



Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the *Unionoida Elliptio complanata* and *Villosa lienosa*.

Benjamin Marie, Jaison Arivalagan, Lucrèce Mathéron, Gérard Bolbach, Sophie Berland, Arul Marie, Frédéric Marin

► To cite this version:

Benjamin Marie, Jaison Arivalagan, Lucrèce Mathéron, Gérard Bolbach, Sophie Berland, et al.. Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the *Unionoida Elliptio complanata* and *Villosa lienosa*.. *Journal of the Royal Society Interface*, 2017, 14 (126), pp.20160846. <10.1098/rsif.2016.0846>. <hal-01470764>

HAL Id: hal-01470764

<https://hal.science/hal-01470764v1>

Submitted on 27 Mar 2017

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.



HAL Authorization

Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the Unionoida *Elliptio complanata* and *Villosa lienosa*

Benjamin Marie^{1,*}, Jaison Arivalagan¹, Lucrèce Mathéron², Gérard Bolbach², Sophie Berland³, Arul Marie¹, Frédéric Marin⁴

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

²UMR 7203 CNRS/UPMC/ENS/INSERM Laboratoire des Biomolécules, Institut de Biologie Paris Seine, Université Pierre et Marie Curie, Paris – France

³UMR 7208 CNRS/MNHN/UPMC/IRD Biologie des Organismes Aquatiques et Ecosystèmes, Muséum National d'Histoire Naturelle, Paris - France

⁴UMR 6282 CNRS/uB Biogéosciences, Université de Bourgogne Franche-Comté (UB-FC), Dijon – France

*Author for correspondence

Benjamin Marie

Email: bmarie@mnhn.fr

Subject area: biomaterials, biochemistry, evolution

Keywords: biomineralization, shell nacre, proteomics, bivalve, organic matrix proteins, calcium carbonate.

24 The formation of the molluscan shell nacre is regulated to a large extent by a matrix of
25 extracellular macromolecules that are secreted by the shell forming tissue, the mantle. This so
26 called “calcifying matrix” is a complex mixture of proteins, glycoproteins and
27 polysaccharides that is assembled and occluded within the mineral phase during the
28 calcification process. The fine molecular basis that regulates nacre formation remains to be
29 better characterized.

30 Noticeable advances in expressed tag sequencing of the freshwater mussels, such as *Elliptio*
31 *complanata* and *Villosa lienosa*, provide a pre-requisite to further characterize the bivalve
32 nacre proteins by a proteomic approach. In this study, we have identified a total of 48
33 different proteins from the nacre insoluble matrices, 31 of which are common of the two
34 species. Few of these proteins, such as PIF, MSI60, CA, shematrin-like, Kunitz-like, LamG,
35 chitin-binding containing proteins, together with A-, D-, G-, M- and Q-rich proteins, appear to
36 be analogues, if not true homologs, of proteins previously described from the pearl oyster or
37 the edible mussel nacre matrices, thus constituting a remarkable list of deeply-conserved
38 nacre protein set. This work constitutes a comprehensive nacre proteomic study of non-
39 pteriomorphid bivalves that concretely gives us the opportunity to describe the molecular
40 basis of deeply conserved biomineralization toolkit among nacreous shell-bearing bivalves,
41 with regard to proteins associated to other shell microstructures, to that of other mollusc
42 classes (gastropods, cephalopods) and finally, to other lophotrochozoans (brachiopods).

43

1. Introduction

(a) Mollusc shell biomineralization

The molluscan shell is a unique model for biomineralization and constitutes a remarkable example of a natural composite biomaterial synthesized at ambient temperature from polar to tropical environment. It is composed of one or two polymorphs of calcium carbonate (*i.e.* calcite and aragonite), generally assembled in superimposed CaCO₃ layers of different microstructures [1]. These different shell layer textures are produced under the control of specific extracellular matrices that are secreted by different zones of the external mantle epithelium [2]. These ‘cell-free’ matrices, mainly constituted of proteins, glycoproteins and polysaccharides, surround nascent CaCO₃ crystals and are incorporated into the shell crystalline architecture in growth. Although the organic matrix is quantitatively a minor component of the shell (0.1 to 5 wt-%), it is thought to play a central role in the formation and shaping of these biomineral structures [3-4]. Indeed, these matrix components have been shown to be involved in the organization of a spatially prearranged scaffold, the nucleation, the orientation and growth of calcium carbonate crystals and the inhibition of their growth [5]. It has also been demonstrated that the nacre matrix controls the calcium carbonate polymorph [6] and that it exhibits cell-signalling properties [7-9].

(b) Nacre

Because of its extraordinary toughness, commercial value in jewelry and remarkable biocompatibility properties, nacre, so called the mother of pearl, is by far the most studied molluscan shell microstructure. Moreover, many authors consider it as the reference model for understanding how molluscs control the fine deposition of calcium carbonate layers [10]. From a structural viewpoint, nacre is constituted by a regular superimposition of polygonal flat aragonitic tablets of about 0.5 µm thickness, embedded in a peripheral thin organic matrix [11]. Nacre constitutes also a fascinating object for the investigation of evolutionary processes, as it appeared early in the fossil record, during the Cambrian times [12-13], and kept almost unchanged until now. At present it is distributed in four mollusc classes, comprising cephalopods, gastropods, monoplacophorans and bivalves. In this latter group, nacre is widespread in paleotaxodont, pteriomorphid, paleoheterodont and anomalodesmata clades [14]. One key question consists in determining if similar shell nacre textures, present in these different taxa, are produced by homologous sets of protein assemblages, or not. Such information would provide us a better appreciation on the recruitment of nacre matrix proteins

in the Cambrian, on the evolutionary constraints exerted on these proteins, and finally, on the open debate on the unique or multiple inventions of nacre among molluscs.

(c) Nacre proteins

Since the initial description of the primary structure of a mollusc shell protein in 1996 [15], numerous nacre associated-proteins have been described so far mainly from three principal models, the pearl oyster [2;16] and the edible mussel for bivalves [17-18], and the abalone for gastropods [19]. Comparison of their respective nacre biomineralization toolkits from the pearl mussel *Pinctada maxima* and the abalone *Haliotis asinina* underlined unexpectedly very few sequence similarities, suggesting either that the bivalves and gastropods acquired independently the capacity to synthesize nacre, or that these proteins present rapid evolution rate, especially in abalone, leading to functional conservation of respective nacre shell matrix protein (SMP) sets with only limited primary structure similarities [20]. In bivalves, most of the published data for nacre proteins comes from the Pteriomorpha sub-class [2;16-18]; only a limited amount of molecular data has been published so far on the other nacre-bearing bivalve groups [21-23]. To complete this patchy picture, we have developed a proteomic investigation on two nacreous bivalves that belong to the Palaeoheterodonta class, and to the Unionoida order, *Elliptio complanata* and *Villosa lienosa* [24]. Our molecular investigations are supported by the newly available large transcriptomic databases constructed in parallel from the two studied species [25-26], and are then providing strong support for the identification of SMP set of unionoid nacre. The present investigation gives us the opportunity to determine if bivalve nacles are made by molecular events that are controlled and conserved by the evolution processes, since the Phanerozoic times.

Our high-throughput high-output shotgun approach has allowed us to identify 48 different proteins from the nacre insoluble matrices, from which 31 are common to the two models. While recent data suggest a rapid evolution of shell matrix proteins leading to their extreme diversity [20], our work emphasizes on the contrary a deeply conserved biomineralization toolkit within nacre-bearing bivalves.

2. Results and Discussion

(a) Protein composition of unionoid nacre matrices

109 In order to investigate the nacre protein set in both *Elliptio complanata* and *Villosa lienosa*,
110 two closely related freshwater bivalves (Unionoida; Unionidae), the non-fractionated
111 insoluble matrices derived from respective shell nacre powder were analysed using a XL
112 Orbitrap nanoLC-MS/MS, after digestion with trypsin enzyme (Figure S1). For both samples,
113 the peak list generated from the MS/MS spectra was directly interrogated against the
114 unionoids EST database, comprising 195,192 nucleotide sequences using MASCOT
115 proteomic search engine (Version 2.1). Using this approach, we were able to enumerate 48
116 different proteins from the two nacre matrices (Figure 1; Supplementary table S1). More than
117 80% of these proteins (40 on the 48 identified in total) have been identified with high
118 confidence, because they are present in both species or are identified by more than one
119 peptide. Interestingly, 31 of the identified proteins have been detected in both *E. complanata*
120 and *V. lienosa* testifying of the similarities of the nacre matrix compositions of these two
121 species. We noticed that all conceptually-translated expressed sequence tag (EST) sequences
122 that match to the MS/MS peptides present a signal peptide, when sequence exhibits complete
123 N-terminus, suggesting that these proteins are all secreted by the mantle epithelia through
124 classical cellular secretion pathway. This list of nacre proteins represents among the largest of
125 nacre SMP sets described so far, comparable to published SMP set from *Pinctada* and *Mytilus*
126 models, which count 35 and 42 proteins, respectively [16-18]. While the present results
127 confirm the presence of previously described protein domains such as Pif, CA, Kunitz-like,
128 WAP, M-rich, Q-rich, A-rich, or chitin-binding domains [21;23-24], in unionoid shell nacre
129 matrix, they also contribute to consistently extend the list of proteins associated to nacre
130 microstructure.

