



HAL
open science

A minimal molecular toolkit for mineral deposition? Biochemistry and proteomics of the test matrix of adult specimens of the sea urchin *Paracentrotus lividus*.

Konstantinos Karakostis, Isabelle Zanella-Cléon, Françoise Immel, Nathalie Guichard, Philippe Dru, Thierry Lepage, Laurent Plasseraud, Valeria Matranga, Frédéric Marin

► To cite this version:

Konstantinos Karakostis, Isabelle Zanella-Cléon, Françoise Immel, Nathalie Guichard, Philippe Dru, et al.. A minimal molecular toolkit for mineral deposition? Biochemistry and proteomics of the test matrix of adult specimens of the sea urchin *Paracentrotus lividus*.. Journal of Proteomics, 2016, 136, pp.133-144. 10.1016/j.jprot.2016.01.001 . hal-01283937

HAL Id: hal-01283937

<https://hal.science/hal-01283937v1>

Submitted on 11 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

A minimal molecular toolkit for mineral deposition? Biochemistry and proteomics of the test matrix of adult specimens of the sea urchin *Paracentrotus lividus*.

Konstantinos Karakostis *[a, b], Isabelle Zanella-Cléon [c], Françoise Immel [d], Nathalie Guichard [d], Philippe Dru [e], Thierry Lepage [e, f], Laurent Plasseraud [g], Valeria Matranga [a], Frédéric Marin *[d]

[a] Dr. K. Karakostis, Dr. V. Matranga:

Cell Stress and Environment Research Group, Consiglio Nazionale delle Ricerche
Istituto di Biomedicina e Immunologia Molecolare "Alberto Monroy"
Via Ugo La Malfa 153 -90146 Palermo, Italy

[b] INSERM UMR S 1162

Institut de Génétique Moléculaire, Hôpital Saint-Louis
27 Rue Juliette Dodu – 75010 Paris, France

[c] Centre Commun de Microanalyse des Protéines

Institut de Biologie et Chimie des Protéines (IBCP)
7, passage du Vercors - 69367 LYON cedex 07

[d] Dr. F. Marin, Dr. F. Immel, N. Guichard :

UMR CNRS 6282 Biogéosciences
Université de Bourgogne Franche-Comté (UBFC), 6 boulevard Gabriel – 21000 Dijon, France

[e] UMR 7009 CNRS/UPMC – Laboratoire de Biologie du Développement,

Observatoire Océanologique de Villefranche-sur-Mer, BP 28
Quai de la Darse - 06234 Villefranche-sur-Mer cedex, France

[f] IBV – Institut de Biologie Valrose,

CNRS UMR7277, Inserm U1091, UNS Université Nice Sophia Antipolis
Parc Valrose
06108 Nice cedex 2 France

[g] Dr. L. Plasseraud :

ICMUB, UMR CNRS 6302, Université de Bourgogne Franche-Comté (UBFC), Faculté des Sciences Mirande, 9 Avenue A. Savary, F-21000 Dijon, France

* Corresponding authors:

Konstantinos Karakostis: chem898@yahoo.gr; konstantinos.karakostis@inserm.fr

Frédéric Marin: Frederic.marin@u-bourgogne.fr

Key-words:

Sea urchin, biomineralization, calcite, proteomics, calcifying matrix, carbonic anhydrase, C-type lectin

Abstract

The sea urchin endoskeleton consists of a magnesium-rich biocalcite comprising a small amount of occluded organic macromolecules. This structure constitutes a key-model for understanding the mineral – organics interplay, and for conceiving *in vitro* bio-inspired materials with tailored properties. Here we employed a deep-clean technique to purify the occluded proteins from adult *Paracentrotus lividus* tests. We characterized them by 1- and 2D -electrophoreses, ELISA and immunoblotting, and using liquid chromatography coupled with Mass Spectrometry (nanoLC-MS/MS), we identified two metalloenzymes (carbonic anhydrase and MMP), a set of MSP130 family members, several C-type lectins (SM29, SM41, PM27) and cytoskeletal proteins. We

demonstrate the effect of the protein extract on the crystals, with an *in vitro* crystallization assay. We suggest that this small set of biomineralization proteins may represent a ‘minimal molecular crystallization toolkit’.

Significance

Biominerals exhibit often superior chemical properties, when compared to their inorganic counterparts. This is due *pro parte* to the proteins that are occluded in the mineral. However, the limited available studies on biomineralization have not yet succeeded in identifying a minimal set of proteins directly involved in the formation of the biomineral *in vivo* and sufficiently required for *in vitro* precipitation. Indeed, the high number of proteins identified by high-throughput screening in the recent years does not encourage the possibility of recreating or tailoring the mineral *in vitro*. Thus, the identification of biomineralization proteins involved in protein-mineral interactions is highly awaited. In the present study, we used the sea urchin, *Paracentrotus lividus* (*P. lividus*), to identify the native proteins directly taking part in protein-mineral interactions. We employed an improved deep-clean technique to extract and purify the native occluded skeletal matrix proteins from the test and identified them by the highly sensitive technique of microLC-MS/MS. We show that this minimal set of proteins has a shaping effect on the formation of biocalcite *in vitro*. This work gives insights on the biomineralization of the sea urchin, while it paves the way for the identification of biomineralization proteins in other biomineralizing systems. Understanding the ‘biologically controlled mineralization’ will facilitate the *in vitro* formation and tailoring of biominerals in mild conditions for applications in medicine and materials science.

Introduction

Echinoderms represent a large phylum of marine invertebrates that comprise echinoids (sea urchins and sand dollars), asteroids (sea stars), ophiuroids (brittle stars), crinoids (feather stars) and holothuroids (sea cucumbers). Although presumably existing since the Late Precambrian [1, 2], they acquired the capacity to mineralize a skeleton during the Precambrian-Cambrian transition [3]. Indeed, similarly to other phyla, during one of the major biological radiations - the famous ‘Cambrian explosion’- echinoderms set up a complex gene regulatory network (GRN) involved in the biomineralization process. Some fundamental aspects of this network have been elucidated in sea urchins, the best-studied echinoderm class so far: analyses on the development of the sea urchin embryo, performed during the last decade, have shown that skeletogenesis is regulated by a subset of the GRN [4-6] and starts in privileged extracellular spaces of the gastrula embryo, 24 h after egg fertilization [7]. This process is mediated by a set of biomineralization proteins specifically expressed by a cellular population devoted to skeleton formation – the primary mesenchyme cells (PMCs) – that are themselves activated by signaling cues from overlying ectodermal cells [8]. In

the adult, *i.e.*, after metamorphosis, the biominerals are deposited as microscopic bony plates forming a three-dimensional meshwork, the stereome [9]. The stereome of the adult *P. lividus* is composed of different calcified structures including the test, the teeth that are parts of the Aristotle's lantern and the spines. The test is characterized by a typical microporous calcitic microstructure [9].

Similarly to other metazoan biominerals, what makes the sea urchin test peculiar is the presence of occluded macromolecules - mostly consisting of proteins, representing about 0.1% w/w of the dry weight of the mineralized structure [10]. These proteins modulate the mineral deposition process [11]. During the last two decades, a few biomineralization-related proteins that are abundant in the calcified sea urchin test have been individually characterized by classical molecular biology and biochemistry approaches [12]. More recently, high-throughput screening techniques, such as transcriptomics and proteomics, have allowed the identification of complete skeletal matrix proteomes, also referred as 'skeletalomes'. In particular, proteomic studies on the adult test, spines and tooth of the sea urchin *Strongylocentrotus purpuratus* (*S. purpuratus*) have revealed more than one hundred proteins within the organic matrix, including several C-type lectins, metalloproteases, small acidic proteins, collagens and enzymes such as carbonic anhydrase [13-15]. However, these analyses were neither complemented nor confirmed by studies aiming at defining the molecular function of the identified proteins and their connection to the process of mineral deposition. The high number of identified proteins raises questions about the putative existence of a 'minimal protein toolkit' required for modulating mineral deposition. This debate is central in biomineralization studies, since it determines whether it is realistic or not to consider synthesizing abiotically biocomposites that perfectly mimic natural biominerals, only with a limited set of genetically engineered skeletal proteins.

Beyond the functional and applied aspects addressed above, the large number of proteins identified by proteomics in general raises fundamental concerns about the interpretation of the results and their integration into a coherent and meaningful molecular model. Indeed, for the accurate identification of the proteins from the skeletal matrix of sea urchin tests by proteomics, high attention should be paid on the *a priori* definition of the functional- and spatial-based protein target group (*i.e.*, skeletal matrix). To this purpose, an optimized preparation of the samples prior to the proteomic analysis constitutes a crucial step that ensures the purification and analysis of the target protein group. This issue was raised through a recent controversy about the coral skeletal matrix, where it was found that some of the 'skeletal' proteins identified by proteomics were more likely to be cellular contaminants rather than being involved in biomineralization [16, 17].

