, 2012) from the server http://octopus.obs-vlfr.fr/ and MASCOT searches were directly performed against nucleotide sequences. Two batches of separate preparations of ASM and AIM were used. LC-MS/MS data generated by each protein extract were analysed separately by Dr. Zanella-Cleon, Service de Spectrometrie de Masse, FR3302, UMS3444/US8
Institut de Biologie et Chimie des Proteines (IBCP), University of Lyon, (France).

Analysis of the acetic acid-insoluble protein matrix (AIM) by FTIR

FTIR spectra were recorded from dry lyophilized purified protein samples. IR spectra for the tablets containing the sample, were recorded in a Fourier transform infrared spectrophotometer with KBr pellets. The absorption bands were assigned on the basis of previous spectra descriptions available in the literature.

Results

Extraction of test matrix proteins

The stereome of adult *P. lividus,* involves several calcified structures including test, teeth and spines. By removing the teeth and the spines as well as the superficial organic material from the tests with incubation in sodium hypochloride (Fig.74), we were able to subsequently extract the occluded matrix proteins from the tests by treatment with acetic acid. The soluble fraction of the matrix (ASM) containing matrix proteins, representing approximately 0.005% of dry weight was separated from the insoluble matrix (AIM) containing a mixture of insoluble matrix proteins, polysaccharides and inorganic material. The AIM represented approximately 0.2% w/w.

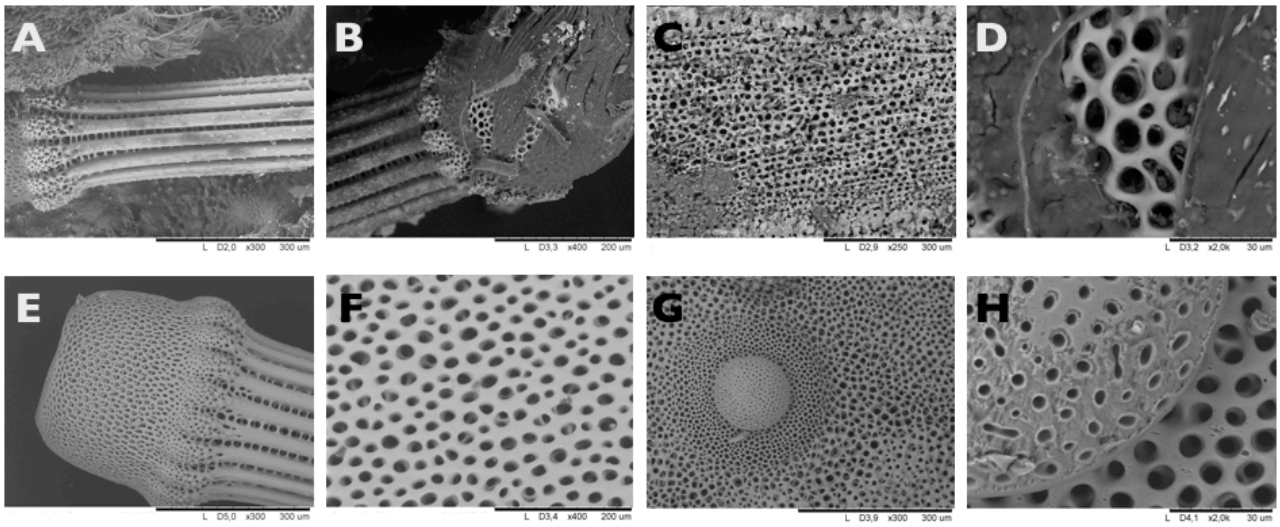


Figure 74. Examination of *P. lividus* adult tests and spines by Scanning Electron Microscopy (SEM). Spines: A,B,E; Tests: C,D,F,G,H. Upper Row: before treatment and Lower Row: after treatment with sodium hypo-chloride and extensive rinsing with water.

Protein Matrices Analyses by SDS-PAGE

The ASM and AIM fractions of the occluded matrix proteins extracted from clean tests were analysed by gel electrophoresis. An indicative assay was used to identify calcium binding proteins and the Alcian Blue stain marked the glyco-conjugated proteins.

One dimensional electrophoresis under denaturing conditions was employed to analyse the occluded protein matrix. Different types of stains were used for the visualization of the bands. Silver staining was proved more efficient than Coomassie Blue as more bands were identified in the ASM sample migrated in SDS-PAGE and stained with silver (Fig.75). Both the ASM and AIM displayed few discrete prominent bands, at 55 kDa, at 25 kDa and at 22 kDa (Fig.75). The gel stained with Alcian Blue, showed a few glycoconjugated protein bands at approximately 40 kDa and 70 kDa in the insoluble fraction (Fig.76). The indicative method of StainsAll identified various protein bands that presumably bind calcium, both in the ASM and AIM fractions. Indeed, protein bands at 10 kDa, 23 kDa and 70 kDa were stained blue. Additionally, prominent bands at 30 kDa and 40 kDa in the AIM, were also stained blue (Fig.76).

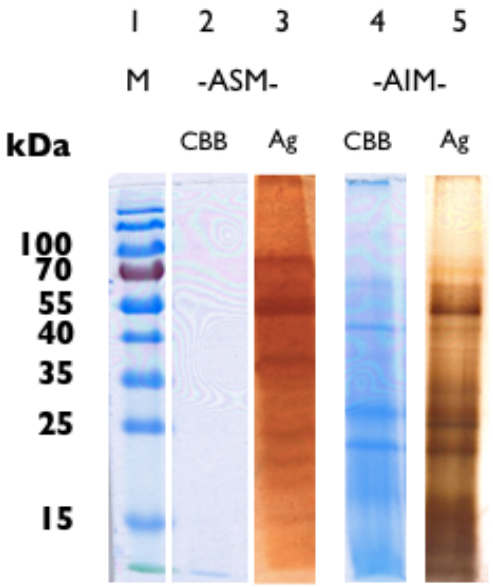


Figure 75. SDS-PAGE fractionation of acid-soluble (ASM) 2,3) and acid-insoluble (AIM) (4,5) shell matrix proteins from *Paracentrotus lividus* adult tests. Molecular weight markers (in kDa) are indicated on the left (Lane 1). Following electrophoresis under denaturing condition, proteins were stained with CBB (Lanes 2,4) or silver stain (Lanes 3,5). Approximately 50 µg of protein material were applied per well. (Karakostis *et al.*, in preparation).

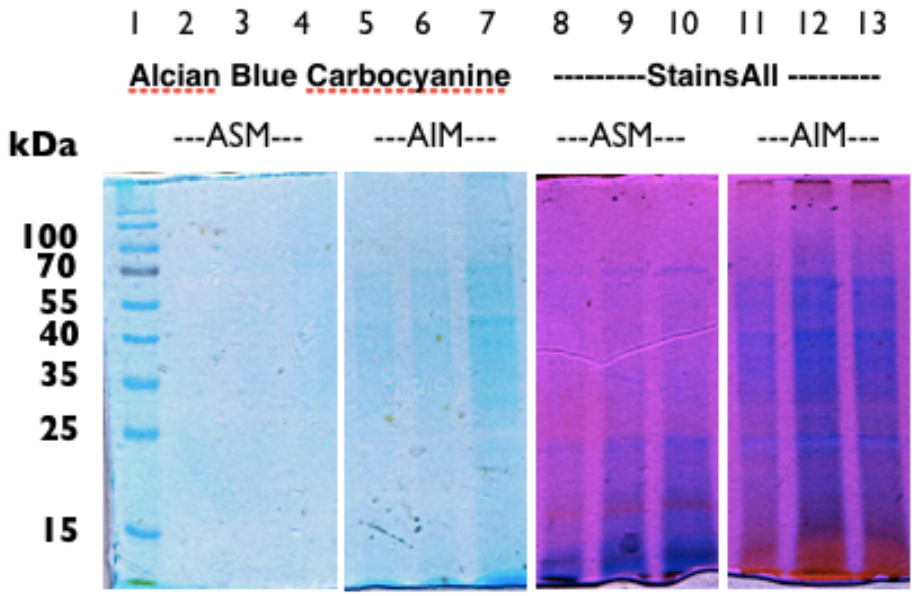


Figure 76. SDS-PAGE fractionation of acid-soluble (ASM) (Lanes 2-4 and Lanes 8-9) and acid-insoluble (AIM) shell matrix proteins (Lanes 5-7 and Lanes 11-13) from *Paracentrotus lividus* adult tests. Molecular weight markers (in kDa) are indicated on the left. Following electrophoresis under denaturing condition, proteins were stained with Alcian Blue (Lanes 1-7) or StainsAll (Carbocyanine, Lanes 8-13). 50 µg of protein material were applied per well. Alcian Blue stain involving copper, identifies heavily glycosilated proteins and glyco-conjugates. Stains-All or carbocyanine stains calcium-binding proteins.

The ASM and the AIM were analysed separately by 2D-electrophoresis (2DE). Isoelectric focusing was followed by gel electrophoresis and gels were stained by CBB and Silver. As expected the protein spots were better visualized when stained by silver. In the soluble fraction, a group of at least six alcaline proteins were visualized at around 70 kDa and 60 kDa and a single acidic spot at 30 kDa (Fig.77). The AIM mainly included a group of four slightly acidic proteins at 20 kDa and 30 kDa as well as an acidic protein spot at 130 kDa (Fig.78). The protein batches were sequenced by nano-LC-MS/MS. The preliminary screening of the identified sequences, identified various proteins in the occluded matrix test proteome, including the embryonic *PI-CA*, SM32, SM29, PM27 and MSP130 related 2 as well as lithostathine, the metalloprotease MMP17 and the fibroblast growth factor receptor 2 (FGF).

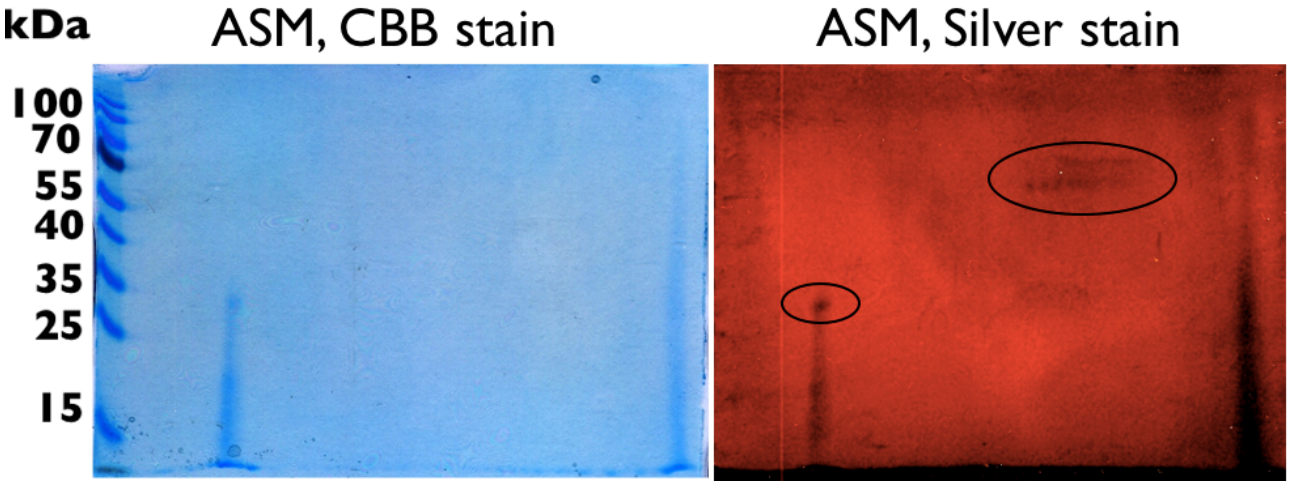


Figure 77. 2DE SDS-PAGE fractionation of acid-soluble (ASM) shell matrix proteins from *Paracentrotus lividus* adult tests. Isoelectric focusing (IEF) was followed by gel electrophoresis under denaturing condition. Molecular weight markers (in kDa) are indicated on the left. Gels were stained with CBB (left gel) or silver stain (right gel). 150 µg of protein material was applied per well. Protein groups identifies are marked.

