The shell-forming proteome of Lottia gigantea reveals both deep conservations and lineage-specific novelties
Marie, B.; Jackson, D.J.; Sequeira dos Ramos Silva, P.; Zanella-Cléon, I.; Guichard, N.; Marin, F.

Published in:
The FEBS Journal

DOI:
10.1111/febs.12062

Citation for published version (APA):
The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties

Benjamin Marie¹*, Daniel J. Jackson², Paula Ramos-Silva¹,³, Isabelle Zanella-Cléon⁴, Nathalie Guichard¹ and Frédéric Marin¹

¹ UMR 6282 (ex 5561) CNRS Biogéosciences, Université de Bourgogne, Dijon, France
² Courant Research Centre Geobiology, Georg-August University of Göttingen, Göttingen, Germany
³ Section of Computational Science, University of Amsterdam, Amsterdam, the Netherlands
⁴ IFR 128 BioSciences Gerland-Lyon Sud, Université de Lyon, Lyon, France

**Keywords**
biomineralization; evolution; mantle; mollusc; shell matrix proteins; proteome

**Correspondence**
B. Marie, UMR 6282 (ex 5561) CNRS Biogéosciences, Université de Bourgogne, Dijon, France
Fax: + 33 (0)3 80 39 63 87
Tel: + 33 (0)3 80 39 63 72
E-mail: bmarie@mnhn.fr

F. Marin, UMR 6282 (ex 5561) CNRS Biogéosciences, Université de Bourgogne, Dijon, France
Fax: + 33 (0)3 80 39 63 87
Tel: + 33 (0)3 80 39 63 72
E-mail: frederic.marin@u-bourgogne.fr

*Present address*
UMR 7245 CNRS MCAM, MNHN, Paris, France

(Received 9 October 2012, revised 1 November 2012, accepted 7 November 2012)

doi:10.1111/febs.12062

Proteins that are occluded within the molluscan shell, the so-called shell matrix proteins (SMPs), are an assemblage of biomolecules attractive to study for several reasons. They increase the fracture resistance of the shell by several orders of magnitude, determine the polymorph of CaCO$_3$ deposited, and regulate crystal nucleation, growth initiation and termination. In addition, they are thought to control the shell microstructures. Understanding how these proteins have evolved is also likely to provide deep insight into events that supported the diversification and expansion of metazoan life during the Cambrian radiation 543 million years ago. Here, we present an analysis of SMPs isolated from the CaCO$_3$ shell of the limpet *Lottia gigantea*, a gastropod that constructs an aragonitic cross-lamellar shell. We identified 39 SMPs by combining proteomic analysis with genomic and transcriptomic database interrogations. Among these proteins are various low-complexity domain-containing proteins, enzymes such as peroxidases, carbonic anhydrases and chitinases, acidic calcium-binding proteins and protease inhibitors. This list is likely to contain the most abundant SMPs of the shell matrix. It reveals the presence of both highly conserved and lineage-specific biomineralizing proteins. This mosaic evolutionary pattern suggests that there may be an ancestral molluscan SMP set upon which different conchiferan lineages have elaborated to produce the diversity of shell microstructures we observe nowadays.

**Database**
Novel protein sequences reported in this article have been deposited in Swiss-Prot database under accession nos. B3A0P1–B3A0S4

**Introduction**

Over the last ~543 million years, molluscs have evolved a wide variety of mineralized shell structures to serve a range of biological functions. The evolutionary success of this morphological innovation is reflected in their presence in almost every ecological niche on the planet. The broad morphological diversity of the
100 000+ species of shell-bearing molluscs [1] extends to a tremendous diversity of mineralogical textures found within the shell, including ‘prismatic’, ‘nacreous’, ‘foliated’, ‘cross-lamellar’, ‘granular’ and ‘homogeneous’ structures [2–5]. Despite this morphological and mineralogical diversity, all molluscan shells are synthesized by a deeply conserved mechanism; they are the result of the secretory activity of an evolutionarily homologous tissue known as the mantle which extrudes inorganic ions and/or amorphous mineral precursors, together with an extracellular organic matrix. All these ingredients self-assemble very precisely in an acellular medium at the interface between the mantle epithelium and the mineralization front. The organic matrix is incorporated into, and surrounds nascent CaCO₃ crystals during the shell layer deposition.

Although the organic matrix represents only a fraction of the total shell weight (usually between 0.1 and 5% w/w), it is known to be essential for both controlling shell formation [6], and for imparting many of the remarkable physical properties (such as fracture resistance) on the mature biomineral. The biochemical characteristics of the organic matrix, usually purified and studied following decalcification of the shell, indicate that it is comprised of a heterogeneous set of macromolecules including mainly proteins, together with variable amounts of polysaccharides and, to a lesser extent, lipids and pigments [7–15].

The protein fraction of this organic matrix has been the subject of much research [16,17]. Since the elucidation of the full-length primary structure of nacrein [18], the first molluscan shell matrix protein (SMP) to be described (from the pearl oyster), the number of SMPs appearing in public sequence databases has gradually increased. More recently, various high-throughput sequencing approaches based on the screening of mantle-derived cDNA libraries and next-generation sequencing methodologies such as RNA-seq, have been employed increasing this rate of discovery [19–22]. Although these DNA- and RNA-based techniques have significantly increased the number of shell-forming candidate protein sequences, they must be cross-referenced with alternative methods in order to identify true shell-forming proteins. Proteomic analyses focused on the characterization of organic material extracted directly from the shell, combined with the interrogation of mantle-derived nucleic acid datasets constitutes one such approach. This strategy has led to the description and robust identification of numerous novel SMPs from various molluscan species [23–27].

One key question concerning the evolution of the Mollusca is whether the diversity of extant shell structures, most of which appeared early during the evolution of this phylum [3,28,29], are in fact constructed from similar SMP assemblages, i.e. whether they truly share a common origin. There is little evidence for the existence of homologous SMPs shared within and between the various bivalve and gastropod models studied to date [20,23,24,27].

In this study, we employed a proteomic approach to investigate the SMPs of an emerging model for biomineralization [30,31], the giant limpet Lottia gigantea. The significant advantage of conducting such a proteomic investigation on SMPs of L. gigantea is that this is the first mollusc for which a draft genome and significant expressed tag sequence (EST) resources are publicly available. We describe the primary structure of 39 SMPs associated with the calcified shell, and based on conserved motifs we discuss the putative functions of these proteins in the calcifying matrix. We also search for homologues of these SMPs in other conchiferan mollusces, and discuss possible scenarios of molecular evolution of SMP genes and the origin of cross-lamellar shell structures.

