Shell matrix proteins of the clam, *Mya truncata:* Roles beyond shell formation through proteomic study

Jaison ARIVALAGAN[1,2], Benjamin MARIE[1], Victoria A SLEIGHT[3], Melody S CLARK[3], Sophie BERLAND[2] and Arul MARIE[1*]

[1]UMR 7245 CNRS/MNHN Molécules de Communications et Adaptations des Micro-organismes, Sorbonne Universités, Muséum National d'Histoire Naturelle, Paris – France

[2]UMR 7208 CNRS/MNHN/UPMC/IRD Biologie des Organismes Aquatiques et Ecosystèmes, Sorbonne Universités, Muséum National d'Histoire Naturelle, Paris - France

[3]British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge, CB3 0ET, UK

Abstract

Mya truncata, a soft shell clam, is presented as a new model to study biomineralization through a proteomics approach. In this study, the shell and mantle tissue were analysed in order to retrieve knowledge about the secretion of shell matrix proteins (SMPs). Out of 67 and 127 shell and mantle proteins respectively, 16 were found in both shell and mantle. Bioinformatic analysis of SMP sequences for domain prediction revealed the presence of several new domains such as fucolectin tachylectin-4 pentraxin-1 (FTP), scavenger receptor, alpha-2-macroglobulin (α2 M), lipocalin and myosin tail along with previously reported SMP domains such as chitinase, carbonic anhydrase, tyrosinase, sushi, chitin binding. Interestingly, these newly predicted domains are attributed with molecular functions other than biomineralization. These findings suggest that shells may not only act as protective armour from predatory action, but could also actively be related to other functions such as immunity. In this context, the roles of SMPs in biomineralization need to be looked in a new perspective.

1. Introduction

Calcium carbonate is the most abundant biogenic mineral in the earth's crust in terms of its quantity and production by different taxa. Organisms that belong to certain taxa use calcium carbonate to form thick exoskeletons, a physical barrier, which protects them from pathogens and predators. Bivalves are one among these taxa, which use organic components to nucleate, arrange and shape calcium carbonate crystals to form this exoskeleton (Lowenstam, Heinz Adolf and Weiner, 1989). In the early 19th century, this organic matter was referred to as conchiolin (Fremy, 1855). Later, investigations revealed the presence of glyco- and sclero- proteins with higher abundance of acidic amino acids in conchiolin (Gordon and Carriker, 1980). The microstructure analysis of the shells often revealed two different calcium carbonate crystals, calcite and aragonite, arranged in several distinct layers (Boggild, 1930). In shell, chitin contributes the framework at the biological hierarchy for calcification and moreover it dictates the orientation of the mature crystals (Ehrlich, 2010).

Even though the amount of proteins in the shells is < 5%, they play a key role in biomineralization and are found in the shell matrix space and are referred to as shell matrix proteins (SMPs). For the past 50 years, technical advancements have enabled the researchers to undertake detailed investigations to identify and characterize these SMPs. Various studies revealed that the SMPs are secreted and transported to the shell from the epithelial cells of the outer mantle (Addadi et al., 2006). The SMPs identified among several species of bivalves are Lustrin A, N16, perlucin, mucoperlin, perlustrin, MSI60, prismalin-14, MSI31, nacrein, caspartin, and recently many others (Marin, 2012). It has been demonstrated that these proteins help in nucleation, scaffold arrangement, growth and inhibition of calcium carbonate biominerals (Mann, 2001). However, the identified SMPs still accounts to a very less percentage of the total SMPs that might be present in the bivalve shells.

Historically, in China 'Mu li', shell powder from oysters was used as a medicine. Recent investigations using scallop shell powder have also showed their ability to inhibit viruses (Thammakarn et al., 2014). The soluble matrix protein from the shell of *Pinctada maxima* showed osteogenic potency on human mesenchymal stem cells (Green et al., 2015) and recent works have proved the potential for scallop powder to be used in dermo-cosmetic applications (Latire et al., 2014). These studies indicate that the proteins in the shell matrix might not solely be involved in the formation of hard biogenic crystal structures but might also exhibit novel functions other than biomineralization or indicate the possibility of discovering new proteins.

Mya truncata (common name- blunt gaper) is an arctic soft-shelled clam and it serves as the main source of food for seals, walruses (Fisher and Stewart, 1997) and also as an important protein source for some indigenous people in northern parts of Europe (Amap, 1998). *M.truncata* is a filter feeder and is used as a monitoring species for the hydrocarbon level in arctic oceans (Humphrey et al., 1987). Recently, it has been found that the population of *M.truncata* shows a steep decline due to climate change and in fact, Wildlife Action Plan places it among three species of greatest conservation need. It should be pointed out that *M.truncata* belongs to taxon Heteroconchia, which diverged early from a sister branch leading to the Pteriomorphia lineage around 513 million years ago (Plazzi and Passamonti, 2010). Most of the commercially important and most studied bivalve species falls under

Pteriomorphia e.g. *Mytilus* spp., *Crassostrea* spp. and *Pinctada* spp. But in Heteroconchia only *Panopea generosa* (common name: geoduck) and *Mya arenaria* (common name: sand gaper) are recognized for their economic interest. So far less than ten bivalve species have been studied by proteomics and this does not include any Heteroconchia species. In this work, we have investigated the shell proteomics of *M.truncata* to gain insight about the diverse possible functions of SMPs.