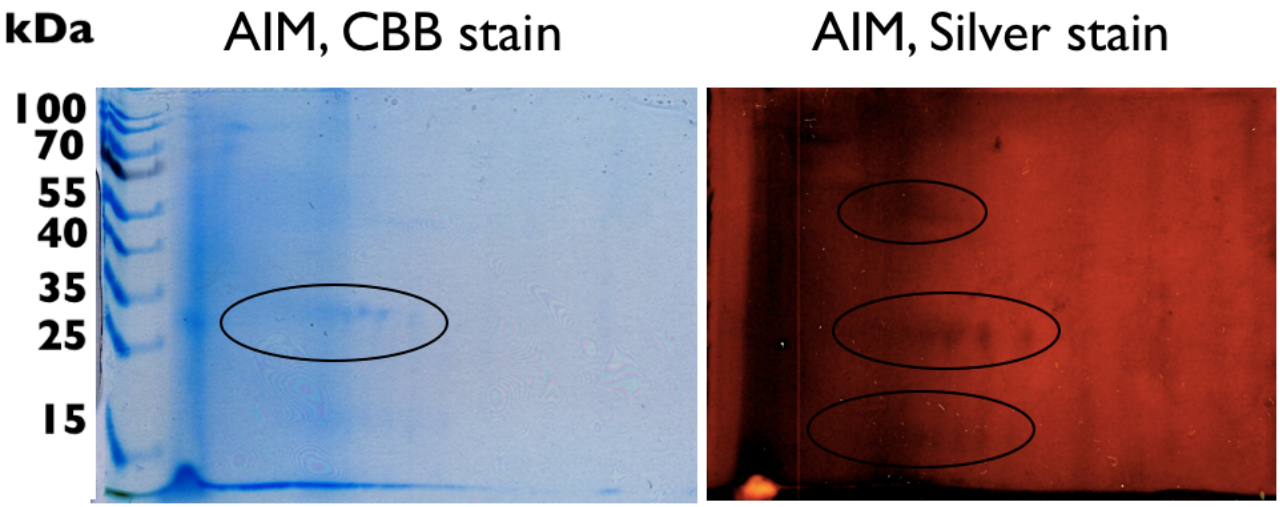


Figure 78. 2DE SDS-PAGE fractionation of acid-insoluble (AIM) shell matrix proteins from *Paracentrotus lividus* adult tests. Isoelectric focusing (IEF) was followed by gel electrophoresis under denaturing condition. The mass of molecular weight markers (in kDa) is indicated on the left. Gels were stained with CBB (left) or silver stain (right). 150 µg of protein material were applied per well. Protein groups identifies are marked.

ASM protein Matrix Analysis by ELISA

Two antibodies cross-reacted with the ASM when tested by ELISA. Anti-SM30 and anti-SM50 from *S. purpuratus* showed a strong reactivity, suggesting the presence of each protein in the extract. These results complied with the proteomic analysis (Fig.79).

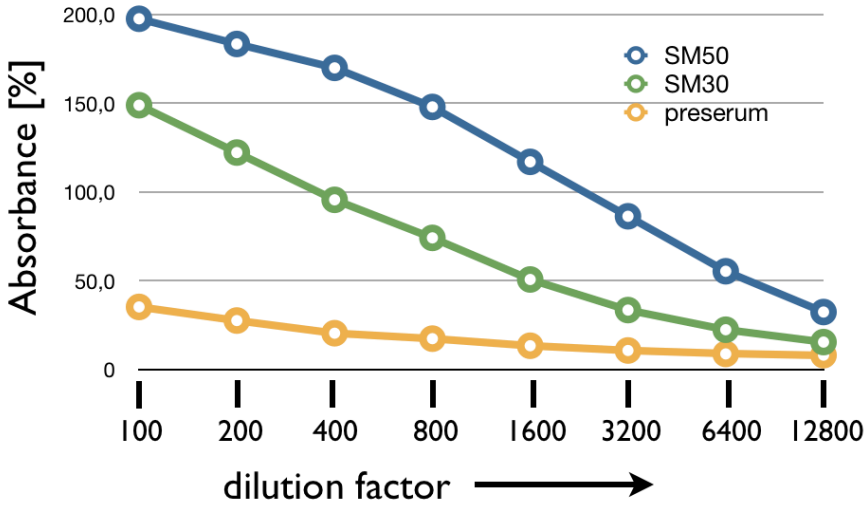


Figure 79. ELISA of adult test ASM of *P. lividus* (2 mg/ml) with serially diluted polyclonal antibodies to SM30 and SM50. As negative control, the pre-immune serum was tested.

Protein Matrix Analysis by Immunoblotting (Western Blotting)

The ASM and AIM fractions of the test matrix were analysed by immunoblotting. Polyclonal antibodies against recombinant proteins from embryonic carbonic anhydrase and galectin-8, cross-reacted with the extracted matrix. CA was predominately found at the AIM. but also in the soluble fraction (Fig.80). Galectin-8 was identified in both the ASM and AIM but most abundantly in the ASM (Fig.80).

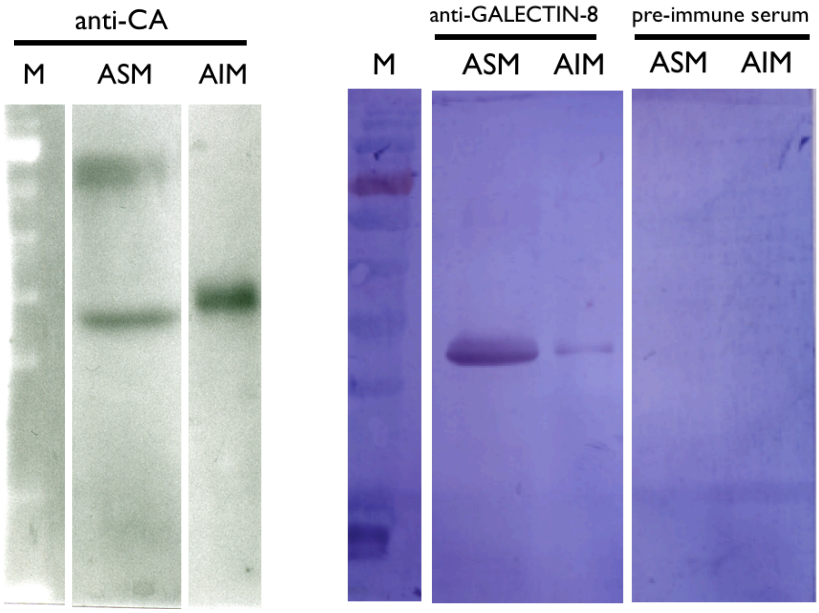


Figure 80. SDS-PAGE fractionation followed by Western Blot. Acid-soluble (ASM) and acid-insoluble (AIM) matrix proteins from *Paracentrotus lividus* adult tests cross-reacted with anti-CA antibody (left) and anti-Gal8 (right). Molecular weight markers (in kDa) are indicated on the left. Anti-CA antibody was used at a dilution of 1:500. Film was developed for 4min. Anti-Gal8 antibody was used at a dilution of 1:1000. Film was developed for 90 sec. 20 μg of protein material were applied per well .

***In vitro* interaction of the ASM with calcium carbonate**

The acetic acid soluble fraction (ASM) test extract was assayed for its effect on the formation of calcium carbonate *in vitro* by the diffusion method. Increasing concentrations of the whole ASM extract were tested, from 5 μg/ml to 100 μg/ml. In control blank experiments no protein was added resulting in production of rhombohedrons of calcite. ASM modified the crystals in a dose-dependent manner. Polycrystalline aggregates of various sizes were observed in presence of ASM at concentrations from 5 μg/ml το 25 μg/ml. At 50 μg/ml, large polycrystalline aggregates and cylinder-like crystals were observed. At concentrations as high as 100 μg/ml, only cylinder-like crystals were formed. No inhibition of the crystal formation, at any concentration, was recorded (Fig.81).

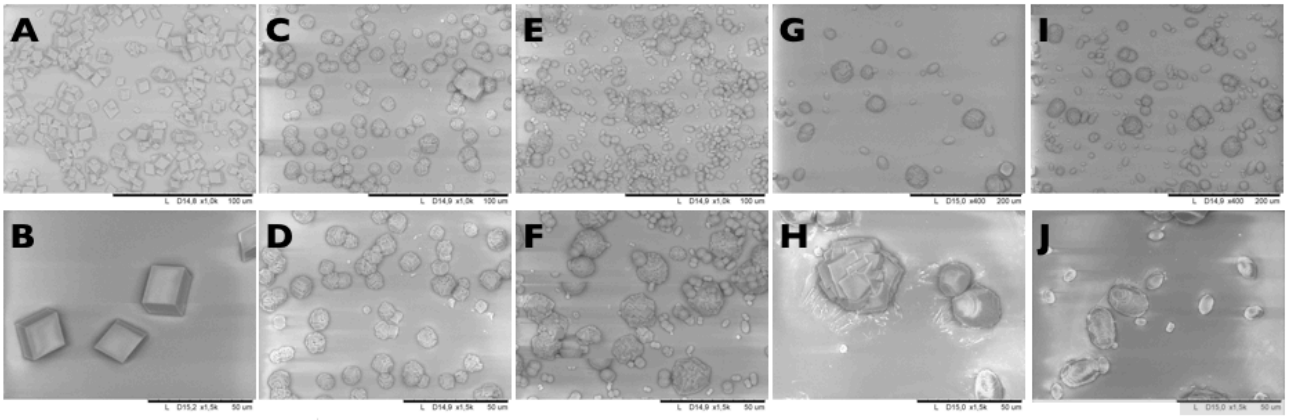


Figure 81. *In vitro* crystallization of calcium carbonate in the presence of ASM. A,B) Blank experiment (no proteins). C,D) 5 μg/ml ASM proteins; E,F) 25 μg/ml ASM proteins; G,H) 50 μg/ml ASM proteins; I,J) ASM proteins 100 μg/ml. (Karakostis *et al.*, in preparation).

