The shell matrix and microstructure of the Ram's Horn squid: molecular and structural characterization.

Morgane Oudot, Pascal Neige, Ira Ben Shir, Asher Schmidt, Jan M. Strugnell, Laurent Plasseraud, Cédric Broussard, René Hoffmann, Alexander Lukeneder, Frédéric Marin

The shell matrix and microstructure of the Ram's Horn squid: molecular and structural characterization

Morgane Oudot[1], Pascal Neige[1], Ira Ben Shir[2], Asher Schmidt[2], Jan M. Strugnell[3,4], Laurent Plasseraud[5], Cédric Broussard[6], René Hoffmann[7], Alexander Lukeneder[8] and Frédéric Marin[1*]

Authors affiliations:

[1] UMR CNRS 6282 Biogeosciences, University of Burgundy - Franche-Comté, 6 Boulevard Gabriel, 21000 DIJON, France.

[2] Schulich Faculty of Chemistry and Russell Berrie Nanotechnology Institute Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel.

[3] Centre for Sustainable Tropical Fisheries and Aquaculture and College of Science and Engineering, James Cook University, Townsville, Qld, 4810, Australia

[4] Department of Ecology, Environment and Evolution, School of Life Sciences, La Trobe University, Kingsbury Drive, Melbourne, Vic. 3086, Australia

[5] Institute of Molecular Chemistry, ICMUB UMR CNRS 6302, University of Burgundy -Franche-Comté, 9 Avenue Alain Savary, Dijon, France.

[6] 3P5 Proteomic Platform, Cochin Institute, 22 Rue Méchain, Paris, France. 3P5 Proteomic Platform, University of Paris, Cochin Institute, INSERM, U1016, CNRS, UMR8104, F-75014 PARIS, France

[7] Branch Paleontology, Institute of Geology, Mineralogy, and Geophysics, Dpt Earth Sciences, Ruhr-Universität Bochum, Universitätsstrasse 150, 44801 Bochum, Germany.

[8] Natural History Museum Vienna, Burgring 7, 1010 Vienna, Austria.

[*] To whom correspondence should be sent.

Abstract

Molluscs are one of the most diversified phyla among metazoans. Most of them produce an external calcified shell, resulting from the secretory activity of a specialized epithelium of the calcifying mantle. This biomineralization process is controlled by a set of extracellular macromolecules, collectively defined as the organic matrix. In spite of several studies, these components are mainly known for bivalve and gastropod classes. In the present study, we investigated the physical and biochemical properties of the internal planispiral shell of the understudied Ram's Horn squid *Spirula spirula* (Cephalopoda, Decabrachia, Spirulida). Scanning Electron Microscope investigations of the shell reveal a complex microstructural organization, with septa sandwiched into the shell wall, in the form of a bevel. The saccharides constitute a quantitatively important moiety of the matrix, as shown by Fourier-transform infrared spectroscopy. Solid-state nuclear magnetic resonance spectroscopy identified *β*-chitin and additional polysaccharides, for a total amount of 80% of the insoluble fraction. Proteomics was applied to both soluble and insoluble matrices and *in silico* searches were performed, first on heterologous metazoans models, and secondly, on an unpublished transcriptome of *Spirula spirula*. In the first case, several peptides were identified, some of them matching with tyrosinase, chitinase 2, protease inhibitor, or immunoglobulin. In the second case, 38 hits were obtained, including transferrin, a serine protease inhibitor, matrilin, different histone-like, a α–2-macroglobulin or a putative heme-binding/calcium-binding protein. The very few similarities with known molluscan shell matrix proteins suggest that *Spirula spirula* uses a unique set of shell matrix proteins for constructing its internal shell. The absence of similarity with closely related cephalopods such as the cuttlefish *Sepia* demonstrates that there is no obvious phylogenetic signal in the skeletal matrix of cephalopods.

Introduction

Molluscs, including chitons, oysters, mussels, snails, slugs, octopus and squids, constitute one of the most diversified phyla in the animal kingdom on the basis of shape, anatomy, size, development, habitat or number of living and fossil species (Bouchet et al., 2016; Mora et al., 2011; Ponder et al., 2019; Vinther, 2015). Many molluscs possess an external calcified shell, which protects and anchors the soft tissues of the organisms. This acellular multi-layered mineralized exoskeleton results from the secretory activity of a specialized

epithelium that delineates the mantle, the ciliated organ that coats the inner surface of the shell (Checa, 2018; Simkiss and Wilbur, 1989). The shell fabrication requires specific cellular and molecular machinery: the mineral deposition is indeed controlled by an extracellular organic matrix, which is subsequently occluded. Since its discovery as a key-player in regulating shell biomineralization, the organic matrix has been the subject of a large number of biochemical analyses focused on its protein moieties (Crenshaw, 1972; Hare, 1963; Marin and Luquet, 2004; Rusenko et al., 1991; Sarashina et al., 2006; Weiner, 1983). Most of these studies concentrated their efforts on bivalves and gastropods, the two modern largest mollusc classes. Far less studied are the cephalopods. They represent however an important class, with about 800 living and more than 10 000 fossil species (http://cephbase.eol.org/; Nishiguchi and Mapes, 2008). The fossil record indicates that the first indisputable cephalopods appeared in the Middle Cambrian (Kröger et al., 2011) but they may have been present as early as the Lower Cambrian with *Tannuella elinorae* (~530 Ma, Brock and Paterson, 2004; Kröger et al., 2011; Vinther, 2015). Since then, cephalopods experienced a complex evolutionary history with important Ordovician, Devono-Carboniferous, Triassic and Cenozoic radiations and major crises, including the end-Permian and end-Cretaceous mass extinctions (Brayard et al., 2009; Murray, 1985; Nishiguchi and Mapes, 2008). From a "biomineralization" viewpoint, cephalopods have a singular evolutionary trajectory (Klug et al., 2019), with a general trend to shell size reduction and internalization (cuttlefish bone), a drastic mineralization decrease (the organic squid pen), leading ultimately to complete loss of skeletal structures, in the most recently appeared cephalopods, *i.e.*, octopuses. Although the clade is commonly recognized as monophyletic, several uncertainties remain regarding the phylogenetic relationships among the major lineages of extant coleoid cephalopods, including that of the internally shelled sepiids (cuttlefishes) and *Spirula* (Lindgren and Anderson, 2018; Strugnell et al., 2017; Tanner et al., 2017). In addition, the molecular mechanism of shell formation remains poorly known for these taxa.

In this general framework, we focused on the Ram's Horn squid *Spirula spirula* (Linné, 1758). This species is a small-sized mesopelagic cephalopod (Coleoidea, Decabrachia), rarely observed in its natural environment. It is considered as a monospecific genus (Haring et al., 2012; Neige and Warnke, 2010) and even a monospecific order, whose shell exhibits peculiar characteristics. Some of them are reminiscent of the distantly related *Nautilus*, such as planispiral coiling and a chambered shell where chambers are separated by concave septa and communicate via a siphuncle, a primordial morphology (plesiomorphic character) within cephalopods. However, the siphuncle of *S. spirula* is in proximal position, while that of *Nautilus*

is more centered in the septa. Furthermore, in contrast to *Nautilus sp.*, the shell of *S. spirula* is fully internal and its microstructure is not nacreous (as it is in all nautiloids), both characters being considered as derived. These features, among others, are shared with some other coleoids, such as cuttlefishes and relatives (Kröger et al., 2011; Strugnell et al., 2017).

Even though the biology of *S. spirula* remains largely elusive (Ohkouchi et al., 2013), it is quite common to find shells on beaches of the Atlantic, Indian and West Pacific oceans (Bruun, 1943, Clarke, 1970) and different studies have investigated their microstructure (Cuif *et al.*, 1983, Bandel and Boletzky v., 1979, Dauphin, 1976, 1977). Nevertheless, the shell biochemistry has been explored only very superficially (Dauphin, 1996; Dauphin and Marin, 1995; Degens et al., 1967). The aim of this work is to determine the biochemical characteristics of the shell matrix of *S. spirula*, with a view to understanding its multiple molecular functions in biomineralization. The second objective of our study is to compare our data with that obtained from better known representatives of the class Cephalopoda and, by extension, of the phylum Mollusca, in order to draw macroevolutionary conclusions.

Material and Methods

Shell sampling

All *Spirula spirula* shell samples used in this study were collected on the beach. They were provided by four of us (R. H., J. M. S., P. N. and A. L.). For subsequent analyses, we used complete shells (Fig. 1) or almost complete, with the exception of the first chamber, frequently

Figure 1: Spirula spirula. A, specimen from East Australian coast (Townsville, Queensland). B, specimen from Thailand. Note that the first specimen still possesses its initial chamber, while it is missing in the second one.

114 missing. Two batches, consisting of dozens of shells were collected in two locations: the
115 beaches of Lanzarote (Playa de Famara), Canary Islands (R. H., sampling in November 2005),
116 and the beaches of Tamandaré, Pernambuco State, Brazil (sampling in July 2007 and donation
117 to P. N.). Additional material in lesser quantities came from East Australian coast (Townsville,
118 Queensland) (J. M. S.) and from Maldives Islands, Thailand, and Canary Islands (A. L.).
119

120 *Shell embedding and microstructure analysis*

121 Shell embedding was performed according to the manufacturer's instructions (Fluka):
122 briefly, benzoyl peroxide (catalyst) was dissolved in the LR White resin monomer solution (9.9
123 g per 500 mL) for 24 hours, at room temperature, under constant stirring. Complete dried
124 specimens were immersed in sample vials containing the resin solution at room temperature,
125 subsequently placed for a few hours in a desiccator under vacuum. The preparations were
126 incubated at 60 °C to allow polymerization (three days). The resin blocks were sliced with a
127 saw microtome (Leica SP1600). Shell sections were mirror-polished (0.05 µm aluminium oxide
128 paste) and slightly etched with EDTA solution (1% wt/vol) for two minutes in an ultrasonic
129 bath, before being rinsed, dried and carbon or gold-coated. The samples were observed with a
130 Hitachi TM-1000 tabletop SEM or with a JEOL JSM-IT 100 InTouchScope™.

131 *Shell matrix extraction*

132 The shells were coarsely powdered with a mortar and bleached in a 10-times diluted
133 sodium hypochlorite solution (Merck, 6-14% active chlorine, ref 105614) for 17 to 20 hours.
134 After several rinsing steps in ultrapure water, the powder was dried at 37 °C and finely ground
135 manually then divided into two batches. The first one was immediately weighed and decalcified
136 overnight at 4 °C by adding progressively cold dilute acetic acid (10% vol/vol) at a flow rate of
137 0.1 mL every 5 sec, under constant stirring (according to the protocol described by Pavat et al.,
138 2012). The second batch was bleached in sodium hypochlorite solution (same concentration)
139 for additional 17-20 hours, rinsed, dried then weighed and decalcified similarly to the first
140 batch.
141 For each batch, the clear solution was centrifuged (3900 G, 30 min). The supernatant
142 containing the acid soluble matrix (ASM) was filtered (5µm, Millipore) on a Nalgene filtration
143 device before being ultrafiltered for volume reduction on an Amicon cell (400 mL; cutoff 10
144 kDa). The solution (10-15 mL) was then extensively dialyzed against Milli-Q water for three

days (6 water changes) and lyophilized. In parallel, the precipitate (*i.e.* pellet) containing the acid insoluble matrix (AIM) was rinsed via 5 cycles of resuspension in Milli-Q water - centrifugation, each supernatant being added to the ASM. Finally, the AIM pellet was lyophilized. All the lyophilisates were weighed on a precision balance (Sartorius Quintix35-1S) three to five times, and the mean value was calculated. In total, three extractions from the Canaries samples were performed, and two from the Brazil ones, leading each time to four batches of extracts, both ASM ($ASM1_b$ $ASM2_b$, $ASM1_c$ $ASM2_c$) and AIM ($AIM1_B$ $AIM2_B$, $AIM1_C$ $AIM2_C$).

Fourier Transform Infra-Red spectroscopic characterization (FT-IR)

FT-IR spectroscopy was performed to identify the shell mineralogy of *S. spirula*, as well as the overall chemical characteristics of the extracted matrices (ASMs and AIMs). In both cases, minute amounts of samples (shell powder or freeze-dried chips (<1 mg) of ASM and AIM) were analysed with an FT-IR Bruker Alpha spectrometer (Bruker Optics Sarl, Marne-la-Vallée, France) fitted with an Attenuated Total Reflectance (ATR) ALPHA-P device equipped with a mono-reflection diamond crystal in the 4000-375 cm^{-1} wavenumber range (24 scans at a spectral resolution of 4 cm^{-1}). The qualitative assignment of absorption bands was performed by comparison with previous spectra descriptions, achieved by our group or available in the literature. An additional spectrum was acquired from commercial crab chitin (Sigma, C9752).

