The shell-forming proteome of Lottia gigantea reveals both deep conservations and lineage-specific novelties

Marie, B.; Jackson, D.J.; Ramos-Silva, P.; Zanella-Cléon, I.; Guichard, N.; Marin, F.

DOI
10.1111/febs.12062

Publication date
2013

Document Version
Final published version

Published in
The FEBS Journal

the FEBS Journal

The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties

Benjamin Marie[1,*], Daniel J. Jackson[2], Paula Ramos-Silva[1,3], Isabelle Zanella-Cléon[4], Nathalie Guichard[1] and Frédéric Marin[1]

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 form 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

Abbreviations
AIM, acid-insoluble matrix; ASM, acid-soluble matrix; BMSP, blue mussel shell protein; CA, carbonic anhydrase; EF-hand, calcium-binding motif (E-helix-loop-F-helix) in a "hand" configuration; EGF, epidermal growth factor; EST, expressed sequence tag; IGF–BP, insulin-growth factor-binding protein; kbp, kilo-base pair; LamGL, laminin G-like; LUSP, *Lottia* uncharacterized shell protein; RLCD, repeated low-complexity domain; SCP, secreted cysteine-rich protein; SMP, shell matrix protein; WAP, whey-acidic protein; ZP, zona pellucida.

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 $CaCO_3$ 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 molluscs, 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 Lottiidae [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 myostracum 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

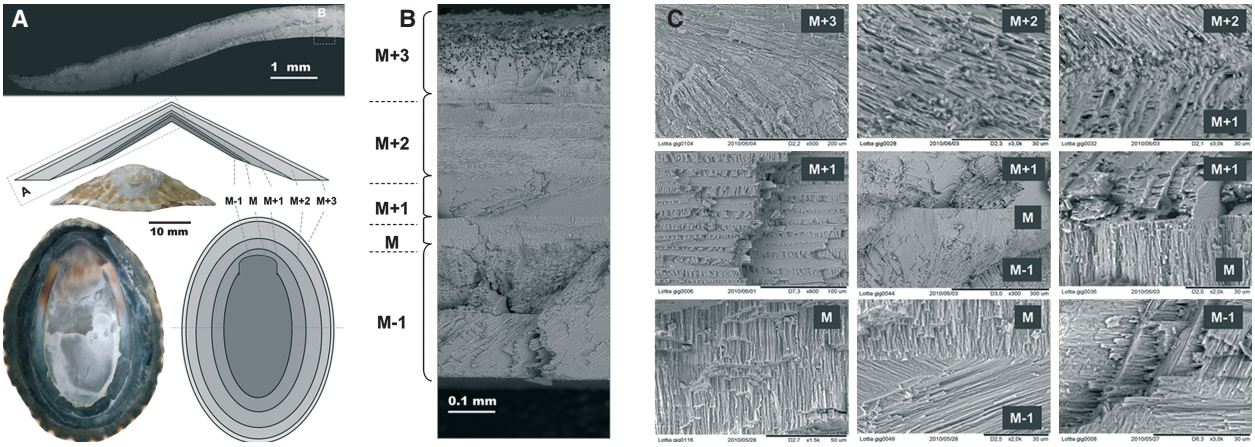

Fig. 1. Shell layers of the giant limpet *L. gigantea*. (A) Low magnification SEM view of a transverse cross-section of the shell, and schematic representations of the different layers. (B) SEM of the cross-sectional area (boxed area in A) showing the five calcified shell layers (M + 3, M + 2, M + 1, M, M–1). (C) SEM detailing the different calcified layers. The outermost M + 3 layer consists of calcitic large irregular spherulitic and prismatic structures. The M + 2 layer consists of aragonitic small microneedle prisms. The M + 1 and M–1 layers possess a characteristic cross-lamellar structure. The M layer, the myostracum, contains large prismatic aragonite structures perpendicular to the shell surface.

minute fragments that were thoroughly decontaminated with sodium hypochlorite. Following decalcification 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 protein identification. The rest of the AIM SDS/PAGE

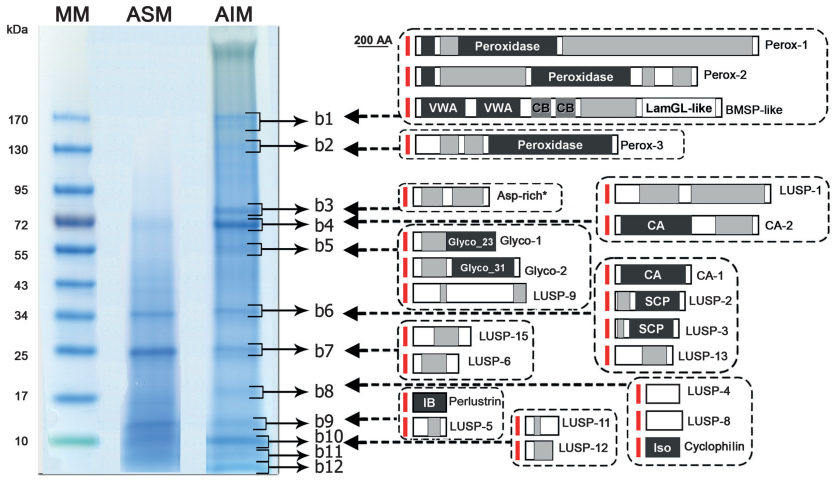

Fig. 2. Main shell matrix proteins of *L. gigantea*. SDS/PAGE separation of acid-insoluble and acid-soluble SMPs. ASM and AIM SMPs were separated on a 4–15% gradient SDS/PAGE gel under denaturing conditions and stained with Coomassie Brilliant Blue. The 12 most intensely stained bands of the AIM (b1–12) were excised for further analysis by MS/MS. A schematic representation of the identified proteins is shown on the right. Grey shaded domains indicate RLCDs. The Asp-rich protein (indicated by *) is likely to possess extensive glycosylation. Red bars indicate signal peptide sequences as determined by SIGNALP 3.0.

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 p*I*), 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 supernumerary 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 biomineralization 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.