In vitro interaction of recombinant proteins with calcium carbonate

Purified recombinant carbonic anhydrase (rCA) and galectin-8 (rGal-8) from the *P. lividus* embryo were assayed for their effect on the precipitation of calcium carbonate *in vitro* by the method of diffusion. Increasing protein concentrations were tested, from 5 μg/ml to 750 μg/ml, modifying the crystals in a dose-dependent manner. In control blank experiments no protein was added resulting in production of rhombohedrons of calcite.

In presense of rCA variations in the shape of the crystals were observed. 150 μg/ml of rCA induced precipitation of predominantly amorphous rounded calcium carbonate, while at higher concentrations (500 μg/ml), rounded calcite crystals were formed. At 750 μg/ml, cylinder-like crystals of a length diameter of 10 μm, were observed. This effect was similar to the effect obtained using 100 μg/ml ASM proteins. No inhibition or significant effect on the size of the crystals, nor aggregates were observed in presence of rCA (Fig.82).

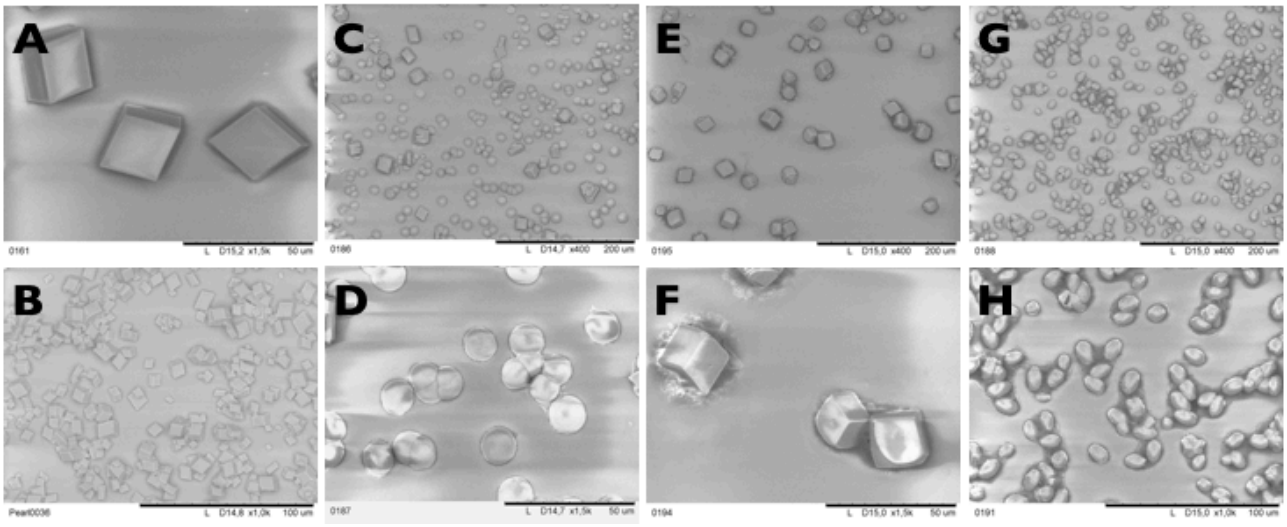


Figure 82. *In vitro* crystallization of calcium carbonate in the presence of recombinant CA. A,B) Blank experiment (no protein). C,D) [protein]= 15 µg/100µl; E,F) [protein]= 50 µg/100µl; G,H) [protein]= 75 µg/100µl. The scale bar of the images A, C, E, G is 200 µm. The scale bar of the images B, D, F, H is 50 µm.

The effect rGal-8 on calcium carbonate precipitation was dose-dependent. Large polycrystalline aggregates were formed in presence of 5 µg/ml. At 10 µg/ml, a remarkable modification effect was recorded. Most of the produced minerals were small amorphous precipitates while few rounded polycrystalline aggregates, were also observed. Interestingly, only small polycrystalline aggregates were not formed at higher concentrations of rGal-8 (750 µg/ml). No inhibition effect was observed in presence of rGal-8 (Fig.83).

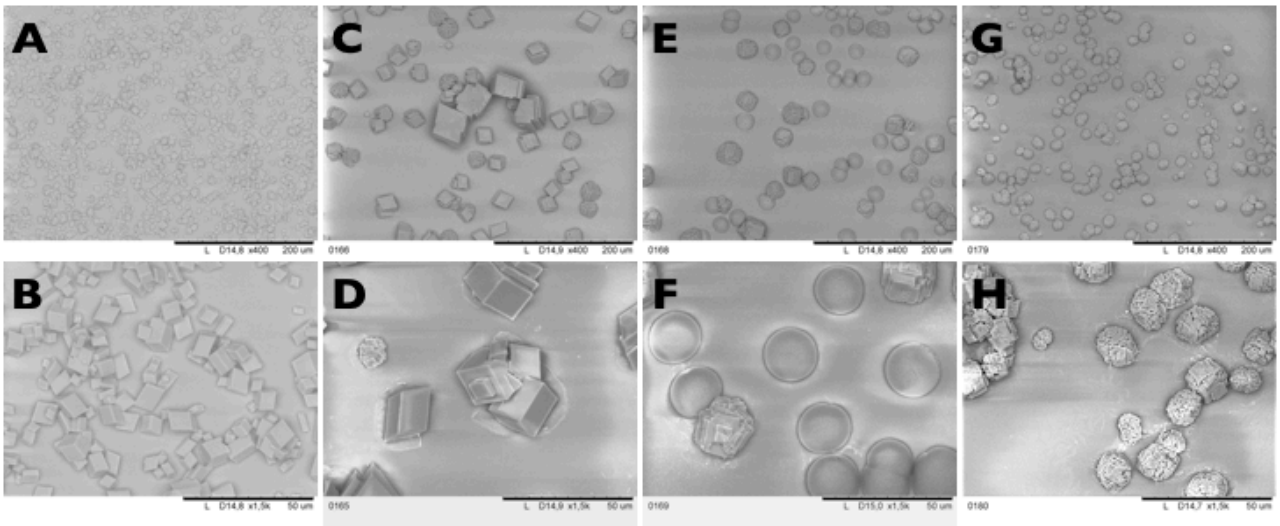


Figure 83. *I In vitro* crystallization of calcium carbonate in the presence of recombinant Galectin-8 (*rPl-GAL8*). A,B) Blank experiment (no protein). C,D) 0.5 µg/100µl rgal; E,F) 10 µg/100µl *rPl-GAL8*; G,H) 75 µg/100µl *rPl-GAL8*. The scale bar of the images A, C, E, G is 200 µm. The scale bar of the images B, D, F, H is 50 µm.

Characterization of the inorganic compounds of the AIM, by FTIR

The acetic acid-insoluble matrix was expected to contain insoluble proteins, carbohydrates, glyco-conjugated proteins, metals and traces of calcium carbonate. For the identification of the inorganic components as well as for the qualitative potential interactions occurring among proteins and carbohydrates, FTIR spectroscopy was used. There was no appreciable change in the positions of the characteristic bands of the drug along with the IR spectrum during the investigation. Since there is no change in the nature and position of the peaks in the formulation, it can be concluded that the samples were stable. The compounds of the AIM were identified by FTIR (Fig.84). The characteristic peaks of amines and carboxylates, corresponding to proteins as well as peaks indicating the presence of carbohydrates, were observed. The presence of carbohydrates is evidenced by the C-H vibrational stretch bands at 2924 cm^{-1} (asymmetric) and at 2852 cm^{-1}. The IR spectrum also shows the corresponding C-H bend bands in the 1100 cm^{-1} until 1250 cm^{-1} region. The broad band at 3200 cm^{-1} is representative of the O-H group from the alcohol groups of sugars. Protein amide group is also identified from the C=O group and N-H bands at 1635 cm^{-1} and 1536 cm^{-1}, respectively. No evidence of aldehyde group was found at 1700 cm^{-1}, but the band at 2800 cm^{-1} could represent interacting aldehyde.

It is noted that the second batch, contained a single extra characteristic peak at 875cm^{-1}, corresponding to undissolved $CaCO_3$, suggesting that the dissolving process in the second batch was not complete.

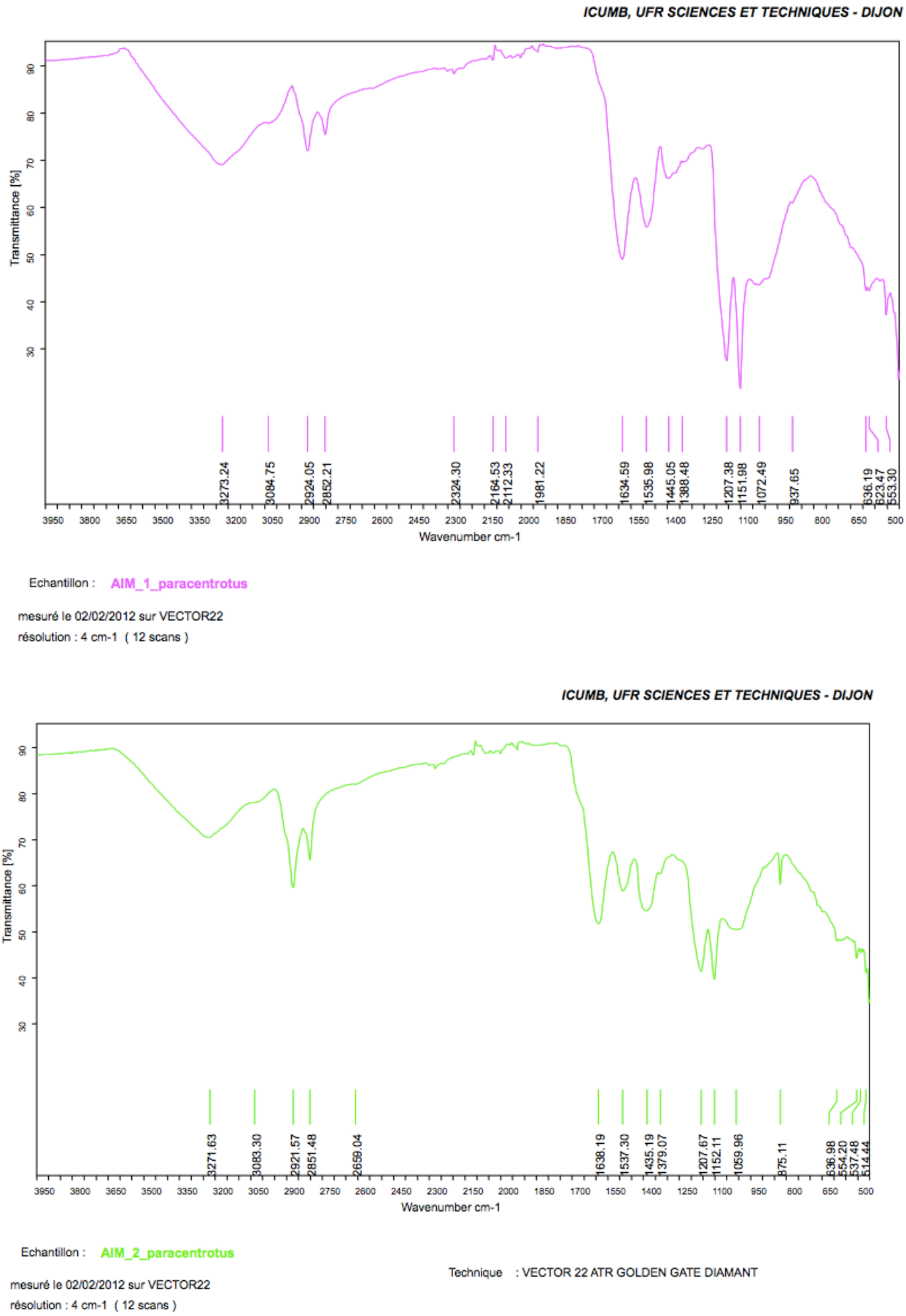


Figure 84. FTIR spectra of the AIM (Batches 1 and 2)