Solid State Nuclear Magnetic Resonance (SS-NMR)

^{13}C (75.4 MHz) and ^{31}P (121.85 MHz) Magic Angle Spinning (MAS) NMR measurements were carried out on 300 MHz solid state NMR Bruker *AVANCE* III spectrometer using a 4 mm triple-resonance MAS NMR probe with zirconia rotors. Samples were spun at 10,000 ± 2 Hz. Two excitation techniques were employed: DE - direct excitation, and CP - Cross Polarization (CP), both using rotor matched chemical shift echo (Stejskal et al., 1977). All experiments employed: 5.0 μs π/2, 10.0 μs π pulse widths, and an echo interval of 100 μs identical to the rotor period T_R; TPPM ^1H decoupling (Bennett et al., 1995) with rf field strength of 100 kHz. The Hartmann-Hahn CP employed 50 kHz rf level for the X-channel (^{13}C, ^{31}P) and 50-70 kHz ramped rf level for the ^1H channel with 1 and 2 ms duration (contact time, ct) for ^{13}C and ^{31}P respectively. Experiments were repeated using relaxation delays of 3 sec (^{13}C) acquiring 8k and 40k transients (Canaries and Brazil samples, respectively) and 3 sec for ^{31}P acquiring 2k transients. Complementary ^{13}C DE echo MAS NMR measurements were carried

out for the bigger Canaries sample acquiring 64 transients with 200 sec relaxation delay (fully relaxed).

Two *Spirula spirula* AIM1 extracts - from the Canaries (16.8 mg, $AIM1_C$) and from Brazil (6.0 mg, $AIM1_B$), and a reference chitin standard (Sigma, ref. C9752) were subjected to solid-state MAS NMR measurement. CP MAS is the method of choice to effectively detect organic matter as it takes advantage of the abundant hydrogens as a source of signal, so highlighting all nearby carbons (below 10 Å) and yielding spectra with high sensitivity (high signal to noise ratio, S/N). Although the quantitative DE MAS spectra are inherently of lower sensitivity, here they were obtained to validate that all peaks (species) are accounted for in the CP spectra.

Mono-dimensional PAGE

The shell extracts were analyzed by conventional mono-dimensional denaturing SDS-PAGE (Bio-Rad, Mini Protean III gels), on precast gradient gels (Mini-PROTEAN TGX Gel 4-15% acrylamide, 90 x 70 mm, Bio-Rad) or on homemade gels (12% polyacrylamide). Prior to migration, the samples were treated as follows: lyophilisates (both ASM and AIM) were dissolved in 2X Laemmli Sample Buffer (LSB) to a final matrix concentration of 5 µg/µL. Both preparations were heat-denatured at 100 °C: 5 min. for the ASM and 10 min. for the AIM. Note that the AIMs are not fully solubilized by the Laemmli treatment: only the soluble fractions, referred as Laemmli-Soluble AIMs or LS-AIMs, can be visualized on gels. The preparations were cooled down and shortly centrifuged before being applied on the top of the gel. Subsequently, gels were stained by silver nitrate (Morrissey, 1981), carbocyanine, *i.e.*, "Stains-all" (Campbell et al., 1983) and with Bio-Rad blue dye. Other staining methods, such as PAS for glycoproteins and Alcian blue for polyanionic substances, were applied too, but did not produce significant results.

Protein-calcium interaction

The effect of the *Spirula spirula* ASM on the growth of calcium carbonate crystals was tested *in vitro*, as previously described (Albeck et al., 1993) with some modifications (Pavat et al., 2012). In this assay, crystal growth is obtained by slow diffusion of NH_4HCO_3 vapor into the $CaCl_2$ solution. To this end, a 10 mM solution of $CaCl_2$ solutions was prepared fresh and filtered (5 µm pore size) and used for serial dilutions of ASM in the range [0.25-16 µg/well,

i.e., 1.25 to 80 µg/mL]. The solutions were incubated in a 16-well culture slide (Lab-Tek, Nunc/Thermo Scientific, Rochester, NY, USA, 200 µL per well), which was covered with a plastic cover that had been pierced in the middle of each well to allow vapor diffusion. The whole system was sealed using parafilm "M" laboratory film (Neenah, WI, USA). Finally, the slide was placed in a desiccator containing ammonium bicarbonate crystals that was subsequently closed and placed under vacuum. In parallel, blank controls were performed without ASMs.

After incubation, the slides were carefully dried by gently removing the solution using a blunt-ended syringe needle connected to a vacuum pump. Slides were dis-assembled and the glass plate was observed using a TM1000 tabletop scanning electron microscope (Hitachi) in back-scattered electron mode. To guarantee the reproducibility of the results, the crystallization assay was performed 5 times, each time with a blank control.

Enzyme-Linked Lectin Assay (ELLA)

Screening of *Spirula spirula* ASMs by Enzyme-Linked Lectin Assay (ELLA) was performed using a set of 21 biotinylated lectins (Vector Labs, Kits I to III, ref. BK-1000, BK-2000, BK-3000 respectively). The binding preferences and specificities of each lectin are indicated in Kanold et al. (2015), which compiles data from different bibliographical sources.

The test was carried out in 96-well microplates (MaxiSorp, NuncTM, Rockslide, Denmark): the *Spirula* ASM diluted in TBS buffer (200 ng/well) was incubated for 90 min at 37 °C, immediately followed by three washing steps with TBS/Tween20. Wells were blocked with 1X concentrated Carbo-free blocking solution (Vector Labs, ref. SP-5040) for at least 30 min at 37 °C. The diluted lectins (in TBS/Tween20) were incubated into the microplate for 90 min at 37 °C, at the following dilution factors: 200 times for kits I and II, and 100 times for kit III. After three washes with TBS/Tween20, the microplates were incubated with a diluted solution of Avidin-Alkaline Phosphatase conjugate (1:70000, ref. Sigma A7294, St. Louis, MO, USA) for 90 min at 37 °C, for the detection of bound biotinylated lectins. They were then washed five times with TBS/Tween20 and revealed with a substrate solution containing aqueous diethanolamine (9.7% vol/vol, pH adjusted to 9.8 with HCl) and phosphatase substrate (0.5 mg/mL, pNPP Tablet, Sigma, St. Louis, MO, USA). The microplates were read at 405 nm every 15 minutes (BioRad Model 680). A check of the background signal without ASM, lectin or Avidin-AP showed no reaction with the substrate solution. This test was performed with quadruplicates of each lectin and the results were normalized and translated in percentage of

reactivity by subtracting the background and considering the highest response as 100%. The test was performed three times.

RNA extraction, transcriptome sequencing and assembly

Tissue samples, collected from a *Spirula spirula* individual captured from the Coral Sea, Australia, were stored in RNAlater (Qiagen) at -80ºC. RNA was extracted using an RNeasy Mini Kit (Qiagen) following the manufacturer's protocol. The sample was disrupted using a desktop homogenizer (IKA). RNA quality and quantity were measured using a Bioanalyzer (Agilent). The transcriptome was sequenced on a single lane of an Illumina HiSeq2000 by the Australian Genome Research Facility (AGRF). The transcriptome was assembled from all sequencing data using the assembly program Trinity (Version 2.2, Grabherr et al., 2011). Default parameters were used with data normalization. The *Spirula spirula* transcriptome assembly included 89,625 contigs and had an average contig length of 625.91 nucleotides. A reference protein database containing 16,405 amino acid sequences was constructed by predicting proteins from the transcriptome using TransDecoder followed by clustering highly similar sequences (95% identity) with cdhit. Sequencing information is available at NCBI under the BioProject (reference to come).

Proteomic analysis of the shell extracts and subsequent in silico analysis

MS/MS analyses were conducted on the four unfractionated bulk matrices, ASM and AIM, which were digested in-gel, after a short migration in an 8 % acrylamide gel stained with Coomassie blue. In-gel digestions were carried out with trypsin. Samples were destained twice with a mixture of 100 mM ammonium bicarbonate (ABC) and 50 % (v/v) acetonitrile (ACN) for 30 min at 22 °C and then dehydrated using 100 % ACN for 15 min, before being reduced with 25 mM ABC containing 10 mM DTT for 1 h at 56 °C and alkylated with 55 mM iodoacetamide in 25 mM ABC for 30 min in the dark at 22 °C. Gel pieces were washed twice with 25 mM ABC and dehydrated (twice, 20 min) with 100 % ACN. Gel cubes were incubated with sequencing grade modified trypsin (Promega, USA; 12.5 ng/µl in 40 mM ABC with 10 % ACN, pH 8.0) overnight at 37 °C. After digestion, peptides were extracted twice with a mixture of 50 % ACN – 5 % formic acid (FA) and then with 100 % ACN. Extracts were dried using a vacuum centrifuge Concentrator plus.

For MS and MS/MS ORBITRAP, analyses were performed using an Ultimate 3000 Rapid Separation Liquid Chromatographic (RSLC) system (Thermo Fisher Scientific) online with a Q-Exactive Plus hybrid quadrupole Orbitrap mass spectrometer (Thermo Fisher Scientific). Briefly, peptides were dissolved in 10 µL of 10 % ACN - 0.1 % trifluoroacetic acid (TFA). Then, peptides were loaded and washed on a C_{18} reverse phase pre-column (3 µm particle size, 100 Å pore size, 75 µm i.d., 2 cm length). The loading buffer contained 98 % H_2O, 2 % ACN and 0.1 % TFA. Peptides were then separated on a C_{18} reverse phase resin (2 µm particle size, 100 Å pore size, 75 µm i.d., 25 cm length) with a 1 hour gradient from 99 % A (0.1 % FA and 100 % H_2O) to 90 % B (80 % ACN, 0.085 % FA and 20 % H_2O).

The mass spectrometer acquired data throughout the elution process and operated in a data dependent scheme with full MS scans acquired with the Orbitrap, followed by up to 10 MS/MS HCD spectra. Mass spectrometer settings were full MS (AGC: $3x10e^6$, resolution: $7x10e^4$, m/z range 400-2000, maximum ion injection time: 100 ms); MS/MS (AGC: $1x10e^5$, maximum injection time: 100 ms, isolation windows: 4m/z Da, dynamic exclusion time setting: 15 s). The fragmentation was permitted for precursors with a charge state of 2, 3, 4 and up. For the spectral processing, the software used to generate .mgf files is Proteome discoverer 1.4.

Database searches were carried out using Mascot version 2.5 (MatrixScience, London, UK) on 'Other Metazoa' proteins (2,282,777 sequences) from NCBInr databank containing 67,337,701 sequences; 24,122,812,982 residues (April 2018) (http://www.ncbi.nlm.nih.gov/). A second search was performed with the 16405 sequences of the protein-translated transcriptome built from the mantle tissue of *Spirula spirula*. In each case, the enzyme specificity was trypsin's and up to one missed tryptic cleavage was tolerated. The precursor mass tolerance was set to 4 ppm and the fragment mass tolerance to 20 mmu for Q-Exactive data. Carbamidomethylation of cysteins and oxidation of methionines were set as variable modifications. Hits with human keratines were ignored.

The hits found with the *Spirula spirula* transcriptome were analysed according to different bioinformatic tools: the identified proteins were blasted against the NCBI dataset "Mollusca", using either BLASTp or Smart BLAST at NCBI; the descriptive protein parameters (molecular weight, isoelectric point, amino acid composition) were assessed with ProtParam and the presence of signal peptide, with SignalP, both tools sheltered by the Expasy Bioinformatics Resources Portal in proteomics (https://www.expasy.org/proteomics). Sequences from the *Spirula spirula* database that were identified by proteomics were all checked for the presence of conserved domains with CD-Search at NCBI (Lu et al., 2020).

Results

Shell microstructures

The different shell microstructures identified in the *S. spirula* are shown in Figure 2. SEM observations confirmed that the shell wall is constituted of two layers: inner and outer. However, we noticed that the internal one is irregular semi-prismatic (Fig. 2C), while the external one appears as tuberculated and regular semi-prismatic (Fig. 2D), with well-developed spicular crystals. Between these two layers, we can observe a thin blackish layer. Moreover, the septa show a lamello-fibrillar pattern (Fig. 2E), which differs from a nacreous microstructure *sensu stricto*. We also noticed that the septa progressively insert into the shell wall in the form of a bevel. An observation made at higher magnification revealed that the internal layer progressively changes into the septa (Fig. 2E) with very gradual transition from