131

132 The proteomic analysis of these two unionoid nacre matrices reveals a diversity of SMP
133 structures that can be broadly categorized into 5 main protein groups: ECM (ExtraCellular
134 Matrix), RLCD-containing (Repetitive-Low Complexity Domain), enzyme, immunity-related,
135 and 'uncharacterized'. Although our protein list may not be exhaustive - *i.e.* other minor
136 protein matrix components remain to be characterized - we assume that it contains most of the
137 abundant nacre SMPs: our approach indeed can identify most of the analogs/homologs of the
138 main nacre SMP described in pteriomorphid models (Supplementary table S1). Many of these
139 proteins, such as Pif, MSI60, CA, shematrin-like, Kunitz-like, LamG, chitin-binding
140 containing proteins, together with A-, D-, G-, M- and Q-rich proteins, appear to be analogues,
141 if not true homologs, of proteins previously described from the pearl oyster or the edible

142 mussel nacre matrices, and thus, constitute a remarkable set of deeply-conserved nacre
143 proteins. Indeed, although only genetic information on the gene synteny would reliably
144 support an hypothesis of homology relation between two proteins/genes of two organisms, we
145 assume because of sequence of high sequence identities and also because of the specificity of
146 some SMP domain architecture, that some of these proteins retrieved in these unionoid
147 organisms would be true homologs of those observed in the pteriomorphid models.

148

149 (b) Extracellular matrix-related (ECM) proteins

150 Among the different unionoid nacre proteins that exhibits ECM protein signature (comprising
151 collagen, fibronectin, laminin, chitin-binding and other protein- and polysaccharide-
152 interacting domains), we identified different Pif-like isoforms, various chitin-binding
153 containing proteins, a Ca-binding protein, a calcium-dependent (C-type) lectin, a mytilin-3
154 ortholog, together with previously unreported proteins for nacre such as a collagen-alpha-like,
155 a MSP130-like, DMP-like, and a ANF_receptor (Figure 2; Supplementary table S1).

156 *Pif-like and Pif-related proteins.* The unionoid nacre matrix exhibits different forms of Pif-
157 like proteins. Pif-177 (C7G0B5) was first described in the nacre of the pearl oyster [27], and
158 homologous forms, BMSP-220 and BMSP-like, were subsequently detected in the shell
159 matrices of the blue mussel *Mytilus galloprovincialis*, and of the giant limpet *Lottia gigantea*,
160 respectively [28-29]. We have identified here at least 2 different Pif-like isoforms, called Pif-
161 like 1 and 2, that exhibit both, from N- to C-terminus, the following succession of domains:
162 one VWA, three CBD2, one RLCD and one LamG. This domain arrangement is characteristic
163 of the Pif-177/BMSP-200/BMSP-like family [27-29]. Considering that the protein scores
164 obtained for the Mascot identification of Pif-like 1 is among the highest obtained here, this
165 supports the idea that the Pif/BSMP protein members are present in very high amount in the
166 nacre matrix and that Pif/Pif-like proteins must certainly play a key function in mollusc shell
167 formation [27-29].

168 Interestingly, three other proteins exhibit a domain organisation that partially resembles that
169 of Pif-like proteins: UNP-2 (1 VWA, 2 long repeats and 1 LamG domains), UNP-3 (1 VWA,
170 16 CCP and 3 CBD domains), and UNP-5 (2 CBD2 and 1 LamG domains). Furthermore,
171 various unionoid nacre proteins, called UNP-1, UNP-4, UNP-5, UNP-6 and UNP-7, bear
172 characteristic chitin- or polysaccharide-binding domains (CBD2, Chitin_binding_3 or lectin
173 C-type). The presence of these various VWA and chitin-binding domains suggest that these
174 proteins might participate in structuring the chitin/protein nacre scaffold. In addition, we

175 identified other ECM-related proteins, such as DMP-like, MSP130-like, mytilin3-like, ANF-
176 receptor containing or Ca-binding proteins, for which homologs have previously been
177 described from other proteins associated with biomineralizations [30-32]. The involvement of
178 these proteins in biomineralization processes is discussed below.

179

180 (c) Repeated low-complexity domain (RLCD) containing proteins

181 Our proteomic investigation also highlights the presence of a large set of 16 RLCD-containing
182 proteins (supplementary table S1) for which similar proteins (that are genuine orthologous for
183 some of them) were previously observed in *Pinctada* nacre, namely MSI60-like, shematin,
184 G-, Q-, Y-, D- and M-rich proteins [2]. In addition, we identified new features such as a
185 byssal protein-like, the homologs of which have been detected some years ago in the byssus
186 of the blue mussel [33].

187 *MSI60-like*. This protein fragment presents high sequence similarities with MSI60, presenting
188 characteristic repeats of G-rich and poly-A blocks (45% identity and E-value 3.1E-55), and
189 initially described as a framework protein in the shell of *Pinctada fucata* [34]. MSI60 is a
190 nacre-specific insoluble protein that exhibits poly-A and poly-G blocks giving MSI60 a
191 certain similarity with hydrophobic spider and worm silk fibroins. More recently, another
192 protein (has_CL10Contig2) with global similarities but no direct homology to MSI60
193 sequence has also been identified in the shell nacre of the abalone *Haliotis asinina* [19]. This
194 suggests once more that such G- and A-rich proteins with hydrophobic silk-fibroin-like
195 domains are key-players in nacre formation, in both bivalves and gastropods.

196 *Shematin-like*. Only one protein corresponding to a shematin-like protein has been identified
197 in the two unionoids according to one peptide that gives a heterologous match with the EST
198 sequence UTP_var00312 available from the brook floater *Alasmidonta vericosa*, another
199 unionoid bivalve. This protein presents strong sequence similarity (72% identity and E-value
200 5E-155) with silkmapin a nacre protein from the unionoid *Hyriopsis cumingii* [35], that
201 present sequence similarity with shematin proteins previously retrieved from *Pinctada* shell.
202 The overall primary structure of shematrins is characterized by a signal peptide, G-G-Y-repeat
203 motif, together with other more variable G- and Y-rich repeated low-complexity domains
204 (RLCDs) and KKKY N-termini [36], that are also observed in this novel unionoid putative
205 member. Shematin-likes constitute one of the main SMP family, found in both nacre and
206 prismatic layer of the shell of *Pinctada spp.*, and their corresponding transcripts are highly
207 expressed in the mantle [2;20]. Surprisingly, their homologous sequences lack in the TSA

208 databases of *E. complanata* and *V. lienosa*. We assume that the matching peptide observed in
209 the two unionoid models may correspond to a genuine shematin-like protein, similar to the
210 one derived from *A. vericosa* EST. The apparent absence of shematin-like transcripts in the
211 *E. complanata* and *V. lienosa* databases call for explanations at two levels: a) Restricted
212 temporal expression [37], outside the 'window' of tissue sampling. Indeed, previous
213 investigations have highlighted that key SMPs could be underrepresented or absent from a
214 mantle EST database for such a reason [29]. b) Spatial expression of the shematin transcripts
215 restricted to the mantle; their absence in the transcriptomes of *E. complanata* and *V. lienosa*
216 may be simply due to tissue sampling for EST construction, which was not particularly
217 focused on the mantle [25;26].

218

219 *Byssal protein-like, N-, T-, Q-, G- and M-rich proteins.*

220 Our proteomic investigation has detected in the shell nacre of *V. lienosa* a sequence, which
221 exhibits global similarities with a byssal protein-like from the mussel *Mytilus*
222 *galloprovincialis* (33% identity). This protein exhibits repeat domains that are enriched in G,
223 D, N and Y residues (Supplementary figure S2; supplementary table S1). Byssus is a silk-like
224 acellular thread that is secreted by the foot of various bivalves, and that firmly anchors the
225 organism to its substratum owing to strong adhesive distal plaques. A very recent proteomic
226 investigation [38] has shown that several proteins constituting the byssal acellular structures
227 (tyrosinase, A2M, EGF and VWA containing proteins) are common with those of the shell
228 layers, claiming for similarities of the formation and the evolution of these two structures.