Chapter VII: Discussion

Skeletogenesis in the sea urchin embryo is a biological process which involves the synthesis and secretion of spicule matrix proteins from the PMCs. Proper spicule morphogenesis requires the interaction of the PMCs with the overlying ectoderm, via signaling cues and growth factors, such as the VEGF, FGF, univin and probably many other actors not yet discovered (Zito *et al.*, 2003; Duloquin *et al.*, 2007; Rottinger *et al.*, 2008). Recently the branching effect of VEGF and the aggregation effect of FGF have been exploited by *in vitro* crystallization assays (Knapp *et al.*, 2012). It is apparent that the deposition of the mineral is a regulated event. Gene regulatory networks (GRNs) guide spiculogenesis as part of the development of the embryo. Several types of proteins serve in a variety of roles for the skeletogenesis: (i) Transcription factors of the GRN, growth factors and signaling biochemical pathways regulate gene expression. (ii) Proteins directing the metabolic activity (ie: protein channels), provide an over-saturated ionic micro-environment, facilitating the precipitation of the biomineral. (iii) Proteins with specific physico-chemical characteristics, having a strong direct morphogenic influence on the mineral/crystal structure, serve as bio-templates and stabilize the mineral forms by protein-mineral interactions. These patterning-promoting proteins are controlling biomineralization which requires the presence of acidic proteins and lectins. **Acidic proteins** have long been known to play a major role in carbonate biomineralization by inducing the exclusive precipitation of a unique $CaCO_3$ polymorph, as in the case of the rich in aspartic acid residues, Aspein from mollusk (Takeuchi *et al.*, 2008). Acidic proteins are effective crystal-modulators, implicated in directing the formation of crystal polymorphs and morphologies in biogenic minerals (or biomineralized material) (Addadi *et al.*, 1989; Fu *et al.*, 2005). Indeed, the biomineralization of calcified hard tissues appears to be dependent on acidic protein residues (i.e. aspartic acid, glutamic acid, and phosphorylated serine) more than basic residues (i.e. arginine and lysine) (Addadi and Weiner, 1985; Chu *et al.*, 2011). Yet the important effect of basic residues has also been shown (Gong *et al.*, 2012; Masica *et al.*, 2010). Furthermore, the high content in magnesium of the bio-calcitic sea urchin structures (Matranga *et al.*, 2011) has been correlated with the presence of acidic proteins rich in aspartic acid, such as UTMP16, a protein found in *L. variegatus* embryo and adult tooth (Alvares *et al.,* 2009). Lastly, the fact that proteins with limited sequence similarity but comparable functional features are present in the sea urchin and in other distant phyla, including mammals, suggests that these small acidic proteins constitute an example of evolutionary convergence (Costa, Karakostis *et al.*, 2012). Moreover, Echinoderms have a large set of **lectins** which are glycoproteins able to bind mono- and disaccharides (L. Courtney Smith *et al.*, 2010), serving as biomineralizing scaffolds. A comprehensive study of skeletogenesis and biomineralization in the sea urchin, involves the characterization of all the above biological processes and protein types.

In the present study, we focused on the identification and characterization of biomineralization genes from *P. lividus* (Fig.85) with the purpose to initiate functional studies on proteins guiding the formation of the bio-calcite in *P. lividus*. This task required

the selection of genes believed to be involved in biomineralization of other species or biological systems, including five proteins proposed to be involved in protein-mineral interactions. These included the acidic proteins p16 and p19. The high amount of acidic amino acid residues, such as aspartic acid in Pl-P16 and glutamic acid in Pl-P19, as well as their expected phosphorylation which confers a total acidic charge, are common features in a wide variety of proteins involved in biomineralization processes (Alvares *et al.*, 2009). Additionally, SM30a and SM50 are two C-type lectins playing different roles in biomineralization as their biochemical properties (pI) and expression profiles are different (Kitajima *et al.*, 1996). Finally, the lectin advillin and the membrane protein tetraspanin, known to have a role in signaling or be involved in protein-protein interactions, were selected in this study.

Homologous sequences were identified by EST data mining or in some cases by molecular techniques, sequenced and deposited in the EMBL bank. The cloning of the CDS allowed the *in silico* sequence and functional analysis and the characterization of each gene at the transcription level unraveling their role in biomineralization of the *P. lividus* embryo. Variations in the sequences (as shown by phylogenetic analysis) and the expression profiles were observed when compared with orthologues from the literature. The small amount of the sequence variations as well as the high similarity of the functional domains found when compared the CDSs from *P. lividus* with orthologues from other sea urchin species, indicated that p16, p19, advillin, tetraspanin, sm30 and sm50 are well conserved in the sea urchin. The conservation of the functional domains of each one of the biomineralization proteins analysed here by the phylogenetic analyses, suggests that Biomineralization employs similar strategies throughout different species and that the biomineralization proteins involve well-conserved functional domains. Furthermore, a biochemical redundancy is noticed in groups of proteins such as the presence of a lectin domain in SM30 and SM50. Even though the primary sequences are well-conserved, temporal variations were observed in the gene expression. RT-PCR and RNA gel blot analysis showed different temporal expression patterns when compared with orthologues from *H. tuberculata* and *H. erythrogramma* (Love *et al.*, 2007) and *S. purpuratus* (Cheers and Ettensohn, 2005). A similar spatial expression pattern of *Pl-advillin, Pl-p16* and *Pl-p19,* involving only small variations was identified during skeletogenesis. Transcripts in *H. tuberculata* pluteus were uniformly localized in all the PMCs, while in *P. lividus* there was an increased expression in the anterolateral rods and the scheitel. Additionally, cells at the ventral transverse rod are also labelled. On the contrary, the expression profile of *tetraspanin* shows differences among the sea urchin species. While in *H. tuberculata* it is expressed from early stages and it is localized at the ectoderm of the pluteus arm-tip, in *P. lividus*, *tetraspanin* is initially expressed in all the ectodermal cells of the prism and pluteus with no specific sub-territorial localization. This could be attributed to variations in the tetrapsanin isoforms or to functional diversity. Altogether, these differences suggest the need of additional functional studies for each gene. The CDSs were cloned in expression vectors for the preparation of a molecular tool-set including recombinant proteins expressed in E. *coli* and specific antibodies aiming to the functional characterization of the

proteins individually and in sets, in *in vitro* assays which reveal the activity of each protein but also monitor its biological role in development and in skeletogenesis.

During embryo development, *Pl-p16, Pl-p19, Pl-advillin* and *Pl-tetraspanin* mRNA, described here for the first time in *P. lividus*, are regulated differently since they show dissimilar patterns of expression. Indeed, while *Pl-p19* and *Pl-advillin* mRNA appear to be expressed in almost all PMCs from the mesenchyme blastula to late pluteus, *Pl-p16* mRNA expression becomes limited to a subset of PMCs close to the regions of active skeletal rod growth. The sites of high expression of *Pl-p16* mRNA include the tips of postoral and anterolateral rods and the scheitel at the pluteus stage, while *Pl-advillin* is mainly expressed in a subset of the PMCs at prism stage, which form the postoral and longitudinal chain. On the other hand, *Pl-tetraspanin*, is localized in the ectoderm. Such diversities in the spatial expression patterns suggest a range of biosynthetic capacities of each individual PMC in different regions of the embryo, but also different roles for each protein in skeletogenesis. It is noteworthy to mention the unusual distribution of *Pl-p19* mRNA within the PMCs syncytium of *P. lividus* embryos (See Results section of Chapter III). This localization pattern includes both the cell bodies and the filopodial cytoplasm close to the ventrolateral clusters at the late gastrula stage and to the scheitel at the late pluteus stage. Such a distribution of mRNA is observed for the first time (*i.e.* the one in the filopodial cytoplasm), suggesting an accumulation of mRNA molecules where there is major request of specific protein synthesis for spicule development. It is well known that PMCs fuse together during gastrulation, forming a syncytial population of cell bodies connected by thin cytoplasmic cables. The synthesis of spicules takes place in the confined space formed from such cytoplasmic cables, thereby the mineralized spicules remain covered by the cytoplasmic sheath of the fused PMCs. They are enveloped by the PMCs syncytium with plasma membranes in close contact to the extracellular spicule. In the presence of fixatives, such as WMISH procedure requires, the spicule is de-mineralized and a fibrous organic structure with the overall shape of the spicule can be visualized. The detected diversities in the spatial expression patterns, while suggesting a range of biosynthetic capacities of each individual PMC in different regions of the embryo, might also imply different roles for each protein in skeletogenesis. The *Pl-p16* spatial expression during skeletogenesis described in this study is similar to that found in two other sea urchin species (Illies *et al.*, 2002; Cheers and Ettensohn, 2005). In fact, the fundamental role of P16 in skeletal rods elongation has been demonstrated in *S. purpuratus* and *L. variegatus* by means of knockdown expression experiments. A few different functions have been proposed for P16: i) putative receptor of signals required for skeletogenesis; ii) regulator of calcium and magnesium transport to PMCs; iii) direct role in biomineral deposition, as suggested by its acidic nature (Cheers and Ettensohn, 2005). P16 has been hypothesized to establish the high magnesium columns that fuse the calcite plates together enhancing the mechanical strength of the mineral (Alvares *et al.*, 2009). Further functional studies are needed to characterize the P16 role in biomineralization. Less studied is the role of P19 in skeletogenesis as no functional assays have been described so far. Since the p19 coding region lacks a signal sequence and transmembrane domains, it is supposed that the protein remains localized in the PMCs

cytoplasm. However, taken into account the potential interaction of the acidic residues of p19 with calcium cations, the observed localization of the *Pl-p19* mRNA in the cell bodies and filopodial cytoplasm of all PMCs might suggest a role for Pl-P19 protein in the regulation of the circumference of the spicules, referring to the linear growth distance around the edge of the closed curve of the cylindrical-shaped spicule. This view is in agreement with the proposed hypothesis that elongation of spicules depends on PMCs located at their tips, while the increase in the width (circumference) depends on PMCs lying along their length (Wilt, 2005). Interestingly, it has been demonstrated that the knockout of the *dmp*1 gene produced mice with defects in tooth mineralization (Qin *et al* 2007). Furthermore, it was demonstrated that DMP1 post-translational modifications, including its phosphorylation/dephosphorylation, proteolytic processing, and glycosylation, control the rate of dentin biomineral nucleation and the extent of growth (Gericke *et al.*, 2010). Thus, the properties of the C-terminus part of DMP1 are reminiscent of UTMP19 (corresponding to Sp-P19 short form), a component of the sea urchin tooth found in *L. variegatus* (Alvares *et al.*, 2009).