Function	Protein name (domain)	Fraction (band/MW)[a]	Calc. molecular mass and p*I*	SP	Lgig Acc.[b]	EST Tissue	SWP Acc.	Mann *et al.* [33]
Enzyme	Peroxidase-like 1 (Peroxidase, G-rich, K-rich domains)	AIM (b1–b12) ASM	191 kDa p*I* = 5.0	Yes	162078	Mantle	B3A0P1	Min.
	Peroxidase-like 2 (Peroxidase, D-rich, G-rich domains)	AIM (b1–b12)[c] ASM	158 kDa p*I* = 8.4	Yes	162082	Mantle	B3A0P3	Min.
	Peroxidase-like 3 (Peroxidase domain)	AIM (b2 ~ 120 kDa) ASM	117 kDa p*I* = 7.6	Yes	162084	Mantle	B3A0Q8	Min.
	CA–1 (CA domain)	AIM (b6 ~ 35 kDa) ASM	42 kDa p*I* = 6.5	Yes	172265	Mantle	B3A0P2	Maj.
	CA–2 (CA, D-rich domains)	AIM (b4 ~ 70 kDa) ASM	69 kDa p*I* = 5.9	Yes	174599	Mantle	B3A0Q6	Min.
	Glycosidase 1 (Glyco_hydro_23 domain)	AIM (b5 ~ 55 kDa) –	43 kDa p*I* = 5.0	Yes	158966	Mantle	B3A0P5	Maj.
	Glycosidase 2 (Glyco_hydro_31, DE-rich domains)	AIM (b5 ~ 55 kDa) –	57 kDa p*I* = 4.8	Yes	174920	Not found	–	Min.
	Cyclophilin (Pro-isomerase domain)	AIM (b8 ~ 18 kDa) –	21 kDa p*I* = 4.8	Yes	151175	Mantle	B3A0R0	Abs.
Extracellular matrix-related	BMSP-like[d] (2 von Willebrand A, 2 CBM_14, LamGL domains)	AIM (b1 ~ 160 kDa) ASM	173 kDa p*I* = 8.5	?	New ORF[d]	Larvae	B3A0P4	Maj.
	Uncharacterized protein 2 /LUSP–2 (SCP domain)	AIM (b6 ~ 35 kDa) ASM	35 kDa p*I* = 9.1	Yes	162766	Mantle	B3A0P7	Maj.
	Uncharacterized protein 3 /LUSP-3 (SCP domain)	AIM (b6 ~ 35 kDa) ASM	33 kDa p*I* = 9.6	Yes	162768	Mantle	B3A0P8	Maj.
	Uncharacterized protein 17 /LUSP–17 (2 EGF-like, ZP domains)	AIM –	53 kDa p*I* = 4.7	Yes	235548	Mantle	B3A0R6	Maj.
	Uncharacterized protein 24 /LUSP–24 (2 EGF-like, ZP domains)	AIM –	50 kDa p*I* = 4.8	Yes	167423	Mantle	B3A0S3	Abs.
	Uncharacterized protein 14 /LUSP–14 (Chitin_bind_3 domain)	– ASM	89 kDa p*I* = 7.5	Yes	174428	Larvae	–	Abs.
	Uncharacterized protein 20 /LUSP–20 (3 CBM_14, LamGL domains)	– ASM	81 kDa p*I* = 6.9	Yes	175684	Mantle	–	Min.
	Perlustrin (IGF–BP domain)	AIM (b9 ~ 13 kDa) –	11 kDa p*I* = 4.0	Yes	174064	Mantle	B3A0Q9	Maj.
RLCD-containing	Uncharacterized protein 1 /LUSP–1 (G-, Q-, M-rich RLCDs)	AIM (b4 ~ 70 kDa) ASM	68 kDa p*I* = 9.2	Yes	174645	Larvae	B3A0P6	Abs.

Table 1. (Continued).

Function	Protein name (domain)	Fraction (band/MW)[a]	Calc. molecular mass and p*I*	SP	Lgig Acc.[b]	EST Tissue	SWP Acc.	Mann *et al.* [33]
	Uncharacterized protein 11 /LUSP–11 (M- and G-rich domains)	AIM (b10 ~ 10 kDa) ASM	13 kDa p*I* = 10.8	Yes	174538	Mantle	B3A0R1	Maj.
	Uncharacterized protein 12 /LUSP–12 (M- and G-rich domains)	AIM (b10 ~ 10 kDa) ASM	13 kDa p*I* = 9.8	Yes	174534	Mantle	B3A0R2	Maj.
	Uncharacterized protein 5 /LUSP-5 (A-rich domains)	AIM (b9 ~ 13 kDa) ASM	23 kDa p*I* = 10.4	Yes	173828	Mantle	B3A0Q0	Maj.
	Uncharacterized protein 7 /LUSP-7 (PTTGGQ-repeat domains)	AIM ASM	28 kDa p*I* = 5.4	Yes	167535	Mantle	B3A0Q2	Maj.
	Uncharacterized protein 22 /LUSP–22 (Q-rich repeat domains)	– ASM	45 kDa p*I* = 6.9	Yes	169441	Mantle	B3A0S0	Min.
	Uncharacterized protein 25 /LUSP–25 (M-rich repeat domains)	AIM –	18 kDa p*I* = 9.9	Yes	237996	Larvae	B3A0S4	Min.
Acidic	Asp-rich protein (D-rich domains)	AIM (b3 ~ 80 kDa) ASM	42 kDa p*I* = 3.5	Yes	163374	Mantle	B3A0Q3	Maj.
	Uncharacterized protein 23 /LUSP–23 (D- and E-rich repeat domains)	AIM –	22 kDa p*I* = 3.6	Yes	172797	Larvae	B3A0S2	Min.
	Uncharacterized protein 9 /LUSP–9 (no characterized domain)	AIM (b5 ~ 55 kDa) –	29 kDa p*I* = 3.7	Yes	166269	Mantle	B3A0Q7	Abs.
	Uncharacterized protein 10 /LUSP–10 (D- and Q-rich repeat domains)	AIM –	68 kDa p*I* = 3.8	Yes	163637	Mantle	–	Min.
Ca-binding	EF-hand containing protein-1 (2 EF-hand domains)	– ASM	23 kDa p*I* = 6.5	Yes	157030	Mantle	B3A0Q5	Maj.
	EF-hand containing protein-2 (2 EF-hand domains)	– ASM	12 kDa p*I* = 4.7	Yes	231427	Mantle	B3A0R9	Min.
Protease inhibitor	Perlwapin (5 WAP domains)	– ASM	43 kDa p*I* = 7.9	Yes	239121	Mantle	B3A0S1	Abs.
Unknown	Uncharacterized protein 4 /LUSP-4 (no characterized domain)	AIM (b8 ~ 18 kDa) –	17 kDa p*I* = 8.8	Yes	168547	Mantle	B3A0P9	Maj.
	Uncharacterized protein 6 /LUSP-6 (no characterized domain)	AIM (b7 ~ 25 kDa) ASM	20 kDa p*I* = 9.6	Yes	167317	Mantle	B3A0Q1	Maj.
	Uncharacterized protein 8 /LUSP-8 (no characterized domain)	AIM (b8 ~ 18 kDa) ASM	18 kDa p*I* = 10.1	Yes	228268	Mantle	B3A0Q4	Maj.
	Uncharacterized protein 13 /LUSP–13 (no characterized domain)	AIM (b6 ~ 35 kDa) ASM	26 kDa p*I* = 8.5	Yes	234885	Mantle	B3A0R3	Maj.
	Uncharacterized protein 15 /LUSP–15 (no characterized domain)	AIM (b6 ~ 35 kDa) ASM	37 kDa p*I* = 6.0	Yes	167515	Mantle	B3A0R4	Min.