Results

The shell of L. gigantea

Like other Lottidae [32], the shell of L. gigantea is a multilayered organomineral structure (Fig. 1). The thin nonmineralized outermost periostracum is comprised of only organic components. The rest of the shell is highly calcified and is composed of five distinct layers, named according to their position relative to the periostracum layer (M): M + 3 (outermost), M + 2, M + 1, M and M – 1 (innermost). The outermost M + 3 layer is calcitic and consists of an assemblage of large irregular spherulitic and prismatic structures composed of a mosaic of granular submicron grains [30]. The M + 2 layer consists of aragonitic small microneedle prisms, stacked obliquely to the surface. The M + 1 and M – 1 layers possess a characteristic cross-lamellar construction consisting of complicated hierarchical aragonite structures with first, second- and third-order lamellae [31]. The M layer contains large prismatic aragonite structures that are perpendicular to the shell surface.

In order to remove all potential bacterial, protein and soft tissue contaminants, and to investigate only proteins that are intimately associated with the mineral phase (e.g. SMPs), aragonitic shell layers of L. gigantea (M + 2, M + 1, M and M–1) were carefully cleaned with mechanical abrasion of the periostracum and the outermost M + 3 layer, and crushed into...
minute fragments that were thoroughly decontami-
nated with sodium hypochlorite. Following decalcifica-
tion of this powdered shell material with cold acetic
acid (5% at 4 °C), we subsequently extracted SMPs
associated with the combined aragonitic layers (M + 2,
M + 1, M and M–1). Proteins associated with the
acid-insoluble matrix (AIM) represented ~0.5% of dry
powdered shell weight, whereas the proteins associated
with the acid-soluble matrix (ASM) represents only
0.05%.

**Lottia gigantea** shell matrix proteins

When analysed using 1D denaturing SDS/PAGE,
ASM and AIM proteins displayed few discrete bands
(Fig. 2). ASM and AIM protein banding patterns
shared few components, such as the prominent AIM
bands that were found around 35, 25 and 13 kDa.
Twelve gel bands (b1–b12) were excised from the AIM
SDS/PAGE and analysed by LC-MS/MS for pro-
tein identification. The rest of the AIM SDS/PAGE
profile (without b1–b12 bands) was similarly analysed, without supplementary protein identification. Unfractionated ASM and AIM proteins were also analysed by LC-MS/MS following cleavage by trypsin. Peak lists generated from the MS/MS spectra were directly interrogated against the draft genome assembly Lotgi1_GeneModels_AllModels_20070424_aa (http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html) using MASCOT software. The resulting data were investigated manually and filtered in order to remove redundant protein entries. In this manner, we could unambiguously identify 39 SMPs (Table 1 and Appendix S1). The full-length or partial sequences of 34 of these 39 SMPs are also present in L. gigantea EST datasets, and have now been deposited into the Swiss-Prot database (accession numbers B3A0P1–B3A0S4). We notice that almost all conceptually translated genomic sequences that match our MS/MS peptides possess a predicted signal peptide (Table 1 and Appendix S1). This indicates that these bioinformatically predicted proteins are likely to represent their entire N-terminus and to be genuinely secreted by the mantle.

Our proteomic analysis of L. gigantea SMPs reveals a diversity of SMP structures that can be broadly categorized into one of the following seven classes: repetitive low-complexity domain-containing (RLCD), extracellular matrix-related, enzymes, acidic (low predicted pI), calcium-binding, protease inhibitor and finally orphan proteins with no identifiable domains (Table 1). Although this list of SMPs is not exhaustive (indeed other proteins are known to be present in the L. gigantea shell matrix, see [33]), we believe it is likely to contain most of the abundant SMPs of the aragonitic shell layers because we were able to identify the predominant SDS/PAGE protein bands with a striking match between the expected and observed molecular masses (Fig. 2 and Appendix S1). Indeed, most, if not all, of the peptide analysed from the bands corresponded to the identified proteins. Furthermore, most of the SMPs we identified appear to be the predominant SMPs in Mann et al.’s dataset [33] (Fig. S1). Interestingly, some of these SMPs (e.g. peroxidase-1, -2, and -3, LUSP–1 and -9) were not, or were only partially, detected by Mann et al. (Fig. S1). For example, we were able to identify three full-length peroxidases (Table 1) that were a minor fraction of the Mann et al. dataset. In addition, LUSP–1, which appears to be one of the main components of the L. gigantea AIM (Fig. 2) was not detected by Mann et al. (Fig. S1). These differences may reflect genuine biological variation in the organic contents of the shells of L. gigantea, because Mann et al. investigated the whole shell layers (comprising the calcitic M + 3 layer, together with other aragonitic layers), whereas we restricted our analysis to the aragonitic shell layers (M + 2, M + 1, M and M–1), and/or may this be the result of subtle differences in shell cleaning, matrix extraction and analysis methods.

**RLCD-containing SMPs**

One of the most striking results of our analysis is the qualitative abundance (at least 13 of 39) of proteins possessing blocks of similar or identical amino acids (Table 1; Fig. 3). These RLCD-containing proteins can be subdivided into three groups.

The first group possess, in addition to RLCDs, conserved enzymatic domains such as peroxidase, carbonic anhydrase (CA) or glycosidase domains (Fig. 3A). Lgi-peroxidase–1 and -2, contain recognizable RLCDs rich in the following amino acids: aspartic acid, lysine, glycine, serine, proline, arginine and glutamic acid. We also detected an RLCD domain rich in Gly and Glu within glycosidase–2. Similarly, the CA–2 protein possesses supernumerary Asp- and Glu-rich domains in its C-terminus. Several previously described SMPs also combine such RLCDs with enzymatic domains. For example, the CA domain of nacrein (first isolated from the pearl oyster *Pinctada fucata*) is split by the insertion of a RLCD rich in Gly and Asn [18]. This super-numerary RLCD domain of nacrein has been proposed to regulate the activity of the CA domain, acting as an inhibitor of the precipitation of calcium carbonate [34]. It is possible that these RLCDs, embedded within or adjacent to enzymatically functional domains, may be responsible for conferring on these protein isoforms their specificity for biominalization purposes. However, this hypothesis awaits further investigation.

Glutamine-rich domains characterize the second group of RLCD-containing proteins (Fig. 3B). We identified six L. gigantea uncharacterized shell proteins (LUSP) with high Gln contents, some of which had additional RLCDs rich in other residues. SMPs rich in Gln have also been found in bivalves, for example MPN88 was previously characterized from the oyster *Pinctada margaritifera* [19], but to date no clear function has been attributed to such Gln-rich SMPs. Interestingly, vertebrate teeth contain various Gln-rich proteins belonging to the secreted calcium-binding phosphoprotein families, including amelotin, amelogenin and enamelin. Secretory calcium-binding phosphoproteins are believed to interact with calcium ions and regulate mineralization processes in vertebrates [35].

The third group of RLCD proteins contains three members, none of which exhibit any sequence similarity.
Table 1. *Lottia gigantea* shell matrix proteins. For complete data including MASCOT scores, BLAST results and sequence details see Table S1. Maj., major protein; Min., minor protein; Abs., absent from the detected protein list.