The C-type lectins SM30 and SM50 are well-documented in other species. In this study, we determined for the first time in *Paracentrotus lividus*, the temporal expression profile of *Pl-SM30a* and *Pl-SM50* during embryo development, by comparative qPCR. Additionally, *in silico* characterization of each primary putative protein sequence, revealed for the first time new characteristics of the non-C-type lectin domains of each protein. Both *Pl-SM30a* and *Pl-SM50* contain abundant amount of weakly hydrophobic amino acids glycine and proline.

Pl-SM30a involves an N-terminal Gly-rich region and a high number of proline residues. The repetitive nature of the glycine-rich domains is likely to allow the formation of a b-sheet structure. The GRP often follows the motif $(Gly\text{-}X)_n$ where X is often glycine or Ala, Ser, Val, His, Phe, Tyr and Glu, without interfering with the structure. Glycine repeats tightly bound to polysaccharides/carbohydrates, facilitate cell adhesion between neighboring cells in various biological systems, promoting tissue formation. As an example, plant cell walls which are composed mainly of polysaccharides and glycine-rich and proline-rich proteins (GRPs and PRPs), promote lignification which is essential for the structural integrity of plant cell walls (Ringli *et al.*, 2001). Gly-rich peptides are known to promote protein-protein interactions. Protein cross-linking determines the hardness and stiffness of several hard tissues of marine invertebrates. As an example, the tissue of beak of the jumbo squid (Dosidicus gigas) mainly consists of chitin fibers and glycine-rich proteins, offering increased hardness, stiffness and plasticity properties (Miserez *et al.*, 2007). *Pl-SM30a*, comprising a poly-glycine region and a calcium-dependent carbohydrate-binding lectin domain may act as a cross-linking agent promoting the deposition of the calcitic biomineral.

Similarly, *Pl-SM50*, the most abundant occluded matrix protein in the sea urchin spicules, is a C-type lectin with an extensive proline-rich region. Indeed, the putative amino-acidic sequence of *Pl-SM50*, involves several proline residues and proline repeats of PGM. Proline residues, due to their bulkiness of the side chain, produce a structure-breaker effect (a-helix and b-sheet-breaker), promoting an extended protein structure. Additionally,

Prolines serve as binding agents in protein-protein interactions. Due to their hydrophobic surface, they bind well to flat hydrophobic surfaces such as aromatic rings. Proline residues are more poorly solvated than other aa, fewer water molecules are needed around the amide bond, which results in stronger interactions with other solutes. Proline-rich peptides are, as a consequence, highly soluble in water. Thus, proline is a favored ligand entropically and enthalpically. PGM repeats are found also in other biological systems, as for example peptides which bind on formins, a group of proteins involved in actin polymerization, as well as in heavily glycosylated proteins, such as mucin which is documented to create an extensive network of membrane-anchored extended chains coating and lubricating the epithelial layer. It should be mentioned that collagen from connective tissues is rich in proline repeats. The pro-rich region of *Pl-SM50*, may interact with other proteins, influencing the solubility. This function is coupled or promoted by the calcium-dependent carbohydrate-binding activity of the C-type lectin domain.In a biological process similar to the function of mucin mentioned above, SM50 may create a network serving as functional template for the deposition of the biomineral. Indeed, Gong *et al.*, 2012 showed that SM50 stabilizes a hydrated form of amorphous calcium carbonate ($ACC{\cdot}H_2O$) *in vitro*, acting as an inhibitor of phase transition.This type of regulation of the mineral phase transition by organic molecules is shared by numerous $CaCO_3$-mineralizing organisms. In conclusion , *Pl-SM30a* and *Pl-SM50*, may play a very important role in the deposition of the mineral, functioning as cross-linking agents with other proteins, forming a mechanistic network of glycoproteins, carbohydrates and calcium.

Apart from the proteins which are involved in protein-mineral interactions and thus have a direct effect on the deposition of the biomineral (P16, P19, SM30a and SM50), other proteins regulate skeletogenesis *in vivo*, facilitating signaling. The role of advillin and tetraspanin in skeletogenesis, is not well documented. Pl-*advillin* is a membrane bound extracellular protein, involving six GELS (Gelsolin domains) and one headpiece domain. The protein includes calcium-binding and actin binding residues (Bazari *et al.*, 1988). Indeed, the six-repeats are Ca^{2+}-dependent, while the headpiece domain is responsible for cross-linking and bundling and is Ca^{2+}-independent (Friederich *et al.*, 1999). In biomineralization, advillin is believed to be involved in filopodial retraction of PMCs, via actin severing and capping (Love *et al.*, 2007). In this work, the high sequence similarity with orthologous sequences in other sea urchin species and the tempo-spatial mRNA expression profile has been presented. Advillin is a well conserved gene, indicating its significance in the development of the sea urchin embryo. Additionally, a sequence (Reference number: Glean3_01874), showing 72.9% similarity with the embryonic advillin, was identified in the protein matrix of the tooth of adult *S. purpuratus*, indicating the importance of this protein in adult biomineralization. Finally, *Pl-tetraspanin*, as a member of tetraspanins involving a distinct family of proteins with characteristic structural features, contains four transmembrane domains: a small outer loop, a larger outer loop, a small inner loop and short cytoplasmic tails. Tetraspanins act as linker proteins; they associate and organize other proteins and lipids (*i.e.* gangliosides and cholesterol) into a network with varying features (ie. resistance to solubility in non-ionic detergents). The role of the transmembrane domains in biomineralization has been demonstrated in corals where

proteins with transmembrane domains may have their extracellular part enzymatically cleaved and subsequently, become incorporated into the skeletal matrix (Ramos-Silva *et al.*, 2013).

The tetraspanin web can associate with other membrane proteins such as integrins as well as with signaling enzymes (protein kinase C), forming a mechanistic framework which provides lateral signaling propagation (Stipp *et al.*, 2003). Thus, *Pl-tetraspanin* may be involved in skeletogenesis as a protein linker which promotes the regulation of proteins affecting the solubility and over-saturation of ions, within the micro-environment. Love *et al.*, 2007 showed that tetraspanin is expressed in proximity with advillin and carbonic anhydrase in the pluteus arms of the embryo. Even though there are apparent variations both in the expression levels of these genes and in localization, among the sea urchin species, these expression patterns indicate a combination of gene activity in the morphogenesis of larval arms which is in accordance with changes in ectoderm and PMC differentiation programs. Although in early stages the gene expression patterns of *Pl-CA, Pl-advillin* and *Pl-tetraspanin* are similar to their orthologues from *H. erythrogramma* and *H. tuberculata*, a clear divergence of regulation is apparent after the gastrulation, resulting in differences in the spatial expression in pluteus. Indeed, the localization pattern of the Pl-CA mRNA showed variations throughout development, when compared with patterns of orthologues from others sea urchin species. Pl-CA is expressed exclusively in the PMCs of the ventrolateral clusters of gastrula embryos, the site of initial calcium carbonate deposition, and at a subset of the PMCs located mainly in the body rod of prism embryos; while in pluteus CA is expressed in all the PMCs. The localization of Pl-CA in pluteus comes as a variation when compared to expression patterns of CA homologues previously described in *H.tuberculata* (Love *et al.*, 2007) where CA mRNA was found in a subset of PMCs localized at the leading edge of both sets of growing arms and in a ring of PMCs forming skeleton at the posterior apex. In *H.erythrogramma,* carbonic anhydrase was indicated to have a broader expression pattern that includes ectoderm, in addition to the skeletal mesenchyme. Furthermore, the temporal expression pattern and the overall amount of the mRNA copy number of CA, also shows variations among different sea urchin species. As monitored by RT-PCR, CA from *H.tuberculata* was previously documented to be initially expressed in blastula and become increased in gastrula and pluteus, while in *H.erythrogramma,* expression was initially observed at low levels in gastrula and gradually increased until the pluteus (Love *et al.*, 2007). Here we show that Pl-CA is first detectable at the mesenchyme blastula stage, maintaining similar levels throughout prism and then rapidly increases to reach a maximum level of expression at the pluteus stage. Even our results indicate that Pl-CA is a PMC-specific gene, similarly to other sea urchin species, it becomes obvious that there is an apparent variation of the tempo-spatial expression profile of CA orthologues, implying divergent evolutionary routes and regulation by upstream genes. Similar variations were observed for tetraspanin and advillin. These data provide with valuable information about the evolutionary novelties among the sea urchin species. It is worth emphasizing that these genes, notably carbonic anhydrase, are required for the normal development of the embryo. Carbonic anhydrase is an essential metalloenzyme that regulates many physiological functions. Its activity is

based on the catalysis of the reversible hydration of CO_2, according the reaction: H_2O + $CO_2 \leftrightarrow HCO_3^- + H^+$, facilitating the interaction of the produced bicarbonate anion (HCO_3^-) and the deriving carbonate (CO_3^{2-}) anions, with Ca^{2+} cations, enhancing the $CaCO_3$ precipitation by supersaturation of these ions at the mineralization site. The precipitation of amorphous calcium carbonate (ACC) within PMCs is intrinsically linked to pH regulation (Stumpp *et al.*, 2012). The calcification process is fueled by bicarbonate (HCO_3^-) supply, which is partially absorbed from the seawater (40%) and partially generated from metabolic CO_2 (60%) (Stumpp *et al.*, 2012). Embryos treated with acetazolamide, a specific inhibitor of CA, resulted in the inhibition of skeleton formation, in a dose-dependent manner. CA was identified by ELISA, Western Blot and by the proteomic analysis of the protein matrix extracted from the adult mineralized tests, indicating the importance of this protein in biomineralization of the adult sea urchin. Indeed the the mouse polyclonal antibody we developed against the recombinant Pl-CA from embryo, cross-reacted against protein lysates from embryos and against the occluded protein matrix of adult tests. This result suggests that Pl-CA is expressed in embryo and in adult but it does not exclude the possibility that other CA isoforms demonstrating high similarity are expressed in adult. It should be noted that the mass of the identified proteins from embryo and adult corresponds to the predicted full-length Pl-CA from embryo. Even though CA from sea urchin was identified both in the embryo and in adult, the assumption that it is enzymatically active was previously based on *in silico* analysis without supportive experimental data. Here, for the first time, we produced a recombinant version of the CA domain of Pl-CA and demonstrated its esterase activity. The esterase activity assay is a simple assay which is commonly used to measure the esterase activity of CAs. It is noted that certain CAs belonging to the beta family do not exhibit esterase activity but they are active in the anhydrase assay. In the case of Pl-CA, our results demonstrate that it exhibits an esterase activity *in vitro.* The activity was found lower compared to bovine CA but comparable to the activity of CA from pearls of the freshwater mussel *Hyriopsis cumingii* (Natoli *et al.*, 2010). It should be mentioned that the recombinant protein prepared in bacteria and used in the activity assay, lacks of all the putative post-translational modifications which does not exclude the possibility that the natural Pl-CA has a higher enzymatic activity. Additionally to the esterase activity, a remarkable effect of recombinant Pl-CA on the morphology of calcite crystals was found which was qualitatively similar to the effect of the acetic-acid soluble protein extract, implying a central role of Pl-CA as a crystal modulator (Karakostis *et al.*, in preparation). It is perhaps important to highlight that Pl-CA bears an N-terminal Glycine-rich domain, known to be involved in protein-protein interactions, which is not present in CAs from other species, as shown by the phylogenetic analysis, but it is conserved in both *P. lividus* and *S. purpuratus*. Indeed, even though Pl-CA exhibits high similarity with the human cytosolic CAII, it involves the additional N' terminal gly-rich region (GPQVGGHPGRGFNWGGHGGNGNGAGGGGGGGGGGAGAGGGGNGWAGWG SWWGGNGWG) with an upstream signal sequence and a cleavage site, indicating that Pl-CA is a secreted protein. Phylogenetic analysis of the full-length protein sequence, indicated that the glycine-rich domain is only found in the sea urchin species.