Table 1. (Continued).

Function	Protein name (domain)	Fraction (band/MW)[a]	Calc. molecular mass and p/	SP	Lgig Acc.[b]	EST Tissue	SWP Acc.	Mann *et al.* [33]
	Uncharacterized protein 16 /LUSP–16 (no characterized domain)	AIM –	23 kDa p/ = 9.8	Yes	158439	Mantle	B3A0R5	Min.
	Uncharacterized protein 18 /LUSP–18 (no characterized domain)	– ASM	55 kDa p/ = 5.7	Yes	167518	Mantle	-	Maj.
	Uncharacterized protein 19 /LUSP–19 (no characterized domain)	– ASM	27 kDa p/ = 9.3	Yes	173200	Mantle	B3A0R7	Maj.
	Uncharacterized protein 21 /LUSP–21 (no characterized domain)	– ASM	16 kDa p/ = 9.7	Yes	157064	Mantle	B3A0R8	Maj.

[a] Bands (b1–b12) were excised from AIM SDS/PAGE, as described in Fig. 2. [b] The Lotgi1_GeneModels_AllModels_20070424_aa complete genomic database used for the proteomic searches was downloaded from *Lottia gigantea* v0.1 genome project website (http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html); [c] We notice that both peroxidase-like 1 and peroxidase-like 2 proteins, or their fragmented peptides, were detected in all MS/MS analysed SDS/PAGE bands from b1 to b12 (Table S1). [d] For this protein a new ORF model was deduced from the genomic sequence (Fig. S3). MW, molecular weight; SP, signal peptide; Lgig acc., *Lottia gigantea* genomic database accession number; SWP acc., Swiss-Prot accession number.

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 $CaCO_3$ biomineralization is unknown. However, we have noticed that the shell matrices of the gastropod *Haliotis 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 $CO_2 + H_2O \leftrightarrow HCO_3^- + H^+$. 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, *Lgi*-CA–1 and *Lgi*-CA–2, in *L. gigantea* shell AIMs and ASMs. Both of these proteins possess a highly conserved α–CA domain in addition to a Gly- and Glu-rich RLCD present in the C–terminus of *Lgi*-CA–2 (Fig. 4A). Their CA domains possess the conserved active residues known from well-studied α–CAs [46], suggesting that these two *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 non-glycosylated 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 *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].

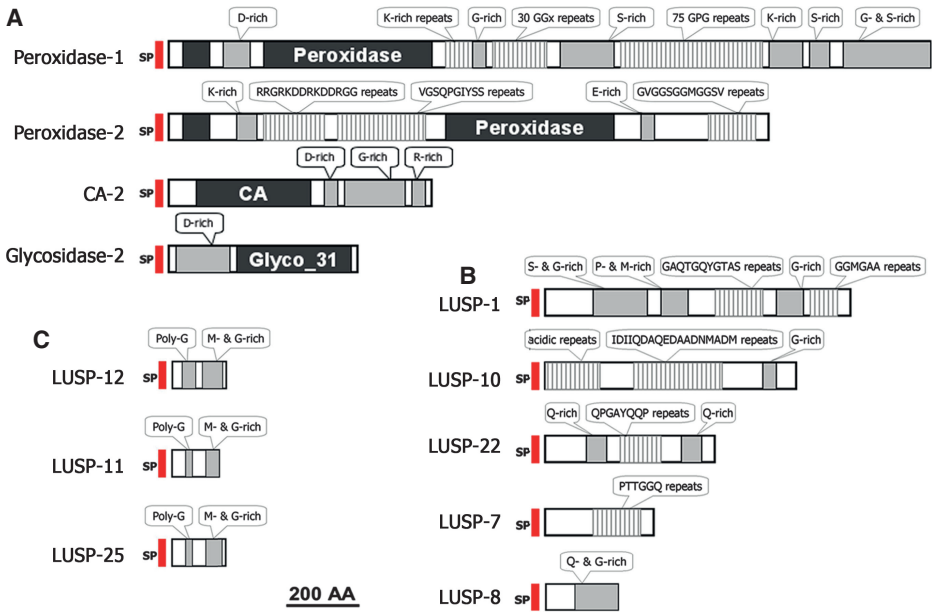

Fig. 3. Schematic summary of *L. gigantea*'s RLCD-containing SMPs. Schematic representations of the primary structure of RLCD-containing SMPs isolated from the shell of *L. gigantea*. (A) RLCD domains of peroxidase–1 and peroxidase–2. (B) LUSP–1, -7, -8, -10 and -22 possess noticeable Q-rich repeats 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