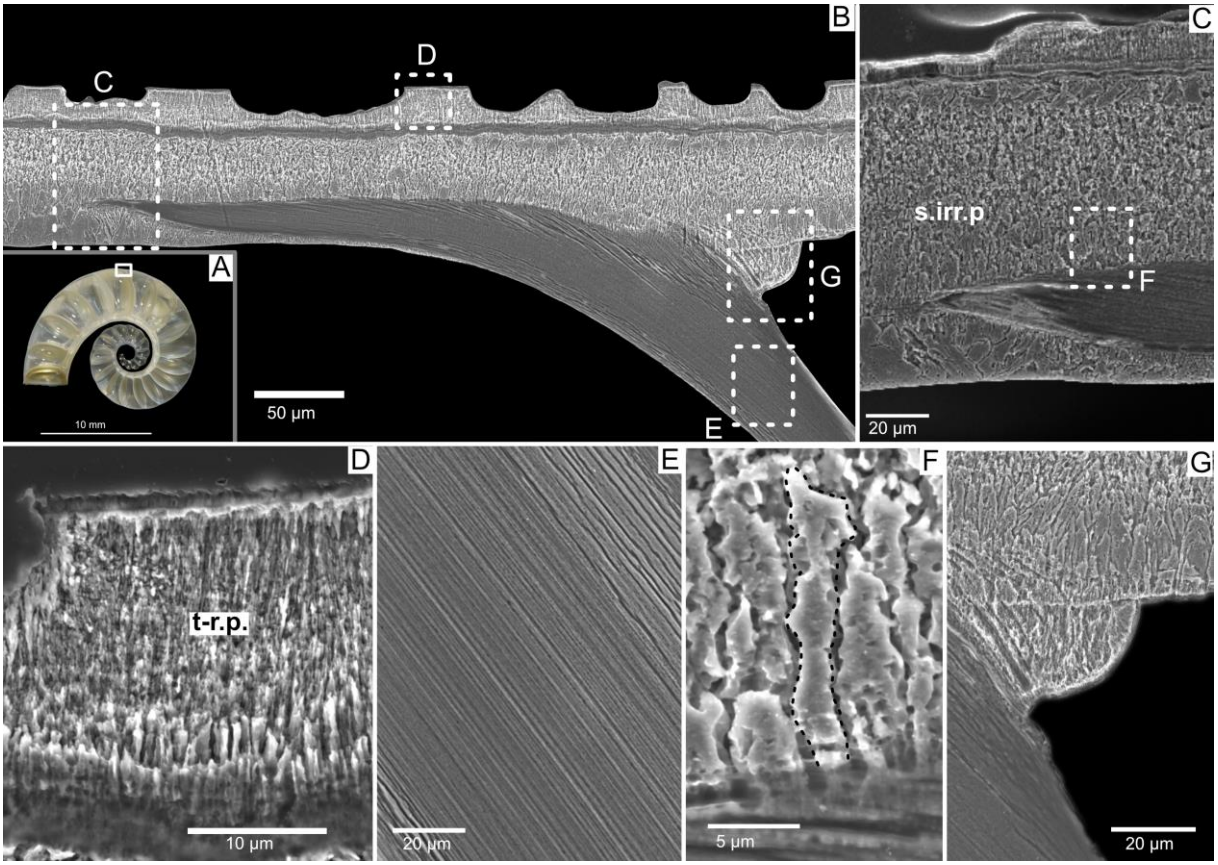

Figure 2: Shell microstructures of Spirula spirula, in the insertion zone of a septum into the shell wall. A) General view, equatorial section. B) General view of the septum insertion in the outer wall. Note the beveled shape of the whole structure. C) Zoom on the very beginning of the septum into the shell outer wall, showing its organization into an internal unit, with an irregular semi-prismatic layer (s.irr. p). D) Zoom on the external unit of the shell wall, showing its tuberculated and regular semi-prismatic structure (t-r.p). E) Zoom on the lamello-fibrillar structure of the septum. F) Zoom on the contact zone between the septum and the internal shell unit, at the very beginning of the insertion, showing the transformation from the irregular semi-prismatic structure into the lamello-fibrillar one. G) Zoom on the supraseptal annular ridge, observed on the aboral side, dorsal corner, of the septum. In spite of a thin delimitation between the annular ridge and the internal unit, one notices a microstructural continuity.

the prisms to the lamello-fibrillar aspect of the septa. Finally, we noticed a little bulge in the aboral side of the septa, in the top corner between them and the shell wall (Fig. 2F).

Quantification of the organic matrix in the shell

Matrix quantification of the last extraction is shown in Table 1. However, we did not detect any significant variations between the different extractions, in term of absolute quantities of matrices, both ASM and AIM. Thus, we can consider that this quantification reflects accurately the amount of organics in the shell of *S. spirula*. After a single bleaching step, the composition of the matrix is dominated by acetic acid-insoluble components (AIM) that represent between two-third and four-fifth of the total matrix (around 70%). The latter accounts for about 1% of the shell powder weight (0.7 and 1.09%, respectively). The second bleaching has a drastic effect on the AIM, since the weight values of this matrix collapse to unquantifiable amounts for the Brazil samples. The effect is less pronounced for the ASM, the absolute quantity of which is divided by a factor 2 to 2.5.

FT-IR spectroscopy

FT-IR (ATR) spectra acquired from eight extracts of AIM and ASM fractions are shown

		Initial weight (g)	ASM extract (mg)	%ASM	AIM extract (mg)	%AIM
Canary Islands	Batch 1	2.0832	5.36	0.25%	22.6	1.09%
	Batch 2	2.0315	2.00	0.1%	0.35	0.01%
Brazil	Batch 1	1.6093	6.24	0.39%	11.2	0.7%
	Batch 2	1.6103	3.98	0.24%	N/A	N/A
standard deviation = ± 0.02 mg						

Table 1: Summary of the quantity and proportion of each organic shell matrix fraction (ASM, AIM) after on (batch 1) or two (batch 2) bleaching treatments. Separate data were obtained for samples of two different geographical sources, Canary Islands and Brazil.

in Figure 3. Two additional spectra, corresponding to shell powder and commercial chitin (α-type), respectively, are also depicted at the top of the figure. For more details and to facilitate

the comparison, the spectra exclusively focus on the range of 2000-400 cm^{-1} (full FT-IR data are depicted in the supplementary Figure 1). Firstly, we verified that the mineralogy of the samples was fully aragonitic - characterized in particular by the 700-713 cm^{-1} double peak (ν_4 internal vibration modes of CO_3^{2-}) (Balmain et al., 1999) - by checking different zones of the shell (initial chamber, one random intermediate chamber, last preserved chamber). For clarity, only the spectrum acquired from the powder of the last chamber is shown in Fig. 3 (top spectrum-shell powder). In the 2500-4000 cm^{-1} range (not shown in Fig. 3), the shell matrix samples exhibit one broad band located at 3271 cm^{-1}, which is usually attributed to the amide A group [ν(N−H)] (Kong and Yu, 2007). In addition, one low amplitude enlarged band centered at 2932 cm^{-1} can be observed, and usually assigned to the ν(C–H) stretching vibrations.

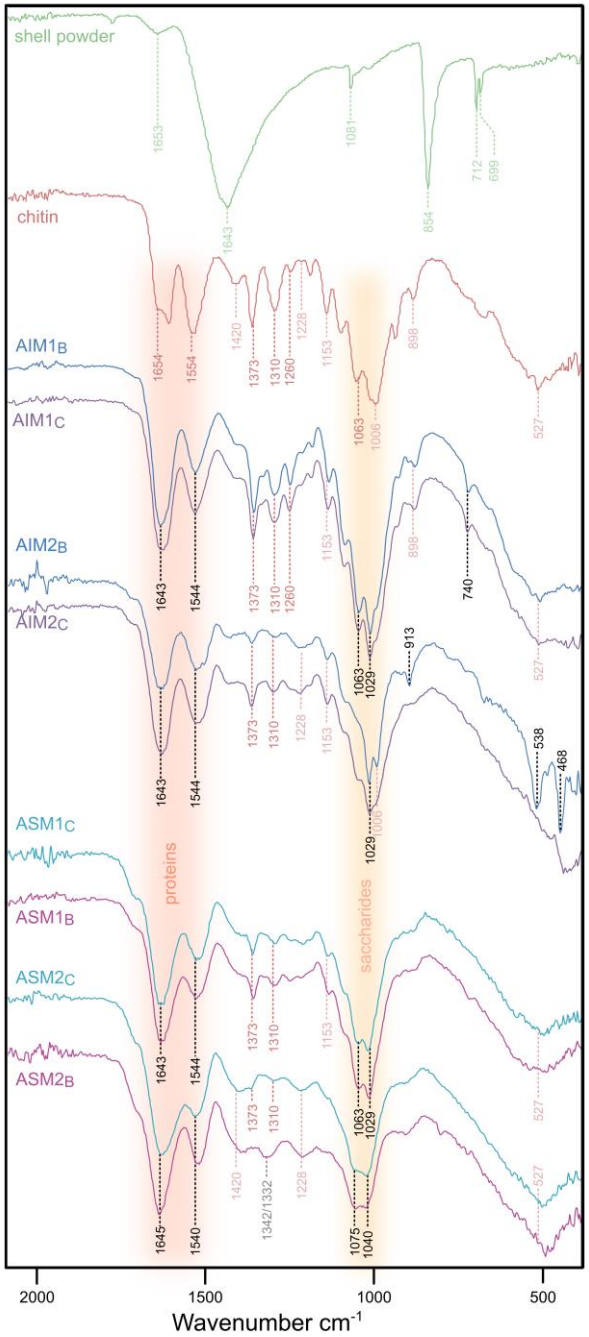

Figure 3: FT-IR spectra of the shell powder of Spirula spirula, of commercial α-chitin and of the different organic fraction of the shell matrix (AIM1, AIM2, ASM1, ASM2) from both localities (i.e. Canaries and Brazil). The "domains" of the characteristic absorption bands of protein and saccharidic moieties are highlighted, respectively, in pink and yellow. The values of the main absorption peaks are indicated under each of them, and colored in pink when identical to those of the commercial chitin.

All extracts (8) show comparable fingerprints and a strong similarity in the distribution of their absorption bands. To facilitate comparisons, we have compiled the FT-IR profiles in pairs of samples from Brazil (B) and Canary Islands (C): $AIM1_B$/$AIM1_C$, $AIM2_B$/$AIM2_C$, $ASM1_C$/$ASM1_B$ and $ASM2_C$/$ASM2_B$. Unambiguously, all spectra exhibit, in particular, the presence of amide absorption bands at 1638/43 and 1535/43 cm^{-1}, ascribed to the Amide I [ν(C=O)] and II [ν(C−N)] bands, respectively, that characterize protein moieties. In addition, AIM and ASM

fractions are also characterized by a high density of absorption bands between 1500 and 1000 cm^{-1}, very similar to those observed for commercial α-chitin. The most significant are located at 1420, 1373, 1310 and 1260 cm^{-1}, and can be attributed to $\delta(CH_2)$, $\delta_s(CH_3)$, δ(C–H) and amide I bands, respectively. Moreover, absorptions of stretching vibrations of saccharidic moieties are clearly revealed by the presence of intense absorptions at 1153, 1063 and 1006 cm^{-1} corresponding to ν_s(C–O–C), ν_{as}(C–O–C), and ν(C–O), respectively (Kaya et al., 2014) . The spectra from both origins – Canary Islands and Brazil – are very similar. It is interesting to note that the $AIM2_B$ spectrum exhibits additional bands, sharp and of medium intensities, between 1000 and 400 cm^{-1} (located at 913, 538, 468 cm^{-1}). To the best of our knowledge, and based on the existing bibliography, these bands suggest the presence of phosphate groups (PO_4^{3-}) (Radev et al., 2012). Thus, the band at 913 cm^{-1} can be assigned to the stretching P–O vibrations and those at 538 and 468 cm^{-1} can be related to the bending O–P–O vibrations of phosphate groups (Jastrzbski et al., 2011). At this stage of investigations, the reason of this variance from other samples remains still undetermined. However, in recent studies on other shells of marine animals, such as *Pinctada margaritifera* (Schmitt et al., 2018) and *Porites australiensis* (Takeuchi et al., 2018), we already made the same spectroscopic observations on AIM fractions leading to comparable suppositions about the presence of PO_4^{3-} groups.

When compared to AIMs, ASMs show a decrease in intensity of the amplitude of their absorption bands in the 1200-1420 cm^{-1} range. It is the same for the saccharidic bands of ASMs also showing a relative intensity decrease in comparison to that of amide absorptions. Finally, we observe that the second bleaching treatment alters the amplitude of the absorption bands in the 1200-1420 cm^{-1} range, without modifying the overall appearance of spectra.

Solid State Nuclear Magnetic Resonance

The ^{13}C CP MAS NMR spectra of the two *Spirula spirula* extracts (AIM1) and of α-chitin standard are shown in Figure 4. The spectra from both origins – Canary Islands and Brazil – are closely identical, and are dominated by the characteristic chitin peaks. The latter are superimposed on a much weaker set of partially resolved peaks between 20 and 55 ppm, which are consistent with small residual proteinaceous content. The merged C_3 and C_5 peaks, the broadened methyl peak, and the chemical shifts of C_1 and C_4 peaks (103.9 and 83.9 ppm) in the ^{13}C CP MAS spectra of the *S. spirula* samples are consistent with β-chitin (Jang et al., 2004; Kaya et al., 2017).

The chitin peaks exhibited by both *S. spirula* extracts deviate from those of the neat chitin peaks, showing broadening and reversed relative C_6 and C_2 peak intensities (increased and decreased, respectively); these deviations represent increased structural heterogeneity due to modification and/or disruption of local environments possibly as a result of the extraction procedure. Closer inspection of the spectra reveals additional partially resolved peaks (102.6, 86.3, and 71.5 ppm respectively; supplementary Fig. 2) that appear as shoulders of the C_1, C_4 and C_5 peaks; these suggest contributions from other polysaccharides. Given the increased linewidths, the co-occurrence of the other chitin polymorphs cannot be ruled out. In conclusion, the spectra imply that over 80% of the insoluble organic extract content (AIM1) is chitinous/saccharidic. This determination is further substantiated by the lower S/N DE MAS spectrum shown in Figure 4.