Unfortunately, the currently available sequence data bases involve the full-length homologues only from *S. purpuratus.* The partial ESTs from other sea urchin species do not give information on the conservation of the gly-rich region in all the sea urchin orthologues. This gly-rich domain could be considered as a supernumerary domain, enriched in glycine residues. In biomineralization studies, since the discovery of the nacrein in mollusk nacre, there is a growing body of evidence that CAs with supernumerary domains are important molecular actors which probably exhibit more than one enzymatic function. Additionally, it is perhaps worth-noting that the gly-rich region exhibits similarity with proteins of the fibrillarin family which are related to the autoimmune disease scleroderma. gly-rich regions are known to be involved in protein-protein interactions for diverse biological and biochemical processes (Mangeon, *et al.*, 2010) and in interactions with RNA (Rogelj, *et al.*, 2012). Furthermore, they are known to play a role in membrane targeting, regulating different functional and activation mechanisms (Salem *et al.*, 2004). Thus, a dual function of *Pl-CA*, both as an enzyme, catalyzing the hydration of CO_2 and as a regulator or co-enzyme of other proteins is hypothesized. Apart from Pl-CA, also SM30 and SM50 contain respectively glycine-rich and proline-rich regions respectively which potentially facilitate protein-protein interactions. It would be of great interest to identify the protein-protein interactions with Pl-CA and analyze the biological role of these interactions in the development of the embryo. Furthermore, protein-carbohydrate interactions, may also be crucial for the effect of the biomineralization proteins in shaping the crystal faces. In this manner, biomineralization proteins act as cross-linking agents, forming a mechanistic network of glycoproteins, carbohydrates and calcium. Various carbohydrates, together with insoluble and aggregated proteins, including Pl-CA, were indeed identified by FTIR spectroscopy and Western Blot, from the acetic acid-insoluble matrix extracted from *P. lividus* adult tests. It becomes clear that a biological matrix, involving acidic proteins, glycoproteins, enzymes and carbohydrates is required for the deposition of the skeleton in the sea urchin. The results of this work, not only confirm the presence of various embryonic lectin proteins in the occluded matrix protein of the adult tests (SM32, SM29, PM27) but also of depolymerizing proteins such as the MMP17 and lithostathine. Furthermore, the identification of Pl-CA and MSP130-related-2, indicates a precipitation of bicarbonate coupled with calcium precipitation during the mineralization of the adult test. These findings demonstrate the similar players employed in the sea urchin embryo and in the adult mineralization and paves the way for ongoing biomineralization studies in *P. lividus.*

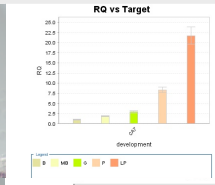
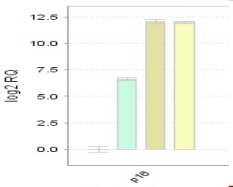
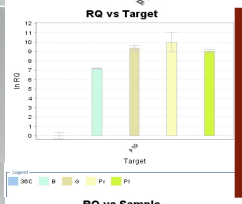
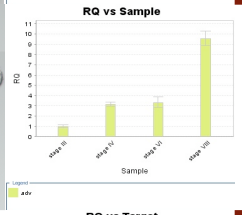
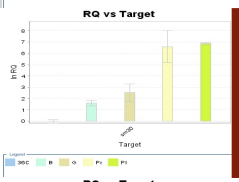
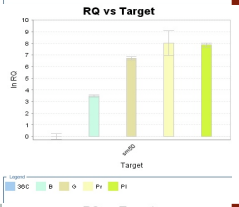
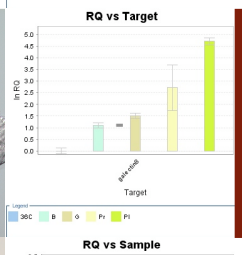
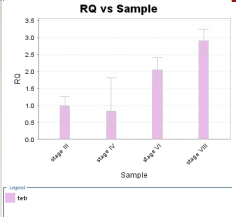
	Pl cDNA	Accession No	Mass (kDa)	p*I*	mRNA embryo localization (WMISH)	qPCR	info
1	*ca*		48	6.8	PMCs		extracellular carbonic anhydrase
2	*p16*	FR693763	17	3,7	PMCs		acidic membrane protein
3	*p19*	FR693764	19	4,8	PMCs		acidic cytoplasmic protein
4	*advillin*	FR693766	93	8,6	PMCs		lectin
5	*sm30*	FR716470	34	6,0	PMCs		C-type lectin
6	*sm50*		31	9,1	PMCs		C-type lectin
7	*galectin-8*	FR716469	35	9,4	foregut		lectin
8	*tetraspanin*	FR693765	30	3,9	Ectoderm		membrane protein

Figure 85: Summary of the identification and characterization of transcripts. From left to right column: mRNA name, Accession number, pI, tempo-spatial expression profile in *P. lividus* embryo and characteristics of the putative protein.

Peer-reviewed articles

The results of the work described in this thesis, involved novel and useful information for the research community; therefore a part of the results was already published in peer-reviewed journals, at the time of the submission of this thesis. Another part has been prepared to be submitted, shortly. The following list, outlines these articles.

1. Echinoderms as blueprints for biocalcification: regulation of skeletogenic genes and matrices, **2011**, Prog Mol Subcell Biol. 52:225-48, V. Matranga, R. Bonaventura, C. Costa, K. Karakostis, A. Pinsino, R. Russo, F. Zito

2. Phylogenetic analysis and expression patterns of *p16* and *p19* in *Paracentrotus lividus* embryos, **2012**, Dev Genes Evol. C. Costa & K. Karakostis, Francesca Zito, V. Matranga (*C. Costa and K. Karakostis contributed equally to the work and should be considered co-first authors.)

3. Cellular and molecular bases of biomineralization in sea urchin embryos, **2013**, Cah. Biol. Mar., V. Matranga, A. Pinsino, R. Bonaventura, C. Costa, K. Karakostis, C. Martino, R. Russo, F. Zito

4. Galectin-8 from *Paracentrotus lividus*, expressed in the archenteron and secondary mesenchyme cells of the embryo and in adult, is a lactose-specific galectin involved in cell adhesion, ***in preparation***, K. Karakostis, C. Costa, F. Zito, V. Matranga.

5. Molecular characterization of a newly identified carbonic anhydrase from *Paracentrotus lividus*, involved in biomineralization, ***in preparation***, K. Karakostis, C. Costa, F. Zito, F. Brümmer and V. Matranga.

6. Proteomic analysis of biomineralization proteins of the organic matrix from the adult test of the sea urchin *Paracentrotus lividus*, ***in preparation***, K. Karakostis, N. Guichard, F. Brümmer, V. Matranga, F. Marin.

Bibliography

ADDADI L *et al*., 1989. Structural and stereochemical relations between acidic macromolecules of organic matrices and crystals. *Connective Tissue Research,* 21 (1-4), 1-4.

ADDADI L and WEINER S, 1985. Interactions between Acidic Proteins and Crystals: Stereochemical Requirements in Biomineralization. *Proceedings of the National Academy of Sciences of the United States of America,* 82 (12), 4110-4114.

ADOMAKO-ANKOMAH A and ETTENSOHN CA, 2011. P58-A and P58-B: Novel proteins that mediate skeletogenesis in the sea urchin embryo. *Developmental Biology,* 353 (1), 81-93.

AKASAKA K *et al*., 1994. Genomic organization of a gene encoding the spicule matrix protein SM30 in the sea urchin *Strongylocentrotus purpuratus*. *The Journal of Biological Chemistry,* 269 (32), 20592-8.

ALTSCHUL SF, *et al*., 1990. Basic Local Alignment Search Tool. *Journal of Molecular Biology,* 215 (3), 403-410.

ALVARES K *et al*., 2009. Echinoderm phosphorylated matrix proteins UTMP16 and UTMP19 have different functions in sea urchin tooth mineralization. *The Journal of Biological Chemistry,* 284 (38), 26149-26160.

AMANO M *et al*., 2003. The ST6Gal I sialyltransferase selectively modifies N-glycans on CD45 to negatively regulate galectin-1-induced CD45 clustering, phosphatase modulation, and T cell death. *The Journal of Biological Chemistry,* 278 (9), 7469-75.

ARMSTRONG JM *et al*., 1966. Purification and properties of human erythrocyte carbonic anhydrases. *The Journal of Biological Chemistry,* 241 (21), 5137-49.

BARONDES SH, *et al*., 1994. Galectins: a family of animal beta-galactoside-binding lectins. *Cell,* 76 (4), 597-8.

BARONDES SH *et al*., 1994. Galectins. Structure and function of a large family of animal lectins. *The Journal of Biological Chemistry,* 269 (33), 20807-10.

BARONDES SH, 1997. Galectins: A Personal Overview. *Trends in Glycoscience and Glycotechnology,* 9 (45), 1-7.

BAZARI WL *et al*., 1988. Villin Sequence and Peptide Map Identify Six Homologous Domains. *Proceedings of the National Academy of Sciences of the United States of America,* 85 (14), 4986-4990.

BLOM N, GAMMELTOFT S and BRUNAK S, 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology,* 294 (5), 1351-1362.

BUCHAN DWA *et al*., 2010. Protein annotation and modelling servers at University College London. *Nucleic Acids Research,* 38 (SUPPL. 2), W563-W568.

CHEERS MS and ETTENSOHN, CA, 2005. P16 is an essential regulator of skeletogenesis in the sea urchin embryo. *Developmental Biology,* 283 (2), 384-396.