In the present paper, we aimed at circumscribing the aforementioned question by focusing on the proteins of the skeletal matrix that are the most tightly linked to the mineral phase. To this end, we

applied a drastic cleaning process of the skeletal test tissues, prior to the extraction of the skeletal matrix. We further characterized biochemically the bulk acetic acid-soluble (ASM) and – insoluble (AIM) matrices, and tested the ability of the ASM to interact with calcium carbonate, by the *in vitro* crystallization assay. In addition, we applied proteomics on these extracts and obtained about ten proteins that can arguably be considered as the most strongly linked to the mineral phase. We suggest that this set may represent the ‘minimal toolkit’ involved in modulating mineral deposition.

Materials and Methods

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

Adult sea urchins (*P. lividus*) were collected from the Northwestern coast of Sicily (Mediterranean Sea) and kept in aquaria, with circulating seawater obtained from the collection site. Tests of adult *P. lividus* were mechanically separated from spines and teeth and cleaned by brushing under running water. Superficial organic contaminants were removed from the surface of the test fragments with incubation for 24 h in bleach (0.26% NaOCl), under constant agitation, at room temperature and were subsequently thoroughly rinsed with Milli-Q water and dried at 25 °C. The purity of the samples was investigated with a tabletop scanning electron microscope (SEM, Hitachi TM1000) and samples were crushed into approximately 1 mm² fragments and subsequently into a fine powder, in liquid nitrogen, with a mortar grinder (pulverisette 2 model, Fritsch), which was subsequently sieved (> 200 µM).

All protein extractions were performed at 4 °C, as described in Osuna-Mascaro *et al.* [18]. Briefly, the organic constituents were obtained by overnight decalcification of the biomineral (25 g) in cold acetic acid (5% v/v) at 4 °C that was added by an automated burette (Titronic Universal, Schott, Mainz, Germany) at a flow rate of 100 µL every 5 s. The solution (final pH around 4.2, final volume about 1 L) was centrifuged at 3900 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 on a Nalgene filtration device (5 µm mesh), and its volume, reduced by ultrafiltration in an Amicon ultra-filtration cell through a Millipore® membrane (YM10; 10 kDa cut-off). The resulting concentrated solution (about 10 mL) was extensively dialyzed against Milli-Q water (with several water changes), freeze-dried and weighed. Two consecutive extractions were performed, referred to as batch 1 and batch 2.

Skeletal matrix analysis on 1D gels.

The extracted ASM and AIM were separated on 12% acrylamide 1D mini-gels (Mini Protean III, Bio-Rad) under denaturing conditions. Samples were heat-denatured (100°C for 3 min) in the Laemmli buffer containing SDS and mercapto-ethanol. For the ASM, 20 µg to 40 µg of matrix were applied on top of the gel. For the AIM, the matrix was suspended in Laemmli buffer and heat-denatured (5 min), before being centrifuged. Only the supernatant was used for SDS-PAGE. This new fraction was referred as the LS-AIM (Laemmli-soluble, acetic acid insoluble). The gels were stained with Bio-Safe Coomassie G250 (CBB, Bio-Rad), Silver Nitrate [19], Alcian Blue [20] and carbocyanine [21] (also called “Stains-all”).

Protein Matrix Analysis on 2D gels.

The ASM and AIM were analysed on a 2D-gel Protean IEF cell (Bio-Rad), according to the manufacturer's instructions. ASM (170 µg in 80 µl) and AIM (1 mg in 125 µl) were rehydrated in the rehydration buffer. Only a part of AIM was solubilized. 7 cm linear pH 3–10 immobilised pH gradient (IPG) strips were re-hydrated overnight, and IEF was processed (250 V for 15 min, followed by 8000 V until 10,000 Vh). The strips were transferred in the equilibration buffer, before being rinsed in TGS buffer, positioned on top of precast gradient gels (Mini-Protean III, 4–12% polyacrylamide), and fixed in place with an overlay solution of 0.5% agarose / TGS (w/v). The second dimension electrophoresis was performed in standard conditions and the gels were subsequently stained with Bio-Safe Coomassie G250 (CBB) and with silver nitrate.

***In vitro* calcium carbonate precipitation assay.**

In vitro precipitation of CaCO₃ was performed by slow diffusion of ammonium bicarbonate vapors into calcium carbonate chloride solution, containing or not dilute ASM, according to a published protocol [22]. Briefly, 200 µL of 10 mM CaCl₂ solutions containing serial dilutions of the ASM, were applied into a 16-well culture slide (BD Falcon, Becton Dickinson, Franklin Lakes, NJ, USA), the cover of which was formerly pierced (1 mm diameter holes). The culture slide was sealed with parafilm, to allow diffusion only through the holes. Blank controls (calcium chloride solution without ASM) were performed. The samples were put in a desiccator containing ammonium bicarbonate crystals. The air was pumped out and the desiccator was stored at 4 °C for 48h. Subsequently, the crystallization solution in each well was discarded by suction with a blunt-ended needle connected to a vacuum pump. The culture slide was dried at 50 °C for 2 hours, disconnected from the plastic wells and subsequently observed directly under a tabletop scanning electron microscope (Hitachi TM-1000), without gold/carbon sputtering. The test was repeated four times.

Proteomic analysis of the ASM and AIM by nanoLC-MS/MS

For both fractions (ASM, AIM), the trypsin digestion was performed in solution. The samples were denatured with 50 μ L of Guanidine 7.2M/Tris 100mM buffer for 15min à 50°C. Reduction was performed with 5 μ L of 100mM Dithiothreitol (Sigma-Aldrich) in Tris100mM for 30min at 50°C. Alkylation was performed with 10 μ L of 200 mM Iodoacetamide (Sigma-Aldrich) in Tris100mM for 30min at room temperature in the dark. Then the solutions were diluted to obtain a 0.8M Guanidine concentration and treated with 1 μ g of trypsin (Sequence grade, Promega) in 10 μ L of 50 mM NH₄HCO₃ at 37°C.

The samples were purified on C18 SpinTips (ProteaBio) according to the manufacturer protocol except elution done by 50 μ L of... and 50 μ L of CH₃CN/H₂O/TFA 80/20/0.1%. Then 1 μ L of sample was injected into the nanoLC-nanoESI-MS/MS system for analysis. Tryptic peptides were analyzed by mass spectrometry performed using a LTQ Velos instrument (Dual Pressure Linear Ion Trap) equipped with a nanospray source (Thermo Fisher Scientific) and coupled to a U3000 nanoLC system (Thermo Fisher Scientific). Using the software Xcalibur 2.1 (Thermo Fisher Scientific), a MS survey scan was acquired over the m/z range 400-1600 in Enhanced resolution mode. The MS/MS scans were acquired in Normal resolution mode over a dynamic m/z range for the 20 most intense MS ions with a charge of 2 or more and with a collision energy set to 35eV. The spectra were recorded using dynamic exclusion of previously analyzed ions for 0.5 min with 50 millimass units (mmu) of mass tolerance. The peptide separation was obtained on a C18 PepMap micro-precolumn (5 μ m; 100 Å; 300 μ m x 5 mm; Dionex) and a C18 PepMap nanocolumn (3 μ m; 100 Å; 75 μ m x 150 mm; Dionex) using a linear 120 min gradient from 0 to 40% of solvent B, where solvent A was 0.1% HCOOH in H₂O/CH₃CN (95/5) and solvent B was 0.1% HCOOH in H₂O/CH₃CN (20/80) at 300 nL/min flow rate.

***In silico* proteomic data analysis**

Protein identification was performed with the MASCOT 2.2.4 Algorithm through the Proteome Discoverer software v1.1 (Thermo Fisher Scientific). The Pliv proteome database, comprising 124032 entries, was used for that purpose. This protein dataset results from the translation of the genome and transcriptome of *Paracentrotus lividus*. The nucleic acid sequences were formerly generated in the frame of the 'Arca di Noe' sea urchin network, a subset of the NOI (Network Of Excellence) "Marine Genomics Europe (MGE)", a consortium comprising 44 institutions from 16 countries within and outside Europe. The different datasets on *P. lividus* are in restricted access on the website of the Villefranche-sur-Mer Marine Laboratory (UMR CNRS-UPMC 7009, see <http://octopus.obs-vlfr.fr/>) but are accessible upon request and will be publicly available soon. Another interrogation was performed with an in-house unpublished non redundant dataset made

from the collection of about 800 proteins associated to calcium carbonate biominerals of several metazoans, from sponges to vertebrates (hen eggshell and fish otoliths). In this dataset, echinoderms are represented by the sea urchins *Strongylocentrotus purpuratus*, *Lytechinus pictus*, *Hemicentrotus pucherrimus*, *Heliocidaris erythrogramma* and *Heliocidaris tuberculata*. The protein hit sequences obtained were further characterized via the following proteomics tools: ClustalW, InterProScan, ProtParam, which are available at www.expasy.org/proteomics/.

Biochemical analysis of the ASM and AIM.