<table>
<thead>
<tr>
<th>Function</th>
<th>Protein name (domain)</th>
<th>Fraction (band/MW)a</th>
<th>Calc. molecular mass and pl</th>
<th>SP</th>
<th>Lgig Acc.b</th>
<th>EST Tissue</th>
<th>SWP Acc.</th>
<th>Mann et al. [33]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enzyme</strong></td>
<td>Peroxidase-like 1 (Peroxidase, G-rich, K-rich domains)</td>
<td>AIM (b1–b12)</td>
<td>191 kDa pI = 5.0</td>
<td>Yes</td>
<td>162078</td>
<td>Mantle</td>
<td>B3A0P1</td>
<td>Min.</td>
</tr>
<tr>
<td></td>
<td>GPCR-like 2 (Peroxidase, D-rich, G-rich domains)</td>
<td>ASM</td>
<td>168 kDa pI = 8.4</td>
<td>Yes</td>
<td>162082</td>
<td>Mantle</td>
<td>B3A0P3</td>
<td>Min.</td>
</tr>
<tr>
<td></td>
<td>Peroxidase-like 3 (Peroxidase domain)</td>
<td>ASM</td>
<td>117 kDa pI = 7.6</td>
<td>Yes</td>
<td>162084</td>
<td>Mantle</td>
<td>B3A0Q8</td>
<td>Min.</td>
</tr>
<tr>
<td></td>
<td>CA–1 (CA domain)</td>
<td>AIM (b6–35 kDa)</td>
<td>42 kDa pI = 6.5</td>
<td>Yes</td>
<td>172265</td>
<td>Mantle</td>
<td>B3A0P2</td>
<td>Maj.</td>
</tr>
<tr>
<td></td>
<td>CA–2 (CA, D-rich domains)</td>
<td>AIM (b4–70 kDa)</td>
<td>69 kDa pI = 5.9</td>
<td>Yes</td>
<td>174599</td>
<td>Mantle</td>
<td>B3A0Q6</td>
<td>Min.</td>
</tr>
<tr>
<td></td>
<td>Glycosidase 1 (Glyco_hydro_23 domain)</td>
<td>AIM (b5–55 kDa)</td>
<td>43 kDa pI = 5.0</td>
<td>Yes</td>
<td>158966</td>
<td>Mantle</td>
<td>B3A0P5</td>
<td>Maj.</td>
</tr>
<tr>
<td></td>
<td>Glycosidase 2 (Glyco_hydro_31, DE-rich domains)</td>
<td>AIM (b5–55 kDa)</td>
<td>57 kDa pI = 4.8</td>
<td>Yes</td>
<td>174920</td>
<td>Not found</td>
<td>–</td>
<td>Min.</td>
</tr>
<tr>
<td></td>
<td>Cyclophilin (Pro-isomerase domain)</td>
<td>AIM (b8–18 kDa)</td>
<td>21 kDa pI = 4.8</td>
<td>Yes</td>
<td>151175</td>
<td>Mantle</td>
<td>B3A0R0</td>
<td>Abs.</td>
</tr>
<tr>
<td><strong>Extracellular</strong></td>
<td>BMSP-like (2 von Willebrand A, 2 CBM_14, LamGL domains)</td>
<td>AIM (b1–160 kDa)</td>
<td>173 kDa pI = 8.5</td>
<td>?</td>
<td>New ORFd</td>
<td>Larvae</td>
<td>B3A0P4</td>
<td>Maj.</td>
</tr>
<tr>
<td><strong>matrix-related</strong></td>
<td>Uncharacterized protein 2 (/LUSP-2 (SCP domain))</td>
<td>AIM (b6–35 kDa)</td>
<td>35 kDa pI = 9.1</td>
<td>Yes</td>
<td>162766</td>
<td>Mantle</td>
<td>B3A0P7</td>
<td>Maj.</td>
</tr>
<tr>
<td></td>
<td>Uncharacterized protein 3 (/LUSP-3 (SCP domain))</td>
<td>AIM (b6–35 kDa)</td>
<td>33 kDa pI = 9.6</td>
<td>Yes</td>
<td>162768</td>
<td>Mantle</td>
<td>B3A0P8</td>
<td>Maj.</td>
</tr>
<tr>
<td></td>
<td>Uncharacterized protein 17 (/LUSP-17 (2 EGF-like, ZP domains))</td>
<td>AIM</td>
<td>53 kDa pI = 4.7</td>
<td>Yes</td>
<td>235548</td>
<td>Mantle</td>
<td>B3A0R6</td>
<td>Maj.</td>
</tr>
<tr>
<td></td>
<td>Uncharacterized protein 24 (/LUSP-24 (2 EGF-like, ZP domains))</td>
<td>AIM</td>
<td>50 kDa pI = 4.8</td>
<td>Yes</td>
<td>167423</td>
<td>Mantle</td>
<td>B3A0S3</td>
<td>Abs.</td>
</tr>
<tr>
<td></td>
<td>Uncharacterized protein 14 (/LUSP-14 (Chitin_bind_3 domain))</td>
<td>–</td>
<td>89 kDa pI = 7.5</td>
<td>Yes</td>
<td>174428</td>
<td>Larvae</td>
<td>–</td>
<td>Abs.</td>
</tr>
<tr>
<td></td>
<td>Uncharacterized protein 20 (/LUSP-20 (3 CBM_14, LamGL domains))</td>
<td>–</td>
<td>81 kDa pI = 6.9</td>
<td>Yes</td>
<td>175684</td>
<td>Mantle</td>
<td>–</td>
<td>Min.</td>
</tr>
<tr>
<td></td>
<td>Perlustrin (IGF–BP domain)</td>
<td>–</td>
<td>11 kDa pI = 4.0</td>
<td>Yes</td>
<td>174064</td>
<td>Mantle</td>
<td>B3A0Q9</td>
<td>Maj.</td>
</tr>
<tr>
<td><strong>RLCD-containing</strong></td>
<td>Uncharacterized protein 1 (/LUSP-1 (G- Q- M-rich RLCDs))</td>
<td>AIM (b4–70 kDa)</td>
<td>68 kDa pI = 9.2</td>
<td>Yes</td>
<td>174645</td>
<td>Larvae</td>
<td>B3A0P6</td>
<td>Abs.</td>
</tr>
<tr>
<td>Function</td>
<td>Protein name (domain)</td>
<td>Fraction (band/MW)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Calc. molecular mass and pI</td>
<td>SP</td>
<td>Lgig Acc.&lt;sup&gt;b&lt;/sup&gt;</td>
<td>EST Tissue</td>
<td>SWP Acc.</td>
<td>Mann et al. [33]</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------</td>
<td>-------------------------------</td>
<td>-----------------------------</td>
<td>----</td>
<td>---------------------</td>
<td>------------</td>
<td>----------</td>
<td>------------------</td>
</tr>
<tr>
<td>Uncharacterized protein 11 /LUSP-11 (M- and G-rich domains)</td>
<td>Uncharacterized protein 11 /LUSP-11 (M- and G-rich domains)</td>
<td>A1M (b10~10 kDa)</td>
<td>13 kDa</td>
<td>Yes</td>
<td>174538</td>
<td>Mantle</td>
<td>B3A0R1</td>
<td>Maj.</td>