229 In addition, mollusc matrix proteins rich in Q have also been found in bivalves: for example
230 MPN88 was previously characterized from the pearl oyster *Pinctada margaritifera* [39]. In
231 the shell of the gastropod *Lottia*, at least 4 different matrix proteins present similar features
232 [29]. Interestingly, vertebrate teeth contain various Q-rich proteins belonging to the secreted
233 calcium-binding phosphoprotein family, which are believed to interact with calcium ions and
234 regulate the tooth mineralization process [40]. In parallel, it has been proposed that the
235 numerous Q residues of various silk-like fibroins, which also present numerous A- and G-rich
236 domains, are implicated in the tensile mechanical properties of these proteins [41]. Few
237 proteins with M- and G-rich composition were identified among the RLCD-containing shell
238 proteins listed in this study. Proteins with similar enrichments in M and G residues have also
239 been recently described in the matrix of *Pinctada* [42], of *Haliotis* [19] and of *Lottia* [29].

240 However, the functional characterization of these characteristic domains in mollusc SMPs
241 remains to be done.

242 In the present case, in spite of overall sequence similarities observed within these several
243 RLCD protein categories identified in different models, it remains difficult to clearly outline
244 whether they are truly homolog considering limitation of standard Blastp tools for low-
245 complexity sequences, the fast rate of evolution of these domains [41;43] and that fortuitous
246 low-complexity sequence similarities can be obtained between non-homologous SMPs that
247 have evolved completely independently.

248

249 (d) Other bivalve shell-conserved proteins

250 *Carbonic anhydrase*. Carbonic anhydrase (CA) is also a ubiquitous enzyme that is essential
251 for calcium carbonate biomineral formation. It has been found in the organic matrices of
252 various metazoan skeletons [44-46]. In molluscs, shell-specific form of CA, exhibiting G, N
253 and D-rich LCDs, have been isolated first from the shell matrix of the pearl oysters *Pinctada*
254 sp. [15]. In our study, the CAs detected in the two unionoid nacre do not possess similar
255 supernumerary RLC domains, suggesting that matrix-associated CAs do not necessarily
256 require such domains for being occluded in calcium carbonate microstructure or for exhibiting
257 affinity with calcium carbonate mineral surfaces. At last, our results confirm the presence of
258 CA in a non-pteriomorph bivalve shell as previously suggested [21;23]. Taken together, our
259 findings reemphasize the central role played by extracellular CAs in CaCO₃ biomineralization
260 processes.

261 *Chitin-interacting enzymes*. Among the SMPs, we identified two enzymes, namely chitinase 3
262 and chitobiase, which both exhibit glyco_hydro_18 and glyco_hydro_20 domains. These
263 domains belong to the glycoside hydrolases family that hydrolyses the glycosidic bond
264 between carbohydrates or between a carbohydrate and other components. Indeed, chitin is
265 known to represent a major polysaccharide in mollusc shells. It is involved in constituting a 3-
266 dimensional framework, in association with silk-like and acidic proteins in which nacre tablet
267 crystals grow [10]. The identification of chitin-interacting enzymes, coupled with the
268 detection of several chitin-binding domains in bivalve nacre SMPs strongly suggests an
269 important role of chitin-cleaving enzymes in the fine molecular rearrangement of the chitin
270 framework following the secretion of the nacre macromolecular ingredients at the vicinity of
271 nascent growing mineral tablets.

272 *Tyrosinase*. The first mention of the involvement of a tyrosinase in bivalve shell formation

273 concerned the deposition of the periostracum and of the prismatic layer in the pearl oyster
274 [47] In particular, it was shown by *in situ* hybridization that the zone of the highest expression
275 of tyrosinase was the mantle edge from which these two layers are secreted. Recently, this
276 enzyme has also been detected in lesser amount in nacre shell structure [16]. Tyrosinase is an
277 oxidase that controls the production of protein melanogenesis *via* cross-linking by
278 hydroxylation of the phenol groups present in tyrosine residues and their conversion to
279 quinones. Biomineral-associated tyrosinases contribute to periostracum tanning [47]. They
280 may be also involved in the maturation and hardening of the biomineral-hydrogel (that
281 precedes crystallization) in prismatic as well as in nacreous shell layers [10].

282 *Thioredoxin (TRX) and superoxide dismutase (SOD)*. Two enzymes involved in antioxidant
283 processes, TRX and SOD, were also detected in the shell nacre of *V. lienosa*. Besides the
284 identification of various isoforms of peroxidases in *Lottia* shell matrix [29], there are to date
285 only a couple of examples of such redox enzymes in mollusc shell matrices [17;48]. In the
286 present study, TRX and SOD were detected only via a limited number of peptides, suggesting
287 that they may be minor component of the nacre matrix. Their putative function in shell
288 biomineralization remains enigmatic. We cannot exclude the possibility that they are
289 contaminants: SOD, for example, constitutes the main hemolymph protein, in charge of
290 oxygen transport in bivalves [49]; its accidental occlusion in shell calcified tissues is not an
291 unlikely scenario.

292 *Immunity-related proteins*. Finally, we detected, in both *Elliptio* and *Villosa*'s nacre matrices,
293 four proteins with potential immunity-related functions. They exhibit indeed Kunitz-like
294 (KU), or whey acidic protein (WAP) or macroglobulin protease inhibitor domains (Figure 1;
295 supplementary table S1). The presence of such protease inhibitors in an acellular biomineral
296 is, at first sight, puzzling. However our finding does not constitute a single observation, since
297 the presence of such immunity-associated domains in shell proteins has been already noticed
298 in different metazoan calcium carbonate biominerals: the shells of abalones [19;50], of pearl
299 oysters [2;51-52], of the giant limpet [29] and, at last, the chicken egg shells [46]. One has
300 proposed that these molecules are involved either in immunological defence mechanisms, or
301 in the regulation of the matrix maturation, by controlling cleavage processes of other matrix
302 proteins [53-54].

303

304 (e) Inherited set of nacre and shell matrix proteins in bivalves

305 One of the most striking results of our study was the detection, in *Elliptio* and *Villosa*'s shell
306 matrices, of various proteins that exhibit high sequence homologies with SMPs extracted from
307 the nacreous shell layers of *Pinctada* and/or *Mytilus* bivalves. This constitutes a very
308 consistent evidence of the existence of a set of ancient, *i.e.* Palaeozoic, nacre proteins shared
309 between the various nacre-forming bivalve taxa (Figure 3).

310 More than two decades ago, when molluscan nacre matrices were solely analyzed for their
311 amino acid compositions after hydrolysis of the bulk, the similarity of their compositions
312 characteristically marked by high contents in glycine, alanine and aspartate residues,
313 suggested identical molecular mechanisms for controlling the deposition of nacre tablets in
314 the different nacre-bearing molluscs [55]. But, since the publication of the first full primary
315 structure of a nacre protein, in 1996 [15], the increasing number of discovered SMPs has point
316 out an extraordinary diversity of nacre-associated proteins among the different taxa [8;9]. This
317 suggested on the contrary the use of different protein toolkits for nacre deposition. This view
318 seemed to prevail, in particular with the finding that the bivalve *Pinctada* and the gastropod
319 *Haliotis* possess absolutely dissimilar sets of shell-building genes [19;20]. This abrupt
320 conclusion was however counterbalanced later by one of our studies that clearly demonstrated
321 that the nacre of the edible mussel *Mytilus* and the non-nacreous shell microstructure of the
322 limpet *Lottia*, contains number of proteins that are similar - if not homologous - to that of
323 *Pinctada*'s nacre [2;29;30]. The present set of data supports and reinforces the "deep origin
324 scenario" by revealing a shared protein toolkit in the nacre of bivalve species that belong to
325 three clades, Mytiloida, Pterioda (both belonging to Pteriomorphia subclass) and Unionoida
326 (the main order of the Palaeoheterodonta subclass) that diverged in the Palaeozoic times [56].

327 The list of shared nacre proteins between the three main bivalve models comprises: MSI60,
328 Pif, CA, macroglobulin, VWA-, LamG-, CBD2- and Kunitz-like-containing proteins, along
329 with various RLCD protein exhibiting A-, D-, G-, Q- and M-rich domains. This short list
330 potentially represents the inherited ancestral toolkit - or at least a part of it - required for nacre
331 formation. Other proteins appear to be shared by only two species out of three, and are
332 probably related with taxon specificity adaptations. One has to notice that some of the
333 proteins listed on figure 3 have been also identified in the prismatic layers of the three
334 bivalves, suggesting that their molecular functions might not be exclusively restricted to nacre
335 deposition.