2. Materials and methods

2.1 Sample collection

M.truncata individuals thriving in shallow waters (less than 25 m depth) were diver collected from Oban, Scotland (56.4120° N, 5.4720° W). Three individuals of same size (9×4 cm) were used for proteomic experiments. The shells were cleaned with water to remove adhering matters and the mantle tissues and shells were separated. Shells were again rinsed with water and air-dried. Mantle tissues were flash frozen using liquid nitrogen.

2.2 Protein extraction and mass spectrometry analysis

Prior to the extraction of proteins from the shells, they were bathed in diluted sodium hypochlorite (NaOCl) (5-10%), to remove the periostracum and further washed with water and dried. To remove any other adhering organic impurities, the shells were polished using a dremel tool and were cut in to small pieces, powdered and graded with a mesh of pore size 250 μm. Decalcification was performed on the shell powder (3 g) using cold acetic acid (5% for 1 hr and 10% overnight). Acid soluble (ASM) and acid insoluble (AIM) matrices were recovered separately through centrifugation (14,000 rpm, 20 min, 4°C). AIM was washed 5 times with milli-Q water before freeze-drying. Similarly ASM was filtered via 10 KDa filter (Sartorius, VIVASPIN 20), washed with water and freeze-dried.

Flash frozen mantle tissue (250 mg) was crushed in a mortar and pestle along with a homogenization buffer (8 M urea, 100 mM tetraethyl ammonium bromide (TEAB) and 0.1% sodium dodecyl sulphate (SDS)). The resulting solution was centrifuged (15,000 rpm, 20 min, 4°C) to separate protein fractions from cellular debris and lipids. The supernatant was collected and filtered via 10 kDa filters (Sartorius, VIVASPIN 20) to enrich the proteins and the filtrate washed with milli-Q water. Finally, the filtrate was transferred and freeze-dried.

Initially the ASM, AIM and mantle samples were treated with 30 μL of 8 M urea for an hour at 37°C. Then, 100 μL of 10 mM (final conc) dithiothreitol (DTT) in 100 mM TEAB was added (1 h, 37°C), followed by 15 mM (final concentration) of iodoacetamide (1 h, 37°C). The required amount of trypsin was added to AIM (10 μg), mantle (10 μg), ASM (5 μg) and BSA (5 μg) for overnight digestion at 37°C. After digestion, quantification of the peptides was performed using the QuantiPro™ High Sensitivity Protein Assay Kit (Sigma-Aldrich, France) with bovine serum albumin (BSA) as a standard. Finally, an equal amount of peptides from ASM and AIM were pooled for further analysis using a liquid chromatography (LC) coupled to a mass spectrometer (MS).

Briefly, a nano LC system (Dionex Ultimate 3000, France) coupled to a LTQ Orbitrap XL (Thermo Fisher Scientific, France) mass spectrometer was operated in positive ion mode. Approximately 1 µg of peptide was concentrated on a 5 mm C_{18} trap column (5 µm, 100 A° pore, 300 µm i.d) and then the peptides were separated on a 50 cm nano-column packed with C_{18} phase (3 µm, 75µm i.d.) at a flow rate of 300 nL.min^{-1} using the following gradient: 1% solvent B (98% ACN, 0.1% formic acid) to 40 % B in 180 min, 40% B to 60 % B in 2 min, 60% B for 28 min. All the MS spectra were acquired on the Orbitrap and the MS/MS spectra in the linear ion trap. The survey scans were performed with the following parameters: 300 – 2000 *m/z*, resolution 30,000 AGC target 2x10^5, maximum injection time 100 ms. Twenty most intense precursors were selected for data dependent CID fragmentation scans with the following parameters: minimum intensity 500, isolation window 2 Da, normalized collision energy 35%, AGC target 5000, maximum injection time 100 ms. Dynamic exclusion was enabled (repeat count 1, duration 80 s).

2.3 Protein identification and analysis

Combined MS/MS spectra from three individuals were used to identify proteins from transcriptome database using the in-house version of MASCOT (version 2.1) and on-line research can be carried out at http://www.matrixscience.com. *M.truncata* transcriptome database was obtained from British Antarctic Survey (Sleight et al., 2016) and contains 20,106 contigs with an average read length of 675 bp. Database search parameters are: carbamido-methylation of cysteine (fixed modification) and oxidation of methionine and de-amidation (D, N) as variable modifications. Peptide mass error was 10 ppm and MS/MS tolerance was set to 0.5 Da. SCAFFOLD software was used to further process the results from Mascot. Protein identification was validated with at least two unique peptides (95% probability at both the peptide and protein levels). The contigs identified from both shell and mantle structures by Scaffold were analysed with the BLAST2GO tool, to predict their functions. The putative protein functions were further validated through identification of conserved domains by using the Simple Modular Architecture Research (SMART) tool (http://smart.embl.de/).