</tr>
<tr>
<td>Uncharacterized protein 12 /LUSP-12 (M- and G-rich domains)</td>
<td>Uncharacterized protein 12 /LUSP-12 (M- and G-rich domains)</td>
<td>A1M (b10~10 kDa)</td>
<td>13 kDa</td>
<td>Yes</td>
<td>174534</td>
<td>Mantle</td>
<td>B3A0R2</td>
<td>Maj.</td>
</tr>
<tr>
<td>Uncharacterized protein 5 /LUSP-5 (A-rich domains)</td>
<td>Uncharacterized protein 5 /LUSP-5 (A-rich domains)</td>
<td>A1M (b9~13 kDa)</td>
<td>23 kDa</td>
<td>Yes</td>
<td>173828</td>
<td>Mantle</td>
<td>B3A0Q0</td>
<td>Maj.</td>
</tr>
<tr>
<td>Uncharacterized protein 7 /LUSP-7 (PTTGGG-repeat domains)</td>
<td>Uncharacterized protein 7 /LUSP-7 (PTTGGG-repeat domains)</td>
<td>A1M</td>
<td>28 kDa</td>
<td>Yes</td>
<td>167535</td>
<td>Mantle</td>
<td>B3A0Q2</td>
<td>Maj.</td>
</tr>
<tr>
<td>Uncharacterized protein 22 /LUSP-22 (Q-rich repeat domains)</td>
<td>Uncharacterized protein 22 /LUSP-22 (Q-rich repeat domains)</td>
<td>–</td>
<td>45 kDa</td>
<td>Yes</td>
<td>169441</td>
<td>Mantle</td>
<td>B3A0S0</td>
<td>Min.</td>
</tr>
<tr>
<td>Uncharacterized protein 25 /LUSP-25 (M-rich repeat domains)</td>
<td>Uncharacterized protein 25 /LUSP-25 (M-rich repeat domains)</td>
<td>A1M</td>
<td>18 kDa</td>
<td>Yes</td>
<td>237996</td>
<td>Larvae</td>
<td>B3A0S4</td>
<td>Min.</td>
</tr>
<tr>
<td>Uncharacterized protein 23 /LUSP-23 (D- and E-rich repeat domains)</td>
<td>Uncharacterized protein 23 /LUSP-23 (D- and E-rich repeat domains)</td>
<td>A1M (b3~80 kDa)</td>
<td>42 kDa</td>
<td>Yes</td>
<td>163374</td>
<td>Mantle</td>
<td>B3A0Q3</td>
<td>Maj.</td>
</tr>
<tr>
<td>Uncharacterized protein 9 /LUSP-9 (no characterized domain)</td>
<td>Uncharacterized protein 9 /LUSP-9 (no characterized domain)</td>
<td>A1M (b5~55 kDa)</td>
<td>29 kDa</td>
<td>Yes</td>
<td>166269</td>
<td>Mantle</td>
<td>B3A0Q7</td>
<td>Abs.</td>
</tr>
<tr>
<td>Uncharacterized protein 10 /LUSP-10 (D- and Q-rich repeat domains)</td>
<td>Uncharacterized protein 10 /LUSP-10 (D- and Q-rich repeat domains)</td>
<td>A1M</td>
<td>68 kDa</td>
<td>Yes</td>
<td>163637</td>
<td>Mantle</td>
<td>–</td>
<td>Min.</td>
</tr>
<tr>
<td>Ca-binding</td>
<td>EF-hand containing protein-1 (2 EF-hand domains)</td>
<td>A1M</td>
<td>–</td>
<td>23 kDa</td>
<td>Yes</td>
<td>157030</td>
<td>Mantle</td>
<td>B3A0Q5</td>
</tr>
<tr>
<td></td>
<td>EF-hand containing protein-2 (2 EF-hand domains)</td>
<td>A1M</td>
<td>–</td>
<td>12 kDa</td>
<td>Yes</td>
<td>231427</td>
<td>Mantle</td>
<td>B3A0R9</td>
</tr>
<tr>
<td>Protease inhibitor</td>
<td>Perlwapin (5 WAP domains)</td>
<td>A1M (b8~18 kDa)</td>
<td>43 kDa</td>
<td>Yes</td>
<td>239121</td>
<td>Mantle</td>
<td>B3A0S1</td>
<td>Abs.</td>
</tr>
<tr>
<td>Unknown</td>
<td>Uncharacterized protein 4 /LUSP-4 (no characterized domain)</td>
<td>Uncharacterized protein 4 /LUSP-4 (no characterized domain)</td>
<td>A1M (b8~18 kDa)</td>
<td>17 kDa</td>
<td>Yes</td>
<td>169547</td>
<td>Mantle</td>
<td>B3A0Q9</td>
</tr>
<tr>
<td></td>
<td>Uncharacterized protein 6 /LUSP-6 (no characterized domain)</td>
<td>Uncharacterized protein 6 /LUSP-6 (no characterized domain)</td>
<td>A1M (b7~25 kDa)</td>
<td>20 kDa</td>
<td>Yes</td>
<td>167317</td>
<td>Mantle</td>
<td>B3A0Q1</td>
</tr>
<tr>
<td></td>
<td>Uncharacterized protein 8 /LUSP-8 (no characterized domain)</td>
<td>Uncharacterized protein 8 /LUSP-8 (no characterized domain)</td>
<td>A1M (b8~18 kDa)</td>
<td>18 kDa</td>
<td>Yes</td>
<td>228268</td>
<td>Mantle</td>
<td>B3A0Q4</td>
</tr>
<tr>
<td></td>
<td>Uncharacterized protein 13 /LUSP-13 (no characterized domain)</td>
<td>Uncharacterized protein 13 /LUSP-13 (no characterized domain)</td>
<td>A1M (b6~35 kDa)</td>
<td>26 kDa</td>
<td>Yes</td>
<td>234885</td>
<td>Mantle</td>
<td>B3A0R3</td>
</tr>
<tr>
<td></td>
<td>Uncharacterized protein 15 /LUSP-15 (no characterized domain)</td>
<td>Uncharacterized protein 15 /LUSP-15 (no characterized domain)</td>
<td>A1M (b6~35 kDa)</td>
<td>37 kDa</td>
<td>Yes</td>
<td>167515</td>
<td>Mantle</td>
<td>B3A0R4</td>
</tr>
</tbody>
</table>
with any other proteins. LUSP–11, LUSP–12 and LUSP–25, contain Met- and Gly-rich domains (Fig. 3C). Putative full-length ORFs for these three proteins were deduced from L. gigantea EST and genomic resources. Similar to Gln-rich domains, the significance of Met- and Gly-rich domains in CaCO3 biomineralization is unknown. However, we have noticed that the shell matrices of the gastropod *Haliothis asinina* [23] and the bivalves *P. margaritifera* and *P. maxima* also contain noticeable Met-rich proteins, such as MRNP34 [36].