Western-blot and ELISA were employed for the analysis of the test matrix. Specific polyclonal antibodies were raised and used to identify carbonic anhydrase; and the presence of CTLs was tested using heterologous antibodies raised against SM30 and SM50, from *Strongylocentrotus purpuratus* (gift from Prof. Dr. F. Wilt, University of Berkeley, USA). For the Western blot, ASM and AIM samples were fractionated by size on a denaturing SDS-gel and then electro-transferred on PVDF Immobilon-P membranes (Millipore Corp.) using the Mini-Trans Blot module (Bio-Rad) for 90 min at 120 mA. The membranes were subsequently blocked in 1% gelatin in Tris-buffered saline (TBS) for 30 min and then incubated overnight at 20 °C with the specific polyclonal antibody anti-*Pl-CA* (dilution 1:1000), developed in mice and elicited against recombinant *Pl-CA* protein, following standard Western Blotting procedure [24]. Negative controls were performed with the pre-immune serum.

Conventional ELISA technique (Enzyme-Linked Immuno-Sorbent Assay) was employed for the identification of SM30 and SM50 [25] in the ASM, by using 96-well microplates (Nunc Maxisorb). 200 ng/well of antigen (ASM) were incubated (37°C, 90 min.). After discarding the solution, three-times rinsing (TBS-Tween 20) and blocking with gelatin solution (1% wt/vol in TBS buffer), the preparation was incubated with the first antibody overnight, at 4 °C, thoroughly rinsed (three-times) and incubated 1 h with the secondary antibody at a dilution of 1:30000 (GAR/AP conjugate, Sigma A3687). After thorough rinsing, the staining was obtained by incubation with substrate (aqueous solution of diethanolamine containing p-nitrophenylphosphate). The microplate was read at 405 nm (Bio-Rad multichannel spectrophotometer). Absorbance values were normalized in percentages, with the highest value set as 100%, and the blank, at 0%.

Results

Matrix extraction and purification

The extraction procedure represents a crucial step for the identification of biomineralization-related proteins from matrices. We employed a stringent method involving three key-steps for the

extraction and purification of the matrix, to preclude the presence of contaminants and to guarantee the identification of exclusively native matrix proteins. We used two different sample batches that gave similar yields, suggesting that in either case, the extraction and purification procedure did not result in loss of sample, and ensuring the reproducibility and robustness of the results (Table 1).

The first extraction step comprised a thorough and extended bleaching treatment with sodium hypochlorite, to remove soft-tissue contaminants and was monitored by scanning electron microscopy (SEM, Fig. 1). The second step involved a slow and mild decalcification treatment of the biomineral using acetic acid, to expose the occluded matrix. Finally, the third crucial step involved the thorough separation of the matrix proteins, based on their solubility in acetic acid [26]. Acetic acid-soluble (ASM) and -insoluble (AIM) fractions were obtained by centrifugation followed by ultrafiltration and dialysis of the ASM. Each fraction was finally lyophilized and quantified by direct weighing. The ratio of the soluble matrix to the insoluble one, ranged between 1:7 and 1:10. The combined fractions represented approximately 0.025% w/w of the dry weight of the biomineral (Table 1), confirming previous results from various marine species indicating a range of 0.02% - 0.1% of the occluded proteins of the matrix to the biomineral mass [26, 27]. It is noted that our yield is consistently lower than the yield previously found by Mann *et al.*, [13] (0.8%) on matrix extracted from tests of *S. purpuratus*.

Table 1 Mass and yield of the acetic acid-soluble (ASM) and acid-insoluble (AIM) matrices extracted from each batch of purification (Batch I and II)

Batch	ASM	AIM	ASM/AIM ratio	Total
Batch I 25g	0.60 mg (0.0024 %)	5.5 mg (0.022 %)	0.109	6.1 mg (0.0244 %)
Batch 2 25g	0.70 mg (0.0028 %)	5 mg (0.020 %)	0.140	5.7 mg (0.0228 %)

1D- and 2D-gel electrophoreses of the skeletal matrices

The occluded matrices (ASM and AIM) extracted from clean tests were analyzed by mono-dimensional SDS-PAGE. Gels were subsequently stained with Bio-Safe Coomassie G250 (CBB), silver nitrate (AgNO₃), Alcian blue and carbocyanine, providing constitutive information on the protein patterns of each fraction. CBB stains mainly proteins, while silver nitrate stains

indiscriminately most of the macromolecules; Alcian Blue stains acidic glycoconjugates while carbocyanine stains in blue proteins that putatively bind calcium. The results are shown in Fig. 2. CBB did not stain the ASM, but stained better the AIM for which three prominent bands were visualized above 40 kDa, 25 kDa and below 25 kDa. Several more diffuse bands were distinguished, in particular in the range between 70 kDa and 15 kDa (Fig. 2A). Silver staining was more efficient in staining ASM and AIM than CBB, as more bands were identified (Fig. 2A). The results of the Alcian blue and carbocyanine staining are shown in Fig. 2B. The protein band patterns of the ASM and AIM suggest that the protein content, from a qualitative viewpoint, is different in each sample. However, certain proteins exhibiting identical apparent molecular mass (~25 kDa, ~50 kDa) are found in both the soluble and insoluble matrices. The ASM, stained with Alcian Blue, shows a faint and diffuse smear. A strongly stained smear of the AIM (with two prominent bands at 50 kDa and 70 kDa) is observed. Carbocyanine (Stains-all) stained blue several discrete bands of the AIM, suggesting that they are presumably able to bind calcium. Carbocyanine mostly stained blue the lower molecular range macromolecules of the ASM (below 25 kDa, Fig. 2B, lane 5) and the higher molecular weight range macromolecules of the AIM (until 70 kDa, Fig. 2B, lane 6). In both fractions, macromolecules of 10 kDa, 23 kDa and 70 kDa, were stained.

Additionally, for a detailed characterization of the proteins and the post-translational modifications in each fraction of the matrix, the ASM and AIM were separately analysed by 2DE (Fig. 3). In the case of AIM, an extremely limited part of the matrix was dissolved in the rehydration buffer. As expected, the protein spots were visualised better when stained by silver nitrate. In the ASM, a group of at least six alkaline proteins ($pI = 6$ to 8) of around 60 kDa and 70 kDa and a single acidic spot ($pI = 3$) of 30 kDa, were visualised (Fig.3A). The AIM mainly included a group of four acidic to slightly acidic proteins ($pI = 3$ to 7) of 20 kDa and 30 kDa, as well as a neutral spot ($pI = 7$) of 130 kDa (Fig. 3B).

Proteomics of the test matrices

The biochemical characterization performed in this study was complemented by a proteomic analysis that emphasized the role of a limited number of protein families in the biomineralization of *P. lividus* test. The ASM and AIM test extracts were analysed by nano-LC-MS/MS after trypsin digestion and the peptide sequences were used for screening the Basic Local Alignment Search Tool, using the proteome database of *P. lividus* at <http://octopus.obs-vlfr.fr>. To ensure the reproducibility of the results, we separately analysed two batches of ASM and AIM, numbered batch I, and batch II. The results are outlined in Table 2 and in the Supplementary Table S1, while some of the rough spectra data are also shown in Supplementary Figure S2. Table 2 and Table S1 indicate that both batches gave congruent results, *i.e.*, batch II contains all the protein hits found in

batch I. We observed however that batch II generated additional hits compared to batch I. In all the cases, at least two peptide sequences corresponding to each protein hit were identified, guaranteeing the robustness of the results. In most cases though, each protein sequence was identified by several peptides (*i.e.* up to 14 for contig17238). In ASM I, 17 peptides corresponding to two protein sequences and two contigs were identified. ASM II generated a list of 23 peptides corresponding to 4 protein sequences and two contigs. AIM I generated a list of 57 peptides that corresponded to 6 protein sequences and two contigs. At last, AIM II generated 74 peptides that gave hits with 7 protein sequences and two contigs.

It is worth-mentioning that some protein hits of the ASM batches were also identified in the AIM ones, suggesting that the identified matrix proteins can be in both soluble and insoluble states. This also suggests that the two matrices may exhibit a relatively high degree of similarity, a finding consistent with previous observations made in two mollusk models, namely the freshwater mussel and the nautilus [28, 29].

Overall, 13 different proteins were identified in the occluded matrix of *P. lividus* test, from 9 contigs and 8 sequences (Table 2). They include five lectins, three members of the MSP130 family, two almost identical carbonic anhydrases, one matrix-metalloproteinase, actin and myosin. These proteins can be consequently clustered in four different groups: (I) C-type lectins, a class of glycoproteins that are able to bind mono- and disaccharides [30], (II) proteins of the MSP130 family, (III) metallo-enzymes and (IV) cytoskeletal proteins.