3. Results and discussion

We could identify 194 with at least two unique peptides (minimum peptide and protein threshold were set to 95% probability in SCAFFOLD). Of these, 127 proteins were derived from the mantle and 67 from the shell. SMPs are believed to be secreted by the epithelial cells of the mantle into the extrapallial space and then subsequently integrated to the shell matrix. However, we could only identify 16 proteins that are shared between the mantle and shell (Fig. 1 & table 1, refer to the supplementary tables for list of shell and mantle proteins). In fact, shell protein secretion is a spatio-temporally regulated phenomenon and this along with the analytical workflow employed in this work may partly explain the identification of few shared proteins between the shell and mantle tissue. The shell proteins are grouped into different classes such as biomineralization proteins (shell skeletal, enzymes), immunity related proteins, lipocalin and myosin tail domain containing proteins and proteins of unknown functions (which includes low complexity domain containing proteins) (Fig. 2). The grouping was based on the functions predicted by the domain prediction tool (SMART). The identification of previously reported shell proteins such as chitinase,

carbonic anhydrase, PIF and tyrosinase in this work validates our experimental protocol.

3.1 Biomineralization: Shell skeletal proteins

The SMPs with domains such as Von Willebrand factor A (VWA), sushi, chitin binding-2 and tenascin-R were grouped under shell skeletal. These domains are usually found in the extracellular matrix proteins and are involved in adhesion and protein-protein interaction. For example, the contig 1866 (supplementary table 1) show sequence similarity with PIF, an acidic matrix protein, which induces the formation of aragonite crystals in the shell matrix in *Pinctada fucata* (Suzuki et al., 2009). This protein is expected to contain at least one chitin binding domain and one VWA domain (Suzuki et al., 2009). However, we found that this contig shows high sequence similarity (e-value = 13.7 e^{-31}) to PIF albeit with three chitin binding domains. Similarly, contig 2027 (supplementary table 1) possessing a VWA domain showed high sequence similarity (e-value = 3.13 e^{-74}) with cartilage matrix protein. Interestingly in our data, we could also find fibrinogen domain containing protein (tenascin R partial, table 1) in both mantle and shell. Cartilage matrix protein and fibrinogen domain containing proteins are usually associated with calcium phosphate biomineralization in vertebrates and its presence in *M.truncata* might indicate that some of the domains are conserved both in molluscs and vertebrates supporting the hypothesis of a potentially ancestral common calcification mechanism (Miyamoto et al., 2002).

3.2 Biomineralization: Enzymes

Chitinase-3 found in the shell matrix of *M.truncata* and comprises two chitin-binding domains and a Glyco-18 domain. Previously, this was reported as an SMP in many bivalve species, such as *Crassostrea gigas* (Badariotti et al., 2011, 2007) *Mytilus galloprovincialis* (Weiss and Schönitzer, 2006) and *Pinctada fucata* (Suzuki et al., 2009). As expected, the sequence of chitinase-3 from *M.truncata* shows closer similarity to that found in the freshwater mussel *Hyriopsis cumingii*, a close relative of Heteroconchia species rather than to the phylogenetically distant marine bivalve Pteriomorphia species. Another enzyme frequently identified in the shell matrix of bivalves is carbonic anhydrase, which is involved in carbonate ion supply for shell building. In the case of *M.truncata*, both carbonic anhydrase II and III were found in the shell (supplementary table 1) and mantle (supplementary table 2). Another protein, tyrosinase like protein tyr-3 was identified, which has been shown to be associated in the formation of the periostracum in *Pinctada fucata* (Zhang et al., 2006). Interestingly, it was also identified in the calcitic prism layer of pearl oyster shells (Marie et al., 2012).

Three enzymes phosphoenol pyruvate carbokinase (PEPCK), enolase and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) identified in both shell and mantle (table 1) play important roles in carbohydrate metabolism. PEPCK catalyzes the formation of phosphoenol pyruvate (PEP) from oxaloacetate. Enolase indirectly catalyzes the formation of oxaloacetate from 2-phospho glycerate and vice versa. GAPDH catalyzes the formation of 1,3-bisphospho glycerate from glyceraldehyde 3-phosphate. All these enzymes are directly or indirectly involved in the formation of oxaloacetate. Interestingly, the presence of oxaloacetate was found to increase the

251 shell deposition rate in *Crassostrea virginica* (Wilbur and Jodrey, 1955), and we may
252 suppose that these enzymes could be involved in the shell growth process. However,
253 these three enzymes are predominantly intra-cellular and although their presence in
254 mantle proteome is consistent with the gluconeogenesis pathway, their presence in the
255 shell is quite enigmatic. One plausible explanation could be that these enzymes might
256 have reached the shell matrix space as a by-products of secretion processes or result
257 from cellular debris during shell formation and eventually embedded in the shell
258 (Mann et al., 2008).

259

260 The cupredoxin domain containing protein hephastin like isoform (Supplementary
261 table 1) was reported earlier in the coral *Acropora millepora*, bivalve *Crassostrea
262 gigas* and echinoderm *Strongylocentrotus pupuratus.* These copper dependent
263 ferroxidases are involved in copper and iron metabolism at the membranes. They also
264 catalyse the oxidation of Fe^{2+} to Fe^{3+} at the site of aragonite crystal precipitation
265 (Ramos-Silva et al., 2013). In fact, it should not be surprising to find proteins
266 containing cupredoxin domains in this study, since the shells of *M.truncata* are only
267 made of aragonite crystals.