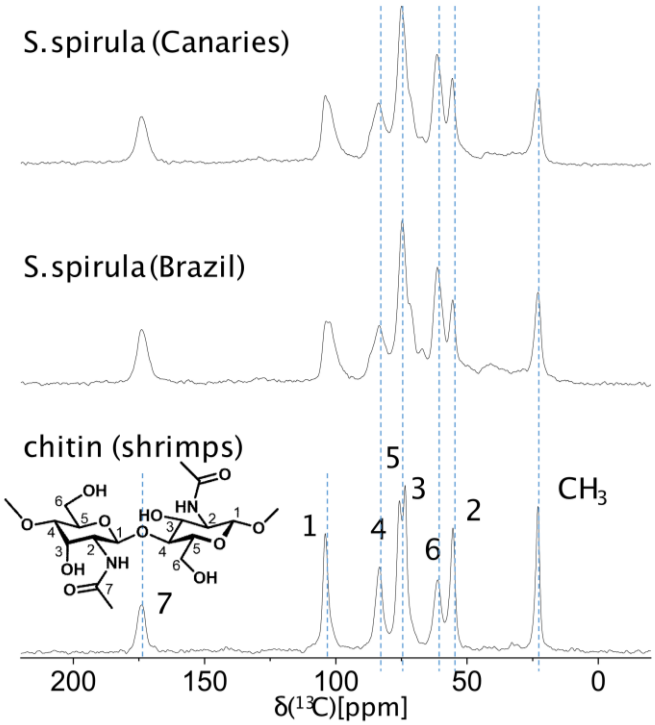

Figure 4: 75.4 MHz 13C CPMAS spectra of acid insoluble extracts (AIM1) of Spirula spirula shells from the Canaries and Brazil, and of α-chitin standard (shrimp) showing that the organic content of AIM1 consists primarily of chitin (the dashed lines denote the characteristic peaks of chitin).

Both *S. spirula* AIM1 extracts show a broad -0.2 ppm phosphate peak in their ^{31}P CP MAS spectra (Figure 5), validating the above IR findings for the AIM2 samples. ^{13}C{^{31}P} CP MAS REDOR measurements applied to AIM1$_C$ showed no REDOR effect (not shown) therefore ruling out phosphorylation of any of the chitin carbons. In the absence of such measurable proximity and in view of the resilience of these P-species under the extraction procedures (AIM1 and the more aggressive of AIM2), we conclude that the observed P-species are inseparable from the proteinaceous content. We have previously reported similar ^{31}P peak for decalcified

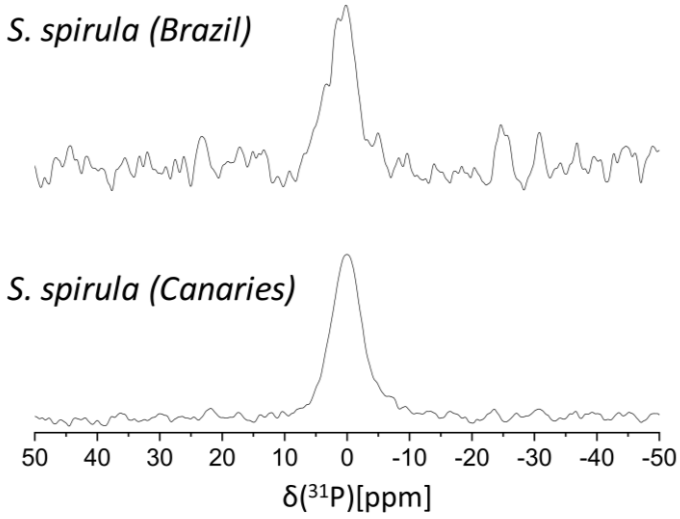

Figure 5: 121.85 MHz 31P CP MAS NMR spectra of the acid insoluble extract (AIM1) of Spirula spirula shells from the Canaries and Brazil. The peaks are centered at -0.2 ppm and are of similar normalized intensities (accounting for sample weights of 16.8 and 6.0 mg, respectively). Each spectrum was obtained acquiring 2k transients.

molluscs shells (Agbaje et al., 2018), and also here we ascribe this peak to represent phosphorylation, most likely of tyrosine and/or serine side chains.

SDS-PAGE of the shell matrix extracts

The results obtained after migrating the eight matrices in denaturing conditions are illustrated in Figure 6. In the silver stained gel (Fig. 6A), all of them are characterized by the abundance of smearing macromolecules along the whole lane, in particular in the LS-AIMs extracts. In addition, discrete bands, either blurred like the ones of low molecular weights (< 17 kDa) in lanes 1, 3, 5 and 7, or thin and tenuous such as in the middle of the lanes 3 and 7, can be observed. We notice that the electrophoretic profiles produce superimposable patterns between similar extracts but from the two locations (such as those of lanes 1 and 5, 3 and 7). The effect of the second bleaching treatment is mainly visible on the ASMs extracts: disappearance of the high molecular diffused black band (above 170 kDa, lanes 1 and 5 *versus* lanes 3 and 7, respectively), and appearance of a set of four bands between 30 and 43 kDa (respectively 32.4, 35.88, 38.69 and 42.05 kDa). The Bio-Rad blue dye (Fig. 6B) stains mainly

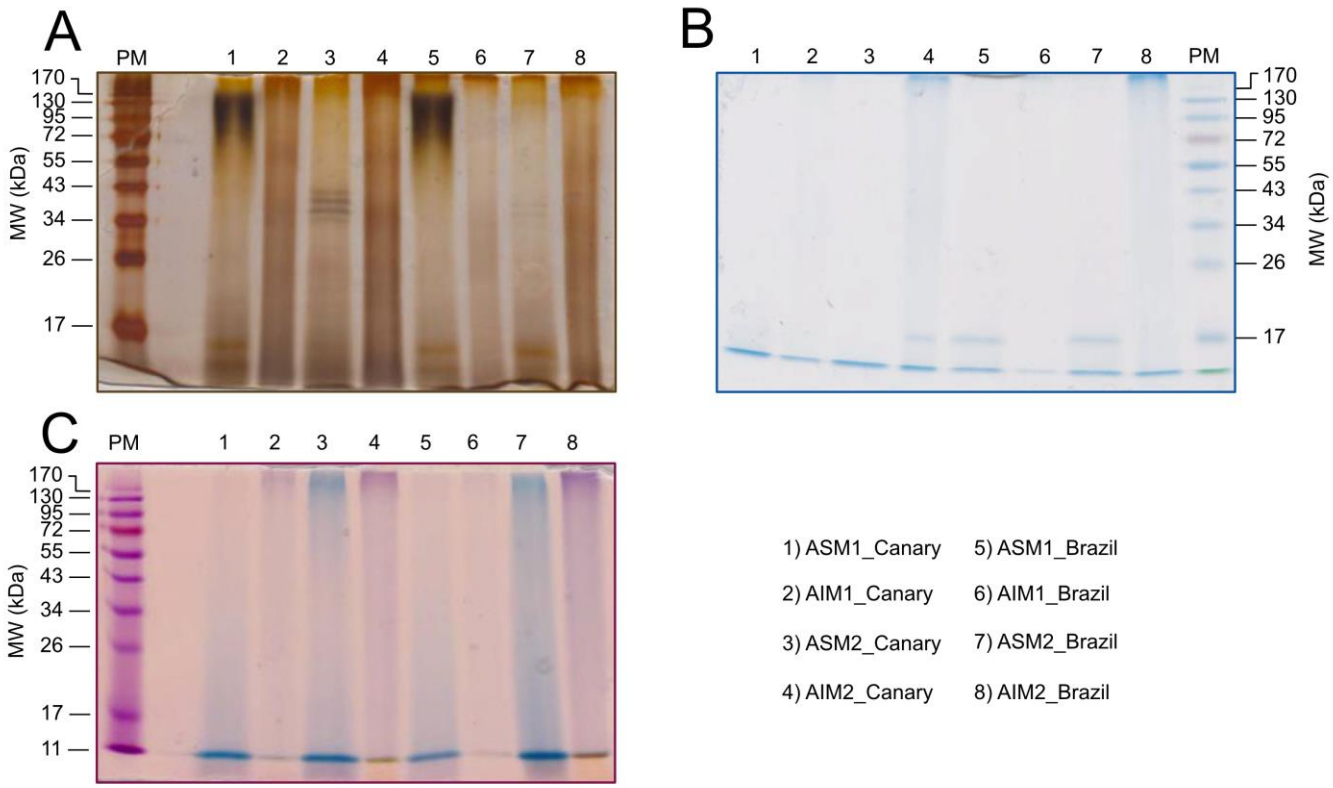

Figure 6: SDS-Page gel electrophoresis on Spirula spirula. A) Silver-staining; B) Bio-rad blue staining; C) Stains-all. PM: standard molecular weight (Fermentas).

low molecular weight macromolecules, in particular the 17kDa band, observed in lanes 4, 5, 7 and 8. Note that lanes 4 and 8 (LS-AIMs) are equally stained along their length, except a very high molecular weight fraction, which is intensely marked. The carbocyanine dye, which is supposed to stain calcium-binding macromolecules (Fig. 6C), discriminates the ASMs from the LS-AIMs: the first ones are stained blue, and the second, purple. None of the discrete bands observed in silver stained gels are visible with carbocyanine, except the migration front, which is intensely dyed in all cases. For all LS-AIMs, one notices in the migration front the superimposition of an additional greenish color (lanes 2, 4, 6, 8) that may suggests the presence of low molecular weight colored components, such as pigments. In some of our other attempts, we could lightly stain in the ASMs two discrete bands, one at 17kDa and another one at very high molecular weight (supplementary Fig. 3).

In vitro crystallization assay

The effect of the addition of the two soluble extracts, ASM1 and ASM2, when tested in the range [0 - 16 µg/well], is shown in Figure 7, which only presents few of the tested concentrations (0, 1, 4 and 16 µg/well). We notice that the effect of the first one (upper panel) is limited at all tested concentrations: we do not observe any significant differences with the blank, up to 4 µg/well. Starting from 8 µg/well (not shown here), the effect becomes

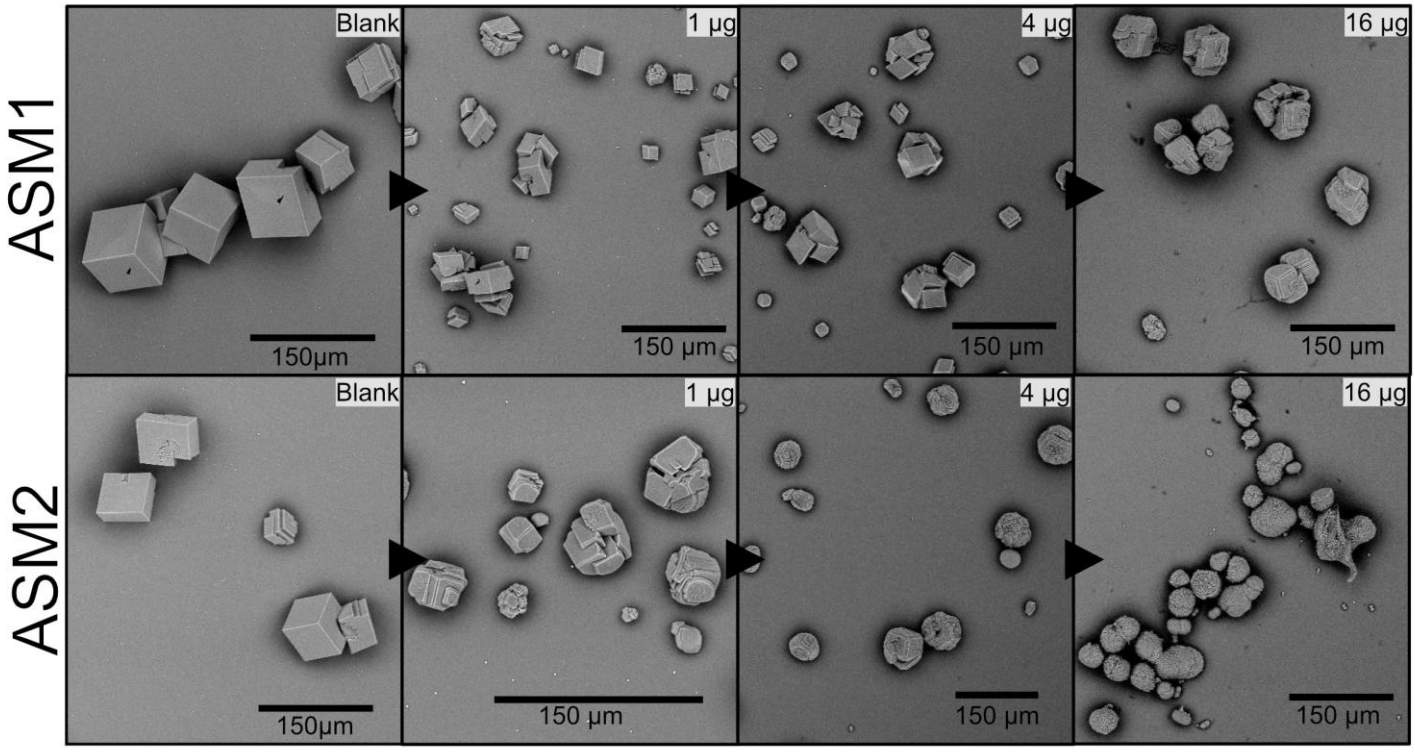

Figure 7: In-vitro crystallization assay performed on both soluble matrix extracts of Spirula spirula, ASM1 (top) and ASM2 (bottom). The black arrows indicate the increasing matrix concentration within each well, from 0 (blank) to 16 µg/well (200 µL per well).

progressively visible, with the formation of grooves on the polycrystalline aggregates edges. At 16 µg/well, the effect is more pronounced with the formation of rounded angles. No signs of crystallization inhibition - characterized by a drastic decrease of the size of polycrystalline aggregates - are recorded with ASM1.

The effect of ASM2 (lower panel) appears much more drastic. At 1 µg/well, grooves are already visible on the crystal faces edges, while the 4 µg condition resembles the 16 µg one of ASM1. At 16 µg, polycrystalline aggregates evolve into complex spherulite, without any significant size changes.