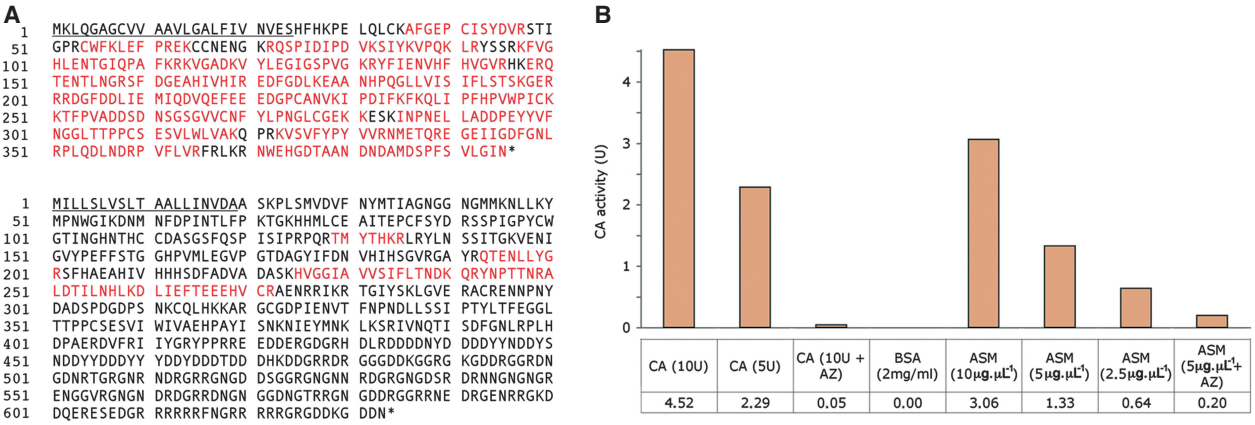

Fig. 4. Structure and activity of two CAs isolated from the shell of *L. gigantea*. (A) Amino acid sequences of CA–1 (upper) and CA–2 (lower). The peptide sequences detected by MS/MS are indicated in red. Signal sequences are underlined. Stars indicate stop codons. (B) CA activity of the ASM derived from *L. gigantea* shells. Commercial CA derived from bovine erythrocytes was used as a positive control and acetozolamide (AZ) was used as a specific inhibitor of carbonic anhydrase activity.

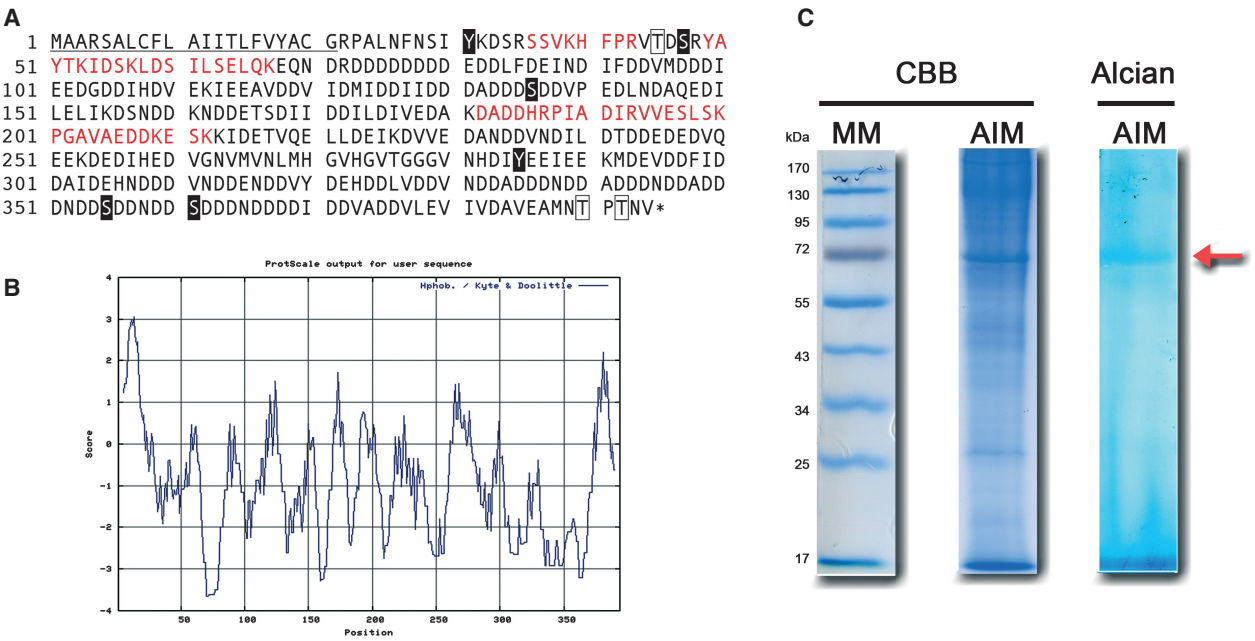

Fig. 5. An Asp-rich SMP isolated from the shell of *L. gigantea*. (A) Amino acid sequence of Asp-rich protein. The peptide sequences detected by MS/MS are indicated in red. The signal peptide sequence is underlined. A star indicates the stop codon. Amino acids boxed in black or white indicate putative phosphorylation or glycosylation sites, respectively. (B) Hydrophobicity 'Kyte and Doolittle' plot of Asp-rich protein suggesting a coiled-coil structure. (C) 12% SDS/PAGE of *L. gigantea* AIM, stained with Coomassie Brilliant Blue or Alcian Blue, pH 1. The molecular mass markers are indicated on the left. The red arrow localizes the 70-kDa band excised for MS analysis that contains the mature acidic Asp-rich glycoprotein.

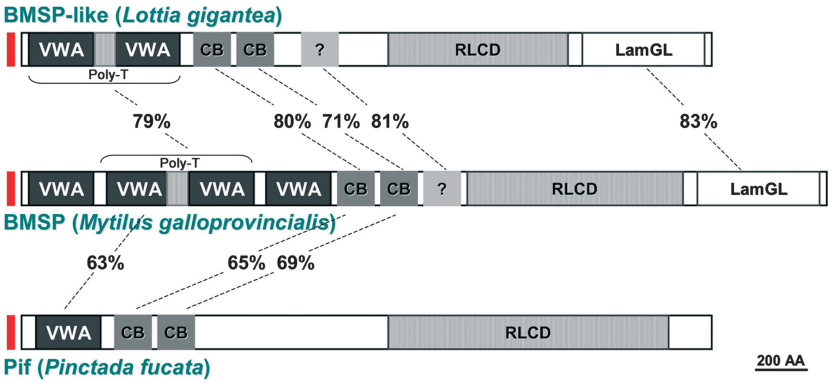

Fig. 6. BMSP-like SMPs isolated from the shell of *L. gigantea*. Schematic representations of the primary structure of *L. gigantea* BMSP-like, *M. galloprovincialis* BMSP and *P. fucata* Pif proteins. von Willanbrand A, peritrophin-A chitin-binding (CB), RLCDs and LamGL domains are indicated. Sequence similarity scores between selected domains are the percentage of amino acid identity.