268

269 *3.3 Immunomodulatory proteins*

270

271 We identified for the first time, a novel lectin family domain: Fucolectin tachylectin-4
272 pentraxin-1 (FTP) was found in the shell (Fig. 3 & supplementary table 1), which is
273 usually related in other organisms and tissues with the function of innate immunity.
274 Transcriptomic analysis of the haemocyte from the snail *Littorina littorea* revealed
275 ten unigenes with FTP domains either singly or associated with other domains. These
276 FTP domains play a potential role in the receptor phase of immune response
277 (Gorbushin and Borisova, 2015). In fact, two contigs 7764 and 9960 (Fig. 3) were
278 found in the *M.truncata* shell proteome, each containing a single FTP domain.
279 Previously, C-type lectins (like FTP, protein belonging to lectin binding protein
280 family) were reported in the spicule matrix of sea urchin (Juneja et al., 2014) and in
281 the shell matrix of mussel (Marie et al., 2011a), but their possible association with
282 immunity was not suggested at that time. We also identified from both the shell and
283 mantle, α2 Macroglobulin-containing proteins (table 1) that are often found in the
284 haemocytes of lower vertebrates and invertebrates and possess functions such as
285 innate immunity, proteinase binding (Armstrong and Quigley, 1999) and receptor
286 mediated endocytosis (Borth, 1992). Although, α2 M is multifunctional, its
287 predominant function is related to immunity. So possibly α2 M is one of the immune
288 proteins present in the shells. Similarly, protein with a scavenger receptor (SR)
289 domain (Fig 3) was found in the shell. SR domain has the function of innate immunity
290 and actively recognizes gram + and gram – bacteria, binds to the lipid cell wall and
291 dissolves it (Fabriek et al., 2009). Earlier, proteins containing this domain were
292 reported in the mantle, adductor, muscle and gonads of scallop *Chlamys farreri* (Liu
293 et al., 2011).

294

295 Although, peptides from the shell match to the contigs 9098 and 10609
296 (supplementary table 1) they do not show homology to any known proteins. For both
297 the contigs, SMART predicts the presence of a single kazal domain. Usually this
298 domain is indicative of serine protease inhibitors, which protect the proteins from
299 degradation by peptidases and proteinases. Usually proteinases are found in
300 haemocytes and are involved in immunity. The proteinase inhibitor produced by the

haemocytes pledged to the shell formation might also have been secreted and trapped in the shell matrix space. In *Pinctada* species, active protease inhibitors have been retrieved in the shell matrix (Bédouet et al., 2007). Previously kazal type domains were identified in the tooth matrix of sea urchin (Mann et al., 2008) and were also reported in shell matrix proteins of *Venerupis philippinarum,* the manila clam (Marie et al., 2011b). Proteins containing domains related to immune functions identified in the shell matrix of *M.truncata* support the idea of a role of the shell in protective mechanisms against infections.

3.4 Myosin-tail-domain proteins

Myosin tail-1 domain containing proteins (contigs 2653 and 11333) (table 1) were found both in the shell and mantle. While their presence in the mantle tissue is not surprising, their identification from the shell matrix is a matter of concern. Historically, myosin is always associated with actin (Jakus and Hall, 1947). Conversely, we could not find any traces of actin in our shell proteome data and reflects the quality of sample cleaning, as presence of actin is attributed to cellular contamination due to incomplete shell cleaning. The presence of actin and myosin in shells is much debated and intracellular proteins such as actins, tubulins and myosin identified from the shell matrix are described as contaminants (Marie et al., 2013). But, Jackson et al (Jackson et al., 2015) found higher concentrations of actin in the shells of *Magellania venosa* and suggested its possible role in biomineralization. Recently the presence of myosin tail proteins were reported in *Mytilus* shells (Liao et al., 2015). Some types of myosin, are not always associated with actins, as in the case of chitin synthase (identified in bivalve species), which contains a myosin domain (Weiss et al., 2006). It should be noted that myosin tails exhibit elastic properties due to their coiled-coil structure (Schwaiger et al., 2002). So, the myosin tail in the shell matrix along with silk fibroin-like proteins (proteins which sandwiches the chitin (Weiner et al., 1984)) might contribute to the mechanical elasticity of the shell.

3.5 Lipocalin-domain proteins

Two novel SPMs with lipocalin domains were found in this study (supplementary table 1) that has not yet been identified in any bivalve shells. Lipocalins are lipid binding proteins with many functions including blood coagulation, immune system modulation, regulation of homeostasis, neural system development, prostaglandin synthesis, complement fixation, regulation of inflammatory process (Greene et al., 2001) and coloration in crustaceans and insects (Wang et al., 2007). Interestingly, phospholipids were found in the prismatic and nacreous layers of *Pinctada* spp. (Farre and Dauphin, 2009; Farre et al., 2011). This identification of lipids in the shell reinforces the observation of lipid binding proteins, yet their biological significance in the shell matrix is unknown.