The lectin sequences identified in the *Paracentrotus lividus* proteome dataset correspond to contigs 6474, 14265, 14621 and 15948 and to the sequence s36065. They all contain a calcium-dependent (C-type) lectin domain either in their N-terminus (contigs 15948, 6474, and 14621), middle (sequence s36065), or C-terminus (contig 14265). A search for homology using BLASTP against the whole UniprotKB database shows that all of them match, at different degrees, to lectins from various sea urchins: for example, contig14265 is 81% identical to the spicule matrix protein SM29 from *Strongylocentrotus purpuratus* (Q8MUL0_STRUPU) and that contig14621 is 73% identical to SM41, the spicule matrix protein from *Hemicentrotus pulcherrimus* (Q26264) but exhibits also similarities with SM50 and SM32 from *Strongylocentrotus purpuratus* (respectively P11994 and Q9MUL1) and with SM34 from *Lytechinus pictus* (Q05904). The sequence s36065 is 86% identical to PM27, the 27-kDa spicule matrix protein from *Strongylocentrotus purpuratus* (Q26616), and 83 and 81 % identical to the two PM27 sequences of *Heliocidaris tuberculata* and *erythrogramma*, respectively (Q95W97_HELTB and Q95W86_HELER). Interestingly, contigs15948 and 6474, identified both in ASM II and AIM I, correspond to two sequences that exhibit similarities with two uncharacterized lectins from *Strongylocentrotus purpuratus*, Sp-C-lectin/PMC1 (W4ZI92_STRPU, database entry Glean3:27906) and Sp-Clect_76 (W4YDJ5_STRPU, database entry Glean3:13825),

respectively. These two lectins are derived from genes SPU_027906 and SPU_013825. Interestingly, in a former study [15], they were identified in the four calcified compartment of *S. purpuratus*, namely the test, the spines, the teeth and the spicules. Contig15948 corresponds to a 480 aa long protein, of 53 kDa theoretical mass and of basic isoelectric point (9.69). Downstream the C-type lectin N terminus (aa 13 to 192) and a short threonine-rich spacer (aa 193 to 216), contig15948 exhibits a long 'repetitive low-complexity domain' (RLCD) almost entirely constituted of a succession of triplets (aa 217 to 449) of the P-x-y type. This RLCD can be subdivided in different sub-domains: one, from aa 226 to 270, is made of perfectly alternating PNR and PGV/F triplets, the second one, from aa 271 to 324, corresponds to P-x-y triplets where X is often G or Q, and a third one, from aa 325 to 357, which exhibits exclusively H residues in x positions. Downstream (aa 358 to 449), the succession of triplets is less perfect and contains intercalating aa residues (one to three) that disrupt the ternary periodicity. The overall sequence of contig15948 is highly enriched in proline residues, which account for more than 21% of the sequence. Contig6474 has a similar primary structure and exhibits a basic theoretical isoelectric point (9.73) though it is much smaller (243 residues, 26 kDa of theoretical mass). It contains a C-type lectin domain at its N-terminus (aa 1 to 126) followed by a long domain (aa 127 to 225) mostly constituted of P-x-y repeats, interspersed from time to time by one (G, S, R, or N) or two (GG, NN) residues. Its overall sequence is enriched in proline (17%) and glycine (14%) residues.

In the proteome of *P. lividus*, the MSP130 protein family is identified by at least six sequences (contigs 17159, 17238 and sequences s8835, s11289, s13733 and s72921). The contig17159 is identical to sequences s72921 and s13733. They exhibit an 80% identity with MSP130 protein from the sea urchin *Heliocidaris tuberculata* (Q9GP17_HELTB) and 78%, with that from *Heliocidaris erythrogramma* (Q9GP19_HELER). The contig17238 exhibits a high degree of similarity to sequence s11289. Both are 72% identical to the MSP130-related-2 protein of *Strongylocentrotus purpuratus* (Q8MUK7_STRPU). Finally, the sequence s8835 is 72% identical to the MSP130-related-1 protein of the same species (Q8MUK8_STRPU).

Concerning the metallo-enzymes, two distinct protein families were identified: at first, two sequences (contig14487 and s75885) encoding for carbonic anhydrase were obtained in the *Paracentrotus lividus* proteome dataset: Although of different length sizes, their overlapping parts are almost identical and differ only by one aa residue (valine *versus* alanine) at the C terminus (10th position upstream the last leucine residue of contig14487). Secondly, another identified sequence, s77447, encodes for a protein that is 87% identical to two uncharacterized matrix metalloproteinases (MMP) of *Strongylocentrotus purpuratus* (W4YD41_STRPU, W4YD42_STRPU). Furthermore, s77447 exhibits lesser similarities with a large set of other matrix metalloproteinases of diverse sources, including an uncharacterized MMP from the giant owl limpet

Lottia gigantea (V4B204_LOTGI), MMP-17 of the edible oyster *Crassostrea gigas* (K1R1S7_CRAGI), MMP-14 of the parasitoid *Fopius arisanus* (A0A0C9QER8_9HYME).

Finally, three sequences encoding cytoskeletal proteins were obtained by blasting peptide data against the proteome dataset of *Paracentrotus lividus*. These sequences include contig20288 that encodes for myosin, and contig20740 and sequence s10890, both encoding for actin.

Table 2 Summary of the peptides identified by nanoLC-MS/MS and prediction of the proteins

Accession No.	Protein similarity	ASM coverage	AIM coverage
Contig14265	C-type lectin SM29	- GCLPGWTNFG NYCFK	- ALPTGGRPGMPMPPPGHLVVTKDMNHNFQHNWLRYTGGGNNKI WMGLAELPSDPESNR
		- LPYDEANNVCR	- LPYDEANNVCR
			- FKPDQPQNAHR
Contig17238	MSP-130-related 2	- SWFNNLDASS TPSAK	- ITAPATDIAQCGFLVAWITR
			- ITDPGTVHVWDMYR
		- AVAEFYPQVFN SK	- WNSLCTGSVGSNPVSLK
		- IGDIDPSK	- GTAAEDPTSGTFTDPEGTISIVRIVKSNRQQQGGQPFPFR
		- TPDNFNYLMVT GK	- AKTPDNFNYLMVTGKKSGLTMYR
		- LGNLQVSK	- FNNMVGMLQIGIR
		- SGLTMYR	- SWFNNLDASSTPSAK
			- LGNLQVSK
			- AVAEFYPQVFNSKSYTPDAQSHGPVQMDQASTKR
			- FSILMENKRIGDIDPSK
			- SYFGLESYLFTADEGAPTTYTMGNKR
			- GTAAEDPTSGTFTDPEGTISIVRIVK
s8835	MSP-130-related 1	- SAYDADMMVA YTAGR	- QGKTMPIYALNRGAAEKSAVDADMMVAYTAGR
			- SAYDADMMVAYTAGR
		- AMNDPTQLGQL R	- TIDFTLFNDPTFAPMLEAQFIR
		- GIGDLDPDSIVFI PAEK	- RRAMNDPTQLGQLR
		- VSGTITMYQIYE IPR	- ADVQDLQNDLR
		- LPLVMVTSRVSGTITMYQIYEIPR	
	- MSYNQLYR	- VAPGNLINR	

- ADVQDLQNDLR
- MSYNQLYR
- GIGDLPDSIVFIPA EK

s13733	MSP-130	<ul style="list-style-type: none"> - FNAPAEVQR - FNAPAEVQR - MTGDLSPESLT FIPPEK
---------------	---------	--

s36065	sp27	<ul style="list-style-type: none"> - GGWFQIGQQC FK - AIDIPFWTGLQN K - MGQWYVR 	<ul style="list-style-type: none"> - AIDIPFWTGLQNKWNQLLERY
---------------	------	--	---

Contig15948	C-type lectin	<ul style="list-style-type: none"> - EDAMVACTEIP EVIGGKLLDITD YDNFK 	<ul style="list-style-type: none"> - EDAMVACTEIP EVIGGKLLDITD YDNFK
--------------------	---------------	--	--

s11289	MSP130 family	<ul style="list-style-type: none"> - WNSLCTGSVG SNPVSLK - GTAAEDPTSGT FTDPEGTISIVR
---------------	------------------	--

Contig14487	Carbonic anhydrase 2	<ul style="list-style-type: none"> - VSTENMYVLR - YNGSLTTPK
--------------------	-------------------------	---

Contig6474	lectin	<ul style="list-style-type: none"> - EATFFCSRYGG SLYALDSPSK 	<ul style="list-style-type: none"> - YGGSLYALDSPSKNR
-------------------	--------	--	---

Contig17159	MSP-130	<ul style="list-style-type: none"> - QPYTGQLGDPGPNTFSQGLEPR - FSLVDGNNPNDPEIFDQVYMF GGR - IFNGAAIHGNPR - MTGDLSPESLTFIPPEK - FNAPAEVQR LKALHVRQPYTGQLGDPGPNTFSQGLEPRHITLDAR E - TRFSLVDGNNPNDPEIFDQVYMF GGR - SDRGPECESLAVGDVQGRK - DGHIGHTYQKLYAERMTGDLSPESLTFIPPEKSLNNKPHLLVTGR
--------------------	---------	---

S72921 = Contig17159	MSP-130	<ul style="list-style-type: none"> - FNAPAEVQR - QPYTGQLGDPGPNTFSQGLEPRHITLDAR - TRFSLVDGNNPNDPEIFDQVYMF GGR - SDRGPECESLAVGDVQGR - DGHIGHTYQKLYAER
---------------------------------	---------	--