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 $CaCO_3$ 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 Ca^{2+}-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]. This

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*-perlwapin contains five WAP domains, the other perlwapins 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 Ca^{2+} 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 nacroprismatic 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 *Halio tis*) [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

B

	RLCD containing						Extracellular matrix							Enzymatic			Ca-interacting			Polysaccharide-interacting						Protease inhibitor			Orphans
	A-rich	G-rich	M-rich	S-rich	Q-rich	V-rich	VWA	SCP	EGF-like	IGF-BP	Cyclophyllin	FN3	Ependymin	Peroxidase	Tyrosinase	Carb. Anhyd.	D-rich	EF-hand	GN-repeat	Glyco_18	Glyco_20	Glyco_23	Glyco_31	Peritrophin A	Chitin-BD_3	LamGL-like	Lectin	WAP	Kunitz-like	TIMP	Unkwnon
Lottia gigantea		■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■	■	■		9
Haliotis spp.	■	■	■	■	■					■			■				■		■								■	■	■		7
Pinctada spp.	■	■	■	■	■	■	■	■	■			■		■	■	■	■	■	■	■	■			■		■			■	■	10

Fig. 8. A comparison of molluscan SMPs isolated from *Lottia, Pinctada* and *Haliotis*. A broad comparison of molluscan SMPs. (A) A summary of the shared and lineage-specific SMPs described to date from *L. gigantea* and various *Pinctada* and *Haliotis* species. Numbers correspond to the number of different SMPs detected to date for each model, for example we can distinguish 32 different SMPs from the 39 different SMPs we have identified for *L. gigantea* (e.g. when considering one entry for the two CAs). (B) After categorizing these SMPs into eight broad categories it is clear that proteins with RLCDs are a common feature of the molluscan shell-forming secretome. Most proteins shared between *L. gigantea* and *Pinctada* species fall into the extracellular matrix category. Grey boxes indicate proteins detected by Mann *et al.* [33] to be minor components of the shell matrix.

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 mm^2. 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 CO_2, in an ice-containing vessel. Four hundred microlitres of phenol red (12.5 $mg{\cdot}L^{-1}$ in 2.6 mM $NaHCO_3$) 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 M Na_2CO_3, 0.412 M $NaHCO_3$) 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 ($CO_2 + H_2O \leftrightarrow HCO_3^- + H^+$). The enzyme unit (EU) activity was calculated according to the following equation: activity units (EU) = $(T_0 - T)/T$; where T and T_0 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.

SDS/PAGE

The fractionation of matrix macromolecules was performed under denaturing conditions by monodimensional SDS/PAGE (Mini-Protean 3; Bio-Rad). One milligram of matrix (both ASM and AIM) was suspended in 200 μL of Laemmli sample buffer [87], heat denatured (10 min, 100 °C) then centrifuged for 1 min at 12 000 ***g***. Ten microlitres of the supernatant, representing a maximum of 50 μg of matrix, were loaded onto gels. Following SDS/PAGE under denaturing conditions (4–15% acrylamide gel), proteins were visualized with Coomassie Brilliant Blue (CBB G-250; Biosafe, Bio-Rad). Alternatively, putative glycosylations were investigated by staining with Alcian Blue 8GX [88], at pH 1 in order to specifically stain sulfated sugars.

Sample preparation for proteomic analysis

An in-gel digestion procedure was performed for 12 predominant protein bands visualized from the electrophoresis gel of the AIM (Fig. 2). These bands were excised from Coomassie Brilliant Blue-stained gels and completely destained by a wash with 400 μL of 50 mM NH_4HCO_3/CH_3CN (50/50) mixture for 15 min at 37 °C. Reduction was performed with 50 μL of 10 mM dithiothreitol in 50 mM NH_4HCO_3 for 15 min at 50 °C. Alkylation was performed with 50 μL of 100 mM iodoacetamide for 15 min at room temperature in the dark. The reagents were taken away and the gel pieces were dried using 100 μL of CH_3CN. Gel pieces were then treated with 0.4 μg trypsin (Sequence grade; Promega, Madison, WI, USA) in 20 μL of 50 mM NH_4HCO_3 for 45 min at 50 °C under 800 rpm agitation. The supernatant was removed and stored. The gel pieces were extracted with 30 μL of H_2O:CH_3CN:HCOOH (68 : 30 : 2) mixture for 30 min at 30 °C. Finally, both supernatant extracts were pooled, dried in a vacuum concentrator and resuspended in 13 μL of 0.1% trifluoroacetic acid.

In-solution digestion of unfractionated *L. gigantea* ASM and AIM was also performed. These samples (0.1 and 1 mg, respectively) were reduced with 50 μL of 10 mM dithiothreitol in 50 mM NH_4HCO_3 for 30 min at 50 °C. Alkylation was performed with 50 μL of 100 mM iodoacetamide in 50 mM NH_4HCO_3 for 30 min at room temperature in the dark. The solution was then treated with 1 μg of trypsin (Sequence grade; Promega) in 10 μL 50 mM NH_4HCO_3 overnight at 37 °C. The sample was dried in a vacuum concentrator and resuspended in 30 μL of 0.1% trifluoroacetic acid and 2% CH_3CN.

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 μm; 100 Å; 300 μm × 5 mm; Dionex), washed for 3 min at 25 μL·min^{-1} with 0.05% trifluroacetic acid/2% acetonitrile, then eluted onto a C_{18} reverse phase column (PepMap100 C_{18}; 3 μm; 100 Å; 75 μm × 150 mm; Dionex). Peptides were separated at a flow rate of 0.300 μL·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 p*I* 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 searche 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 PHYML from the www.phylogeny.fr server [90].

Acknowledgements

We thank Eric Edsinger-Gonzales for providing shell material. BM thanks Jérome Thomas for handling shell pictures of *Lottia* and *Pinctada*. The work of BM was supported by the GDR ADEQUA consortium (Coordinator Nathalie Cochennec-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 INTERRVIE program (2010). DJJ is supported by DFG funding to the University of Göttingen through the German Excellence Initiative.