3.6 Novel Species-specific proteins

Out of 67 shell proteins that were identified (supplementary table 1), 34 proteins did not possess any known functional domains and signal peptides. These proteins are either specific to *M.truncata* or are yet to be discovered in other species. However, among these 34 proteins, 16 are rich in low complexity domains (LCDs). Many proteins involved in biomineralization are characterized by LCDs, which confer them

351 strong static conformations or flexible regions lacking well defined folding structures
352 (Coletta et al., 2010). The presence of many LCDs might also result in the formation
353 of novel protein coding sequences (Toll-Riera et al., 2012).
354
355 **4. Conclusion**
356
357 Shell matrix proteins are usually considered to be specifically involved in
358 biomineralization processes but the presence of many other classes of SMPs and their
359 corresponding domains brings into question the completeness of this assumption. The
360 SMPs containing domains such as immunomodulatory, lipocalin, myosin tail and
361 LCD rich (sections 3.3 to 3.6) might imply its role in different biological functions
362 and not only in biomineralization. Moreover, the occurrences of many LCD domains
363 in these proteins potentially provide the basis for adaptation and the evolution of
364 emerging novel proteins. We suggest that future work on SMPs should take into
365 consideration the functional diversity of SMPs in the context of shell formation.
366

References

Addadi, L., Joester, D., Nudelman, F., Weiner, S., 2006. Mollusk shell formation: A source of new concepts for understanding biomineralization processes. Chem. - A Eur. J. 12, 980–987. doi:10.1002/chem.200500980

Amap, 1998. AMAP assessment report. Arct. Pollut. Issues 859.

Armstrong, P.B., Quigley, J.P., 1999. ?? 2-macroglobulin: An evolutionarily conserved arm of the innate immune system. Dev. Comp. Immunol. 23, 375–390. doi:10.1016/S0145-305X(99)00018-X

Badariotti, F., Lelong, C., Dubos, M.P., Favrel, P., 2011. Identification of three singular glycosyl hydrolase family 18 members from the oyster Crassostrea gigas: Structural characterization, phylogenetic analysis and gene expression. Comp. Biochem. Physiol. - B Biochem. Mol. Biol. 158, 56–63. doi:10.1016/j.cbpb.2010.09.009

Badariotti, F., Thuau, R., Lelong, C., Dubos, M.P., Favrel, P., 2007. Characterization of an atypical family 18 chitinase from the oyster Crassostrea gigas: Evidence for a role in early development and immunity. Dev. Comp. Immunol. 31, 559–570. doi:10.1016/j.dci.2006.09.002

Bédouet, L., Duplat, D., Marie, A., Dubost, L., Berland, S., Rousseau, M., Milet, C., Lopez, E., 2007. Heterogeneity of proteinase inhibitors in the water-soluble organic matrix from the oyster nacre. Mar. Biotechnol. 9, 437–449. doi:10.1007/s10126-007-7120-y

Boggild, O.B., 1930. The shell structure of the mollusks. Det K. Danske Vidensk. Selsk. Skr. Natruvidenskabelig og Math. Afd. Ser. 9 2, 231–326.

Borth, W., 1992. Alpha 2-macroglobulin, a multifunctional binding protein with targeting characteristics. FASEB J. 6, 3345–3353.

Coletta, A., Pinney, J.W., Solís, D.Y.W., Marsh, J., Pettifer, S.R., Attwood, T.K., 2010. Low-complexity regions within protein sequences have position-dependent roles. BMC Syst. Biol. 4, 43. doi:10.1186/1752-0509-4-43

Ehrlich, H., 2010. Chitin and collagen as universal and alternative templates in biomineralization. Int. Geol. Rev. doi:10.1080/00206811003679521

Fabriek, B.O., Bruggen, R. Van, Deng, D.M., Ligtenberg, A.J.M., Nazmi, K., Schornagel, K., Vloet, R.P.M., Dijkstra, C.D., Van Den Berg, T.K., 2009. The macrophage scavenger receptor CD163 functions as an innate immune sensor for bacteria. Blood 113, 887–892. doi:10.1182/blood-2008-07-167064

Farre, B., Brunelle, A., Laprévote, O., Cuif, J.P., Williams, C.T., Dauphin, Y., 2011. Shell layers of the black-lip pearl oyster Pinctada margaritifera: Matching microstructure and composition. Comp. Biochem. Physiol. - B Biochem. Mol. Biol. 159, 131–139. doi:10.1016/j.cbpb.2011.03.001

Farre, B., Dauphin, Y., 2009. Lipids from the nacreous and prismatic layers of two Pteriomorpha Mollusc shells. Comp. Biochem. Physiol. Part B Biochem. Mol. Biol. 152, 103–109.

Fisher, K.I., Stewart, R.E. a, 1997. Summer foods of Atlantic walrus, Odobenus rosmarus rosmarus, in Northern Foxe Basin, Northwest Territories. Can. J. Zool. 75, 1116.

Fremy, E., 1855. Recherches chimiques sur les os. Mallet-Bachelier.

Gorbushin, A.M., Borisova, E. a., 2015. Lectin-like molecules in transcriptome of Littorina littorea hemocytes. Dev. Comp. Immunol. 48, 210–220. doi:10.1016/j.dci.2014.10.007

415 Gordon, J., Carriker, M.R., 1980. Sclerotized protein in the shell matrix of a bivalve mollusc. Mar.
416   Biol. 57, 251–260.