Enzyme-Linked Lectin Assay (ELLA)

All screening experiments performed with a large set of commercial lectins are summarized in Table 2. We analyzed only ASMs: the insolubility of AIMs in standard buffers precluded the use of ELLA. Firstly, we do not detect significant variations of lectin reactivity between Canaries Island extracts and that of Brazil: the few low amplitude variations concern a couple of lectins that either do not react or react very weakly (10-20%), such as LCA or GSL-1. To summarize, we obtain three classes of reactivity: high, intermediate and low. In the first class, one finds LEL, WGA, STL (only for ASMs1) and Jacalin. The three first lectins bind oligomers (dimers and trimers in the case of WGA) of N-acetylglucosamine and are usually considered as chitin-binding lectins. Jacalin binds preferentially D-galactose or oligomers that terminate with galactose residues, as well as N-acetylgalactosamine-D-galactose dimers that very often correspond to the first glycosylation steps of Ser/Thr residues in O-glycosylated proteins. The second category of reactivity comprises ConA, succinylated-WGA and, in a lesser extent, GSL-II, and DSL. ConA binds

	Canary Islands		Brazil	
	ASM1	ASM2	ASM1	ASM2
ConA	XX	XXX	XX	XXX
SBA	-	-	-	-
WGA	XXXX	XXXX	XXXX	XXXX
DBA	-	-	-	-
UEA 1	-	-	-	-
RCA	-	-	-	-
PNA	-	-	-	-
GSL 1	-	X	-	-
PSA	-	-	-	-
LCA	-	-	X	X
PHA-E	-	-	-	-
PHA-L	-	-	-	-
SJA	-	-	-	-
succ-WGA	XX	X	XXX	X
GSL 2	X	X	X	XX
DSL	X	-	XX	-
ECL	-	-	-	-
Jacalin	XX	XXXX	XXX	XXXX
LEL	XXXX	XXXX	XXXX	XX
STL	XXXX	X	XXXX	X
VVA	-	-	-	-

Table 2: Summary of the reactivity pattern of lectins on both soluble fractions (ASM1 and ASM2) for both localities. 'XXXX' indicates a reactivity that exceeds 60%, 'XXX', a reactivity comprised between 40 and 60%, 'XX' between 20-40%, 'X', between 10-20% and '-' for a reactivity below 10%.

specifically to mannose-N-acetylglucosamine dimers and is also considered as a mannose-binding lectin when this residue is part of N-linked oligosaccharides. Succinylated-WGA has similar binding properties as native WGA, but its binding capacity is inhibited by sialic acid. Thus, linking the reactivities of native and succinylated WGAs informs on the degree of sialylation of the tested extract. GSL-II and DSL bind either oligomers (DSL) or monomers (GSL-II) of N-acetylglucosamine, the latter one binding specifically the terminal residues. The third category groups all the other lectins that give weak or no signal with the tested extracts, and will not be discussed further.

Although the reactivity patterns are very similar between ASM1 and ASM2, it is interesting to notice that the second bleaching step modifies the signal given by few lectins, suggesting that the saccharidic groups targeted by the diverse lectins are differentially exposed to sodium hypochlorite: in particular, STL collapses drastically, while, inversely, Jacalin increases significantly. Other minor variations are recorded: Succ-WGA and DSL tend to decrease whereas ConA rises substantially.

Proteomic analysis

		AIM1	ASM1	AIM2	ASM2	TOTAL
A	Total number of hits	47	17	44	33	141
	Number of unique hits	22	5	19	15	61
	Number of shared hits between 2 or more extracts	25	12	25	18	30
B	AIM1		8	21	14	
	ASM1			8	8	
	AIM2				15	
	ASM2					
C	Number of identified peptides	82	20	55	42	199
	Number of hits identified by 2 or more peptides	4	2	5	6	17
	Number of hits identified by 1 peptide	43	15	39	27	124
	Dominant AA residues in the identified peptides	Leu: 13.7	Leu: 13.6	Leu: 10.5	Leu: 10.8	
		Ile: 8.7	Val: 8.9	Ile: 9.6	Ile, Val: 8.4	
		Asp: 8.3	Ala, Arg, Glu: 8.4	Val: 8.8	Thr: 8.1	
		Arg: 7.5	Pro, Ser: 6.3	Asp, Gly: 8.0	Glu: 7.9	
		Ala: 7.2	Asp, Thr: 5.8	Arg: 7.8	Arg: 7.4	
		Glu: 7.0	Ile: 5.2	Ala: 7.1	Ala: 6.9	
		Gly, Val, Ser: 6.4		Ser: 6.8	Asp: 6.7	
		Thr: 6.0		Pro: 6.4	Ser: 6.2	
		Pro: 5.5		Thr: 6.3	Pro: 5.5	
				Glu: 5.8		

Table 3: Overview of the main proteomic results by MASCOT. A) Summary of the identified hits per extract. B) Cross-comparison of shared hits in the 4 extracts. C) Summary of the identified peptides per extract and occurrence of the most represented AA residues (sorted by decreasing order and colored according to their chemical properties: Blue: basic; red: acidic; orange; hydroxylated; green: hydrophobic)

Our proteomic *in silico* investigation on the four extracts (ASM1, AIM1, ASM2, AIM2) was performed on two databases: in the first case, we searched against "Other metazoans" NCBI dataset, and, in the second case, against the *Spirula spirula* transcriptome. The results of the first approach are summarized in Table 3. In addition, Table 4 represents the list of peptides identified for each extract and Table 5,the hits shared by the different extracts . The overall proteomic results are available in the Supplementary Table 1. In total, 141 hits are obtained, and after suppressing duplicates, 91 (Table 3A). This score is particularly low. We notice that the two ASMs extracts yielded less hits than their insoluble counterparts. Unique hits represent about 40-45% of the total number of identified hits, except for ASM1, where it only accounts for 29%. In short, 22 hits are unique to AIM1, 19 to AIM2, 5 to ASM1 and 15 to ASM2. By focusing on shared hits (Table 3B), we notice that AIM1 and AIM2 are the most similar (21 hits), followed by ASM2, which exhibits some similarities with insoluble extract (14 shared hits with AIM1, and 15 with AIM2). ASM1 appears to be the most different extract from the four ones. Table 3C shows the number of identified peptides, combined to the occurrence of the main amino acid residues in each shell extract. It shows that very few hits (less than 20% of the total matches per extract) are identified by two or more peptides, the remaining hits being recognized by only one peptide. A brief survey of the number of identified peptides per extract confirms our previous finding, namely that both AIMs exhibit higher number of peptides than

AIM1 (13959)	AGFAGDDAPR ; GYSFTTTAER ; EISALAPATMK ; QEYDESGPSIVHR ; SYELPDGQVITIGNER ; VAPEEHPVLLTEAPLNPK ; DLYANTVLSGGSTMFPGIADR ; TTGIVMDSGDGVTHTVPIYEGYALPHAILR ; EITALAPSTMK ; VAIVITDGR ; DYAIVDER ; STELLIR ; ISGLIYEETR ; TITLEVEPSDTIENVK ; DEVLGGILR ; LILPGELAK ; GDGTGGLSIYGDR ; IDVADLLR ; SPNQNVQQAAAGALR ; VAVDTSYR ; SPIDLVR ; IDLLIPR ; ITIDLFEK ; IVMSQGTHLYFDHPYEPDPEER ; VSIVITDGR ; LDSDIKPQMKR ; TPTDVPVNLSCGR ; QLVQDSLDSLR ; ANILSILEK ; KLVLLNYER ; YSPVDPIR ; QSIESVSPR ; LGIYEDFGTK ; IVIGLFGK ; IAFLVLK ; DSTLIMQLLR ; ELQDLNER ; LFDQAFGLPR ; DALEELDADR ; KILSIVER ; EGIEAIEDDR ; LDILPLK ; DLVGELADAIR ; QLSGQQYR ; LDTELNTYR ; VLSIILEK ; QLLEEIANLQK ; NIIDLNTEMK ; LMASFER ; LLLTAAISADK ; KINPPYYR ; ATSQTLLDENDITFR ; ATLGLPQPK ; VATVPLSR ; TGVSSEGDTDKR ; TTGIVMDSGDGVTHTVPIYEGYALPHAIIR ; SPLDLVR ; VVMTQASHLYFDHPYEPDPEER ; ASLVLDTGR ; LDLLPLR ; IGGIGTVPVGR ; IADADFR ; IASYLEK ; LGTALPFVR ; DYLQIVDAR ; EGELSDLSR ; KLLSDILR ; LDILPIK ; AQIEIEIR ; LGDVEILR ; IGSLSLPR ; LLLPGELAK ; RIAEDSLLYFR ; ASCSERCVAR ; VATVPSLR ; LAVVITDGR ; EIIDLVLDR ; SSFFVAR ; GDFQLPALR ; QEGAKDLLLR ; ATNLSNFR ; FASFLDK
ASM1 (11712)	AGFAGDDAPR ; GYSFTTTAER ; SYELPDGQVITIGNER ; VAPEEHPVLLTEAPLNPK ; VAIVITDGR ; VATVSLPR ; DYAIVDER ; LVALFDLR ; FQNALLVR ; EDQTEYLEER ; QSIESVSPR ; IVIGLFGK ; LPVLETER ; FDCLLSR ; KLVLLNYER ; QLSGQQYR ; LLLTAAISADK ; IAFLVLK ; EPSPQSPQLHR ; VAVDTSYR
AIM2 (13977)	AGFAGDDAPR ; GYSFTTTAER ; EITALAPSTMK ; QEYDESGPSIVHR ; SYELPDGQVITIGNER ; VAPEEHPVLLTEAPLNPK ; DLYANTVLSGGSTMFPGIADR ; EISALAPATMK ; TTGIVMDSGDGVTHTVPIYEGYALPHAILR ; AVFPSIVGR ; DYAIVDER ; NFDGLDLDWEYPANR ; VAIVITDGR ; IVMSQGTHLYFDHPYEPDPEER ; IALVITDGR ; GDGTGGLSIYGDR ; IDVADLLR ; QLFDQVVK ; IGFPWSEIR ; APDFVFYAPR ; FGGTLADCIR ; STELLIR ; LILPGELAK ; ISGLIYEETR ; EGIEAIEDDR ; IAIVITDGK ; IIIVDSLGR ; GWLEPLLDR ; IDLLIPR ; ITQSNAILR ; VSIVITDGR ; LVALFDLR ; TITLEVEPSDTIENVK ; SETLFLR ; DSTLIMQLLR ; QFGSSIAR ; IPVVLLNYMR ; ASDGVPQVIVVLTDGR ; IATFVLR ; QSIESVSPR ; LDDVHLR ; YSPVDPIR ; IAFLVLK ; IVIGLFGK ; EPSPQSPQLHR ; VAVDTSYR ; LSSPTALSDR ; HALSDAAVR ; VESVASLKER ; VPGVVTPR ; TPTDVPVNLSCGR ; DLVGELADAIR ; SPIDLVR ; LIGWGQIR ; ILITGGAGFVGSHLVDR ; ;
ASM2 (13218)	FESNFNTQATNR ; NTDGSTDYGILQINSR ; DYAIVDER ; LLLPGELAK ; VFLENVIR ; ISGLIYEETR ; VAIVITDGR ; STELLIR ; LILPGELAK ; AGFAGDDAPR ; GYSFTTTAER ; EITALAPSTMK ; QEYDESGPSIVHR ; VAPEEHPVLLTEAPLNPK ; VSIVITDGR ; LPLQDVYK ; YYVTIIDAPGHR ; TITLEVEPSDTIENVK ; EGIEAIEDDR ; EDQTEYLEER ; EISALAPATMK ; LTAFLVR ; TGVSVSSLR ; YSPVDPIR ; ILVPQLDR ; EPSPQSPQLHR ; LSSPTALSDR ; IAFLVLK ; INHDLLR ; GIVHTDLR ; ASDGVPQVIVVLTDGR ; QAAIVNGTPR ; IHSDCAANQQVTYR ; IVIGLFGK ; VAVDTSYR ; KLVLLNYER ; TANLETSLR ; TLINEVDTR ; NFEEDFYR ; LMKFVTAVK ; NIIDLNTEMK ; SPIDLVR ;

Table 4: List of the peptides identified by MASCOT within each organic fraction of the shell matrix of Spirula spirula. For each fraction, the numbers in parentheses correspond to the number of peptides (number of queries) detected in total.

ASMs. In each extract, the most represented AA residues are the hydrophobic ones, with a peculiar emphasis on leucine, isoleucine and valine in all extracts. However, we observe that all hydrophobic residues are more abundant in the insoluble matrices. The two acidic residues (Asp, Glu) account for about 14% of all residues, and the hydroxylated ones, around 13%.

All the peptides identified per extract are listed in Table 4. The shortest ones comprise 7 residues while the longest, 30. One can notice some variation on the identified number of peptides, depending on the bleaching. We observe that there is a slight decrease of the number of peptides from AIM1 to AIM2, and a significant increase from ASM1 to ASM2.