		- SLNNKPHLLVTGR
		- LGNLQVSK
		- AVAEFYQVFNSK
		- SYTPDAQSHGPVSQMDQASTK
		- FSILMENKR
		- IGDIDPSK
Contig14621	SM41	- GWPVNPQNPLAPAGSAPIMKRQAPPTRPGQGGQQIPAGVGPQW DLLAVTEMR
s75885	Carbonic anhydrase 2	- KQSPINIQTRDVVQADLGEFVFDGLQSNFGPQVGGHPGR - GLHGYNAPTRVEVSDNGHTLKVSTENMYVLR - YNGSLTPK
s77447	MMP17	- VLSSAFQVWGDVARLDFNPTFRPKA - GIQSLYGAR - YAVQHGWSK - QVSTFKDNEVFFIAETR
Contig20288	Myosin	- RAINDLQNSKSRLGSENSDLARQLEDLENQLSQLNKTRT - VNGELNQLR
Contig20740	Actin	- AGFAGDDAPR
s10890	Actin	- AVFPSIVGRPRHQGMVMGMGQKDSYVGDEAQSQR

***A posteriori* confirmation of the presence of specific markers in the matrix**

Since our proteomic analysis revealed the presence of C-type lectins (CTLs) and of carbonic anhydrase (CA) in the test matrix, we confirmed their effective presence in the different skeletal extracts by ELISA or by Western-blot, by using heterologous or homologous polyclonal antibodies. Fig. 4B demonstrates the ELISA results performed on the ASM using heterologous polyclonal anti-SM30 and anti-SM50 antibodies (against *S. purpuratus*). Because of its low solubility, AIM could not be tested by this technique. The results showed that the antibodies cross-reacted with the ASM, giving a classical sigmoid response (in function of the antibody concentration), the one obtained with anti-SM50 being stronger. These results confirm our proteomic analysis and indicate the presence of CTL domains in the ASM, without specifying the identity of the targeted proteins. These antibodies were also used in Western blots, but the results (not shown) presented a high background in spite of our multiple attempts.

Fig. 4A shows the Western blot pattern of the matrices, when incubated with an antibody elicited against embryonic *P. lividus* carbonic anhydrase (*Pl-CA*). In both ASM and AIM, this homologous antibody shows that CA is present as a component of the test matrix of *P. lividus*, confirming the proteomic data that identified *Pl-CA* in the two fractions. The band of *Pl-CA*, identified in the AIM was slightly higher (second lane, around 27 kDa) compared to the equivalent band in the ASM (first lane, around 24 kDa), suggesting that this protein might be post-translationally modified (glycosylation), but this possibility was not investigated further. Additionally, a second band at 58 kDa was identified by Western blot in the ASM but not in the AIM suggesting the presence of a CA-related protein of higher molecular weight; we cannot exclude the possibility that this band is a *Pl-CA* dimer, in spite of the denaturing gel conditions (SDS and β -mercaptoethanol).

In conclusion, *Pl-CA* (identified by ELISA, Western Blot and by proteomic analysis LC-MS/MS), SM30 and SM50 (identified by ELISA and LC-MS/MS), are components of the protein matrix extracted from the adult mineralized tests. Furthermore, our results show that *Pl-CA* and CTL-containing proteins, identified in embryonic skeletons, are also expressed in the adult skeleton.

***In vitro* functional crystallization assay: interaction of the ASM with calcium carbonate**

The ASM of the test extract was assayed *in vitro*, by the diffusion method, for checking its effect on the formation of calcium carbonate. Note that, to obtain homogeneous results, this experimental device is miniaturized. Thus, it does not allow producing sufficiently large amount of crystals, in order to collect them, dissolve them and perform proteomics to identify the crystal-occluded proteins. The results are shown in Fig. 5. Increasing concentrations of the whole ASM extract, varying from 5 $\mu\text{g/ml}$ to 100 $\mu\text{g/mL}$, were tested. In control blank experiments, no protein was added, resulting in the production of calcite crystals with the typical rhombohedral habitus. ASM modified the crystals in a dose-dependent manner. Polycrystalline aggregates of various sizes were observed in the presence of ASM at concentrations varying from 5 $\mu\text{g/mL}$ to 25 $\mu\text{g/mL}$. At 50 $\mu\text{g/mL}$, large polycrystalline aggregates and ellipsoidal-like crystals were observed. At concentrations as high as 100 $\mu\text{g/mL}$, exclusively ellipsoidal-like crystals were formed. No inhibition of the crystal formation at any concentration was recorded (Fig. 5). Our results illustrate the strong interaction of the ASM with the mineral. This interaction was dose-dependent, *i.e.*, stronger at higher ASM concentrations.

Discussion

We have identified and characterized components of the test matrix of the sea urchin *Paracentrotus lividus*, a common species distributed throughout the Mediterranean Sea and in the North-Eastern Atlantic Sea. *P. lividus* is a key-model for developmental studies, but molecular data on the

biomineralization process of its skeletal tissues are fragmentary. The occluded matrix macromolecules of its test were extracted by decalcification and characterized by biochemical assays and proteomic analysis. The biochemical characterization comprised mono- and bi-dimensional electrophoreses, Western blots, ELISA and *in vitro* functional crystallization assay, while LC-MS/MS was employed for the proteomic analysis.

The number of proteins identified in this study is significantly lower (by a factor of ten) than what reported in previous findings on *S. purpuratus*, listing a total of 110 different proteins in the matrix of test and spines [13]. One may claim that the paucity of protein hits may be the result of three different combined factors: the quality of the dataset used for the interrogation; technical aspects of the trypsin digestion; the abundance of proteins with RLCs (Repetitive Low Complexity Domains) in the matrix of *Paracentrotus lividus*; The first possibility may be ruled out since the proteome dataset of *P. lividus*, with 124032 entries, is rather complete, and since all the ‘biomineralization proteins’ (*sensu lato*) identified in *S. purpuratus* were identified in *P. lividus* proteome. Secondly, about the technical aspects of trypsin digestion, one often asserts that bulk digestion (‘in-solution’) have a lesser efficiency on protein mixtures than in-gel digestion [30]. However, this type of digestion may miss low molecular weight and high molecular weight proteins, while in-solution digestion takes in account all the proteins whatever their mass; furthermore, in our case, the digestion procedure has been optimized. About the third factor, it is known indeed that calcium carbonate biominerals contain numerous proteins with Repetitive Low-Complexity Domains [31], and that RLCs-containing proteins of the ‘biomineralization type’ (*i.e.*, with abundance of Gly, Ser, Pro, Asp, Ala, Gln) have statistically less chance to be cleaved by trypsin than ‘standard’ proteins, or will generate very long peptides which cannot be accurately analyzed by MS-MS. As many proteins of the skeletal matrix of *S. purpuratus* have homologues in the matrix of *P. lividus*, it is however rather unlikely that the two matrices have very dissimilar amounts of biomineralization RLC-containing proteins, that would explain the discrepancy.

We believe that the stringent extraction conditions that we employed, including the overnight bleaching treatment, is probably the main factor explaining the discrepancy between the low number of proteins identified in our case, in comparison to the high number identified in former studies [13]. Our treatment keeps intact the occluded proteins that are tightly bound to the mineral phase, degrading other proteins, in particular, superficial ‘adsorbed’ proteins that may or may not be involved in regulating the bio-calcification process. We think that an accurate cleaning of the biomineralized structure is required for avoiding contaminants that may randomly adsorb onto the mineral surfaces, but have probably little or no role in the regulation of the mineral deposition. On the other hand, we cannot exclude that a drastic bleaching treatment may degrade some proteins that are required for modulating mineral deposition but that are not occluded in the mineral phase,

just simply adsorbed. To the ‘stringency factor’, we cannot rule out another factor, the high insolubility of the AIM. AIM represents indeed about 90% of the matrix, and its high insolubility properties may preclude its proper digestion by trypsin, resulting in the limited release of peptides, and subsequently, in partial identification of key proteins present in this fraction. However, we obtained similar results (*i.e.*, few hits) with the ASM fractions, suggesting that the solubility parameter does not play important role in the number of identified proteins. Finally, since our purpose was to keep stringent conditions for the identification of the matrix components, we used strict threshold values for filtering proteomic data.

We identified a dozen proteins that can be categorized in four different groups. It is important to highlight that all the identified proteins have also been found in the sea urchin embryo [15]. This suggests that these proteins have a crucial role at different stages of the sea urchin development, from the embryo to the adult. Moreover, previous studies on the embryo offer insights on the functional aspects of each protein in the adult mineralization. We suggest that the small subset of the proteins found in the adult tests as described in this work, has an exclusive and direct role in biomineralization and satisfies the functional requirements of biomineralization. These groups are: (I) C-type lectins (CTLs), (II) proteins of the MSP130 family, (III) metallo-enzymes, and (IV) cytoskeletal proteins.