417 Green, D.W., Kwon, H.-J., Jung, H.-S., 2015. Osteogenic Potency of Nacre on Human Mesenchymal
418   Stem Cells. Mol. Cells 38, 267.

419 Greene, L.H., Chrysina, E.D., Irons, L.I., Papageorgiou, A.C., Acharya, K.R., Brew, K., 2001. Role of
420   conserved residues in structure and stability: tryptophans of human serum retinol-binding
421   protein, a model for the lipocalin superfamily. Protein Sci. 10, 2301–2316. doi:10.1110/ps.22901

422 Humphrey, B., Boehm, P., Hamilton, M., Norstrom, R., 1987. The fate of chemically dispersed and
423   untreated crude oil in Arctic benthic biota. Arctic 40, 149–161.

424 Jackson, D.J., Mann, K., Häussermann, V., Schilhabel, M.B., Lüter, C., Griesshaber, E., Schmahl, W.,
425   Wörheide, G., 2015. The Magellania venosa Biomineralizing Proteome: A Window into
426   Brachiopod Shell Evolution. Genome Biol. Evol. 7, 1349–1362.

427 Jakus, M.A., Hall, C.E., 1947. Studies of actin and myosin. J. Biol. Chem. 167, 705–714.

428 Juneja, P., Rao, A., Cölfen, H., Diederichs, K., Welte, W., 2014. Crystallization and preliminary X-ray
429   analysis of the C-type lectin domain of the spicule matrix protein SM50 from Strongylocentrotus
430   purpuratus. Acta Crystallogr. Sect. F Struct. Biol. Commun. 70, 260–262.

431 Latire, T., Legendre, F., Bigot, N., Carduner, L., Kellouche, S., Bouyoucef, M., Carreiras, F., Marin,
432   F., Lebel, J.M., Galéra, P., Serpentini, A., 2014. Shell extracts from the marine bivalve Pecten
433   maximus regulate the synthesis of extracellular matrix in primary cultured human skin
434   fibroblasts. PLoS One 9. doi:10.1371/journal.pone.0099931

435 Liao, Z., Bao, L., Fan, M., Gao, P., Wang, X., Qin, C., Li, X., 2015. In-depth proteomic analysis of
436   nacre, prism, and myostracum of Mytilus shell. J. Proteomics 122, 26–40.

437 Liu, L., Yang, J., Qiu, L., Wang, L., Zhang, H., Wang, M., Vinu, S.S., Song, L., 2011. A novel
438   scavenger receptor-cysteine-rich (SRCR) domain containing scavenger receptor identified from
439   mollusk mediated PAMP recognition and binding. Dev. Comp. Immunol. 35, 227–239.
440   doi:10.1016/j.dci.2010.09.010

441 Lowenstam, Heinz Adolf and Weiner, S., 1989. On biomineralization. Oxford University Press.

442 Mann, K., Poustka, A.J., Mann, M., 2008. In-depth, high-accuracy proteomics of sea urchin tooth
443   organic matrix. Proteome Sci. 6, 33. doi:10.1186/1477-5956-6-33

444 Mann, S., 2001. Biomineralization: Principles and Concepts in Bioinorganic Materials Chemistry,
445   Oxford chemistry masters.

446 Marie, B., Joubert, C., Tayalé, A., Zanella-Cléon, I., Belliard, C., Piquemal, D., Cochennec-Laureau,
447   N., Marin, F., Gueguen, Y., Montagnani, C., 2012. Different secretory repertoires control the
448   biomineralization processes of prism and nacre deposition of the pearl oyster shell. Proc. Natl.
449   Acad. Sci. U. S. A. 109, 20986–91. doi:10.1073/pnas.1210552109

450 Marie, B., Le Roy, N., Zanella-Cléon, I., Becchi, M., Marin, F., 2011a. Molecular evolution of mollusc
451   shell proteins: Insights from proteomic analysis of the edible mussel mytilus. J. Mol. Evol. 72,
452   531–546. doi:10.1007/s00239-011-9451-6

453 Marie, B., Ramos-Silva, P., Marin, F., Marie, A., 2013. Proteomics of CaCO3 biomineral-associated
454   proteins: How to properly address their analysis. Proteomics 13, 3109–3116.
455   doi:10.1002/pmic.201300162