All the hits identified in 2, 3 or 4 extracts are summarized in Table 5. Of particular relevance are the hits with specific enzymes that are putatively involved in biomineralization, such as "putative tyrosinase-like protein Tyr-3", a "probable chitinase 2", and a "serine protease K12H4.7 precursor". Other candidates of interest are proteins that exhibit calcium-binding properties (CBR-MUP-4 protein, transmembrane cell adhesion receptor MUA-3 precursor, and a hypothetical protein HELRODRAFT_161577) and an immunoglobulin protein (IgGFc-binding protein). Five identified proteins are thus found into the four matrices extracts: CBR-MUP-4 from *Caenorhabditis briggsae*, putative tyrosinase-like from *Crassostrea gigas*, IgGFc-

Protein hits			AIM1	ASM1	AIM2	ASM2
Id	Name	Organism				
gi|268575272	C. briggsae CBR-MUP-4 protein	*Caenorhabditis briggsae* - nematode				
gi|405964315	Putative tyrosinase-like protein tyr-3	*Crassostrea gigas* - pacific oyster				
gi|405970334	IgGFc-binding protein	*Crassostrea gigas - pacific oyster*				
gi|268570024	C. briggsae CBR-CYN-5 protein	*Caenorhabditis briggsae* - nematode				
gi|158297124	AGAP008053-PA, partial	*Anopheles gambiae str. PEST* - african malaria mosquito				
gi|284999807	beta actin	*Idiosepius paradoxus* - nothern pygmy squid				
gi|323230	polyprotein [Bovine viral diarrhea virus 1-Osloss]					
gi|510849995	core histone H2A/H2B/H3/H4	*Ancylostoma ceylanicum* - nematode				
gi|156359527	predicted protein	*Nematostella vectensis* - starlet sea anemone				
gi|170578661	Transmembrane cell adhesion receptor mua-3 precursor, partial	*Brugia malayi* - filarial nematode worm				
gi|675868340	hypothetical protein HELRODRAFT_161577	*Helobdella robusta* - californian leech				
gi|158300608	AGAP012043-PA, partial	*Anopheles gambiae str. PEST* - african malaria mosquito				
gi|768189576	hypothetical protein DICVIV_07943, partial	*Dictyocaulus viviparus* - bovine lungworm				
gi|669305641	hypothetical protein M513_09906	*Trichuris suis* - pig whipworm				
gi|163637123	ribosomal protein S3, partial	*Crassostrea gigas* - pacific oyster				
gi|223016075	actin	*Octopus vulgaris* - common octopus				
gi|524896874	PREDICTED: peptidyl-prolyl cis-trans isomerase B-like	*Aplysia californica* - california sea hare				
gi|170586872	hypothetical protein Bm1_33780	*Brugia malayi* - filarial nematode worm				
gi|33321997	gamma-glutamyl phosphate reductase, partial	*Lactobacillus delbrueckii subsp. Lactis* - bacteria				
gi|676489609	hypothetical protein LOTGIDRAFT_236342	*Lottia gigantea* - owl limpet				
gi|748378928	homeobox domain protein, partial	*Ancylostoma duodenale* - nematode				
gi|1360640	14-3-3 zeta protein	*Xenopus laevis* - African clawed frog				
gi|156544592	PREDICTED: alpha-L-fucosidase-like	*Nasonia vitripennis* - jewel wasp				
gi|193591718	PREDICTED: signal peptide, CUB and EGF-like domain-containing protein 1	*Acyrthosiphon pisum* - pea aphid				
gi|542607731	carbomoylphosphate synthase, partial	*Culicoides immaculatus* - diptera				
gi|307203534	Probable chitinase 2	*Harpegnathos saltator* - Jerdon's jumping ant				
gi|358337009	monocarboxylate transporter 14, partial	*Clonorchis sinensis* - chinese liver fluke				
gi|585686849	PREDICTED: collagen alpha-3(VI) chain isoform X1	*Elephantulus edwardii* - cape elephant shrew				
gi|225718928	serine protease K12H4.7 precursor	*Caligus clemensi* - sea louse				
gi|8272602	82 kDa heat shock protein 3, partial	*Philodina roseola* - rotifer				

Table 5: Summary of the main proteomic hits obtained by MASCOT and their presence/absence (full box/empty box) within each fraction of the organic matrix. The hits mentioned here only represent those shared by, at least, two fractions; the list of the hits that are unique to each extract is available in the supplementary data

binding from *Crassostrea gigas*, CBR-CYN-5 from *Caenorhabditis briggsae* and AGAP008053-PA from *Anopheles gambiae str. PEST*. More strangely, we also find two hits with the cytoskeletal protein actin (found in three extracts), and one hit with a histone, a nuclear protein (also found in three extracts). Nine hits (on 31) are found with either hypothetical or predicted proteins, most of them from invertebrate models, including two molluscs. So far, we do not find any hits corresponding to proteins that exhibit either Low Complexity Domains (LCDs) or Repetitive Low-Complexity domains (RLCDs).

The results of the search against the *S. spirula* transcriptome brought additional results that are summarized in Table 6. In total, we obtained 39 hits: 27 with AIM1, 31 with AIM2, 10 with ASM1 and 12 with ASM2. Among them, 22 are shared: 6 between the 4 extracts, 7 between 3 of them and 9 between 2. In the shared hits, the percentage of coverage varies from almost 60 to 1.4%. Furthermore, 17 hits are unique to one of the four extracts: 6 to AIM1, 9 to AIM2, 1 to ASM1 and 1 to ASM2. It is interesting to note that he two AIMs generate more hits than the ASMs, and the second extracts (AIM2, ASM2), more than the first ones (AIM1, ASM1). Among the 6 hits shared by the four extracts, 4 of them are invariably the ones with the highest peptidic coverage. They include a true transferrin, a serine protease inhibitor (elastase-inhibitor like), a matrilin, and actin. The two other "shared by four" hits are an uncharacterized protein with a heme-binding domain and a peptidyl prolyl cis trans isomerase. Among the 7 hits shared by 3 extracts, four of them exhibit a peptidic coverage higher than 10% of their sequence. They include α-2 macroglobulin, collagen (N-terminus), a protein/nucleic acid deglycase and a histone. The three other proteins are one ubiquitin and two histones. The proteins shared by two extracts are represented by a calcium-binding protein, a peptidyl-prolyl cis-trans isomerase, a chitobiase/sugar-binding protein, a GLIPR1 protein, a transferrin fragment, a collagen N-terminus, and a lamin. The other 17 proteins found in one extract include, among others, a Cys-rich secretory protein, a galectin3 binding-protein, a calcium-binding protein and four enzymes involved in saccharidic moieties modifications (heparan sulfate glucosamine sulfotransferase, N-acetyl-galactosaminidase, chitin deacetylase, UDP-glucuronic acid decarboxylase)

Protein hit * indicates complete sequences		Shell organic matrix extract AIM1	AIM2	ASM1	ASM2	Functional domains and motifs (starting AA - final AA, domain	Putative function in biomineralization
Transferrin*	% cover. *AA / MW /*	X 50.1 *293 (316) / 32.28 / 7.89*	X 59.7	X 36.2	X 35.5	Transferrin (28-311, cl30084)	Iron-binding & incorporation in the shell. Putative bactericidal substance (?)
Serine protease inhibitor / Elastase inhibitor-like	% cover. *AA / MW /*	X 30.7 *101 / 11.07 / 4.74*	X 30.7	X 21.8	X 30.7	Serpin (1-99, pfam00079)	Protection against the degradation of the matrix
Matrilin / Collagen α–4 precursor (N-term)	% cover. *AA / MW /*	X 19.6 *126 / 13.99 / 4.57*	X 19.6	X 9.8	X 18.6	vWF-type A (65-126, cd01450)	"Organizing" protein of the ECM. Interaction with other ECM proteins
Actin*	% cover. *AA / MW /*	X 17.8 *376 / 41.78 / 5.29*	X 17.8	X 14.4	X 13.6	Actin (whole seq., PTZ00281)	Cytoskeletal - Filament formation
Uncharacterized (from *Octopus)*	% cover. *AA / MW /*	X 13.4 *284 / 31.59 / 7.72*	X 7.0	X 7.0	X 7.0	SOUL heme-binding (57-224, pfam04832)	Interaction with heme, iron complexed in porphyrin ring
Peptidyl-prolyl cis-trans isomerase B-like isoform*	% cover. *AA / MW /*	X 3.4 *195 (232) 21.07 / 9.56*	X 3.4	X 3.4	X 3.4	Cyclophilin ABH-like (51-215, cd01926)	Cis-trans isomerization of Pro peptidic bonds. Regulates ECM proteins folding
α–2 macroglobulin (N-term fragment)	% cover. *AA / MW /*	X 16.8 *119 / 13.40 / 9.24*	X 16.8	X 16.8		MG2 (72-119, pfam01835)	Stops or reduces the activity of endopeptidases
Collagen α–1 (XXI) (N-term fragment)	% cover. *AA / MW /*		X 22.5 *147 / 15.59 / 8.05*	X 9.0	X 9.0	vWF-type A (41-120, pfam00092) Poly-T (130-147) in C-term	"Organizing" protein of the ECM. Interaction with other ECM proteins
Protein/Nucleic acid deglycase DJ-1	% cover. *AA / MW /*	X 7.6 *157 / 16.38 / 9.19*	X 21.6	X 7.6		Type-1 glutamine aminotransferase (1-140, cd03135)	Precludes the formation of Maillard products. Redox sensitive chaperone
Histone H4-like*	% cover. *AA / MW /*	X 9.7 *103 / 11.37 / 11.36*	X 9.7		X 17.4	Histone H4 (1-103, PLN00035)	If extracellular, involved in cell lysis may act as antimicrobial agent
Ubiquitin 40S ribosomal protein	% cover. *AA / MW /*	X 8.9 *179 / 20.85 / 9.65*	X 8.9		X 8.9	Ubiquitin-like (23-98, cd01803) Ribosomal S27a (27-169, pfam01599)	Conjugates to Lys residues of other proteins. Protein kinase activation, cell
Histone H2B gonadal-like	% cover. *AA / MW /*	X 7.0 *129 / 14.21 / 10.45*	X 7.0		X 7.0	Histone H2B (39-127, smart00427) K-rich (17-38)	Immunity-related function when extracellular may act as antimicrobial peptide
Histone H3-like*	% cover. *AA / MW /*	X 5.0 *136 / 15.33 / 11.27*	X 5.0		X 5.0	Histone H3 (1-136, PTZ00018)	Cadherin-binding. If extracellular, interacts with Toll-like receptors may act as antimicrobial agent
Peptidyl-prolyl cis-trans isomerase B (partial)	% cover. *AA / MW /*	X 17.0 *141 / 15.34 / 8.97*	X 17.0			Cyclophilin (3-124, cl00197)	Cis-trans isomerization of Pro peptidic bonds. Regulates ECM proteins folding
GLIPR1-like protein 1, isoform X1 *(fragment)*	% cover. *AA / MW /*	X 5.5 *163 / 19.13 / 8.49*	X 9.8			No identified domain	Unknown - Involved in defense mechanism (?). Highly similar to unknown *Octopus*
Melanotransferrin / Transferrin (fragment)	% cover. *AA / MW /*	X 7.6 *118 / 12.95 / 5.65*	X 7.6			Type 2 periplasmic binding fold (9-117, cl21456)	Fe-binding glycoprotein, controls the level of free Fe. Other putative functions (see
Neurocalcin / hippocalcin-like / neuronal Ca sensor*	% cover. *AA / MW /*	X 5.3 *189 / 21.98 / 5.26*	X 5.3			Homeodomain (5-34, smart00389) Ca-binding EF-hand (22-177,	High affinity Ca-binding. Activation or inactivation of target proteins
Unhealthy ribosome biogenesis protein 2	% cover. *AA / MW /*	X 3.9 *180 / 20.31 / 5.20*	X 3.9			No identified domain	Ribosome biogenesis - Unknown if extracellular
Collagen α–6 (N term) / Cartilage matrix protein*	% cover. *AA / MW /*	X 3.6 *389 (421) / 44.02 / 8.14*	X 3.6			vWF-type A (44-217, pfam00092) vWF-type A (243-397,	Typical ECM protein, provides a scaffold. Organizes the matrix, binds to other ECM molecules
14-3-3 protein zeta/delta	% cover.	X 3.5	X 3.5			14-3-3 protein (47-267, pfam00244)	Signal transduction pathway through phosphoprotein-binding