The first group comprises five calcium-dependent lectin domain-containing proteins, also referred as CTLs. They represent a large class of glycoproteins able to bind mono- and di-saccharides [32]. In sea urchin skeletal tissues, they were the first family of proteins to be identified [33] and are usually considered as the most prominent actors of the bio-calcification of this class of echinoderms. For example, PM27, which is closely related to the sequence s36065, is a biomineralization protein, previously extracted from NaOCl-cleaned demineralized embryo spicules from *S. purpuratus* [34] and described as a skeletal protein specifically associated with the mineral growth [35, 36]. PM27 exhibits a glycine-loop structure in its Pro-Gly-Met repeat sequence motif, which is associated to protein elasticity, this parameter being linked to the fracture resistance in biomineralized structures. We suggest that the homolog of PM27 in *P. lividus* may play a key-role in the elasticity of the sea urchin test. The spicule matrix protein SM29 has been previously described in the embryo as a secreted basic protein, which includes a signal sequence, a CTL-domain and a proline-rich region [37]. The SM32 protein is also a secreted, basic protein, closely related to SM50 (~60% amino acid identity to the corresponding CTL domains), which contains twenty-eight PXY repeat elements, where X is usually an asparagine and Y is usually a glutamine. Similarly, SM41 belongs to the SM50 sub-family, which represents the most abundant occluded matrix protein in the sea urchin spicules. The SM50 members exhibit a CTL domain and an extensive proline-rich region, which facilitates the formation of a network serving as functional

template for the deposition of the biomineral. We observed the presence of several CTLs in the test matrix. Accordingly, a similar strategy is employed in the embryo, where PMCs express various proteins containing CTL domains [38]. Indeed, several of these proteins, *i.e.* PM27 and SM30, exhibit different localization and expression patterns in the embryo, suggesting different roles in biomineralization, such as defining the extracellular space (PM27, SM50) or regulating the secretion of the mineral [34]. It is also important to mention that the CTL-containing proteins identified here also contain abundant amounts of weakly hydrophobic amino acid residues, such as glycine and proline. The repetitive nature of the glycine-rich (GRP) domains is likely to allow the formation of β -sheet structures. Glycine repeats, tightly bound to polysaccharides/carbohydrates, facilitate cell adhesion between neighboring cells in various biological systems, promoting protein-protein interactions and tissue formation [39]. Protein cross-linking determines the hardness and stiffness of several hard tissues of marine invertebrates. As an example, the beak of the jumbo squid (*Dosidicus gigas*) mainly consists of chitin fibers and glycine-rich proteins, offering increased hardness, stiffness and plasticity properties [40]. *Pl-SM30*, which contains a poly-glycine region and a CTL-domain, may act as a cross-linking agent to promote the deposition of calcite. Similarly to the Gly-rich regions, extensive proline-rich regions and proline repeats (PGM) serve as binding domains in protein-protein interactions. Due to their hydrophobic surface, they bind well to flat hydrophobic surfaces or heavily glycosylated proteins, thus promoting the solubility of the protein complex in water. In accordance, Gong *et al* [41] showed that the proline-rich protein SM50, which exhibits a high similarity (above 80%) with the SM32 and SM41 proteins identified here, stabilizes a hydrated form of amorphous calcium carbonate ($\text{ACC}\cdot\text{H}_2\text{O}$) *in vitro*. From a general viewpoint, this type of regulation of the mineral phase transition by organic molecules is commonly encountered in numerous CaCO_3 -mineralizing organisms [41].

The second protein group comprises proteins of the MSP-130 family [42]. MSP-130 is a biomineral forming mesenchyme-specific cell surface glycoprotein that was represented by several peptides in all the four fractions (ASM and AIM, batches I and II). MSP-130 has been identified as a PMC-specific protein: it comprises two glycine-rich regions toward the amino-terminal end [43] and binds calcium [42, 44]. The *mSP130* mRNA was localized in PMCs associated with the rapidly growing tips of the spicular skeleton, in the pluteus larva [35, 45]. Ettensohn [46], showed that MSP-130 proteins are present in several invertebrate deuterostomes and in molluscs; in addition, related proteins are found in bacteria and in several algae, suggesting that horizontal gene transfer occurred for *mSP130*, allowing the evolution of biomineralization in metazoans. Accordingly, Livingston *et al.* [47] have proposed that pre-vertebrates and echinoderms might have utilized a shared genetic “toolkit” to carry out biomineralization. In accordance, Cameron *et al.* [48] identified three *mSP-130* genes, a matrix metalloproteinase and a carbonic anhydrase involved in

biomineralization in the *S. kowalevskii*-translated genome, underlining the importance of these proteins as well as the evolutionary relationship between the genetic biomineralization programs in echinoderms and hemichordates. In conclusion, MSP-130 is a key-player of biocalcification of the sea urchin, although its exact role is still unclear. It is a well-conserved protein that exhibits homologues in various biomineralizing systems. In the present study, we identify MSP-130 and MSP-130-related-1 and -2 proteins as major compounds of the *P. lividus* test matrix.

The third group of proteins corresponds to two metalloenzymes: one member of the matrix metalloproteinase family (MMP) and, one carbonic anhydrase (*Pl-CA*). These zinc-containing proteins were identified in both the ASM and the AIM. Matrix metalloproteinases (MMPs) and carbonic anhydrases have conserved functions in the biomineralization process of echinoderms, metazoans [49] and vertebrates [47, 50-52]. Two hundred and forty metalloprotease putative genes were identified in the *S. purpuratus* genome and authors claimed that their activity is linked to developmental functions, including the modification of the extracellular matrix, as well as the regulation of signaling pathways or modulation of the cellular adhesive properties [53]. Mann et al. [13] have also identified several MMPs in *S. purpuratus* proteome and speculated that they take part in the maturation of spine and test matrix proteins. Interestingly, MMP activity was shown to be required for the genesis of spicules in the sea urchin embryo, by promoting the delivery of calcium carbonate and matrix proteins to the sites of spicule growth [54]. Interestingly, in mammals, MMPs were shown to play an essential role in modeling/remodeling the skeletal and extra skeletal connective tissues [55]; and their activity was shown to depend on the concentration of manganese [56]. Here, we identify one MMP in the occluded test matrix as an important player of sea urchin biomineralization. Additionally, we showed by MS-MS analysis that this protein is present in both soluble and insoluble forms.

Besides MMP, carbonic anhydrase (CA) was also identified in our study. It was previously characterized in the sea urchin embryo as a PMC-specific biomineralization protein involving a glycine-rich N-terminal domain that potentially facilitates protein-protein and protein-mineral interactions; and a CA-specific domain which retains all the crucial residues required for carbonic anhydrase activity (Karakostis *et al.*, submitted in Marine Biotechnology, MBT-15-0118). In that manuscript, we proposed a dual function of *Pl-CA* in embryo biomineralization, acting both as an enzyme catalyzing the hydration of CO₂ and as a regulator or co-enzyme of other proteins, by promoting protein-carbohydrate interactions. Here, we confirm that *Pl-CA* is present in both the ASM and AIM of the adult tests. CA has been described in various systems as an essential regulator of calcification. By catalyzing the reversible hydration of CO₂, it promotes the saturation of bicarbonate (HCO₃⁻) and carbonate (CO₃²⁻) anions at the mineralization site, which interact with Ca²⁺ cations, to form CaCO₃ minerals. Indeed, 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%) [57]. The importance of CA activity in skeletogenesis was shown by inhibition experiments, leading to perturbation of the skeleton formation in the embryo [58, 59]. Indeed, inhibition of CA blocks skeletogenesis and echinochrome production in *P. lividus* and *H. tuberculata* embryos [60]. Previous analysis on *S. purpuratus* has also identified CA in the proteome of test and spines [13]. Here, we identified the embryonic *Pl*-CA as a component of the skeletal matrix of adult tests, emphasizing the importance of this protein in sea urchin biomineralization during all developmental stages.

The fourth group is represented by the cytoskeletal proteins, myosin and actin. These two proteins have already been identified in the skeletal matrix of different metazoans [61-63]. Their presence in calcified tissues is often interpreted as cellular contaminations [64]. In the present case, their persistence, in spite of our thorough bleaching, is puzzling and may have a high significance in the context of biomineralization. In a milestone paper, Weiss and coworkers [65] showed indeed that a chitin synthase identified in the calcifying tissue of a bivalve exhibits a myosin motor head domain. In their mollusk model, they suggested for the first time a functional role for 'cytoskeletal forces' in the precise controlled mineral deposition process performed outside the cells, establishing a link between the cytoplasm and the extracellular matrix. Whether this also occurs for sea urchins should be tested.