456 Marie, B., Trinkler, N., Zanella-Cleon, I., Guichard, N., Becchi, M., Paillard, C., Marin, F., 2011b.
457   Proteomic Identification of Novel Proteins from the Calcifying Shell Matrix of the Manila Clam
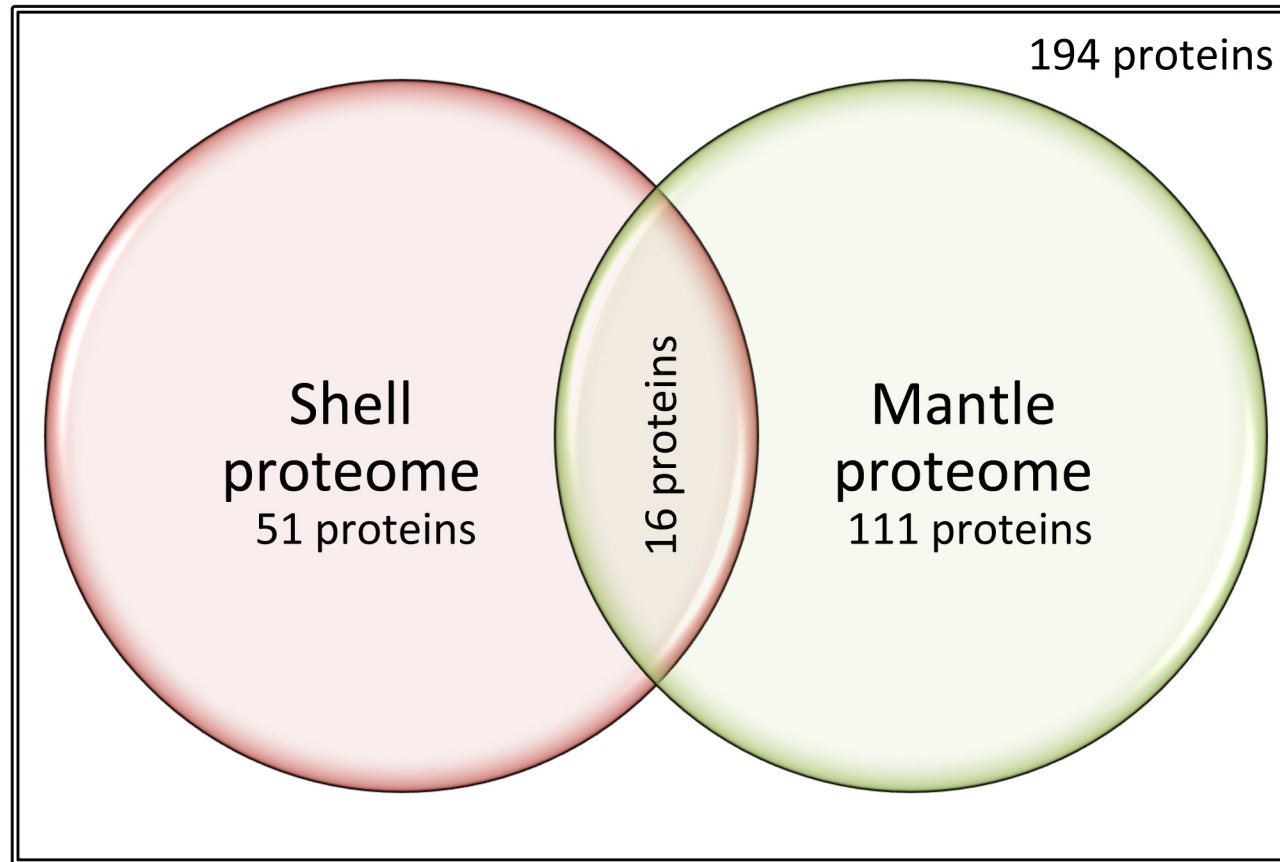458   Venerupis Philippinarum. Mar. Biotechnol. 13, 955–962. doi:10.1007/s10126-010-9357-0

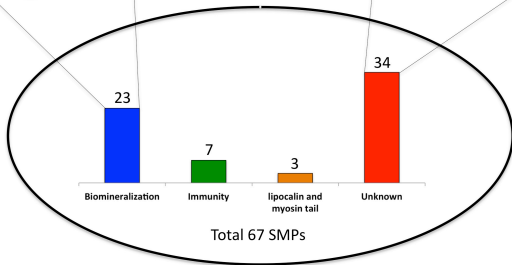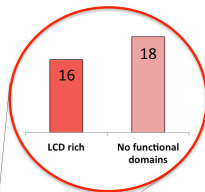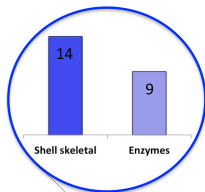459 Marin, F., 2012. The formation and mineralization of mollusk shell. Front. Biosci. doi:10.2741/S321

460 Miyamoto, H., Hamaguchi, M., Okoshi, K., 2002. Analysis of genes expressed in the mantle of oyster
461   Crassostrea gigas. Fish. Sci. 68, 651–658.

462 Plazzi, F., Passamonti, M., 2010. Towards a molecular phylogeny of Mollusks: Bivalves' early
463   evolution as revealed by mitochondrial genes. Mol. Phylogenet. Evol. 57, 641–657.
464   doi:10.1016/j.ympev.2010.08.032

465 Ramos-Silva, P., Kaandorp, J., Huisman, L., Marie, B., Zanella-Cléon, I., Guichard, N., Miller, D.J.,
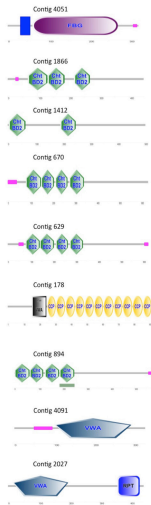466   Marin, F., 2013. The skeletal proteome of the coral acropora millepora: The evolution of