### Peroxidases

We detected three different peroxidase-domain-containing proteins in *L. gigantea* shell matrices. Peptides of the RLCD-containing *Lgi*-peroxidase–1 and -2 were detected in all MS/MS experiments derived from SDS/PAGE bands b1 to b12 (Fig. 2 and Table S1). This suggests that *Lgi*-peroxidase–1 and -2 are either extremely abundant in the shell matrix, and/or are cleaved into a wide range of peptide lengths after being secreted from the mantle and incorporated into the calcifying shell matrix. Interestingly, all three *Lgi*-peroxidases cluster together in our phylogenetic reconstruction (Fig. S2). Because no other peroxidase breaks this strongly supported clade (posterior probability 0.98), these three limpet peroxidases may have been produced by two gene duplication events in an ancestor that directly gave rise to the *Lottia* lineage. In addition, these three *L. gigantea* peroxidase-encoding genes are all located on the same genomic scaffold (sca_32; Table S2) within 157 kbp of each other.

Interestingly, a similar peroxidase (H2A0M7) has been recently retrieved from the shell matrix of the prismatic layer of the pearl oyster *P. margaritifera* [27]. Peroxidases catalyse the oxidation of many aromatic amines and phenols by hydrogen peroxide. These enzymes have long been associated with molluscan shell formation [37]. The function of such peroxidases within the calcifying shell matrix, or even whether they exhibit peroxidase activity once secreted by the mantle, is unknown. One hypothesis would be that these enzymes act in the same way as the melanogenic peroxidase found in the ink gland of the cuttlefish *Sepia officinalis*, serving to cross-link proteins [38]. Biomineral-associated peroxidases might therefore be involved in biomineral–hydrogel formation via protein matrix framework assembly [39]. Similar functional activity is thought to be mostly provided by two tyrosinases in the *Pinctada* shell matrix [40].
Carbonic anhydrases

CA is a ubiquitous metalloenzyme found in animals, plants and bacteria which catalyses the reversible hydration of carbon dioxide, according to the equation \( \text{CO}_2 + \text{H}_2\text{O} \leftrightarrow \text{HCO}_3^- + \text{H}^+ \). This enzyme is believed to be essential for biomineral formation because bicarbonate, the product of the catalytic process, can directly react with calcium ions to form calcium carbonate. Furthermore, CA has been found in the organic matrices of various metazoan skeletons [41–45]. We detected two different CAs, \( \text{Lgi-CA} -1 \) and \( \text{Lgi-CA} -2 \), in \( \text{L. gigantea} \) shell AIMs and ASMs. Both of these proteins possess a highly conserved \( \alpha \)–CA domain in addition to a Gly- and Glu-rich RLCD present in the C–terminus of \( \text{Lgi-CA} -2 \) (Fig. 4A). Their CA domains possess the conserved active residues known from well-studied \( \alpha \)–CAs [46], suggesting that these two \( \text{Lottia} \) CAs are active enzymes. In support of this, we were able to significantly detect a specific CA activity in the ASM fraction (Fig. 4B).

Asp-rich, low \( p_I \) proteins

Another group of proteins that emerged from our analyses were the acidic Asp-rich proteins ‘Asp-rich’ (Fig. 5A), LUSP–23, LUSP–9 and LUSP–10 with predicted \( p_I \) values of 3.5, 3.6, 3.7 and 3.8, respectively (Table 1). According to Coomassie Brilliant Blue-stained SDS/PAGE gels, the abundant protein ‘Asp-rich’ (which also has the lowest predicted \( p_I \)) has an apparent molecular mass of 80 kDa (Fig. 5). In contrast to this, the predicted molecular mass for the nonglycosylated mature form is only 42 kDa. A likely explanation for this discrepancy is the observation that this band was intensively stained with the cationic dye Alcian Blue, suggesting that ‘Asp-rich’ bears extensive acidic polysaccharide moieties. The hydrophobicity ‘Kyte and Doolittle’ [47] plot of the ‘Asp-rich’ protein suggests that it might also exhibit a coiled-coil structure (Fig. 5B).

The presence of such unusually acidic proteins in the molluscan shell matrix is known from the pioneering work of Meenakshi et al. [48], Crenshaw [7] and Weiner and Hood [49], and has been further confirmed by several investigations [50–54]. However, because of the technical challenges of isolating and purifying these acidic proteins, reports of their primary sequence are rare [55–58]. To our knowledge, the Asp-rich protein detected here, together with MSP–1 extracted from the calcitic foliated layer of \( \text{Patinopecten} \) shell [55], is one of the most acidic molluscan SMPs described to date. Although there are several theoretical models regarding the function that these acidic proteins play in the process of shell formation [59], to date only a few in vivo functional studies that have tested these theories [56].

Fig. 3. Schematic summary of \( \text{L. gigantea} \)’s RLCD-containing SMPs. Schematic representations of the primary structure of RLCD-containing SMPs isolated from the shell of \( \text{L. gigantea} \). (A) RLCD domains of peroxidase–1 and peroxidase–2. (B) LUSP–1, -7, -8, -10 and -22 possess noticeable Q-rich repeats or domains. (C) LUSP–11, -12 and -25 exhibit both M- and G-rich domains. Each protein sequence possesses a signal sequence indicated by a red bar. RLCDs are indicated in light grey, with specific repeats indicated by small white boxes.
Blue mussel shell protein (BMSP)-like

Our MS/MS analyses of a Coomassie Brilliant Blue-stained band with an apparent molecular mass of ~160 kDa (b1 in Fig. 1) identified peptides on the genomic scaffold sca_149. After re-evaluating this genomic locus with an ORF-finding tool (Fig. S3), we identified a protein with a calculated molecular mass of 173 kDa. Two molluscan SMPs shared sequence similarity with this novel *L. gigantea* protein: BMSP-220 (derived from the blue mussel *Mytilus galloprovincialis*; G1UCX0); and Pif-177 (derived from *P. fucata*; C7G0B5) (Fig. 6). These proteins all possess von Willebrand A, peritrophin-A chitin-binding and...
RLCD domains. The *M. galloprovincialis* BMSP [60] and the *L. gigantea* BMSP-like proteins also possess a laminin G-like (LamGL) domain and a poly(T) domain between the von Willebrand A domain and the LamGL domain. The *M. galloprovincialis* BMSP and the *L. gigantea* BMSP-like proteins also share the highest sequence similarity in these domains. The *P. fucata* Pif protein was recently shown to bind both CaCO₃ and chitin, and by RNAi to play a role in nacre formation in vivo [56]. Given that *L. gigantea* does not form nacre it will be interesting to determine the function of the *L. gigantea* BMSP-like protein.