Our proteomic analysis has neither identified collagens nor acidic proteins. Mann and coworkers [13] detected collagens, owing to peptides carrying the Gly-x-y triplet signature typical of the triple helical domains found in this protein family. None of the peptides identified here has such a signature. More intriguing is the case of acidic proteins, which are usually considered to be significant players in biocalcification [66]. Due to their negative charge and to their hydrophilicity, they strongly interact with calcium in solution, with calcium carbonate mineral surfaces, and are prone to be easily occluded in the mineral during its growth. In the sea urchin *S. purpuratus*, P16 and P19, two acidic proteins that are PMC-specific [37], were identified as potential candidates for such interaction with calcium carbonate. However, we did not detect their putative homologs in the test matrix of *P. lividus*. Neither did we evidence the presence of other Asp-rich or Glu-rich proteins. Besides the possibility that such a protein class is not at all present in the calcified tissues of sea urchins, another possible explanation of its absence is that the RLCD-containing proteins of the Asp-rich or Glu-rich type contain very few arginine or lysine cleavage sites, as explained above; they would not be properly cleaved by trypsin, and the resulting peptides would not be correctly ionized and analyzed by MS-MS. We cannot exclude such a technical bias in our analyses.

Conclusions

We have identified and characterized components of the test matrix of the sea urchin *Paracentrotus lividus*, a key-model for developmental studies. The matrix macromolecules were extracted from the biomineral (tests) after thorough cleaning and complete decalcification. They were subsequently analyzed with both biochemical assays and proteomic analyses. We identified five proteins exhibiting a C-type lectin domain, two metalloenzymes, including *Pl-CA* and the depolymerizing protein MMP, a set of proteins of the MSP130 family and finally, proteins of the cytoskeleton. Although none of these proteins is particularly acidic, we demonstrated that they collectively have a true affinity for the calcite of the test of *P. lividus* and that they exhibit a prominent shaping effect on crystal formation. Thus, they may represent a ‘minimal toolkit’ for handling calcium carbonate biomineralization in sea urchins.

Acknowledgements

This work has been fully supported by the EU-ITN BIOMINTEC Project, contract number 215507 to V. Matranga. The first author, K. Karakostis, has been the recipient of a Marie Curie ITN program fellowship in the frame of the aforementioned project, which allowed three short-term scientific missions in Biogeosciences research unit, Dijon, France. The authors thank Prof. Dr. F. Wilt, Graduate School Division of Cell and Developmental Biology, University of Berkeley, USA for providing SM30 and SM50 antibodies. The authors also thank the researchers of different European marine biology institutions who contributed to the construction of the *Paracentrotus lividus* libraries; they include P. Cormier, J. Morales, I. Arnone and P. Oliveri. Additional support of the work of K. Karakostis in F. Marin’s lab includes INTERRVIE projects (INSU, CNRS).

Figure 1. Examination of *P. lividus* adult tests by Scanning Electron Microscopy (SEM). A and B: tests with organic contaminants. C and D: Tests cleared off contaminants.

Figure 2. Gel electrophoresis of the test extracts. A: 1D SDS-PAGE fractionation of acid-soluble (ASM) and acid-insoluble (AIM) shell matrix proteins from *P. lividus* adult tests, stained with CBB (Lanes 1-3) or AgNO₃ (Lanes 4-5). Molecular weight markers (in kDa) are indicated on the left. B: 1D SDS-PAGE fractionation of ASM and AIM stained with Alcian Blue (Lanes 1-3) or StainsAll (Carbocyanine, Lanes 4-6. Alcian Blue stain involving copper, identifies heavily glycosylated proteins and glyco-conjugates. Bands that are stained blue with Stains-All are potentially calcium-binding proteins.

Figure 3: 2DE SDS-PAGE fractionation of the shell matrix proteins from *P. 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). A: 2DE SDS-PAGE fractionation of ASM; B: 2DE SDS-PAGE fractionation of AIM. Most prominent spots are indicated by circles.

Figure 4. *A posteriori* analysis of the shell matrix. A: SDS-PAGE fractionation followed by Western Blot with antibody against PI-CAN. ASM and AIM from *P. lividus* adult tests cross-reacted with the anti-CA Ab. Anti-CA antibody was used at a dilution of 1:1000. Film was developed for 4min. B: ELISA with heterologous antibodies against proteins involving C-type lectin domains (SM30 and SM50). Both antibodies cross-react with the ASM.

Figure 5. *In vitro* crystallization of calcium carbonate in the presence of ASM. A,B) Blank experiment (no proteins). C,D) 5 µg/ml of ASM; E,F) 25 µg/ml of ASM; G,H) 50 µg/ml of ASM; I,J) 100 µg/ml of ASM. Photos were taken by SEM at 400X (A,C,E,G) and at 1500X (B,D,F,H).

References

- [1] Swalla BJ, Smith AB. Deciphering deuterostome phylogeny: molecular, morphological and palaeontological perspectives. *Philos Trans R Soc Lond B Biol Sci.* 2008;363:1557-68.
- [2] Blair JE, Shah P, Hedges SB. Evolutionary sequence analysis of complete eukaryote genomes. *BMC Bioinformatics.* 2005;6:53.
- [3] Bottjer DJ, Davidson EH, Peterson KJ, Cameron RA. Paleogenomics of echinoderms. *Science.* 2006;314:956-60.
- [4] Etensohn CA. Lessons from a gene regulatory network: echinoderm skeletogenesis provides insights into evolution, plasticity and morphogenesis. *Development.* 2009;136:11-21.
- [5] Sharma T, Etensohn CA. Activation of the skeletogenic gene regulatory network in the early sea urchin embryo. *Development.* 2010;137:1149-57.
- [6] Oliveri P, Tu Q, Davidson EH. Global regulatory logic for specification of an embryonic cell lineage. *Proceedings of the National Academy of Sciences of the United States of America.* 2008;105:5955-62.
- [7] Davidson EH. The sea urchin genome: where will it lead us? *Science.* 2006;314:939-40.
- [8] Cheers MS, Etensohn CA. P16 is an essential regulator of skeletogenesis in the sea urchin embryo. *Developmental biology.* 2005;283:384-96.
- [9] Smith AB. *Stereom microstructure of the echinoid test.* London: Palaeontological Association; 1980.
- [10] Wilt FH, Killian CE, Hamilton P, Croker L. The dynamics of secretion during sea urchin embryonic skeleton formation. *Exp Cell Res.* 2008;314:1744-52.
- [11] Wilt FH. Developmental biology meets materials science: Morphogenesis of biomineralized structures. *Developmental biology.* 2005;280:15-25.
- [12] Gilbert PU, Wilt FH. Molecular aspects of biomineralization of the echinoderm endoskeleton. *Progress in molecular and subcellular biology.* 2011;52:199-223.

- [13] Mann K, Poustka AJ, Mann M. The sea urchin (*Strongylocentrotus purpuratus*) test and spine proteomes. *Proteome science*. 2008;6:22.
- [14] Mann K, Poustka AJ, Mann M. In-depth, high-accuracy proteomics of sea urchin tooth organic matrix. *Proteome science*. 2008;6:33.
- [15] Mann K, Wilt FH, Poustka AJ. Proteomic analysis of sea urchin (*Strongylocentrotus purpuratus*) spicule matrix. *Proteome science*. 2010;8:33.
- [16] Ramos-Silva P, Marin F, Kaandorp J, Marie B. Biomineralization toolkit: the importance of sample cleaning prior to the characterization of biomineral proteomes. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110:E2144-6.
- [17] Marie B, Ramos-Silva P, Marin F, Marie A. Proteomics of CaCO₃ biomineral-associated proteins: how to properly address their analysis. *Proteomics*. 2013;13:3109-16.
- [18] Osuna-Mascaro A, Cruz-Bustos T, Benhamada S, Guichard N, Marie B, Plasseraud L, et al. The shell organic matrix of the crossed lamellar queen conch shell (*Strombus gigas*). *Comparative biochemistry and physiology Part B, Biochemistry & molecular biology*. 2014;168:76-85.
- [19] Morrissey JH. Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. *Analytical biochemistry*. 1981;117:307-10.
- [20] Wall RS, Gyi TJ. Alcian blue staining of proteoglycans in polyacrylamide gels using the "critical electrolyte concentration" approach. *Analytical biochemistry*. 1988;175:298-9.
- [21] Campbell KP, MacLennan DH, Jorgensen AO. Staining of the Ca²⁺-binding proteins, calsequestrin, calmodulin, troponin C, and S-100, with the cationic carbocyanine dye "Stains-all". *The Journal of biological chemistry*. 1983;258:11267-73.
- [22] Pavat C, Zanella-Cleon I, Becchi M, Medakovic D, Luquet G, Guichard N, et al. The shell matrix of the pulmonate land snail *Helix aspersa maxima*. *Comparative biochemistry and physiology Part B, Biochemistry & molecular biology*. 2012;161:303-14.
- [23] Marie B, Zanella-Cleon I, Le Roy N, Becchi M, Luquet G, Marin F. Proteomic analysis of the acid-soluble nacre matrix of the bivalve *Unio pictorum*: detection of novel carbonic anhydrase and putative protease inhibitor proteins. *Chembiochem : a European journal of chemical biology*. 2010;11:2138-47.
- [24] Towbin H, Staehelin T, Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America*. 1979;76:4350-4.
- [25] Clark MF, Adams AN. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *The Journal of general virology*. 1977;34:475-83.
- [26] Weiner S, Hood L. Soluble protein of the organic matrix of mollusk shells: a potential template for shell formation. *Science*. 1975;190:987-9.
- [27] Swift DM, Sikes CS, Wheeler AP. Analysis and function of organic matrix from sea urchin tests. *Journal of Experimental Zoology*. 1986;240:65-73.
- [28] Marie B, Luquet G, Bedouet L, Milet C, Guichard N, Medakovic D, et al. 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*. 2008;9:2515-23.
- [29] Marie B, Marin F, Marie A, Bedouet L, Dubost L, Alcaraz G, et al. Evolution of nacre: biochemistry and proteomics of the shell organic matrix of the cephalopod *Nautilus macromphalus*. *Chembiochem : a European journal of chemical biology*. 2009;10:1495-506.