References

1 Ponder W & Lindberg DR (2008) Phylogeny and Evolution of the Mollusca. University of California Press, Berkley, 469 p.
2 Bøggild O (1930) The shell structure of the mollusks. *K Dan Vidensk Selsk Skr Naturvidensk Math Afd* **9**, 233–326.
3 Taylor J, Kennedy WJ & Hall A (1973) The shell structure and mineralogy of the Bivalvia, Part II, *Lucinacea–Clavagellacea*. *Bull Brit Mus Zool* **22**, 253–294.
4 Carter JG (1990) Skeletal Biomineralization: Patterns, Processes and Evolutionary Trends. Van Nostrand Reinhold, New York.
5 Chateigner D, Hedegaard C & Wenk HR (2000) Mollusc shell microstructures and crystallographic textures. *J Struct Geol* **22**, 1723–1735.
6 Mann S (2001) Biomineralization: Principles and Concepts in Bioinorganic Materials Chemistry. Oxford University Press, New York.
7 Crenshaw MA (1972) The soluble matrix from *Mercenaria mercenaria* shell. *Biomineralization* **6**, 6–11.
8 Weiner S & Traub W (1984) Macromolecules in mollusc shells and their functions in biomineralization. *Phil Trans R Soc Lond* **304**, 425–434.
9 Lowenstam HA & Weiner S (1989) On Biomineralization. Oxford University Press, New York.
10 Keith J, Stockwell S, Ball D, Remillard K, Kaplan D, Thanhauser T & Sherwood R (1993) Comparative analysis of macromolecules in mollusc shells. *Comp Biochem Physiol B* **105**, 487–496.
11 Levi-Kalisman Y, Falini G, Addadi L & Weiner S (2001) Structure of the nacreous organic matrix of a bivalve mollusk shell examined in the hydrated state using cryo-TEM. *J Struct Biol* **135**, 8–17.
12 Bédouet L, Schuller MJ, Marin F, Milet C, Lopez E & Giraud M (2001) Soluble proteins of the nacre of the giant oyster *Pinctada maxima* and of the abalone *Haliotis tuberculata*: extraction and partial analysis of nacre proteins. *Comp Biochem Physiol B* **128**, 389–400.
13 Marie B, Luquet G, Pais De Barros J-P, Guichard N, Morel S, Alcaraz G, Bollache L & Marin F (2007) The shell matrix of the freshwater mussel *Unio pictorum* (Paleoheterodonta, Unionoida). Involvement of acidic polysaccharides from glycoproteins in nacre mineralization. *FEBS J* **274**, 2933–2945.
14 Marie B, Marin F, Marie A, Bédouet L, Dubost L, Alcaraz G, Milet C & Luquet G (2009) Evolution of nacre: biochemistry and proteomics of the shell organic matrix of the cephalopod *Nautilus macromphalus*. *ChemBioChem* **10**, 1495–1510.
15 Farre B & Dauphin Y (2009) Lipids from the nacreous and prismatic layers of two pteriomorpha mollusc shells. *Comp Biochem Physiol B* **152**, 103–109.
16 Marin F, Luquet G, Marie B & Medakovic D (2008) Molluscan shell proteins: primary structure, origin, and evolution. *Curr Top Dev Biol* **80**, 209–276.
17 Marin F, Le Roy N & Marie B (2012) The formation and mineralization of mollusc shell. *Front Biosci* **4**, 1099–1125.
18 Miyamoto H, Miyashita T, Okushima M, Nakano S, Morita T & Matsushiro A (1996) A carbonic anhydrase from the nacreous layer in oyster pearls. *Proc Natl Acad Sci USA* **93**, 9657–9660.
19 Joubert C, Piquemal D, Marie B, Manchon L, Pierrat F, Zanella-Cléon I, Cochennec-Laureau N, Gueguen Y