**Epidermal growth factor and sona pellucida domain-containing SMPs**

We also detected two similar proteins (LUSP–17 and LUSP–24) each containing two epidermal growth factor (EGF)-like domains, and one zona pellucida (ZP) domain in their C-termini. Although separate EGF-like and ZP domains are commonly encountered in organic matrix proteins associated with calcification processes [20,61,62], the presence of both domains in one protein is more uncommon. Previous proteomic investigations have described one similar protein from the shell matrix of the pacific oyster *Crassostrea gigas* [25], and two from the *Pinctada* shell matrix [27]. A sequence alignment of these latter proteins with the two EGF-like SMPs of *L. gigantea* is presented in Fig. 7, and illustrates the strong conservation of each domain. LUSP–17 and LUSP–24 are also located on the same genomic scaffold (sca_66). This, in combination with their high degree of sequence identity (79%), strongly suggests that they originated from a gene duplication event.

EGF-like domains are involved in a wide variety of functions such as protein/protein recognition, protein aggregation, molecular signalling or Ca²⁺-binding ability [63]. ZP domains are present in a range of extracellular filament or matrix proteins from a wide variety of eukaryotic organisms, and are characterized by eight conserved cysteine residues, which are involved in protein polymerization processes [64]. Furthermore, the urine-secreted protein, uromodulin (Tamm–Horsfall protein, Q91X17) that exhibits three EGF domains and one ZP domain can potentially contribute to colloid osmotic pressure and modulates formation of supersaturated salts and their crystals [65]. Such similar functions could easily be credited to the EGF- and ZP-containing SMPs and be integrated into a theoretical model of calcified shell biomineralization. However, these hypotheses await validation by functional experiments.

**Other SMPs**

**Cyclophilin**

We also detected a protein in the *L. gigantea* shell matrix presenting sequence similarities with cyclophilins (Fig. S4). Cyclophilins are peptidyl–prolyl isomerases that are believed to mostly facilitate protein folding. In mice, the absence of expression of cyclophilin B has been shown to induce severe osteogenesis imperfecta [66]. Although the specific role of this enzyme in calcium carbonate mineralization is not known, Jackson and co-workers [20] described a cyclophilin gene highly expressed in the nacre forming cells of the pearl oyster *P. maxima*.

**Glycosidases**

Two different glycosidase-related proteins were also detected in *Lottia’s* shell matrix (Fig. S5). The first, named *Lgi*-glycosidase–1, contains a characteristic glycosyl_hydrolase_23 domain and shares significant
sequence similarity with patent lysozyme 2 proteins described from other molluscs [67]. The second, named Lgi-glycosidase-2, possesses Asp- and Glu-rich domains and a conserved glyco_hydro_31 domain. Chitin and other insoluble polysaccharides are major nonprotein components of molluscan shells [68–70]. In classical models of mollusc shell biomineralization, these molecules form a framework of parallel layers between which silk-like and acidic proteins are sandwiched [11,59]. Lottia’s glycosidase SMPs might reasonably be expected to modify this chitin/polysaccharide framework during biomineral formation.

**EF-hand containing proteins**

We also identified two short proteins containing two EF-hand domains (Fig. S6). Similar proteins have been described previously from the shell matrix of the bivalves Pinctada and Venerupis [26,27]. Two consecutive EF-hand domains are known to bind Ca^{2+} ions with high affinity [71], and are observed in many extracellular matrix proteins, such as calmodulins, troponin–C or S–100, often in association with other domains.

**Secreted cysteine-rich protein-like proteins**

Two similar L. gigantea SMPs, LUSP-2 and LUSP-3, contain characteristic secreted cysteine-rich protein (SCP) domains (Fig. S6). We also found that the genes encoding LUSP-2 and LUSP-3 are located at adjacent genomic loci (sca_35). Interestingly, three additional SCP-domain-containing proteins, which were not detected in our MS/MS analyses, are also present on this scaffold (Fig. S7). Interestingly, a similar SCP-containing protein has recently been described from the nacre of P. margaritifera [27]. Because SCP domains have also been described in association with a variety of extracellular matrix proteins, no clear function has been yet assigned to such domains in the context of biomineralization.

**Perlustrin**

We also detected a protein in the shell matrix of L. gigantea with sequence similarities to perlustrin, a protein containing an insulin-like growth factor binding protein (IGF–BP) domain first isolated from the nacre of the abalone Haliotis laevigata (Fig. S8) [72].

---

**Fig. 7.** Two EGF-like SMPs isolated from the shell of L. gigantea. A sequence alignment of EGF-like proteins retrieved from the shell of L. gigantea (B3A0R6 and B3A0S3) against C. gigas (P86785), and P. maxima (P86953 and P86954). Signal peptides (yellow), EGF-like (green) and ZP (blue) domains are highlighted. Stars at the end of each sequence indicate a stop codon.
**Lgi-IGF–BP** is characterized by a pattern of 12 conserved Cys residues. Interestingly, vertebrate bone matrix contains IGF–BPs, which are involved in bone formation, possess an effective affinity for growth factors of the insulin type, and function by modulating IGF metabolism.

**Perlwapin**

We detected in *Lottia*’s shell matrix one protein containing five whey-acidic protein (WAP) domains, and with high overall sequence similarity to the perlwapin family (Fig. S8). WAP consists of two ‘four-disulfide core’ domains that are present in various serine-proteinase inhibitors. Perlwapin proteins, containing such WAP domains, have been identified in *Haliotis* [23,73,74] and from the shell of the blue mussel *M. galloprovincialis* [24]. However, whereas Lgi-perlwapis from the species listed above possess only oen to three WAP domains.

**Orphans**

Nine other *L. gigantea* SMPs do not display any sequence similarity with previously described proteins, or possess recognizably conserved domains (Table 1). These proteins were categorized as orphans. Comparative metazoan genome analyses suggest that every taxonomic group contains 10–20% of these so-called ‘orphan’ or ‘taxonomically restricted’ genes. Such genes are thought to underlie mechanisms that can support the generation of morphological novelties [75]. Interestingly, all molluscan shell matrices broadly investigated at the ‘omic’ level (genomic, transcriptomic or proteomic) contain such orphan proteins. The presence of such orphans may reflect the evolvability of the molluscan shell matrix, suggesting that the appearance of such new proteins within the SMP set could potentially be related to modification of the biomineral structure through evolutionary time. Perhaps more than any other, this class of biomineral-associated proteins highlights the need for *in vivo* gene function assays to be developed for molluscan biomineralizing systems.