- [30] Choksawangkar W, Edwards N, Wang Y, Gutierrez P, Fenselau C. Comparative study of workflows optimized for in-gel, in-solution, and on-filter proteolysis in the analysis of plasma membrane proteins. *Journal of proteome research*. 2012;11:3030-3034.
- [31] Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R, Roksar DS, Montagnani C, Joubert C, Piquemal D, Degnan BM. Parallel evolution of nacre building gene sets in molluscs. *Molecular biology and evolution*. 2010;27:591-608.
- [32] Smith LC, Ghosh J, Buckley KM, Clow LA, Dheilly NM, Haug T, et al. Echinoderm immunity. *Advances in experimental medicine and biology*. 2010;708:260-301.
- [33] Benson S, Sucov H, Stephens L, Davidson E, Wilt F. A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. I. Authentication of the cloned gene and its developmental expression. *Developmental biology*. 1987;120:499-506.
- [34] Urry LA, Hamilton PC, Killian CE, Wilt FH. Expression of spicule matrix proteins in the sea urchin embryo during normal and experimentally altered spiculogenesis. *Developmental biology*. 2000;225:201-13.
- [35] Harkey MA, Klueg K, Sheppard P, Raff RA. Structure, expression, and extracellular targeting of PM27, a skeletal protein associated specifically with growth of the sea urchin larval spicule. *Developmental biology*. 1995;168:549-66.
- [36] Wustman BA, Santos R, Zhang B, Evans JS. Identification of a "glycine-loop"-like coiled structure in the 34 AA Pro,Gly,Met repeat domain of the biomineral-associated protein, PM27. *Biopolymers*. 2002;65:362-72.
- [37] Illies MR, Peeler MT, Dechtiaruk AM, Ettensohn CA. Identification and developmental expression of new biomineralization proteins in the sea urchin *Strongylocentrotus purpuratus*. *Development genes and evolution*. 2002;212:419-31.
- [38] Kitajima T, Tomita M, Killian CE, Akasaka K, Wilt FH. Expression of spicule matrix protein gene SM30 in embryonic and adult mineralized tissues of sea urchin *Hemicentrotus pulcherrimus*. *Development, growth & differentiation*. 1996;38:687-95.
- [39] Bruce Alberts AJ, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter. *Molecular Biology of the Cell* 2007.
- [40] Miserez A, Li Y, Waite JH, Zok F. Jumbo squid beaks: inspiration for design of robust organic composites. *Acta biomaterialia*. 2007;3:139-49.
- [41] Gong YU, Killian CE, Olson IC, Appathurai NP, Amasino AL, Martin MC, et al. Phase transitions in biogenic amorphous calcium carbonate. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109:6088-93.
- [42] Anstrom JA, Chin JE, Leaf DS, Parks AL, Raff RA. Localization and expression of msp130, a primary mesenchyme lineage-specific cell surface protein in the sea urchin embryo. *Development*. 1987;101:255-65.
- [43] Parr BA, Parks AL, Raff RA. Promoter structure and protein sequence of msp130, a lipid-anchored sea urchin glycoprotein. *The Journal of biological chemistry*. 1990;265:1408-13.
- [44] Kabakoff B, Hwang SP, Lennarz WJ. Characterization of post-translational modifications common to three primary mesenchyme cell-specific glycoproteins involved in sea urchin embryonic skeleton formation. *Developmental biology*. 1992;150:294-305.
- [45] Bonaventura R, Zito F, Costa C, Giarrusso S, Celi F, Matranga V. Stress response gene activation protects sea urchin embryos exposed to X-rays. *Cell Stress Chaperones*. 2011;16:681-7.
- [46] Ettensohn CA. Horizontal transfer of the msp130 gene supported the evolution of metazoan biomineralization. *Evolution & development*. 2014;16:139-48.

- [47] Livingston BT, Killian CE, Wilt F, Cameron A, Landrum MJ, Ermolaeva O, et al. A genome-wide analysis of biomineralization-related proteins in the sea urchin *Strongylocentrotus purpuratus*. *Developmental biology*. 2006;300:335-48.
- [48] Cameron CB, Bishop CD. Biomineral ultrastructure, elemental constitution and genomic analysis of biomineralization-related proteins in hemichordates. *Proceedings Biological sciences / The Royal Society*. 2012;279:3041-8.
- [49] Le Roy N, Jackson D, Marie B, Ramos-Silva P, Marin F. The evolution of metazoan α -carbonic anhydrases and their roles in calcium carbonate biomineralization. *Frontiers in Zoology*. 2014;11:1-16.
- [50] Rafiq K, Cheers MS, Etensohn CA. The genomic regulatory control of skeletal morphogenesis in the sea urchin. *Development*. 2012;139:579-90.
- [51] Krane SM, Inada M. Matrix metalloproteinases and bone. *Bone*. 2008;43:7-18.
- [52] Wuthier RE, Lipscomb GF. Matrix vesicles: structure, composition, formation and function in calcification. *Frontiers in bioscience*. 2011;16:2812-902.
- [53] Angerer LM, Angerer RC. Patterning the sea urchin embryo: gene regulatory networks, signaling pathways, and cellular interactions. *Curr Top Dev Biol*. 2003;53:159-98.
- [54] Ingersoll EP, Wilt FH. Matrix metalloproteinase inhibitors disrupt spicule formation by primary mesenchyme cells in the sea urchin embryo. *Developmental biology*. 1998;196:95-106.
- [55] Holmbeck K, Bianco P, Caterina J, Yamada S, Kromer M, Kuznetsov SA, et al. MT1-MMP-deficient mice develop dwarfism, osteopenia, arthritis, and connective tissue disease due to inadequate collagen turnover. *Cell*. 1999;99:81-92.
- [56] Russo R, Pinsino A, Costa C, Bonaventura R, Matranga V, Zito F. The newly characterized PI-jun is specifically expressed in skeletogenic cells of the *Paracentrotus lividus* sea urchin embryo. *The FEBS journal*. 2014;281:3828-43.
- [57] Stumpp M, Hu MY, Melzner F, Gutowska MA, Dorey N, Himmerkus N, et al. Acidified seawater impacts sea urchin larvae pH regulatory systems relevant for calcification. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109:18192-7.
- [58] Mitsunaga K, Akasaka K, Shimada H, Fujino Y, Yasumasu I, Numanoi H. Carbonic anhydrase activity in developing sea urchin embryos with special reference to calcification of spicules. *Cell differentiation*. 1986;18:257-62.
- [59] Chow G, Benson SC. Carbonic anhydrase activity in developing sea urchin embryos. *Exp Cell Res*. 1979;124:451-3.
- [60] Zito F, Koop D, Byrne M, Matranga V. Carbonic anhydrase inhibition blocks skeletogenesis and echinochrome production in *Paracentrotus lividus* and *Heliocidaris tuberculata* embryos and larvae. *Development, growth & differentiation*. 2015;57:507-14.
- [61] Drake JL, Mass T, Haramaty L, Zelzion E, Bhattacharya D, Falkowski PG. Proteomic analysis of skeletal organic matrix from the stony coral *Stylophora pistillata*. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110:3788-93.
- [62] Rahman S, Yamato I, Saijo S, Mizutani K, Ishizuka-Katsura Y, Ohsawa N, et al. Biochemical and biophysical properties of interactions between subunits of the peripheral stalk region of human V-ATPase. *PloS one*. 2013;8:e55704.
- [63] Isowa Y, Sarashina I, Oshima K, Kito K, Hattori M, Endo K. Proteome analysis of shell matrix proteins in the brachiopod *Laqueus rubellus*. *Proteome science*. 2015;13:21.
- [64] Ramos-Silva P, Kaandorp J, Huisman L, Marie B, Zanella-Cleon I, Guichard N, et al. The skeletal proteome of the coral *Acropora millepora*: the evolution of calcification by co-option and domain shuffling. *Molecular biology and evolution*. 2013;30:2099-112.

- [65] Weiss IM, Schonitzer V, Eichner N, Sumper M. The chitin synthase involved in marine bivalve mollusk shell formation contains a myosin domain. FEBS letters. 2006;580:1846-52.
- [66] Marie B, Luquet G, Pais De Barros JP, Guichard N, Morel S, Alcaraz G, et al. The shell matrix of the freshwater mussel *Unio pictorum* (Paleoheterodonta, Unionoida). Involvement of acidic polysaccharides from glycoproteins in nacre mineralization. The FEBS journal. 2007;274:2933-45.