	AA / MW /	*284 / 32.14 / 5.21*		
Chitobiase / β–N–αχετψληεξοσαμινιδασε*	% cover. *AA / MW /*	X 2.6 *860 / 98.40 / 5.38*	X 2.6	Carbohydrate-binding (13-160, pfam03173) Glycosylhydrolase 20 (198-322,	Polysaccharide binding - Chitin scaffold remodeling Oligosaccharide degradation
Lamin / Intermediate filament protein ifa1*	% cover. *AA / MW /*	X 1.4 *577 / 65.56 / 5.60*	X 1.4	Intermediate filament (66-417, pfam00038) Lamin tail, Ig fold (465-569,	Nuclear / cytoplasmic protein. Possible binding of ECM proteins via lamin tail
Uncharacterized / Cys-rich secretory protein	% cover. *AA / MW /*	X 10.3 *106 / 12.28 / 8.06*		Cap: Cys-rich secretory protein (42-103, cl00133)	GAG-binding
α–1–Τυβυλιν (N term)	% cover. *AA / MW /*	X 5.9 *152 / 16.60 / 5.46*		Tubulin α–χηαιν (whole seq., cl30502)	Cytoskeletal protein. Unknown function if extracellular
Poly(U)-specific endoribonuclease-like isoform 3*	% cover. *AA / MW /*	X 4.8 *316 (337) / 35.75 / 9.68*		Somatomedin B (27-65, pfam01033) Endoribonuclease XendoU (82-	If extracellular, polysaccharide-binding (?). Metal ion-binding (?) - Protease
Heparan sulfate gluco-samine sulfotransferase	% cover. *AA / MW /*	X 2.7 *261 / 30.23 / 9.74*		Sulfotransferase (43-260, pfam00685)	Catalyzes the transfer of a sulfo group to glucosamine in heparan
Alpha-N-acetyl-galactosaminidase-like*	% cover. *AA / MW /*	X 2.6 *389 (411) / 43.97 / 5.40*		α–γαλαχτοσιδασε A (30-309, pfam16499) α–Γαλαχτοσιδασε A (312-396,	Glycoside hydrolase (in glyco-lipids/-proteins)
Elongation factor 1-a*	% cover. *AA / MW /*	X 2.4 *463 / 51.11 / 9.03*		Elongation factor 1-a (1-454, PTZ00141)	Involved in protein translation - Binds actin. If extracellular, may regulate cell
Methylmalonic aciduria & homocystinuria type C	% cover. *AA / MW /*	X 12.8 *242 / 28.15 / 9.43*		MMACH type C (7-225, pfam16690)	Catalyzes the reductive dealkylation of cyanocobalamin - Binding of vitamin B12
Ca-binding protein / Calmodulin-like (fragment)	% cover. *AA / MW /*	X 11.7 *111 / 12.65 / 4.57*		Ca-binding EF hand (10-72, cl08302) Ca-binding EF hand (47-109,	High affinity Ca-binding - Calcium cell signaling
Galectin 3 binding protein (fragment)	% cover. *AA / MW /*	X 8.9 *101 / 11.39 / 4.68*		Scavenger receptor Cys-rich (SRCR) (2-93, smart00202) - 6 Cys	Involved in immunity (?)
Chitin deacetylase-like 5 (C-term fragment)	% cover. *AA / MW /*	X 8.9 *179 / 20.64 / 6.34*		Catalytic NodB homology domain of the carbohydrate esterases 4 (1-163, cl15692)	Deacetylation of O- and N-acetylated polysaccharides like chitin
α–4 Collagen-like (*Octopus*) (fragment)	% cover. *AA / MW /*	X 5.3 *207 / 22.58 / 8.31*		Poly-T (1-11) vWF type A (28-190, pfam00092)	Typical ECM protein. Interacts with other ECM proteins
Moesin / Vilin 2 / Radixin *	% cover. *AA / MW /*	X 4.6 *583 / 68.82 / 5.76*		FERM (203-299, cd13194) ERM (7-209, smart00295)	Actin-binding. Signaling pathway regulation
UDP-glucuronic acid decarboxylase 1	% cover. *AA / MW /*	X 4.1 *417 / 47.79 / 9.42*		DUF3573 (24-91, cl13524) UGD (95-398, cd05230)	Decarboxylation of UDP-glucuronic acid to UDP-xylose. Involved in ECM proteoglycan
Arginine kinase*	% cover. *AA / MW /*	X 2.4 *409 / 46.24 / 8.57*		Arginine kinase (68-409, cd07932)	Phosphotransferase: conversion of ATP & arginine to ADP & phosphoarginine - If extracellular,
Eukaryotic translation initiation factor 2 subunit 3	% cover. *AA / MW /*	X 1.6 *489 / 53.26 / 8.84*		Eukaryot. translation initiation factor 2 γ subunit (30-483, PTZ00327)	GTPase activity, GTP-binding, cadherin-binding. Regulation of the extracellular ERK
Histone H2A*	% cover. *AA / MW /*	X 7.0 *128 / 13.42 / 10.58*		Histone H2A (6-121, cl30550)	If extracellular, interaction with Toll-like receptors may act as antimicrobial agent
Elongation factor 1-a (fragment)	% cover. *AA / MW /*	X 7.5 *107 / 12.09 / 7.99*		Elongation factor 1-a (whole seq., cl33168)	Same as above - If extracellular, may mediate cell growth & immune response

Discussion

The present paper describes the shell structure and biochemistry of the Ram's Horn squid *Spirula spirula.* To our knowledge, this work represents the first exhaustive characterization of the shell matrix of this species. We describe both the microstructural features of the shell and the matrix biochemical fingerprint, with a particular emphasis on sugars and protein content.

From a microstructural viewpoint, our investigation confirms that the shell wall is made of two layers, as previously suggested (Dauphin, 1976, 1977; Cuif *et al.*, 1983). However, we do not recognize the homogeneous semi-prismatic pattern described by these authors. We rather describe the internal layer as a succession of block-like/irregular semi-prismatic structures, and the external tuberculated layer as a regular semi-prismatic unit. A thin blackish layer between the two previous units is observable, which has already been noticed (Dauphin, 1976, 1977) and described as purely organic by Cuif *et al.* (1983) after an enzymatic treatment.

We also observe the lamello-fibrillar microstructure of the septa, usually considered as a nacreous layer (Barskov, 1973; Bøggild, 1930; Dauphin, 1976; Mutvei, 1964). Macroscopically, the septa exhibit indeed iridescence properties. However, we feel that the "nacreous" terminology is misleading and we follow the assertion of Erben (1972) who claimed that the nacre of *Spirula* and other coleoids is rather a 'false nacre' and recommended the use of "lamello-fibrillar" terminology. This expression is entirely justified in view of a very recently published work (Lemanis et al., 2020), which precisely visualizes the substructures of the lamello-fibrillar layer, and describes it as the stacking of "thin lamellae that are themselves composed of aragonitic rods that tend to be parallel within a lamella." The lamello-fibrillar terminology was also used by Fuchs et al. (2012) to describe the septal microstructure of ancestral spirulid coleoids. Two patterns are of particular importance: the insertion of the septa in the shell walls, and the presence of what resembles cameral deposit at the back of the insertion zone. The first one, although mentioned in earlier works (Bandel and Boletzky, 1979; Dauphin, 1976; Hoffmann et al., 2018), was neither fully described, nor schematized (Bandel and Boletzky, 1979). We confirm that the septa are sandwiched as a bevel in the internal wall. Such structures were also observed among the fossil relatives of *Spirula* (Doguzhaeva, 1996) and defined as "mural flap" (Fuchs et al., 2013): this character state may be an apomorphy of the Spirulida order. Given the thinness of the *S. spirula* shell, one can assume that a "sandwich" insertion considerably reinforces the overall shell mechanical properties and was consequently selected by evolution. However, Lemanis *et al.* (2020) questioned such hypotheses, demonstrating that mural flaps have “no notable structural function”. The second pattern - deposits in the dorsal septal corner and mentioned by Appelöf (1893), Bøggild (1930) and Dauphin (1976) - may evoke cameral deposits found in nautiloids (Mutvei, 2018). They are

610 defined by different terminologies: "supraseptal annular ridges" (Mutvei, 1964), or "adapical
611 ridge", "mural ridge" and "septal ring" (Lemanis et al., 2020). Their function is also to increase
612 the overall mechanical properties of the shell. However, their set-up appears enigmatic: as noted
613 by Dauphin (1976) and confirmed by our observations, their microstructure is in continuity with
614 the internal wall layer on which they lay. This suggests either a continuous growth from the
615 internal wall, or an epitaxial growth after a growth stop between the secretion of the wall and
616 that of the supraseptal annular ridges. The chronology of the different secretory steps of shell
617 construction is not discussed further here.
618 Beside the microstructure overview, the biochemical characterization of the shell of *S.*
619 *spirula* represents the core of the paper. To limit potential biases due to the geographical source
620 of the samples, we performed our experiments on two batches of specimens from two spots,
621 distant from each other (Canaries Islands, and Brazil). Our investigations revealed that the
622 biochemical fingerprints do not depend on the geographical origin: there is no significant
623 difference between the two batches, whatever technique used. From now, we will no longer
624 consider the provenance of the specimens, but only focus on the biochemical signature.
625 After a superficial bleaching, the total shell matrix of *S. spirula* represents, in all our
626 extractions, about 1% of the shell weight, a value that classifies it among the matrix-rich shells
627 such as nacro-prismatic ones (Palmer, 1992, 1983). Similarly to these latter, a large proportion
628 of the matrix is insoluble in acetic acid, representing around 70% of the organic content. A
629 second thorough bleaching treatment decreases drastically the amount of matrix, in particular
630 of the insoluble fraction. Following the criteria defined primarily by Crenshaw (Crenshaw,
631 1972), this obviously suggests that the shell matrix of *S. spirula* - including the soluble fraction
632 - is predominantly intercrystalline and not protected by the mineral phase during the second
633 bleaching step. Another argument in favor of an intercrystalline position of the matrix is the
634 moderate capacity of the soluble fraction to interact with the *in vitro* precipitation of calcium
635 carbonate and the absence of any sign of inhibition even at high matrix concentration (16
636 µg/well and higher, not shown), these two properties being often associated to acidic, *i.e.* Asp-
637 rich intracrystalline proteins (Albeck et al., 1993). This suggests that most of the soluble
638 fraction of *S. spirula* is not strongly bound to the mineral phase.
639 Our combined biochemical and physical investigations show that the saccharidic
640 moieties constitute a notable fraction of the matrix. In particular, the FTIR spectra reveal high
641 amplitude bands compared to those of proteinaceous moieties (amide), in both ASMs and
642 AIMs. This finding is further substantiated by the SS-NMR spectra acquired for the AIM
643 extracts which identify *β*-chitin and also additional polysaccharides that together constitute

~80% of the insoluble fraction. This percentage is higher than that indicated in earlier works, which found a protein/chitin ratio around two (Degens et al., 1967; Hunt and Nixon, 1981). However, this latter value is certainly overestimated (in favor of proteins) because these authors employed harsh hydrolytic conditions that have likely degraded a part of the saccharide moieties. In our case, the abundance of chitin identified by SS-NMR confirms earlier findings based on monosaccharide analysis, which showed that both ASM and AIM are composed predominantly of the deacetylated form - due to hydrolytic conditions - of N-acetyl-glucosamine, the monomer of chitin (Dauphin and Marin, 1995). Such a composition where sugars predominate is rather unusual for matrices associated with mollusc shells: most of the works published so far on this type of materials tend to show that the saccharidic moieties are often much less abundant than the protein ones (Agbaje et al., 2018; Immel et al., 2016). Similarities of ASMs profiles with chitin, seen by FTIR, are also in agreement with the ELLA test, since the main chitin-binding lectins such as WGA, STL and LEL give the highest reactivity. The presence of chitin in ASMs may appear odd as this polymer is supposed to be particularly insoluble. This suggests different potential explanations that do not mutually exclude each other: 1. Soluble non-crossed-linked chitin may exist *per se* in the ASM, which would be rather unusual. 2. N-acetylglucosamine oligomers (that "mimic" chitin and are consequently targeted by chitin-binding lectins) are bound to soluble protein core. 3. From our discussion above, the chitin framework occupies an intercrystalline position in the biominerals; consequently, this framework is partly cleaved chemically in a random manner by sodium hypochlorite during the bleaching treatment, and then, released in solution during decalcification. Then, the cleaved fragments remain large enough not to be eliminated by subsequent ultrafiltration. It is interesting to notice that if chitin represents about 80% of the AIM, and that AIM itself accounts for about 70% of the whole matrix, this means that chitin represents between 50 and 60% of the matrix: in other words, this means that the matrix is more chitinous than proteinaceous. This character brings it close to the cuttlefish matrix (Stegemann, 1963), and consequently, to the Sepiidae.

Both insoluble extracts, AIM1 and AIM2, of *S. spirula* show phosphorous content that is inferred as being an integral part of the proteinaceous content most commonly through phosphorylation of the abundant tyrosine and/or serine sidechains (13% as shown above). We therefore propose that the P-signals we have found, both via IR and SS-NMR, are likely associated with post-translationally modified amino acids. While phosphates in general, and phosphorylation in particular, are implicated in regulation of biomineralization, their exact

identity and role herein, remain to be studied (Brauer and Sykes, 1984; Lowenstam, 1981; Sviben et al., 2016).

To investigate further the biochemical fingerprint of the matrix, we performed proteomics followed by two *in silico* investigations: one based on NCBI search on heterologous metazoan models, the second one, on the recently acquired transcriptome of *Spirula spirula*. The first approach identified around 200 peptides, from 7 to 30 residues long, but no complete protein sequences. We are fully aware that this relatively small number of peptides - although very informative - gives a very partial picture of its shell proteome and that this picture may be biased. For example, among all the identified peptides - regardless of the matrix extract - the hydrophobic amino acid residues are predominantly represented in our analysis, a property often encountered in mollusc shell matrices (Marin et al., 2008). However, in the single bulk amino acid analysis published so far on the shell matrix of *S. spirula* (Degens et al., 1967), glutamic acid, serine and aspartic acid are more abundant than glycine, alanine and leucine. However, the two sets of data cannot be compared for several reasons: one of them is that Degens and coworkers analysed only the insoluble matrix, obtained after a harsh decalcification with a mixture of HCl/TCA followed by 6N hydrochloric acid hydrolysis.