**Discussion**

**RLCD-containing SMPs**

RLCD proteins are a prominent feature of all shell-forming proteomes studied to date. Most, if not all, of the RLCD-containing SMPs we have detected appear to be lineage-specific proteins, supporting the idea that such biomineralizing proteins have evolved independently in the different molluscan models. Various RLCD-containing proteins are present in a wide range of metazoan-secreted structures, for example silk fibroin [76], the mussel byssus [77] or the insect chorion [78]. Molluscan shell-forming proteins with RLCDs include nacrein and lustrin–A which contain GN- or GS-rich domains [18,74], MSI60 and CL10Contig2 contain poly (G) and poly(A) blocks [79], Pif-177 contains D-rich domains [56], MPN88 contains Q-, M- and G-rich repeated sequences [19], and the Shematrin family bear numerous GY-rich domains [80]. RLCDs are likely to represent regions with intrinsically disordered conformations thought to be structurally unstable [81]. Such domains possess low binding affinity for other organic macromolecules (such as proteins or polysaccharides), but weakly bind mineral surfaces and ions in aqueous phases. Indeed, GY or GN repeats of the nacrein and shematrins have been proposed to weakly bind Ca²⁺ ions [34,80], whereas the D-rich domains of Pif-177 were shown to directly bind aragonitic mineral surfaces [56]. It has also been proposed that the poly(G), poly (A), or poly(S) regions of MSI60, CL10Contig2 or lustrin–A may confer elastomeric properties to the mature biomineral [23,74,79,82]. Given that RLCD proteins are a major component of the protein fraction within a wide range of molluscan shells, it is clear that they are likely to be playing crucial roles in either shell formation, and/or imparting to the shell certain physical properties such as fracture resistance.

**Conservation of SMPs and their evolution**

Given that *L. gigantea* does not form nacre, one of the most surprising results of our study was the detection of various proteins that share high sequence similarities with SMPs previously identified from the nacre-prismatic shells of *Pinctada* bivalves. Figure 8 summarizes the co-occurrence of SMPs known from various molluscan models of biomineralization: bivalves of the genus *Pinctada* [19,27]; abalone (genus *Haliotis*) [23]; and *L. gigantea* (comprising the proteins reported here together with the 23 main shell proteins identified by Mann *et al.* [33]). Protein sequence alignments and overall domain conservation suggest that most of the eight proteins shared between *Lottia* and *Pinctada* (CAs, BMSP and EGF-like in particular) may be true orthologues (Figs 3, 4, 6 and 7 and Fig. S6). For the two proteins shared between *L. gigantea* and the *Haliotids*, IGF–BP (perlustrin) and perlwapin, accurate evolutionary relationships (orthology versus paralogy) are difficult to assign because the sequence based
similarities between these proteins are restricted to amino acid positions that are specific to the IGF–BP and WAP families (Fig. S8).

Counterintuitively, we have found more SMPs shared between *L. gigantea* and bivalve species *Pinctada* than between the gastropods *L. gigantea* and...
abalone. This trend was also independently described in a transcriptomic comparison of the mantle tissues of *L. gigantea*, *P maxima* and *H. asinina* [20]. One potential explanation for these observations is that the shell-forming secretome of the abalone has accumulated more changes since its divergence from a limpet-abalone ancestor than the limpet has since its divergence from a bivalve-gastropod ancestor. Given the fundamental crystallographic differences between the limpet and abalone shells (presence/absence of nacre and crossed lamellae, for example), such a scenario is conceivable. Complicating this issue is the fact that beyond the species and genus level, molluscan shell microstructures are notoriously evolutionary plastic. To a large degree this plasticity must be the result of the evolution of the organic molecules that coordinate deposition of the shell (past ocean chemistries and temperatures would also affect shell evolution). Our molecular data are compatible with the hypothesis of a genuine affiliation between cross-lamellar structures and nacre [3]. However, a well-resolved, robust and taxonomically well-represented phylogenetic tree for the Conchifera is essential before any scenarios of shell evolution can be proposed and then tested. Fortunately, recent genomic efforts are moving towards this goal [83,84]. In addition to such a resource, better taxon sampling of mantle tissue transcriptomes and shell proteomes would allow us to better understand how this shelled diversity has been generated over the last 550 million years.

**Conclusions**

The availability of genome, proteome and transcriptome scale datasets from non-model organisms is enabling more complete assessments of complex biological processes to be performed. Molluscan shell formation is certainly such a process that will benefit from such analyses. By combining a proteomic analysis of SMPs extracted from the shell of *L. gigantea* with a draft genome assembly, we have identified several new biomineralizing proteins, and further characterized several others. Many of these proteins are characterized by apparently lineage-specific arrangements of RLCDs and highly conserved enzymatic domains such as CA, peroxidase and glycosidase. Even when combined with a recent analysis by Mann et al. [33], the complete shell-forming proteome of *L. gigantea* is unlikely to have been described, and further work will probably identify additional components. Indeed, it remains possible that the trypsin hydrolysis of few SMPs generate only peptides of unsuitable length (too short or too long) for MS analysis, being undetectable by classical proteomic approach [85]. However many of the primary shell-forming proteins are likely to be in hand, and it is becoming increasingly clear that the challenge that now faces the field is to characterize the function of these proteins using *in vivo* techniques.

**Materials and methods**

**Sampling**

Fresh *L. gigantea* shells (5–7 cm in length) were collected from the West Pacific coast of the USA (California). Shell microstructure was observed with a scanning electron microscope Philips XL-30 LaB6 under back-scattered electron mode.

**Shell matrix extraction**

The external organic layer, the periostracum, and the outermost M + 3 calcified layer that presents burrowing traces were mechanically remove under cold water in order to avoid shell heating, then the rest of the shell, comprising the M + 2, M + 1, M and M – 1 layers were crushed into fragments of ~1 mm². Any other superficial organic contaminants were removed by incubating shell fragments in NaOCl (1%, v/v) for 24 h, and which were then thoroughly rinsed with water and subsequently ground into a fine powder that was sieved (> 200 μm). All protein extractions were performed at 4 °C, as previously described [52]. Briefly, powdered samples were decalcified overnight in cold dilute acetic acid (5%, v/v), which was slowly added by an automated titrator (Titronic Universal, Mainz, Germany) at a flow rate of 100 μL every 5 s. The solution (final pH 4.2) was centrifuged at 3900 g (30 min). The resulting pellet, corresponding to the AIM, was rinsed six times with MilliQ water, freeze-dried and weighed. The supernatant containing ASM was filtered (5 μm) and concentrated with an Amicon ultra-filtration system on a Millipore® membrane (10 kDa cut-off). The final solution (> 5 mL) was extensively dialysed against 1 L of MilliQ water (six water changes) before being freeze-dried and weighed.