467   calcification by co-option and domain shuffling. Mol. Biol. Evol. 30, 2099–2112.
468   doi:10.1093/molbev/mst109

469   Schwaiger, I., Sattler, C., Hostetter, D.R., Rief, M., 2002. The myosin coiled-coil is a truly elastic
470   protein structure. Nat. Mater. 1, 232–235. doi:10.1038/nmat776

471   Sleight, V.A., Thorne, M.A.S., Peck, L.S., Arivalagan, J., Berland, S., Marie, A., Clark, M.S., 2016.
472   Characterisation of the mantle transcriptome and biomineralisation genes in the blunt-gaper clam,
473   Mya truncata. Mar. Genomics 1–9. doi:10.1016/j.margen.2016.01.003

474   Suzuki, M., Saruwatari, K., Kogure, T., Yamamoto, Y., Nishimura, T., Kato, T., Nagasawa, H., 2009.
475   An acidic matrix protein, Pif, is a key macromolecule for nacre formation. Science 325, 1388–
476   1390. doi:10.1126/science.1173793

477   Thammakarn, C., Satoh, K., Suguro, A., Hakim, H., Ruenphet, S., Takehara, K., 2014. Inactivation of
478   Avian Influenza Virus, Newcastle Disease Virus and Goose Parvovirus Using Solution of Nano-
479   Sized Scallop Shell Powder. J. Vet. Med. Sci. 76, 1277.

480   Toll-Riera, M., Radó-Trilla, N., Martys, F., Albá, M.M., 2012. Role of low-complexity sequences in
481   the formation of novel protein coding sequences. Mol. Biol. Evol. 29, 883–886.
482   doi:10.1093/molbev/msr263

483   Wang, M.R., Zhu, X.J., Yang, J.S., Dai, Z.M., Mahmood, K., Yang, F., Yang, W.J., 2007. Prawn
484   lipocalin: Characteristics and expressional pattern in subepidermal adipose tissue during
485   reproductive molting cycle. Comp. Biochem. Physiol. - B Biochem. Mol. Biol. 147, 222–229.
486   doi:10.1016/j.cbpb.2007.01.003

487   Weiner, S., Traub, W., Parker, S.B., 1984. Macromolecules in Mollusc Shells and Their Functions in
488   Biomineralization [and Discussion]. Philos. Trans. R. Soc. London. B, Biol. Sci. 304, 425–434.
489   doi:10.1098/rstb.1984.0036

490   Weiss, I.M., Schönitzer, V., 2006. The distribution of chitin in larval shells of the bivalve mollusk
491   Mytilus galloprovincialis. J. Struct. Biol. 153, 264–277. doi:10.1016/j.jsb.2005.11.006

492   Weiss, I.M., Schönitzer, V., Eichner, N., Sumper, M., 2006. The chitin synthase involved in marine
493   bivalve mollusk shell formation contains a myosin domain. FEBS Lett. 580, 1846–1852.
494   doi:10.1016/j.febslet.2006.02.044

495   Wilbur, K.M., Jodrey, L.H., 1955. Studies on shell formation. V. The inhibition of shell formation by
496   carbonic anhydrase inhibitors. Biol. Bull. 108, 359–365.

497   Zhang, C., Xie, L., Huang, J., Chen, L., Zhang, R., 2006. A novel putative tyrosinase involved in
498   periostracum formation from the pearl oyster (Pinctada fucata). Biochem. Biophys. Res.
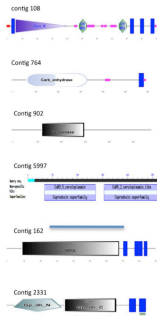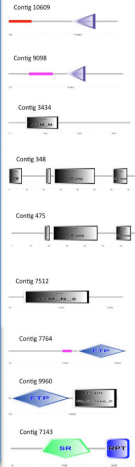499   Commun. 342, 632–639. doi:10.1016/j.bbrc.2006.01.182

500

194 proteins

Shell
proteome
51 proteins

16 proteins

Mantle
proteome
111 proteins

Shell skeletal: 14
Enzymes: 9

LCD rich: 16
No functional domains: 18

Biomineralization: 23
Immunity: 7
lipocalin and myosin tail: 3
Unknown: 34
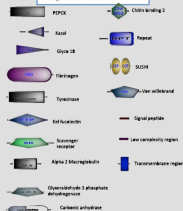
Total 67 SMPs

Shell matrix proteins

Figure 1. Venn diagram representation of the proteins identified in shell and mantle (three individuals pooled)

Figure 2. Grouping of the 67 shell matrix proteins (SMPs) identified in *M.truncata* according to their domain functions predicted by SMART. Proteins under biomineralization are further divided into shell skeletal proteins and enzymes; 16 proteins under "unknown" category contain Low Complexity Domains (LCDs)

Figure 3. Protein domain architecture of SMPs predicted by SMART.

Table 1. List of proteins shared between the mantle and shell of *M.truncata*. Columns three and four denotes the total number of matched peptides from the mantle and shell. Column five shows the homologous hit proteins obtained *via* BLAST.

Supplementary table 1. List of proteins identified in *M.truncata* shell

Supplementary table 2. List of proteins identified in *M.truncata* mantle