CHEN C and LAWRENCE J, 1987. The role of carbonic anhydrase in facilitating the transport of CO_2 in the tooth of *Lytechinus variegatus* (Echinodermata: Echinoidea). *Physiology Comparative Biochemistry and Physiology Part A: Physiology,* 87 (2), 327-331.

CHU X *et al*., 2011. Unique Roles of Acidic Amino Acids in Phase Transformation of Calcium Phosphates. *The Journal of Physical Chemistry B,* 115 (5), 1151-1157.

COSTA C *et al*., 2010. Phylogenetic analysis and homology modelling of *Paracentrotus lividus* nectin. *Molecular Diversity,* 14 (4), 653-665.

DECKER GL and LENNARZ WJ, 1988. Skeletogenesis in the sea urchin embryo. *Development (Cambridge, England),* 103 (2), 231-47.

DODD RB and DRICKAMER K, 2001. Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. *Glycobiology,* 11 (5).

DRICKAMER K, 1988. Two distinct classes of carbohydrate-recognition domains in animal lectins. *The Journal of Biological Chemistry,* 263 (20), 9557-60.

DYRLØV BENDTSEN J *et al*., 2004. Improved Prediction of Signal Peptides: SignalP 3.0. *Journal of Molecular Biology,* 340 (4), 783-795.

ESWAR N *et al*., 2008. Protein structure modeling with MODELLER. *Methods in Molecular Biology (Clifton, N.J.),* 426, 145-59.

ETTENSOHN CA, 2009. Lessons from a gene regulatory network: echinoderm skeletogenesis provides insights into evolution, plasticity and morphogenesis. *Development (Cambridge, England),* 136 (1), 11-21.

FRIEDERICH E *et al*., 1999. Villin function in the organization of the actin cytoskeleton. Correlation of in vivo effects to its biochemical activities in vitro. *The Journal of Biological Chemistry,* 274 (38), 26751-60.

FU G *et al*., 2005. CaCO3 biomineralization: acidic 8-kDa proteins isolated from aragonitic abalone shell nacre can specifically modify calcite crystal morphology. *Biomacromolecules,* 6 (3).

GASTEIGER E *et al*., 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research,* 31 (13), 3784-3788.

GEORGE NC, KILLIAN, CE and WILT FH, 1991. Characterization and expression of a gene encoding a 30.6-kDa *Strongylocentrotus purpuratus* spicule matrix protein. *Developmental Biology,* 147 (2), 334-342.

GERICKE A *et al*., 2010. Different forms of DMP1 play distinct roles in mineralization. *Journal of Dental Research,* 89 (4), 355-359.

GHAZARIAN H, IDONI B and OPPENHEIMER SB, 2011. A glycobiology review: Carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochemica,* 113 (3), 236-247.

GOLOKHVAST KS, PAMIRSKIY IE and PANICHEV AM, 2011. Homology of Bacteria Proteins, Diatoms and Sponges Participating in Biomineralization, and Human Proteins, and other Animals . *Pacific Science Review,* 13 (1), 39-46.

GORELIK E, GALILI U and RAZ A, 2001. On the Role of Cell Surface Carbohydrates and their Binding Proteins (lectins) in Tumor Metastasis. *Cancer and Metastasis Reviews,* 20 (3/4), 245-277.

GUSS KA and ETTENSOHN CA, 1997. Skeletal morphogenesis in the sea urchin embryo: regulation of primary mesenchyme gene expression and skeletal rod growth by ectoderm-derived cues. *Development (Cambridge, England),* 124 (10), 1899-908.

HODOR PG and ETTENSOHN CA, 1998. The Dynamics and Regulation of Mesenchymal Cell Fusion in the Sea Urchin Embryo. *Developmental Biology,* 199 (1), 111-124.

HOFMANN GE, O'DONNELL MJ and TODGHAM AE, 2008. Using functional genomics to explore the effects of ocean acidification on calcifying marine organisms.*Marine Ecology Progress Series,* 373, 219-225.

HOUZELSTEIN D *et al*., 2004. Phylogenetic analysis of the vertebrate galectin family. *Molecular Biology and Evolution,* 21 (7), 1177-87.

HUANG S *et al*., 2005. The phylogenetic analysis of tetraspanins projects the evolution of cell–cell interactions from unicellular to multicellular organisms. *Genomics,* 86 (6), 674-684.

ILLIES MR *et al*., 2002. Identification and developmental expression of new biomineralization proteins in the sea urchin *Strongylocentrotus purpuratus*. *Development Genes and Evolution,* 212 (9), 419-31.

JONES DT, 1999. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology,* 292 (2), 195-202.

KAKEI M and NAKAHARA H, 1996. Aspects of carbonic anhydrase and carbonate content during mineralization of the rat enamel. *Biochimica Et Biophysica Acta,* 1289 (2), 226-30.

KILLIAN CE and WILT FH, 2008. Molecular aspects of biomineralization of the Echinoderm endoskeleton. *Chemical Reviews,* 108 (11), 4463-4474.

KILLIAN CE and WILT FH, 1996. Characterization of the proteins comprising the integral matrix of *Strongylocentrotus purpuratus* embryonic spicules. *The Journal of Biological Chemistry,* 271 (15), 9150-9.

KILLIAN CE, CROKER L and WILT FH, 2010. SpSM30 gene family expression patterns in embryonic and adult biomineralized tissues of the sea urchin, *Strongylocentrotus purpuratus*. *Gene Expression Patterns,* 10 (2-3), 135-139.

KIM J *et al*., 2010. Functional genomic screen for modulators of ciliogenesis and cilium length. *Nature,* 464 (7291), 1048-51.

KITAJIMA T *et al*., 1996. Expression of spicule matrix protein gene SM30 in embryonic and adult mineralized tissues of sea urchin *Hemicentrotus pulcherrimus*. *Development, Growth and Differentiation,* 38 (6), 687-695.

KIYOMOTO M *et al*., 2007. Skeletogenesis by transfated secondary mesenchyme cells is dependent on extracellular matrix-ectoderm interactions in *Paracentrotus lividus* sea urchin embryos. *Development, Growth & Differentiation,* 49 (9), 731-741.

KNAPP RT *et al*., 2012. Recombinant Sea Urchin Vascular Endothelial Growth Factor Directs Single-Crystal Growth and Branching in Vitro. *J. Am. Chem. Soc., 134, 17908–17911*

KYTE J and DOOLITTLE RF, 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology,* 157 (1), 105-132.

LEFFLER H *et al*., 2004. Introduction to galectins. *Glycoconjugate Journal,* 19 (7-9), 7-9.

LI J *et al*., 2010. Application of glutaraldehyde to in-cell Western assay for normalization. *Analytical Biochemistry,* 398 (2), 254-6.

LIVAK KJ and SCHMITTGEN TD, 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{(-\text{Delta Delta C(T)})}$ Method. *Methods (San Diego, Calif.),* 25 (4), 402-8.

LIVINGSTON BT *et al*., 2006. A genome-wide analysis of biomineralization-related proteins in the sea urchin *Strongylocentrotus purpuratus*. *Developmental Biology,* 300 (1), 335-348.

MACDOUGALL M *et al*., 1998. Identification of a novel isoform of mouse dentin matrix protein 1: spatial expression in mineralized tissues. *Journal of Bone and Mineral Research : The Official Journal of the American Society for Bone and Mineral Research,* 13 (3), 422-31.

MAKABE KW *et al*., 1995. Cis-regulatory control of the SM50 gene, an early marker of skeletogenic lineage specification in the sea urchin embryo. *Development -Cambridge-,* 121 (7), 1957.

MANN K, MANN M and POUSTKA AJ, 2008. The sea urchin (*Strongylocentrotus purpuratus*) test and spine proteomes. *Proteome Science,* 6.

MANN K, WILT FH and POUSTKA AJ, 2010. Proteomic analysis of sea urchin *(Strongylocentrotus purpuratus*) spicule matrix. *Proteome Science,* 8.

MARIE B *et al*., 2008. Nacre calcification in the freshwater mussel Unio pictorum: carbonic anhydrase activity and purification of a 95 kDa calcium-binding glycoprotein. *Chembiochem : A European Journal of Chemical Biology,* 9 (15), 2515-23.

MARIE B *et al*., 2007. The shell matrix of the freshwater mussel Unio pictorum (Paleoheterodonta, Unionoida). *FEBS Journal,* 274 (11), 2933-2945.

MARIE B *et al*., 2012. Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. *PNAS, 109, 51*

MATRANGA V *et al*., 2011. Echinoderms as blueprints for biocalcification: regulation of skeletogenic genes and matrices. *Progress in Molecular and Subcellular Biology,* 52, 225-48.

MCGUFFIN LJ, BRYSON K and JONES DT, 2000. The PSIPRED protein structure prediction server. *Bioinformatics (Oxford, England),* 16 (4), 404-5.

MEDAKOVIC D, 2000. Carbonic anhydrase activity and biomineralization process in embryos, larvae and adult blue mussels *Mytilus edulis* L. *Helgoland Marine Research,* 54 (1), 1-6.

MINOKAWA T *et al*., 2004. Expression patterns of four different regulatory genes that function during sea urchin development. *Gene Expression Patterns,* 4 (4), 449-56.

MITSUNAGA K *et al*., 1986. Carbonic anhydrase activity in developing sea urchin embryos with special reference to calcification of spicules. *Cell Differentiation,* 18 (4), 257-262.

MOULIN L *et al*., 2011. Effects of seawater acidification on early development of the intertidal sea urchin *Paracentrotus lividus* (Lamarck 1816). *Marine Pollution Bulletin,* 62 (1), 48-54.

MÜLLER WEG *et al*., 2007. The unique skeleton of siliceous sponges (Porifera; Hexactinellida and Demospongiae) that evolved first from the Urmetazoa during the Proterozoic: a review. *Biogeosciences,* 4 (2), 219-232.

NOWAK TP, HAYWOOD, PL and BARONDES SH, 1976. Developmentally regulated lectin in embryonic chick muscle and a myogenic cell line. *Biochemical and Biophysical Research Communications,* 68 (3), 650-657.

OGAWA T *et al*., The speciation of conger eel galectins by rapid adaptive evolution. *Glycoconj J Glycoconjugate Journal,* 19 (7-9), 451-458.

OKAZAKI K, 1965. Skeleton formation of sea urchin larvae. V. Continuous observation of the process of matrix formation. *Experimental Cell Research,* 40 (3), 585-96.

OZEKI Y, MATSUI T, SUZUKI M, TITANI K, 1991. Amino acid sequence and molecular characterization of a D-galactoside-specific lectin purified from sea urchin (*Anthocidaris crassispina*) eggs. Biochemistry, Mar 5; 30 (9), 2391-4.

OZEKI Y, KATO KH, TITANI K, MATSUI T, 1995. Developmental expression of D-galactoside-binding lectin in sea urchin (*Anthocidaris crassispina*) eggs. Exp Cell Res., Feb; 216 (2), 318-24.