336 By identifying nacre proteins from two non-pteriomorph species and reporting a deeply
337 conserved nacre protein set in bivalves this work adds new evidence that that the calcifying

338 matrix of molluscs is comprised of both deep conservations and lineage-specific protein
339 novelties. The comparison of these bivalve-conserved protein sets to those previously
340 detected in gastropod nacre is confirming the global dissimilarity of the nacre SMP sets of
341 bivalves and of gastropods (at least those of the abalone) [19]. Bivalve and gastropod exhibit
342 different kinds of nacles (respectively, sheet and columnar nacles), which homology still
343 constitute an actual debate that still oppose parcimony principles [14], crystallographic and
344 microstructural textures [11;57], along with paleontological and molecular evidences
345 [12;13;20].

346

347 Figure 4 summarizes all available information on the occurrence of protein domains observed
348 from shell proteome of various mollusc models, comprising nacre and prismatic shell layers
349 of bivalves (Unionoids, *Pinctada* and *Mytilus*), cross-lamellar layers of gastropod (*Lottia* and
350 *Cepaea*), and finally, fibrous layers of brachiopods, taken as an outgroup. This figure shows
351 that several proteinaceous domains, which appear to be shared between the different nacre
352 models, are also detected in bivalve prismatic layers or in the shell of cross-lamellar
353 gastropods. This clearly suggests that these domains are not specifically involved in nacre
354 formation, but that they display general function in shell formation.

355 The brachiopod shell proteomes, retrieved from the CaCO_3 shell of *Magellania venosa* [58]
356 and the CaPO_3 one of *Lingula anatina* [59], respectively, exhibit many similarities with those
357 of molluscs in their respective protein and functional domain compositions. The list of shared
358 domains comprises potentially key contributors of shell formation processes, such as D-, A-
359 and G-rich, VWA, EGF, CCP, peroxidase, tyrosinase, chitinase, CA, KU, A2M and CBD2.
360 Furthermore, Luo and co-workers [59] notice also that genes shared between *Lingula* and
361 molluscs, such as calcium-dependent protein kinase and chitin synthase, exhibit high
362 expression in larvae and mantle, indicating that they may also be involved in brachiopod shell
363 formation. Taken together, these results indicate that a conserved molecular machinery exists
364 for shell biomineralization in brachiopods and molluscs. This assumption needs however a
365 further validation via functional analyses.

366

367 (h) Are metazoan SMPs deeply conserved?

368 The appealing scenario of an ancient and common origin of the biomineralization within the
369 metazoan has been proposed at various occasions [3;60]. This scenario is based on the
370 supposed existence of a shared ancestral metazoan toolkit that was implemented before the

371 main metazoan radiations. But to date, the data that support this view still remain extremely
372 tenuous. For example, no one ever evidenced a universal molecular marker of the skeletal
373 matrices of all calcifying metazoans. Many skeletal matrix proteins belong to families that are
374 not necessarily specific of biomineralization processes, suggesting that many proteins could
375 have been coopted secondarily for calcification purpose. For example, although CA has been
376 observed in various biomineralization models, comprising egg-shell, echinoderm skeleton,
377 mollusc shell or sponge spicules, it constitutes a broad and ubiquitous protein family
378 presenting various members in all organisms, calcifying or not; their phylogenetic analysis
379 shows a very complex evolutionary history, indicating that they could have been recruited
380 several times, in independent manners, in biomineralization structure within the different
381 metazoan clades [61].

382 At the opposite of the CA example, MSP-130 represents a remarkable illustration of a
383 conserved molecular function that may be strictly restricted to biocalcification, but present in
384 distinct metazoan lineages. MSP-130 is a protein, initially identified in the sea urchin
385 *Strongylocentrotus purpuratus* [62], which expression is specific to primary mesenchyme
386 cells [63] - the cells that initiate the skeletal tissues in echinoderms - and that represents one
387 of the main constituent of the calcifying matrix of the different skeletal structures, including
388 the teeth, the spicules, the test or the spines [64]. So far, no other function than those related
389 to biomineralization has been described for this protein. Interestingly, during our previous
390 investigation of the shell proteins of the Pacific oyster *Crassostrea gigas*, we detected a
391 protein, called gigasin-3, which partial sequence exhibits noticeable similarity with *S.*
392 *purpuratus* MSP-130 [31]. Even more recently, the msp130 gene was also identified in a
393 calcified tube-forming polychaete [65]. With the present study, the recognition in nacre
394 matrix of two unionoids of a similar MSP-130-like protein firmly supports the idea that this
395 biomineralization-specific protein or protein family is shared between different metazoan
396 calcifying lineages, reintroducing the scenario of common ancestral mechanism for metazoan
397 biomineralization processes.

398 In addition, we have detected that the two unionoid proteomes contain both a dentin-matrix
399 protein-like (DMP-like), exhibiting a remarkable FAM20C domain (Supplementary figure
400 S3). Indeed, FAM20C is a kinase that phosphorylates the majority of the phosphoproteome in
401 the extracellular matrices of bones and teeth, constituting a key player of vertebrate
402 biomineralization process [66]. Other recent investigations have revealed the presence of
403 FAM20C containing protein in the shell matrix of the giant limpet *Lottia gigantea* [48], and

404 of the brachiopod *Magellania venosa* [58]. The presence of such key-protein, regulating the
405 biomineralization formation of vertebrates, in the shell matrices of molluscs and brachiopods
406 strongly reinforces the idea of common molecular machinery, supporting the scenario of a
407 conserved biomineralization toolkit between different metazoan lineages.

408

409

410 **3. Conclusion**

411 Our study presents proteomic results obtained from the nacre matrices of two unionoid
412 (Paleoheterodonta) bivalves. Our approach is robust and powerful since it allows annotating
413 genes encoding proteins that are undoubtedly shell matrix constituents, therefore, putatively
414 important players in the nacre assembling process. It excludes genes that encode mantle-
415 secreted proteins – recognizable by their signal peptide signature – but which are not occluded
416 within the calcified tissue, *i.e.* proteins, which are not constituents of the shell organic matrix.
417 We are fully conscious that some of these proteins, not detected by our approach, may be
418 important players for calcification as well [67].

419 Our study has macro-evolutionary implications. Nacre is supposed to represent an ancestral
420 shell microstructure, shared by different orders within the bivalve class. Because it remained
421 almost unchanged until now, from a microstructural viewpoint, this suggests that at least a
422 part of the molecular toolkit used for constructing nacre must have been kept conserved
423 within these nacre-forming clades. Our proteomic approach focused on the unionoid bivalves
424 has allowed detecting several genuine sequence homologies with known sequences of nacre
425 proteins of pteriomorph bivalves, but also with other SMPs from gastropod and brachiopod
426 models, unveiling a part of a deeply conserved shared legacy.

427

428

429 **4. Material and methods**

430 (a) Nacre matrix extraction

431 Fresh *Villosa lienosa* (6-8 cm in length) and *Elliptio complanata* (8-10 cm in length) shells
432 were collected from Chewacla Creek near Auburn, Alabama (USA), in April 2012, and from
433 Cree Lake near Cochrane, Ontario (Canada), in spring 2011, respectively. Among the
434 paleoheterodont bivalves, both are representatives of unionoid order, the shell microstructures
435 of which are characterized by the association of aragonitic prisms and nacre. Superficial
436 organic contaminants as well as the periostracum were removed by incubating intact shells in

437 NaOCl (5%, v/v) for 24-48 h [68]. Shells were then thoroughly rinsed with water. The
438 external prismatic layer was mechanically removed and the shell nacre broken into 1-mm
439 large fragments before being ground to fine powder ($> 200\ \mu\text{m}$) that was subsequently
440 decalcified in cold acetic acid overnight (10%; $4\ ^\circ\text{C}$). The insoluble nacre matrix was then
441 collected by centrifugation (15,000 g; 10 min; $4\ ^\circ\text{C}$), and the pellet rinsed six times with
442 MilliQ water by series of resuspension-centrifugation before being freeze-dried and weighed.

443

444 (b) Peptide digest proteomic analyses

445 In-solution digestions of insoluble nacre matrix samples were performed. One mg of the
446 matrices was reduced with 100 μL of 10 mM dithiothreitol (Sigma-Aldrich, France) in 500
447 mM triethylammonium bicarbonate pH 8.0 for 30 min at 57°C . Fifteen μL of iodoacetamine
448 (50 mM, final concentration) was added and alkylation was performed for 30 min at room
449 temperature in the dark, then digestion was performed by adding 10 μg of trypsin (T6567,
450 proteomics grade, Sigma). The samples were incubated overnight at $37\ ^\circ\text{C}$, then centrifuged
451 (30 min, 14,000 g) and 5 μL of the acidified supernatants was injected into the tandem mass
452 spectrometer with an electrospray source coupled to a liquid chromatography system analysis
453 (nanoLC-ESI-MS/MS). The protein digest concentration was prior estimated using a micro-
454 BCA kit (Sigma-Aldrich, USA), with BSA as protein standard.