**CA activity measurement**

The miniaturized colorimetric method developed by Maren [86] was employed for measuring the CA activity (EC 4.2.1.1) of the shell ASM. The experiment was carried out under stabilized flow of CO₂, in an ice-containing vessel. Four hundred microlitres of phenol red (12.5 mg L⁻¹ in 2.6 mm NaHCO₃) were mixed with 200 μL of water and 100 μL of sample. The reaction was initiated by adding 100 μL of freshly made carbonate buffer (0.6 mM Na₂CO₃, 0.412 mM NaHCO₃) and the...
time interval until the colour changed from red to yellow was monitored. This colour change characterizes the pH decrease of the solution (from 8.2 to 7.3), resulting from the production of protons during the reaction catalysed by the CA (CO$_2$ + H$_2$O $\leftrightarrow$ HCO$_3^-$ + H$^+$). The enzyme unit (EU) activity was calculated according to the following equation: activity units (EU) = ($T_o - T$)/$T$; where $T$ and $T_o$ are the reaction times required for the pH change with and without a catalyst, respectively. Acetozolamide was used as a specific inhibitor of the reaction. Commercial bovine CA and BSA were used as positive and negative controls, respectively.

Peptide fractionation and data acquisition

MS was performed using a Q-Star XL nanospray quadrupole/time-of-flight tandem mass spectrometer, nanospray-Qq-TOF-MS/MS (Applied Biosystems, Villebon-sur-Yvette, France), coupled to an online nano liquid chromatography system (Ultimate Famos Switchos from Dionex, Amsterdam, The Netherlands). One microlitre of each sample was loaded onto a trap column (PepMap100 C$_{18}$; 5 $\mu$m; 100 A; 300 $\mu$m × 5 mm; Dionex), washed for 3 min at 25 $\mu$L.min$^{-1}$ with 0.05% trifluoroacetic acid/2% acetonitrile, then eluted onto a C$_{18}$ reverse phase column (PepMap100 C$_{18}$; 3 $\mu$m; 100 A; 75 $\mu$m × 150 mm; Dionex). Peptides were separated at a flow rate of 0.300 $\mu$L.min$^{-1}$ with a linear gradient of 5–80% acetonitrile in 0.1% formic acid over 120 min. MS data were acquired automatically using ANALYST QS 1.1 software (Applied Biosystems). Following a MS survey scan over $m/z$ 400–1600 range, MS/MS spectra were sequentially and dynamically acquired for the three most intense ions over $m/z$ 65–2000 range. The collision energy was set by the software according to the charge and mass of the precursor ion. MS and MS/MS data were recalibrated using internal reference ions from a trypsin autolysis peptide at $m/z$ 842.51 [M + H]$^+$ and $m/z$ 421.76 [M + 2H]$^{2+}$.

MS data analysis

Protein identification was performed using the MASCOT search engine (version 2.1; Matrix Science, London, UK) against protein databases derived from the EST and the genomic libraries of *L. gigantea* comprising 252 091 and 23 851 sequences, and downloaded (March 2010) from the NCBI server (http://www.ncbi.nlm.nih.gov) and the *L. gigantea* genome website (http://genome.jgi-psf.org/Lotgi1/Lotgi1.download ftp.html), respectively. LC-MS/MS data were searched using carbamidomethylation as a fixed modification, and methionine oxidation as a variable modification. The peptide mass and fragment ion tolerances were set to 0.5 Da. Only protein identifications with at least two different peptide hits and/or that were independently obtained from two different samples were considered to be valid. The peptide hits were manually confirmed by the interpretation of the raw LC-MS/MS spectra with ANALYST QS software (Version 1.1). Quality criteria were the peptide MS value, the assignment of major peaks to uninterrupted y- and b-ion series of at least three to four consecutive amino acids and the match with the *de novo* interpretations proposed by the software.
Sequence analysis

Protein sequence identification was performed using BLASTP and TBLASTN analyses performed against Swiss-Prot, GenBank’s nr db and dbEST using the online tool provided by UniProt (www.uniprot.org) and NCBI (http://blast.ncbi.nlm.nih.gov/blast.cgi) servers. Signal peptides were predicted using SIGNALP 3.0 (http://www.cbs.dtu.dk/services/SignalP/), and conserved domains were predicted using SMART (http://smart.embl-heidelberg.de/) and INTERPROSCAN (http://www.ebi.ac.uk/Tools/InterProScan/). Following peptide signal removal, theoretical masses and pI values were determined using the EXPASY PROTPARAM tool (www.expasy.org/tools/protparam.html).

Sequence alignment and phylogenetic analysis

Representative full-length sequences of the major nonvertebrate metazoan peroxidases were selected from the results of a BLAST search performed with the three peroxidases from L. gigantea SMPs, using UniProt and NCBI online tools, against Swiss-Prot, GenBank’s nr db and dbEST. The multiple alignment was created using T-COFFEE 6.85 [89] set to standard parameters. Phylogenetic reconstructions were performed using the maximum likelihood method implemented in PHYLML from the www.phylogeny.fr server [90].

Acknowledgements

We thank Eric Edsinger-Gonzales for providing shell material. BM thanks Jérôme Thomas for handling shell pictures of Lottia and Pinctada. The work of BM was supported by the GDR ADEQUA consortium (Coordinator Nathalie Cochenne-Loreau/Yannick Gueguen), while the work of NG and FM was supported via ANR Accro-Earth (ref. BLAN06-2_159971, Coordinator Gilles Ramstein, LSCE). Additional supports include ITN BIOMINTEC and INTERRIVIE program (2010). DJJ is supported by DFG funding to the University of Göttingen through the German Excellence Initiative.

References


**Supporting information**

Additional supporting information may be found in the online version of this article at the publisher’s website:

**Appendix S1.** Excel file of *Lottia gigantea* shell matrix proteomic dataset.

**Fig. S1.** Comparison of *Lottia gigantea* SMPs identified in this study or by Mann et al. (2012).

**Fig. S2.** Phylogenetic analysis of three peroxidase SMPs.

**Fig. S3.** Reinterpretation of the BMSP-like *L. gigantea* ORF.

**Fig. S4.** Comparison of *L. gigantea's* SMP and various metazoan cyclophilins.

**Fig. S5.** Alignments of *L. gigantea* glycosidase SMPs with homologous sequences.

**Fig. S6.** Alignments of *L. gigantea* SMPs with putative homologous SMP sequences from bivalve shell matrices.

**Fig. S7.** Organization of the gene cluster containing LUSP-2 and -3 on *L. gigantea* genomic scaffold sca_35.

**Fig. S8.** Alignments of *L. gigantea* SMPs with putative homologous SMP sequences from *Haliotis*.

**Table S1.** Proteomic detection of *L. gigantea* peroxidase-1 and -2 in various AIM SDS/PAGE bands.

**Table S2.** Genomic localization of *Lottia gigantea* SMP genes.