POIRIER F, 2002. Roles of galectins *in vivo*. *Biochemical Society Symposium,* (69), 95-103.

PRASAD M, BUTLER WT and QIN C, 2010. Dentin sialophosphoprotein in biomineralization. *Connective Tissue Research,* 51 (5), 404-17.

PUSKÁS LG *et al*., 2000. A periplasmic, alpha-type carbonic anhydrase from Rhodopseudomonas palustris is essential for bicarbonate uptake. *Microbiology (Reading, England),* 146, 2957-66.

QIN C, D'SOUZA R and FENG JQ, 2007. Dentin matrix protein 1 (DMP1): new and important roles for biomineralization and phosphate homeostasis. *Journal of Dental Research,* 86 (12), 1134-41.

RAMOS-SILVA P *et al*., 2012. Novel molluskan biomineralization proteins retrieved from proteomics: a case study with Upsalin. *Chembiochem : A European Journal of Chemical Biology,* 13 (7), 1067-78.

RAMOS-SILVA P *et al*., 2013. The Skeletal Proteome of the Coral Acropora millepora: The Evolution of Calcification by Co-Option and Domain Shuffling. Mol Biol Evol. 30(9): 2099–2112.

RIZZO F *et al*., 2006. Identification and developmental expression of the ets gene family in the sea urchin (*Strongylocentrotus purpuratus). Developmental Biology,* 300 (1), 35-48.

SAITOH M *et al*., 2010. Asymmetric inhibition of spicule formation in sea urchin embryos with low concentrations of gadolinium ion. *Development Growth and Differentiation,* 52 (9), 735-746.

SCHRÖDER HC *et al*., 2007. Silicateins, silicase and spicule-associated proteins: synthesis of demosponge silica skeleton and nanobiotechnological applications. *In:* CUSTÓDIO MR, LÔBO-HAJDU G, HAJDU E, MURICY G, ed. *Porifera research : biodiversity, innovation and sustainability* Rio de Janeiro : Museu Nacional, 2007, pp. 581.

SCOTT K and WEINBERG C, 2004. Galectin-1: a bifunctional regulator of cellular proliferation. *Glycoconjugate Journal,* 19 (7-9), 7-9.

SETO J *et al*., 2004. The localization of occluded matrix proteins in calcareous spicules of sea urchin larvae. *Journal of Structural Biology,* 148 (1), 123-130.

SMITH K, 2000. Prokaryotic carbonic anhydrases. *FEMS Microbiology Reviews,* 24 (4), 335-366.

SO AK *et al*., 2004. A novel evolutionary lineage of carbonic anhydrase (epsilon class) is a component of the carboxysome shell. *Journal of Bacteriology,* 186 (3), 623-30.

SODERGREN E, *et al*., 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science (New York, N.Y.),* 314 (5801), 941-52.

STIPP CS, KOLESNIKOVA TV and HEMLER ME, 2003. Functional domains in tetraspanin proteins. *Trends in Biochemical Sciences,* 28 (2), 106-12.

STUMPP M *et al*., 2012. Acidified seawater impacts sea urchin larvae pH regulatory systems relevant for calcification. *PNAS, vol. 109 no. 44 18192-18197.*

SUN Y *et al*., 2011. Roles of DMP1 processing in osteogenesis, dentinogenesis and chondrogenesis. *Cells, Tissues, Organs,* 194 (2-4), 2-4.

SUPURAN CT, SCOZZAFAVA A, and CONWAY J, 2004. *Carbonic anhydrase : its inhibitors and activators.* Boca Raton: CRC Press.

THOMPSON JD, HIGGINS DG and GIBSON TJ, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research,* 22 (22), 4673-80.

TRIPP BC, SMITH K and FERRY JG, 2001. Carbonic anhydrase: new insights for an ancient enzyme. *The Journal of Biological Chemistry,* 276 (52), 48615-8.

VEIS A *et al*., 2009. Characterization of two distinctly different mineral-related proteins from the teeth of the Camarodont sea urchin *Lytechinus variegatus:* Specificity of function with relation to mineralization. *Frontiers of Materials Science in China,* 3 (2), 163-168.

VEIS A, SABSAY B and WU C, 1991. Phosphoproteins as Mediators of Biomineralization. *ACS Symposium Series,* 444, 1-12.

WANG DG, BRITTEN RJ and DAVIDSON EH, 1995. Maternal and embryonic provenance of a sea urchin embryo transcription factor, SpZ12-1. *Molecular Marine Biology and Biotechnology,* 4 (2), 148-53.

WANG X *et al*., 2012. Silicateins, silicatein interactors and cellular interplay in sponge skeletogenesis: formation of glass fiber-like spicules. *The FEBS Journal,* 279 (10), 1721-36.

WEINER S and ADDADI L, 2011. Crystallization Pathways in Biomineralization. *Annual Review of Materials Research,* 41, 21-40.

WESTBROEK P and Marin F, 1998. A marriage of bone and nacre. Nature, 392: 861-862.

WILT FH *et al*., 2008. The dynamics of secretion during sea urchin embryonic skeleton formation*. Experimental Cell Research,* 314 (8), 1744-1752.

WILT FH, 2005. Developmental biology meets materials science: Morphogenesis of biomineralized structures. *Developmental Biology,* 280 (1), 15-25.

WILT FH, KILLIAN CE and LIVINGSTON BT, 2003. Development of calcareous skeletal elements in invertebrates. *Differentiation,* 71 (4-5), 237-250.

WOJTAS M, DOBRYSZYCKI, P and OŻYHAR A, 2012. Intrinsically Disordered Proteins in Biomineralization. *In:* JONG SETO, ed., *Advanced Topics in Biomineralization.* INTECH Open Access Publisher, 2012, pp. 1-32.

YU Z *et al*., 2006. A novel carbonic anhydrase from the mantle of the pearl oyster (*Pinctada fucata*). *Comparative Biochemistry and Physiology, Part B,* 143 (2), 190-194.

ZHU X *et al*., 2001. A large-scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo. *Development (Cambridge, England),* 128 (13), 2615-27.

ZITO F and MATRANGA V, 2009. Secondary Mesenchyme Cells as Potential Stem Cells of the Sea Urchin Embryo. *In:* RINKEVICH B and MATRANGA V, eds. *Stem Cells in Marine Organisms.* Springer Netherlands, 2009, pp. 187-213.

List of Abbreviations

Ab	Antibody
Abs	Absorbance
ACC	Amorphous Calcium Carbonate
AE	Aboral ectoderm
AIM	Acid-insoluble matrix
AP	Apical plate
APS	Ammonium persulfate
ASM	Acid-soluble matrix
AV	Animal-vegetal
AZM	Acetazolamide
B	Blastula
BC	Blastocoelar cells
bp	Base pair
BCIP	5-bromo-4-chloro-3-indolyl phosphate disodium salt
BSA	Bovine Serum Albumin
°C	Degree Celsius
CA	Carbonic Anhydrase
CBB	Coomassie Brilliant Blue
cDNA	Complementary DNA
CDS	Coding sequence
Cl	Cleavage
cm	Centimeter
CRD	Carbohydrate recognition domain
Ct	Threshold cycle
Da	Dalton
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
DTT	Dithiothreitol
ELISA	Enzyme-linked immunosorbent assay
EN	Endoderm
et al.	*et alii*
Fig.	Figure
FTIR	Fourier transform infrared
g	Gram
G	Gastrula
Gal-8	Galectin-8
Gly	Glycine
h	Hour
His6	6 Histidine redidues
Hp	*Hemicentrotus pulcherrimus*
Ht	*Heliocidaris tuberculata*

ISH	*in situ* hybridization
in vitro	within the glass
in vivo	within the living
IEF	Iso Electric Focusing
Inh.	Inhibitor
kb	Kilo-base pair
kDa	Kilo-Dalton
LB	Lysogeny broth
LG	Late gastrula
Lv	*Lytechinus variegatus*
M	Molar (mol/l)
mA	Milliampere
MB	Mesenchyme blastula
min	Minute
ml	Milliliter
mm	Millimeter
mM	Millimolar
mV	Millivolt
MW	Molecular Weight
NBT	4-Nitroblue Tetrazolium Chloride
ng	Nanogram
Ni	Nickel
Ni-NTA	Nickel-nitrilotriacetic acid
nm	Nanometer
NSM	Non-skeletogenic mesoderm
NTA	Nitrilotriacetic acid
OA	Oral-aboral
OD	Optical density
OE	Oral ectoderm
OLE	Oral-lateral ectoderm
P	Pluteus
PAGE	Poly-Acrylamide-Gel-Electrophoresis
PBS	Phosphate Buffered Saline
PBS-T	Phosphate Buffered Saline-0.1% Tween 20
PC	Pigment cells
PCR	Polymerase chain reaction
pH	*Potentia Hydrogenii*
pI	Isoelectric Point
Pl	*Paracentrotus lividus*
PMC	Primary Mesenchyme Cells
PMSF	Phenylmethanesulfonylfluoride
Pr	Prism
qPCR, Q-PCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
rpm	Revolutions per minute
rPl-CA	Recombinant protein: *Paracentrotus lividus* carbonic anhydrase
RQ	Relative expression ($2^{-\Delta\Delta CT}$)
RT	Room temperature
RT-PCR	Real-time polymerase chain reaction
SDS	Sodium Dodecyl Sulfate
sec	Second
SEM	Scanning Electron Microscopy

Ser	Serine
SM	Spicule Matrix
SMC	Secondary Mesenchyme Cells
SMic	Small Micromeres
Sp	*Strongylocentrotus purpuratus*
TEM	Transmission Electron Microscopy
TEMED	N,N,N',N'-Tetramethylethylenediamine
Thr	Threonine
T_m	DNA melting temperature
Tris	Tris(hydroxymethyl)aminomethane
Tween 20	Polyoxyethylene (20) sorbitan monolaurate
UTR	Untranslated region of mRNA
V	Volt
v/v	Volume/volume (Vol.%)
w/v	Weight/volume
WB	Western Blot
WMISH	whole mount *in situ* hybridizaion
ΔCCt	Comparative Threshold Cycle Method
ΔCt value	Threshold cycle (Ct) value, normalized to the median Ct value
µg	Microgram
µl	Microliter
µm	Micrometer

Acknowledgements

I welcome the people that contributed to this work and thank them for taking part in my life.

I acknowledge my feelings for the people who have shaped my personality.

Special Thanks to..

- Prof. Dr. Franz Brümmer
- Dr. Valeria Matranga
- Dr. Frédéric Marin
- Prof. Dr. Robin Ghosh and Prof. Dr.-Ing. M. Schmidt (Dekan)

- Dr. Francesca Zito, Dr. Rosa Bonaventura, Mauro Biondo, Dr. Antonina Azzolina, Dr. Vladislav Grebenjuk

- Dr. Caterina Costa and the administration personnel

- The 7th EU Framework Program of Marie Curie ITN
- Universität Stuttgart
- CNR of Palermo
- Université de Bourgogne

- Prof. Dr. Robin Fåhræus
- My colleagues
- The 'unknown behind the scene'