455 Protein digests were analyzed using a nano-LC (Ultimate 3000, Dionex) coupled to an ESI-
456 LTQ-Orbitrap (LTQ Orbitrap XL, Thermo Scientific) mass spectrometer [69]. Above 10 μg
457 of AIM digest protein digests solubilized in 10% ACN with 0.1% formic acid were injected in
458 triplicates by the autosampler and concentrated on a trapping column (Pepmap, C_{18} , $300\ \mu\text{m} \times$
459 $5\ \text{mm}$, $5\ \mu\text{m}$ $100\ \text{\AA}$, Dionex) with water containing 10% ACN with 0.1% formic acid (solvent
460 A). After 5 min, the peptides eluted onto a separation column (Pepmap, C_{18} , $75\ \mu\text{m} \times$ 500
461 mm , $2\ \mu\text{m}$ $100\ \text{\AA}$, Dionex) equilibrated with solvent A. Peptides were separated with a 2h-
462 linear gradient of 80% ACN + 0.1% formic acid (solvent B) at flow rate of $200\ \text{nL}\cdot\text{min}^{-1}$. The
463 LTQ-Orbitrap mass spectrometer is outfitted with a nano-ESI electrospray source. Spectra of
464 multiple charged precursor ions (+2 to +4) were collected from m/z 300-2000 at a resolution
465 of 30,000 in the profile mode followed by data dependent CID and/or HCD spectra of the ten
466 most intense ions. A dynamic exclusion window of 60 s was used to discriminate against
467 previously analyzed ions.

468

469 (c) Bioinformatics

470 The MS/MS data were used for database searches using an in house version of the MASCOT
471 search engine (Matrix Science, London, UK; version 2.1) against 138,345 and 46,051 TSA
472 nucleotide sequences derived from *E. complanata* and *V. lienosa* EST libraries [25;26],
473 respectively (plus few hundreds *Alasmidonta* spp. sequences), downloaded in December 2015
474 from the NCBI server (<http://www.ncbi.nlm.nih.gov>). LC-MS/MS data generated by each
475 shell sample and protein band were searched separately, using carbamido-methylation as a
476 fixed modification and methionine oxidation as variable modification. The peptide MS
477 tolerance was set to 10 ppm and the MS/MS tolerance was set to 20 mDa. Proteins with at
478 least two different peptides or identified in both species were considered to be fully valid.
479 Protein identification was performed using BLAST searches against the UniProtKB/Swiss-
480 Prot protein sequence database (www.uniprot.org), retrieving only hit with e-value < 7E-8,
481 and without filtering low complexity region. Signal peptides were predicted using SignalP 4.0
482 (<http://www.cbs.dtu.dk/services/SignalP>) and conserved domains from database models were
483 predicted using external source database SMART (<http://smart.embl-heidelberg.de>). 3-D
484 protein structures were predicted according to sequence homologies using the Phyre 2.0
485 online tool (<http://www.sbg.bio.ic.ac.uk/phyre2>) [70].

486

487

488 **Data accessibility.** Supplementary table S1 containing the complete list and description of all
489 proteins identified in *E. complanata* and *V. Lienosa* nacre matrices by shotgun proteomics has
490 been uploaded as electronic supplementary materials.

491

492 **Abbreviations:** A2M, alpha-2 macroglobulin; APase, alkaline phosphatase; ANF,
493 extracellular ligand binding; BMSP, blue mussel shell protein; C1q, complement complex
494 protein 1; CA, carbonic anhydrase; CBD, peritrophin-A chitin-binding domain; CCP/SUSHI,
495 complement control protein; Chit_bind, chitin-binding domains; CLECT, C-type lectin;
496 Copamox, copper amine oxidase; DMP, dentin matrix protein; EGF, epidermal growth factor;
497 FAM20C, family with sequence similarity 20 - member C; FN3, fibronectin 3; IGF-BP,
498 insulin growth factor-binding protein; KU, Kunitz-like domain; LamG, laminin G; MSP,
499 mesenchym-specific protein; MSI60, matrix silk-like insoluble protein 60; PIF/Pif177,
500 *Pinctada fucata* 177-kDa protein, RLCD, repetitive low-complexity domain; SMP, shell
501 matrix protein; SCP, Cysteine-rich secretory protein; SOD, superoxide dismutase; TIMP,
502 tissue inhibitor of metalloprotease; TRX, thioredoxin; UNP, uncharacterized nacre protein;
503 VWA, von Willebrand factor type A; WAP, whey acidic protein; ZP, zona pellucida.

504

505 **Author's contribution.** B.M., G.B., S.B. & F.M. conception, design and coordination; J.A.,
506 A.M., L.M. preparation of the sample, analysis and interpretation of the data. All author
507 contribute to the preparation and the revision of the manuscript.

508

509 **Competing interests.** The authors have no competing financial interests.

510

511 **Funding.** Jaison Arivalagan is supported by a funding from the Marie Curie European Union
512 Seventh Framework Programme under grant agreement no. 605051.

513

514 **Acknowledgments.** We are especially grateful to Dr. James Stoeckel and Dr. Emily Awad for
515 kindly providing us the *Villosa lienosa* and *Elliptio complanata* shells.

516

517

518 **References**

- 519 [1] Taylor J, Kennedy WJ, Hall A. 1973 The shell structure and mineralogy of the Bivalvia,
520 Part II, Lucinacea–Clavagellacea. *Bull Brit Mus Zool* **22**, 253–294.
- 521 [2] Marie B, Joubert C, Tayale A, Zanella-Cléon I, Belliard C, Cochenec-Loreau N,
522 Piquemal D, Marin F, Gueguen Y, Montagnani C. 2012 Different secretory repertoires
523 control the biomineralization processes of prism and nacre deposition of the pearl oyster
524 shell. *Proc Natl Acad Sci* **109**, 20986-20991.
- 525 [3] Lowenstam HA, S. Weiner S. 1989 On Biomineralization. Oxford University Press, New
526 York.
- 527 [4] Suzuki M, Nagasawa H. 2013 Mollusk shell structures and their formation mechanism.
528 *Can J Zool* **91**, 349-366.
- 529 [5] Mann S. 2001 Biomineralization: Principles and Concepts, in *Bioinorganic Materials*
530 *Chemistry*. Oxford University Press, New York.
- 531 [6] Falini G, Albeck S, Weiner S, Addadi L. 1996 Control of aragonite and calcite
532 polymorphism by mollusk shell macromolecules. *Science* **271**, 67–69.
- 533 [7] Berland S, Delattre O, Borzeix S, Catonne Y, Lopez E. 2005 Nacre/bone interface changes
534 in durable nacre endosseous implants in sheep. *Biomaterials* **26**, 2767-2773.
- 535 [8] Marin F, Luquet G, Marie B, Medakovic D. 2008 Molluscan shell proteins: primary
536 structure, origin, and evolution. *Curr Top Dev Biol* **80**, 209–276.