& Montagnani C (2010) Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: focus on biomineralization. *BMC Genomics* **11**, 613.
20 Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R, Rokshar DS, Montagnani C, Joubert C *et al*. (2010) Parallel evolution of nacre building gene sets in molluscs. *Mol Biol Evol* **27**, 591–608.
21 Gardner LD, Mills D, Wiegand A, Leavesley D & Elizur A (2011) Spatial analysis of biomineralization associated gene expression from the mantle organ of the pearl oyster *Pinctada maxima*. *BMC Genomics* **12**, 455.
22 Kinoshita S, Wang N, Maeyama K, Okamoto K, Nagai K, Hirono I, Asakawa S & Watabe S (2011) Deep sequencing of ESTs from nacreous and prismatic layer producing tissues and a screen for novel shell formation-related genes in pearl oyster. *PLoS One* **6**, e21238.
23 Marie B, Marie A, Jackson DJ, Dubost L, Degnan BM, Milet C & Marin F (2010) Proteomic analysis of the organic matrix of the abalone *Haliotis asinina* calcified shell. *Proteome Sci* **8**, 54.
24 Marie B, Le Roy N, Zanella-Cléon I, Becchi M & Marin F (2011) Molecular evolution of mollusc shell proteins: insights from proteomic analysis of the edible mussel *Mytilus*. *J Mol Evol* **72**, 531–546.
25 Marie B, Zanella-Cleon I, Guichard N, Becchi M & Marin F (2011) Novel proteins from the calcifying shell matrix of the Pacific oyster *Crassostrea gigas*. *Mar Biotechnol* **13**, 1159–1168.
26 Marie B, Trinkler N, Zanella-Cleon I, Guichard N, Becchi M, Paillard C & Marin F (2011) Proteomic identification of novel proteins from the calcifying shell matrix of the Manila clam *Venerupis philippinarum*. *Mar Biotechnol* **13**, 965–962.
27 Marie B, Joubert C, Tayale A, Zanella-Cléon I, Belliard C, Cochennec-Loreau N , Piquemal D, Marin F, Gueguen Y & Montagnani C (2012) Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. *Proc Natl Acad Sci* in press. doi:10.1073/pnas.1210552109.
28 Runnegar B (1985) Shell microstructures of Cambrian mollusc replicated by phosphate. *Alcheringa* **9**, 245–257.
29 Carter JG & Clark GR II (1985) Classification and phylogenetic significance of molluscan shell microstructure. In *Classification and Phylogenetic Significance of Mollusk Shell Microstructures* (Broadhead TW, ed.), pp. 50–71. University of Tennessee Press, Knoxville, Tennessee.
30 Suzuki M, Kameda J, Sasaki T, Saruwatari K, Nagasawa H & Kogure T (2010) Characterization of the multilayered shell of a limpet, *Lottia kogamogai* (Mollusca: Patellogastropoda), using SEM–EBSD and FIB–TEM techniques. *J Struct Biol* **171**, 223–230.
31 Suzuki M, Kogure T, Addadi L & Weiner S (2011) Formation of aragonite crystals in the crossed lamellar microstructure of limpet shells. *Cryst Growth Des* **11**, 4850–4859.
32 Fushigami T & Sasaki T (2005) The shell structure of the recent Patellogasteropoda (Mollusca: Gastropoda). *Paleont Res* **9**, 143–168.
33 Mann K, Edsinger-Gonzales E & Mann M (2012) In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*. *Proteome Sci* **10**, 28.
34 Miyamoto H, Miyoshi F & Kohno J (2005) The carbonic anhydrase domain protein nacrein is expressed in the epithelial cells of the mantle and acts as a negative regulator in calcification in the mollusc *Pinctada fucata*. *Zool Sci* **22**, 311–315.
35 Kawasaki K, Buchanan AV & Weiss KM (2009) Biomineralization in humans: making the hard choices in life. *Annu Rev Genet* **43**, 119–142.
36 Marie B, Joubert C, Tayale A, Zanella-Cléon I, Marin F, Gueguen Y & Montagnani C (2012) MRNP34, a novel methionine-rich protein from the pearl oysters. *Amino Acids* **5**, 2009–2017.
37 Timmermans LPM (1969) Studies on shell formation in mollusks. *Neth J Zool* **19**, 417–523.
38 Gesualdo I, Aniello F, Branno M & Palumbo A (1997) Molecular cloning of a peroxidase mRNA specifically expressed in the ink gland of *Sepia officinalis*. *Biochim Biophys Acta* **1353**, 111–117.
39 Moreira Teixeira LS, Feijen J, van Blitterswijk CA, Dijkstra PJ & Karperien M (2011) Enzyme-catalyzed crosslinkable hydrogels: emerging strategies for tissue engeneering. *Biomaterials* **33**, 1281–1290.
40 Nagai K, Yano M, Morimoto K & Myiamoto H (2007) Tyrosinase localization in mollusk shells. *Comp Biochem Biophys B* **146**, 207–214.
41 Rahman AM, Isa Y & Uehara T (2005) Proteins of calcified endoskeleton: II. Partial amino acid sequences of endoskeletal proteins and the characterization of proteinaceous organic matrix of spicules from the alcyonarian, *Sinularia polydactyla*. *Proteomics* **5**, 885–893.
42 Marie B, Luquet G, Bedouet L, Milet C, Guichard N, Medakovic D & Marin F (2008) Nacre calcification in the freshwater mussel *Unio pictorum*: carbonic anhydrase activity and purification of a 95 kDa calcium-binding glycoprotein. *ChemBioChem* **9**, 2515–2523.
43 Tambutté S, Tambutté E, Zoccola D, Caminiti N, Lotto S, Moya A, Allemand D & Adkins J (2007) Characterization and role of carbonic anhydrase in the calcification process of the azooxanthellate coral *Tubastrea aurea*. *Mar Biol* **151**, 71–83.

44 Jackson DJ, Macis L, Reitner J, Degnan BM & Wörheide G (2007) Sponge paleogenomics reveals an ancient role for carbonic anhydrase in skeletogenesis. *Science* **216**, 1893–1895.
45 Mann K, Poustka AJ & Mann M (2008) In-depth, high-accuracy proteomics of sea urchin tooth organic matrix. *Proteome Sci* **6**, 33.
46 Hewett-Emmett D & Tashian RE (1996) Functional diversity, conservation, and convergence in the evolution of the alpha-, beta-, and gamma-carbonic anhydrase gene families. *Mol Phyl Evol* **5**, 50–77.
47 Kyte J & Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132.
48 Meenakshi VR, Hare PE & Wilbur KM (1971) Amino acids of the organic matrix of neogastropod shells. *Comp Biochem Physiol* **40**, 1037–1043.
49 Weiner S & Hood L (1975) Soluble proteins of the organic matrix of mollusk shells: a potential template for shell formation. *Science* **190**, 987–989.
50 Weiner S (1979) Aspartic acid-rich proteins: major components of the soluble organic matrix of mollusk shells. *Calcif Tissue Int* **29**, 163–167.
51 Weiner S (1983) Mollusk shell formation: isolation of two organic matrix proteins associated with calcite deposition in the bivalve *Mytilus californianus*. *Biochemistry* **22**, 4139–4145.
52 Marin F, Amons R, Guichard N, Stiger M, Hecker A, Luquet G, Layrolle P, Alcaraz G, Riondet C & Woestbroek P (2005) Caspartin and calprismin, two proteins of the shell calcitic prisms of the mediterranean fan mussel *Pinna nobilis*. *J Biol Chem* **280**, 33895–33908.
53 Marin F & Luquet G (2007) Unusually acidic proteins in biomineralization. In Handbook of Biomineralization Vol. 1: The Biology of Biominerals Structure Formation (Baeuerlein E, ed.), pp. 273–290. Wiley-VCH, Weinheim, Germany.
54 Nudelman F, Gotliv B, Addadi L & Weiner S (2006) Mollusk shell formation: mapping the distribution of organic matrix components underlying a single aragonitic tablet in nacre. *J Struct Biol* **153**, 176–187.
55 Sarashina I & Endo K (1998) Primary structure of a soluble matrix protein of the scallop shell: implications for calcium carbonate biomineralization. *Am Miner* **83**, 1510–1515.
56 Suzuki M, Saruwatari K, Kogure T, Yamamoto Y, Nishimura T, Kato T & Nagasawa H (2009) An acidic matrix protein, Pif, is a key macromolecules for nacre formation. *Science* **325**, 1388–1390.
57 Gotliv B-A, Kessler N, Sumerel JL, Morse DE, Tuross N, Addadi L & Weiner S (2005) Asprich: a novel aspartic acid-rich protein family from the prismatic shell matrix of the bivalve *Atrina rigida*. *ChemBioChem* **6**, 304–314.
58 Tsukamoto D, Sarashina I & Endo K (2004) Structure and expression of an unusually acidic matrix protein of pearl oyster shells. *Biochem Biophys Res Commun* **320**, 1175–1180.
59 Addadi L, Joester D, Nudelman F & Weiner S (2006) Mollusk shell formation: a source of new concepts for understanding biomineralization processes. *Chem Rev* **12**, 980–987.
60 Suzuki M, Iwasima A, Tsutsui N, Ohira T, Kogure T & Nagasawa H (2011) Identification and characterisation of a calcium carbonate-binding protein, blue mussel shell protein (BMSP), from the nacreous layer. *ChemBioChem* **16**, 278–287.
61 Inoue K, Takeuchi Y, Miki D & Odo S (1995) Mussel adhesive plaque protein gene is a novel member of epidermal growth factor-like gene family. *J Biol Chem* **270**, 6698–6701.
62 Mann K, Poustka A & Mann M (2010) Phosphoproteomes of *Strongylocentrotus purpuratus* shell and tooth matrix: identification of a major acidic sea urchin tooth phosphoprotein, phosphodontin. *Proteome Sci* **8**, 6.
63 Maurer P & Hohebester E (1997) Structural and functional aspects of calcium binding in extracellular matrix proteins. *Matrix Biol* **15**, 569–580.
64 Jovine L, Darie CC, Litscher ES & Wassarman PM (2005) Zona pellucida domain proteins. *Annu Rev Biochim* **74**, 83–114.
65 Mo L, Hunag HY, Zhu XH, Sapiro E, Hasty DL & Wu WR (2004) Tamm–Horsfall protein is a critical renal defense factor protecting against calcium oxalate crystal formation. *Kidney Int* **66**, 1159–1166.
66 Choi JW, Sutor SL, Lindquist L, Evans GL, Madden BJ, Bergen HR, Hefferan TE, Yaszemski MJ & Bram RJ (2009) Severe osteogenesis imperfect in cyclophilin B-deficient mice. *PLoS Genet* **5**, e1000750.
67 Xue QG, Itoh N, Schey KL, Li YL, Cooper RK & La Peyre JF (2007) A new lysozyme from the eastern oyster (*Crassostrea virginica*) indicates adaptative evolution of I-type lysozymes. *Cell Mol Life Sci* **64**, 82–95.
68 Goffinet G & Jeuniaux C (1979) Distribution and quantitative importance of chitin in Mollusca shells. *Cah Biol Mar* **20**, 341–349.
69 Zentz F, Bédouet L, Almeida MJ, Milet C, Lopez E & Giraud M (2001) Characterization and quantification of chitosan extracted from nacre of the abalone *Haliotis tuberculata* and the oyster *Pinctada maxima*. *Mar Biotechnol* **1**, 36–44.
70 Palpandi C, Shanmugam V & Shanmugem A (2009) Extraction of chitin and chitosan from shell and operculum of mangrove gastropod *Nerita crepidularia* Lamarck. *Int J Med Medic Sci* **1**, 198–205.
71 Kretzinger RH (1976) Calcium-binding proteins. *Annu Rev Biochem* **45**, 239–266.