- 537 [9] Marin F, Le Roy N, Marie B. 2012 The formation and mineralization of mollusc shell.
538 *Front Biosci* **4**, 1099–1125.
- 539 [10] Addadi L, Joester D, Nudelman F, Weiner S. 2006 Mollusk shell formation: a source of
540 new concepts for understanding biomineralization processes. *Chem Rev* **12**, 980–987.
- 541 [11] Checa A, Rodriguez-Navarro A. 2001 Geometrical and crystallographic constraints
542 determine the self-organisation of the shell microstructures in Unionidae (Bivalvia:
543 Mollusca). *Proc Biol Sci* **268**, 771–778.
- 544 [12] Kouchinsky A. 2000 Shell microstructures in Early Cambrian molluscs. *Acta Palaeontol*
545 *Pol* **45**, 119–150.
- 546 [13] Vendrasco MJ, Porter SM, Kouchinsky A, Li G, Fernandez CZ. 2010 New data on
547 molluscs and their shell microstructures from the Middle Cambrian Gowers Formation,
548 Australia. *Palaeontology* **53**, 97–135.
- 549 [14] Carter JG. 1990 Skeletal Biomineralization: Patterns, Processes and Evolutionary Trends.
550 Van Nostrand Reinhold, New York.
- 551 [15] Miyamoto H, Miyashita T, Okushima M, Nakano S, Morita T, Matsushiro A. 1996 A
552 carbonic anhydrase from the nacreous layer in oyster pearls. *Proc Natl Acad Sci* **93**, 9657–
553 9660.
- 554 [16] Liu C, Li S, Kong J, Liu Y, Wang T, Xie L, Zhang R. 2015 In-depth proteomic analysis
555 of shell proteins of *Pinctada fucata*. *Sci Rep* **5**, 17269.
- 556 [17] Gao P, Liao Z, Wang X, Bao L, Fan M, Li X, Wu C, Xia S. 2015 Layer-by-layer
557 proteomic analysis of *Mytilus galloprovincialis* shell. *PLoS ONE* **10**:e0133913.
- 558 [18] Liao Z, Bao L, Fan M, Gao P, Wang X, Qin C, Li X. 2015 In-depth proteomic analysis
559 of nacre, prism, and myostracum of *Mytilus* shell. *J Proteomics* **122**, 26–40.
- 560 [19] Marie B, Marie A, Jackson DJ, Dubost L, Degnan BM, Milet C, Marin F. 2010
561 Proteomic analysis of the organic matrix of the abalone *Haliotis asinina* calcified shell.
562 *Proteome Sci* **8**, 54.
- 563 [20] Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R,
564 Rokhsar DS, Montagnani C, Joubert C, Piquemal D, Degnan BM. 2010 Parallel evolution
565 of nacre building gene sets in molluscs. *Mol Biol Evol* **27**, 591–608.
- 566 [21] Marie B, Zanella-Cléon I, Le Roy N, Becchi M, Luquet G, Marin F. 2010 Proteomic
567 analysis of the acid-soluble nacre matrix of the bivalve *Unio pictorum*: detection of novel
568 carbonic anhydrase and putative protease inhibitor proteins. *ChemBioChem* **11**, 2138–2147.
- 569 [22] Ramos-Silva P, Benhamada S, Le Roy N, Marie B, Guichard N, Zanella-Cléon I,

570 Plasseraud L, Corneillat M, Alcaraz G, Kaandorp J, Marin F. 2012 Novel molluscan
571 biomineralization proteins retrieved from proteomics: a case study with Upsalin.
572 *Chembiochem* **13**, 1067-1078.

573 [23] Berland S, Ma Y, Marie A, Andrieu J-P, Bédouet L, Feng Q. 2013 Proteomic and profile
574 analysis of the proteins laced with aragonite and vaterite mussel *Hyriopsis cumingii* shell
575 biominerals. *Prot Peptide letters* **20**, 1170-1180.

576 [24] Marie B, Aviralagan J, Dubost L, Berland S, Marie A, Marin F. 2015 Unveiling the
577 evolution of bivalve nacre proteins by shell proteomics of Unionoids. *Key Eng Mater* **672**,
578 158-167.

579 [25] Wang R, Li C, Stoeckel J, Moyer G, Liu Z, Peatman E. 2012 Rapid development of
580 molecular resources for a freshwater mussel, *Villosa lienosa* (Bivalvia: Unionidae), using
581 an RNA-seq-based approach. *Freshw Sci* **31**, 695–708.

582 [26] Cornman RS, Robertson LS, Galbraith H, Blakeslee C. 2014 Transcriptomic analysis of
583 the mussel *Elliptio complanata* identifies candidate stress-response genes and an
584 abundance of novel or noncoding transcripts. *PLoS One* **9**:e112420.

585 [27] Suzuki M, Saruwatari K, Kogure T, Yamamoto Y, Nishimura T, Kato T, Nagasawa H.
586 2009 An acidic matrix protein, Pif, is a key macromolecule for nacre formation. *Science*
587 **325**, 1388–90.

588 [28] Suzuki M, Iwasima A, Tsutsui N, Ohira T, Kogure T, Nagasawa H. 2011 Identification
589 and characterisation of a calcium carbonate-binding protein, blue mussel shell protein
590 (BMSP), from the nacreous layer. *ChemBioChem* **16**, 278–287.

591 [29] Marie B, Jackson DJ, Ramos-Silva P, Zanella-Cléon I, Guichard N, Marin F. 2013 The
592 shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-
593 specific novelties. *FEBS J* **280**, 214-232.

594 [30] Marie B, Le Roy N, Zanella-Cléon I, Becchi M, Marin F. 2011 Molecular evolution of
595 mollusc shell proteins: insights from proteomic analysis of the edible mussel *Mytilus*. *J*
596 *Mol Evol* **72**, 531–546.

597 [31] Marie B, Zanella-Cléon I, Guichard N, Becchi M, Marin F. 2011 The calcifying shell
598 matrix of the Pacific oyster *Crassostrea gigas*: proteomic identification of novel proteins.
599 *Mar. Biotechnol* **13**, 1159-1168.

600 [32] Xiao J, Tagliabracci V, Wen J, Kim S, Dixon J. 2013 Crystal structure of the Golgi
601 casein kinase. *Proc Natl Acad Sci* **110**, 10574-10579.

602 [33] Sagert J, Waite J. 2009 Hyperunstable matrix protein in the byssus of *Mytilus*

603 *galloprovincialis*. *J Exp Biol* **212**, 224-2236.

604 [34] Sudo S, Fujikawa T, Nagakura T, Ohkubo T, Sakagushi K, Tanaka M, Nakashima K,
605 Takahashi T. 1997 Structures of mollusc shell framework proteins. *Nature* **387**, 563–564.

606 [35] Liu X, Dong S, Jin C, Bai Z, Wang G, Li J. 2015 Silkmapin of *Hyriopsis cumingii*, a
607 novel silk-like shell matrix protein involved in nacre formation. *Gene* **555**, 217-222.

608 [36] Yano N, Nagai K, Morimoto K, Miyamoto H. 2006 Shematin: a family of glycine-rich
609 structural proteins in the shell of the pearl oyster. *Comp Biochem Physiol* **144**, 254–262.

610 [37] Miyazaki Y, Usui T, Kajikawa A, Hishiyama H, Matsuzawa N, Nishida T, Machii A,
611 Samata T. 2008 Daily oscillation of gene expression associated with nacreous layer
612 formation. *Front Mater Sci China* **2**, 162-166.

613 [38] Qin CL, Pan QD, Qi Q, Fan MH, Sin JJ, Li NN, Liao Z. 2016 In-depth proteomic
614 analysis of the byssus from marine mussel *Mytilus coruscus*. *J Proteomics* **144**, 87-98.

615 [39] Berland S, Marie A, Duplat D, Milet C, Sire J-Y, Bédouet L. 2011 Coupling proteomics
616 and transcriptomics for identification of novel variant forms of mollusc shell proteins: a
617 study with *P. margaritifera*. *ChemBioChem* **12**, 950-961.

618 [40] Kawasaki K, Buchanan AV, Weiss KM. 2009 Biomineralization in humans: making the
619 hard choices in life. *Annu Rev Genet* **43**, 119–142.

620 [41] Catesy J, Hayashi C, Motruik D, Woods J, Lewis R. 2001 Conservation and convergence
621 of spider silk fibroin sequences. *Science* **291**, 2603-2605.

622 [42] Marie B, Joubert C, Tayale A, Zanella-Cléon I, Marin F, Gueguen Y, Montagnani C.
623 2012 MRNP34, a novel methionine-rich protein from the pearl oysters. *Amino Acids* **5**,
624 2009–2017.

625 [43] McDougall C, Aguilera F, Degnan BM. 2013 Rapid evolution of pearl oyster shell matrix
626 proteins with repetitive, low complexity domains. *J R Soc Interface* **10**, 20130041.

627 [44] Tambutté S, Tambutté E, Zoccola D, Caminiti N, Lotto S, Moya A, Allemand D, Adkins
628 J. 2007 Characterization and role of carbonic anhydrase in the calcification process of the
629 azooxanthellate coral *Tubastrea aurea*. *Mar Biol* **151**, 71–83.

630 [45] Jackson DJ, Macis L, Reitner J, Degnan BM, Wörheide G. 2007 Sponge paleogenomics
631 reveals an ancient role for carbonic anhydrase in skeletogenesis. *Science* **216**, 1893–1895.

632 [46] Mann K, Macek B, Olsen J. 2006 Proteomic analysis of the acid-soluble organic matrix
633 of the chicken calcified eggshell layer. *Proteomics* **6**, 3801–3810.

634 [47] Nagai K, Yano,M, Morimoto K, Miyamoto H. 2007 Tyrosinase localization in mollusc
635 shells. *Comp Biochem Physiol* **146**, 207–214.

- 636 [48] Mann K, Edsinger E. 2014 The *Lottia gigantea* shell matrix proteome: re-analysis
637 including MaxQuant iBAQ quantification and phosphoproteome analysis. *Proteome Sci*
638 **12**, 28.
- 639 [49] Oliveri C, Peric L, Sforzini S, Banni M, Viarengo A, Cavaletto M. 2014 Biochemical and
640 proteomic characterisation of haemolymph serum reveals the origin of the alkali-labile
641 phosphate (ALP) in mussel (*Mytilus galloprovincialis*). *Comp Biochem Physiol Part D*
642 *Genomics Proteomics* **11**, 29–36
- 643 [50] Treccani L, Mann K, Heinemann F, Fritz M. 2006 Perlwapin, an abalone nacre protein
644 with three four-disulfide core (whey acidic protein) domains, inhibits the growth of
645 calcium carbonate crystals. *Biophys J* **91**, 2601–2608.
- 646 [51] Liu HL, Liu SF, Ge YJ, Liu J, Wang XY, Xie LP, Zhang RQ, Wang Z. 2007
647 Identification and characterization of a biomineralization related gene PFMG1 highly
648 expressed in the mantle of *Pinctada fucata*. *Biochemistry* **46**, 844–851.
- 649 [52] Bédouet L, Duplat D, Marie A, Dubost L, Berland S, Rousseau M, Milet C, Lopez E.
650 2007 Heterogeneity of proteinase inhibitors in the water-soluble organic matrix from the
651 oyster nacre. *Mar Biotechnol* **9**, 437–449.
- 652 [53] Arivalagan J, Marie B, Sleight VA, Clark MS, Berland S, Marie A. 2016 Shell matrix
653 proteins of the clam, *Mya truncata*: Roles beyond shell formation through proteomic study
654 *Mar Genomics* **27**, 69–74.
- 655 [54] Dombre C, Guyot N, Moreau T, Monget P, Da Silva M, Gautron J, Réhault-Godbert S.
656 2016 Eggsepsins: The chicken and/or the egg dilemma. *Semin Cell Dev Biol* in press.
657 <http://dx.doi.org/gate1.inist.fr/10.1016/j.semcdb.2016.08.019>
- 658 [55] Keith J, Stockwell S, Ball D, Remillard K, Kaplan D, Thannhauser T, Sherwood R. 1993
659 Comparative analysis of macromolecules in mollusc shells. *Comp Biochem Physiol B* **105**,
660 487–496.
- 661 [56] Giribet, G. 2008 Bivalvia In *Phylogeny and Evolution of the Mollusca*. (W. F. Ponder &
662 D. R. Lindberg, ed. Berkeley) Berkeley: University of California Press, pp. 105–141.
- 663 [57] Chateigner D, Ouhenia S, Krauss C, Hedegaard C, Gil O, Morales M, Lutterotti L,
664 Rousseau M, Lopez E. 2010 Voyaging around nacre with the X-ray shuttle: from bio-
665 mineralisation to prosthetics via mollusc phylogeny. *Mater Sci Engin A* **528**, 37–51.
- 666 [58] Jackson DJ, Mann K, Häussermann V, Schilhabel M, Lüter C, Griesshaber E, Schmahl
667 W, Wörheide G. 2015 The *Magellania venosa* biomineralizing proteome: a window into
668 Brachiopod shell evolution. *Gen Bio Evol* **7**, 1349–1362.

- 669 [59] Luo Y, Takeushi T, Koyanagi R, Yamada L, Kanda M, Khalturine M, Fujie M,
670 Yamasaki S, Endo K, Satoh N. 2015 The *Lingula* genome provides insights into
671 brachiopod evolution and the origin of phosphate biomineralization. *Nature Comm* **6**,
672 8301.
- 673 [60] Marin F, Layrolle P, De Groot K, Westbroek, P. 2003 The Origin of metazoan skeleton,
674 in *Biomineralization: Formation, Diversity, Evolution and Application* (Kobayashi and H.
675 Osawa, Eds.), Tokai University Press, Kanagawa. pp. 50–53.
- 676 [61] Le Roy N, Jackson D, Marie B, Ramos-Silva P, Marin F. 2014 The evolution of metazoan
677 α -carbonic anhydrases and their roles in calcium carbonate. *Frontiers Zool* **11**:75.
- 678 [62] Leaf DS, Anstrom JA, Chin JE, Harkey MA, Showman RM, Raff RA. 1987 Antibodies
679 to a fusion protein identify a cDNA clone encoding msp130, a primary mesenchyme-
680 specific cell surface protein of the sea urchin embryo. *Dev Biol* **121**, 29-40.
- 681 [63] Harkey MA, Whietley HR, Whiteley HA. 1992 Differential expression of the msp130
682 gene among skeletal lineage cells in the sea urchin embryo: a three dimensional *in situ*
683 hybridization analysis. *Mech Dev* **37**, 173-184.
- 684 [64] Mann K, Wilt F, Poustka A. 2010 Proteomic analysis of sea urchin (*Strongylocentrotus*
685 *purpuratus*) spicule matrix. *Proteome Sci* **8**, 33.
- 686 [65] Szabo R, Ferrier DE. 2015 Another biomineralising protostome with an msp130 gene
687 and conservation of msp130 gene structure across Bilateria. *Evol Dev* **17**, 195-197.
- 688 [66] Tagliabracci V, Wiley S, Guo X, Kinch L, Durrant E, Wen J, Xiao J, Cui J, Nguyen K,
689 Engel J, Coon J, Grishin N, Pinna L, Pagliarini D, Dixon J. 2015 A single Kinase generates
690 the majority of the secreted phosphoproteome. *Cell* **161**, 1619-1632.
- 691 [67] Marin F, Bundeleva I, Takeuchi T, Immel F, Medakovic D. 2016 Organic matrices in
692 metazoan calcium carbonate skeletons: composition, functions, evolution. *J Struct Biol* in
693 press. DOI: 10.1016/j.jsb.2016.04.006
- 694 [68] Marie B, Ramos-Silva P, Marin F, Marie A. 2013 Proteomics of CaCO₃ biomineral-
695 associated proteins: How to properly address their analysis. *Proteomics* **13**, 3109-3116.
- 696 [69] Arivalagan J, Yarra T., Marie B, Sleight V, Duvernois-Berthet E, Clark M, Marie A,
697 Berland S. 2016 Insight from the shell proteome: biomineralization to adaptation. *Mol Biol*
698 *Evol* in press. DOI: 10.1093/molbev/msw219
- 699 [70] Kelley L, Mezullis S, Yates C, Wass M, STengerg M. 2015 The Phyre2 web portal for
700 protein modelling, prediction and analysis. *Nature protocols* **10**, 845-858.
- 701 [71] Zhang G, Fang X, Guo X, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que

702 H, Xie Y, Holland PW, Paps J, Zhu Y, Wu F, *et al.* 2012 The oyster genome reveals stress
703 adaptation and complexity of shell formation. *Nature* **490**, 49-54.
704 [72] Mann K, Jackson D. 2014 Characterization of pigmented shell-forming proteome of the
705 common grove snail *Cepaea nemoralis*. *BMC Genomics* **15**:249.

706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737

Figure legends

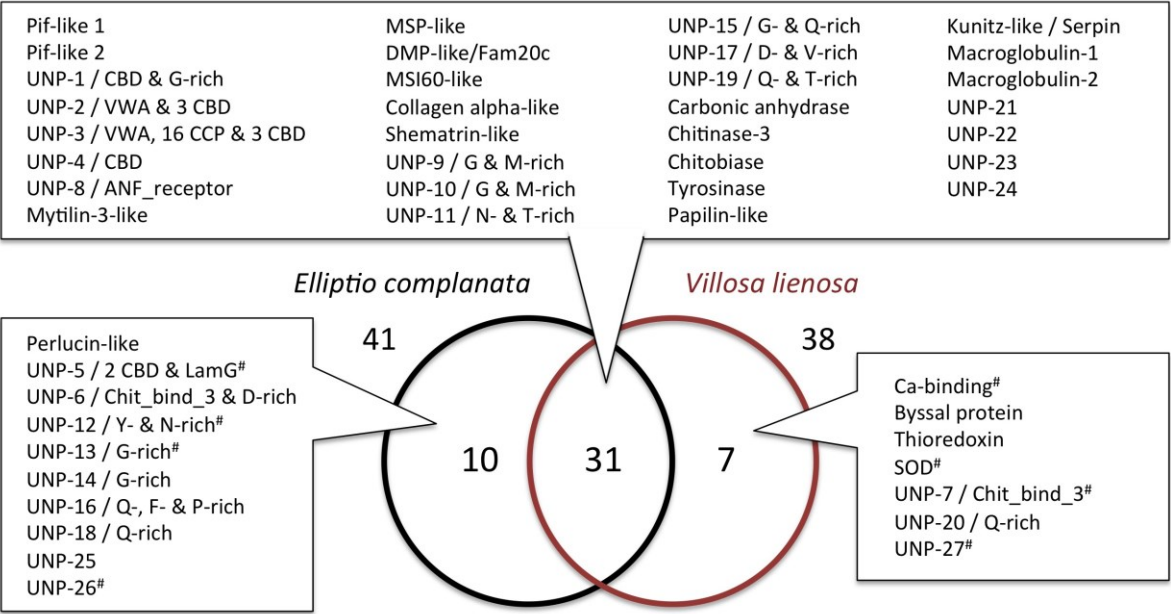
Figure 1. Identification of the nacre matrix proteins from *Elliptio complanata* and *Villosa lienosa* shells using MS/MS analyses. Venn diagram of 41 and 38 proteins identified in both nacre matrices. The hashtag symbol (#) beside some protein names represents cases where the identification is not 100% certain, since it is based only on one peptide and only in one of the two species.