72 Weiss IM, Göhring W, Fritz M & Mann K (2001) Perlustrin, a *Haliotis laevigata* (abalone) nacre protein, is homologous to the insulin-like growth factor binding protein N–terminal module of vertebrates. *Biochem Biophys Res Commun* **285**, 244–249.
73 Treccani L, Mann K, Heinemann F & Fritz M (2006) Perlwapin, an abalone nacre protein with three-four-disulfide core (whey acidic protein) domains, inhibits the growth of calcium carbonate crystals. *Biophys J* **91**, 2601–2608.
74 Shen X, Belcher AM, Hansma PK, Stucky GD & Morse DE (1997) Molecular cloning and characterization of lustrin–A, a matrix protein from shell and pearl nacre of *Haliotis rufescens*. *J Biol Chem* **272**, 32472–32481.
75 Khalturin K, Hemmrich G, Fraune S, Augustin R & Bosch T (2009) More than just orphans: are taxonomically-restricted genes important in evolution? *Trend Genet* **35**, 404–413.
76 Gatesy J, Hayashi C, Motruik D, Woods J & Lewis R (2001) Extreme diversity, conservation, and convergence of spider silk fibroin sequences. *Science* **291**, 2603–2605.
77 Sagert J & Waite JH (2009) Hyperunstable matrix protein in the byssus of *Mytilus galloprovincialis*. *J Exp Biol* **212**, 2224–2236.
78 Waring GL, Hawley RJ & Schoenfejd T (1990) Multiple proteins are produced from *DEC–1* eggshell gene in *Drosophila* by alternative RNA splicing and proteolytic cleavage events. *Dev Biol* **142**, 1–12.
79 Sudo S, Fujikawa T, Nagakura T, Ohkubo T, Sakaguchi K, Tanaka M, Nakashima K & Takahashi T (1997) Structure of mollusc shell framework proteins. *Nature* **387**, 563–564.
80 Yano M, Nagai K, Morimoto K & Miyamoto H (2006) Shematrin: a family of glycin-rich structural proteins in the shell of the pearl oyster. *Comp Biochem Physiol* **144**, 254–262.
81 Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* **27**, 527–533.
82 Smith BL, Schäffer TE, Viani M, Thompson JB, Frederick NA, Kindt J, Belcher AM, Stucky GD, Morse DE & Hansma PK (1999) Molecular mechanistic origin of the toughness of natural adhesives, fibres and composites. *Nature* **399**, 761–763.
83 Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander C, Moroz LL, Lied B *et al.* (2011) Phylogenomics reveals deep molluscan relationships. *Nature* **477**, 452–456.
84 Smith SA, Wilson NG, Goetz FE, Feehery C, Andrade SCS, Rouse GW, Giribert G & Dunn CW (2011) Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* **480**, 364–367.
85 Bédouet L, Marie A, Berland S, Bordenave-Auzoux S, Marie B, Marin F & Milet C (2008) Identification of novel shell matrix proteins in *Haliotis tuberculata* by proteomics using trypsin digestion and protein ladder sequencing. *Mar Biotechnol* **414**, 446–458.
86 Maren TH (1960) A simplified micromethod for the determination of carbonic anhydrase and its inhibitiors. *J Pharmacol Exp* **130**, 26–29.
87 Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
88 Wall RS & Gyi TJ (1988) Alcian Blue staining of proteoglycans in polyacrylamide gels using the 'critical electrolyte concentration' approach. *Anal Biochem* **175**, 298–299.
89 Notredame C, Higgins DG & Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205–217.
90 Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M *et al.* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acid Res* **36**, 465–469.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:

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.