Figure 2. Schematic representation of the primary structure of unionoid extracellular matrix (ECM) nacre proteins retrieved from *Elliptio complanata* and/or *Villosa lienosa*. Each complete protein sequence possesses a signal peptide indicated in red. Typical ECM domains are indicated as boxes on the protein structure. The list of ECM protein domains comprises VWA, CBD2, LamG, Chitin_binding_3, EFh, CLECT, ANF_receptor and FAM20C/DUF1193, together with few RLCs or repeated domains. The “?” symbol indicates missing sequence information.

Figure 3. Nacre SMPs from bivalve models. In total, 48, 35 and 42 various matrix proteins have already been detected in the nacre of the three bivalve models that belong to the Unionidae, Pteriidae and Mytilidae families, respectively. Names correspond to proteins and/or proteinaceous domains identified in the different nacre SMPs. ^p and ^m exponents indicate that the protein was also detected in the prismatic shell layer of either *Pinctada spp.* or *Mytilus spp.*, respectively [2;15;17;18].

Figure 4. Comparison of SMP domain distribution in the shells of various lophotrochozoan models, focused on molluscs and brachiopods. The domains are sorted according to main categories detected in SMPs. * indicates that only genuine abundant SMPs presenting high confidence for their specific identification in the shell matrix, and not being potential considered as contaminant have been mentioned here.

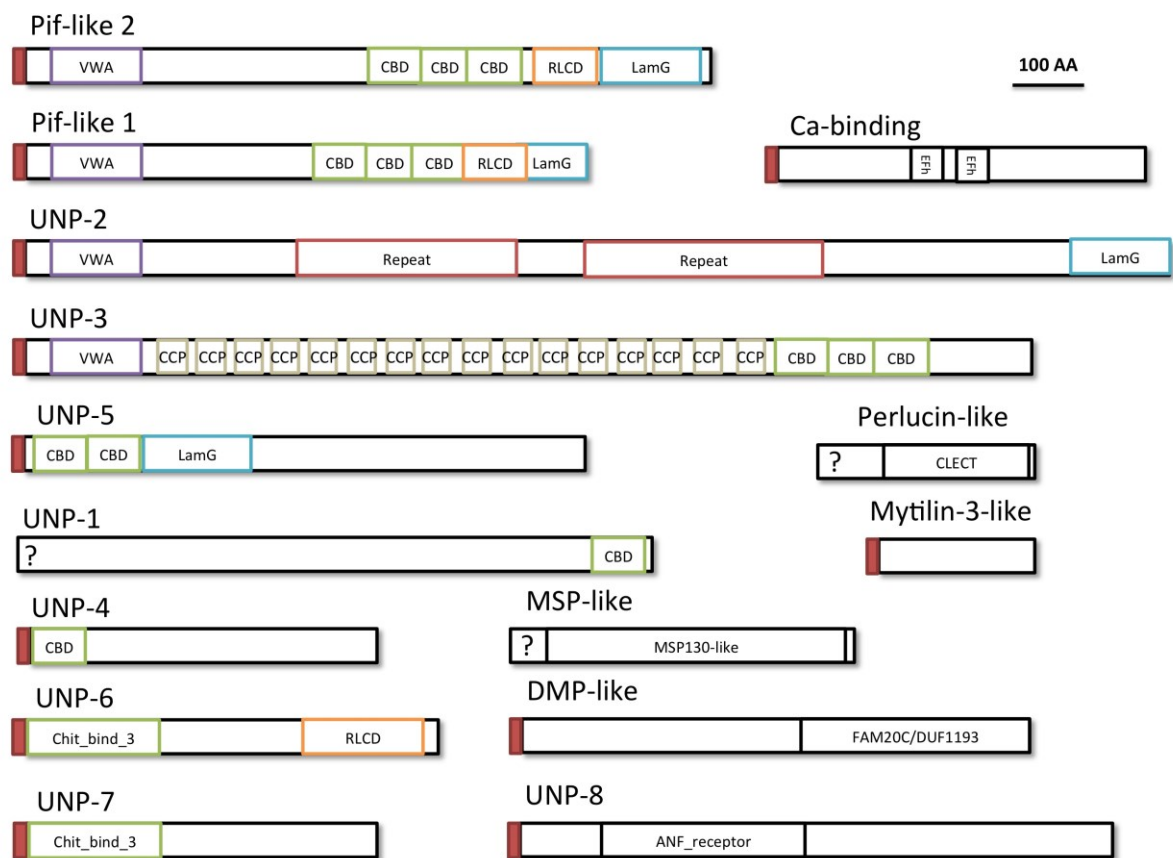
738 Fig. 1



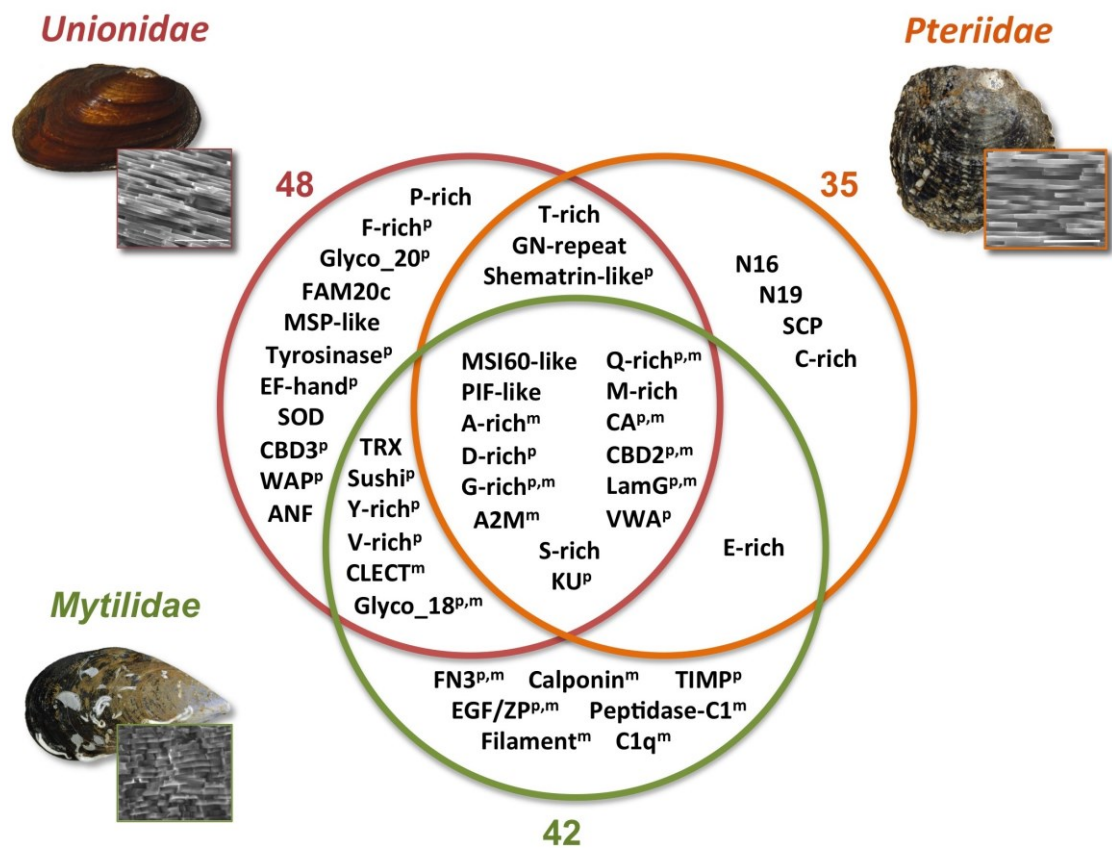
739

740

741



746 Fig. 3



747

748

749

751

[illegible]