Other technical biases, not discussed here, may hamper the interpretation of the proteomic results (see Marin et al., 2016 for more details). Interestingly, the number of identified peptides significantly increased between ASM1 and ASM2. Similarly to what likely happens to chitin during the second bleaching treatment, we suggest that insoluble proteins are randomly cleaved by sodium hypochlorite and consequently, partly solubilized, becoming components of the ASM2 fraction. In total, the overall peptides lead to the identification of 91 protein hits, which constitutes a low score, given the number of peptides. In any event, the peptides recognized several interesting hits, namely specific enzymes, such as tyrosinase, chitinase and serine protease, which constitute a part of the molecular toolkit frequently found in shell matrices (Marie et al., 2012; Marin et al., 2014). Tyrosinase is thought to be biomineralization-linked either by being involved in the melanin biosynthesis or in the periostracum formation (Arivalagan et al., 2016; Nagai et al., 2007; Zhang et al., 2006), while chitinase and serine protease are associated to organic matrix remodelling (Marin et al., 2016; Yonezawa et al., 2016). We also identified a set of sequences that can be related to non-molluscan calcium-binding proteins, with either EF-hand, or EGF-like domains, both belonging to the "high-affinity - low capacity" category. Among the unexpected hits, one finds actin, the main cytoskeletal protein, and histone, a nuclear protein. The actin identified here is identical to that found in the cephalopods to *Idiosepius paradoxus* and *Octopus vulgaris*. The

711 significance of the presence of actin and histone in a shell matrix is discussed below. We also
712 found the putative presence of IgG-Fc binding protein, a candidate that binds immunoglobulins,
713 and which is involved in immune protection (Arivalagan et al., 2016; Harada et al., 1997). We
714 did not detect proteins that exhibit Low Complexity Domains (LCDs) or Repetitive Low
715 Complexity Domains (RLCDs) such as aspartic acid-rich proteins, with one exception: a
716 peptide matching with a hypothetical protein from *Helobdella robusta* (Californian leech) that
717 exhibits a very large C-terminal region composed of D-rich Low Complexity repeats.
718 The second approach we employed, based on the *S. spirula* transcriptome, generated
719 another set of proteins: some of them, like actin, histones or calcium-binding, confirmed the
720 hits found with MASCOT against the "other metazoans" database. Others, not identified before,
721 are relevant in the context of biomineralization. Among the protein list shared with the four
722 extracts, one finds the highly covered transferrin, a serine protease inhibitor, a matrilin and a
723 peptidyl prolyl cis-trans isomerase B-like, in addition to actin and to an unknown protein with
724 a heme-binding domain. Transferrin is a ferritic-ion-binding glycoprotein that controls the level
725 of free iron in biological fluids and its activation requires the presence of an anion, preferably
726 carbonate CO_3^{2-}. At first sight, transferrin might be involved in the incorporation of iron in the
727 shell. However, its presence in the organic matrix may appear odd and the link between iron
728 and calcium carbonate is rather tenuous. Indeed, unpublished data on the iron content of *Spirula*
729 *spirula* shell (Y. Dauphin, pers. comm.) show that the different shell layers contain between 67
730 and 110 ppm of this element, values that do not reflect any exceptional enrichment and fall in
731 the range of most aragonitic mollusc shells (Masuda & Hirano, 1980). We may consequently
732 have to infer other functions than "iron incorporation". Interestingly, transferrin has been
733 identified in the biomineral-associated matrices of two other calcifying systems: the cichlid fish
734 otolith (Weigele et al., 2016) and the hen's eggshell (Nys et al., 1999). In the first model, a true
735 transferrin has been clearly identified together with a transferrin domain-containing protein,
736 OMP-1 (otolith matrix protein-1). It was suggested, without any experimental evidence, that
737 transferrin is associated to the occurrence of ferric ions in the otolith, and to the presence of
738 magnetite particles that are geomagnetic sensors. In the second model, transferrin is suspected
739 to be multifunctional, by interacting with the formation of calcium carbonate and by playing a
740 role in natural immunity (Giansanti et al., 2012). It is also suspected to act as a strong
741 antimicrobial substance, by depriving bacteria of iron (Gautron, personal comm.). It is to note
742 that beside the full-length transferrin present in the four extracts, we also identified a shortened
743 variant of this protein present only in the two AIMs. The two transferrin sequences are 48%
744 identical (69% similar) and may display different functions. In addition to transferrins, we

detected in the four extracts a nameless protein (covered at 7 to 13%) with a SOUL heme-binding domain (pfam04832) which putatively complexes iron. This protein is 45% identical to an unknown protein (NCBI reference sequence XP_014780085.1) from the octopus *Octopus bimaculoides.* To broaden the circle of the relation between iron and biomineralization to other mollusc models, let's mention another protein with analogous functions, *i.e.*, iron binding and storage: ferritin; ferritin was reported to be involved in shell formation of the pearl oyster *Pinctada fucata* but its function was not elucidated (Zhang et al., 2003). All these arguments combined together emphasize the underestimated - but not yet understood - role of iron in calcium carbonate biomineralization.

The elastase inhibitor-like isoform is a member of the SERPIN family (SERine Proteinase INhibitor), a large group of proteins that inhibit the activity of hydrolytic enzymes, such as elastase or cathepsin G. In the last decade, members of this protein family have been recurrently detected as major components of the matrices of multiple calcified tissues, including mollusc shell (Calvo-Iglesias et al., 2017; Marie et al., 2010; Marin et al., 2014). Although the molecular activity has never been firmly tested *in vitro*, protease inhibitors are usually believed to be part of a protective system that prevents the degradation of the calcifying matrix during mineral deposition, by proteolytic enzymes. More generally, they may also be involved in immune functions.

Third candidate of interest, matrilin is an acidic (theoretical pI 4.57) proteoglycan-associated protein and a major component of extracellular matrices (ECM) of various tissues. It comprises a von Willebrand type A domain (vWF-type A, simplified in VWA), commonly found in extracellular matrix proteins. This domain can bind several proteins of the ECM and is involved in cell adhesion (Whittaker and Hynes, 2002). It has been identified in diverse molluscan shell matrix proteins such as PIF (Suzuki et al., 2009) and many others (Arivalagan et al., 2017; Feng et al., 2017). In vertebrate calcified tissues, the VWA-containing matrilin is thought to act as an "organizing" protein of the ECM (Fresquet et al., 2010). In particular, it mediates interactions between collagens and proteoglycans. Its multiple functions are related to its ability to form filaments (Klatt et al., 2000). It is interesting to note that our analysis identified three other VWA domain-containing proteins that exhibit high similarity with diverse α-collagens.

The identified peptidyl prolyl cis-trans isomerase B-like is an enzyme that exhibits a cyclophilin domain, which regulates protein folding by catalysing the cis-trans isomerization of proline imidic peptidic bonds. It has been shown to be involved in collagen maturation in general, but, more specifically, in the maturation of the tooth extracellular matrix during

development in murine model (Pandya et al., 2017). Interestingly, in the shell matrix of *S. spirula*, beside the full-length member, we detected a shortened version of peptidyl prolyl cis-trans isomerase B, with a reasonably good 17% peptidic coverage, and corresponding to 60% of the C-terminus of the first sequence. This fragment is not identical but a variant that exhibits 75% identity (and 89% similarity) with the complete form. Curiously, while the latter is present in the four extracts, the short variant has been detected only in the two AIMs, which suggests that it is cross-linked with other matrix components.

In the protein set identified in three, two or one extract, some other hits are relevant in the biomineralization context. Our biochemical characterization underlines the quantitative importance of the carbohydrate moieties, in particular of chitin. Consequently, the proteomic analysis evidences logically enzymes that are involved, near or far, in the assembly, degradation and modification of extracellular saccharides, and more specifically, in the remodeling of the chitin scaffold: they include chitobiase, heparan sulfate glucosamine sulfrotransferase, a N-acetyl-galactosaminidase, chitin-deacetylase, the UDP-glucuronic acid decarboxylase. They are accompanied by a set of putative carbohydrate-binding proteins, like the Cys-rich secretory protein. The case of deglycase is interesting since this enzyme, identified in three extracts on four, is thought to prevent the formation of Maillard products, *i.e.*, the reaction between sugars and amino acids. Thus, deglycase can be placed in the category of saccharide-interacting enzymes.

Other candidates of importance, the two calcium-binding proteins. Both are acidic (theoretical pI 4.57 and 5.26, respectively) and exhibit the typical EF-hand motif that binds calcium ions with high affinity. These proteins have been often identified in calcium carbonate associated matrices (Feng et al., 2017). They are thought to be part of the extracellular signaling machinery, rather than being involved in crystal nucleation or calcium ion transport for mineralization, a function that is usually attributed to low affinity - high capacity calcium-binding proteins, such as the aspartic-rich proteins, not detected here (Marin et al., 2008).

In Table 6, we identified a series of proteins that are not supposed to be present in the extracellular matrix since they are normally localized in the cytoplasmic or in the nucleus. The first ones include in particular cytoskeletal proteins, such as actin or tubulin, and other cytoplasmic partners that bind them, like elongation factors (actin-binding) or moesin/radixin. Nuclear proteins are represented by different histones (H2A, H2B, H3, H4) and lamin. For cytoskeletal proteins, we cannot exclude that these proteins are all cellular contaminants, in spite of the thorough cleaning of the skeletal tissues by sodium hypochlorite. We cannot rule out neither that a tight link exists between these proteins and the extracellular matrix, several

evidences show it, in particular when vesicles are involved in the process: actin was found to be a component of ECM (Accinni et al., 1983); in cartilage, it is associated to ECM in mineralizing/demineralizing vesicles (Holliday et al., 2020; Rosenthal et al., 2011). In molluscs, Weiss and coworkers evidenced the relationship between the cytoskeletal forces and the shell forming matrix (Weiss et al., 2006), a concept that can be extended to other calcifying systems (Tyszka et al., 2019). For histones, the identification of these very basic (lysine-rich) proteins in a calcium carbonate-associated matrix is not new. It is generally believed that histones or histone-derived peptides may function as antimicrobial agents. This hypothesis has been repeatedly evoked for the eggshell matrix (Réhault-Godbert et al., 2011) and somehow confirmed in another model: Molluskin, an antimicrobial peptide identified in several molluscs, was found to derive from histone H2A (Sathyan et al., 2012). Clearly, the presence of nuclear or cytoplasmic proteins in the calcifying extracellular matrix needs to be reevaluated and explained accurately.

To summarize, the shell matrix of *Spirula spirula*, as seen by the MASCOT analysis of the protein-translated transcriptome of *Spirula spirula*, contains numerous molecular actors, including a minimum of 11 true enzymes and 28 proteins of very different functions. All these 39 members can be categorized as follows: extracelllular matrix proteins (matrilin, collagen-like) together with ECM-binding partners and ECM-interacting enzymes (like cis-trans isomerase) required for ECM protein folding and activation; a set of enzymes that regulate, modify and remodel the saccharidic moieties (chitin deacetylase, N-acetylgalactosaminidase...) together with saccharide-binding proteins that are not enzymes; a set of cytoskeletal proteins (actin, lamin, tubulin) that function in extracellular environment together with their binding partners (elongation factor, moesin). The system includes also proteins involved in cell signaling (14-3-3 protein zeta, ubiquitin, protein with MMACH domain...) and calcium-binding proteins (hippocalcin-like, calmodulin-like). The system possesses also its protective functions at molecular and cellular levels, respectively: at molecular level, these are protease inhibitors on the one hand, and putative bactericidal factors (histones?), immunity-related proteins and possibly proteins that bind iron, on the other hand. Finally, the system contains a set of proteins of totally unknown functions, as exemplified by the unhealthy ribosome biogenesis protein 2 and GLIPR1 protein 1.

Are notably absent from this molecular landscape the following candidates usually found in skeletal matrices: carbonic anhydrase and low complexity domain containing proteins. Carbonic anhydrase (CA) is a key-enzyme in calcium carbonate biomineralization. It reversibly catalyzes the conversion of CO2 to bicarbonate. In the protein-translated transcriptome of *S.*

spirula, we identified three CA sequences, which possess tryptic cleavage sites (data not shown) of which no peptide is present in the matrix proteomic results. This is a strong indication that CA is not incorporated as a shell matrix protein in the Ram's Horn model. This conclusion confirms earlier finding on the shell matrix of the abalone *Haliotis tuberculata* (Le Roy et al., 2014) and shows that CA, depending on the mollusc model is or is not a shell matrix protein. The other absent players of the list are proteins that contain low complexity domains or repetitive low complexity domains. In calcium carbonate biomineralization, theses domains can be of several types, the most known being the acidic, aspartic acid-rich ones, and the hydrophobic "alanine/glycine-rich" domains (Marin et al., 2016). We admit that their underrepresentation in the protein list may be due to technical bias, such domains having few or no trypsin cleavage sites. Clearly, the absence of peptides exhibiting low complexity sequences does not mean that the corresponding proteins are absent from the shell forming-machinery: a separate analysis of the proteome-translated transcriptome (not shown) reveals putative candidates, which - for an unknown reason - are not incorporated to the matrix.

From an evolutionary viewpoint, it is puzzling that most of the peptides obtained do not exhibit any similarity with already known molluscan shell proteins. Among the few hits with molluscan proteins, the majority belongs to a bivalve, the Pacific oyster *Crassostrea gigas* (5 hits in total), while there are only two hits with a cephalopod, the common octopus *Octopus vulgaris*. A comparison of our peptide list with already published data on *Sepia* (Čadež et al., 2017; Le Pabic et al., 2017) and *Nautilus* (Marie et al., 2011, 2009) did not reveal any similarity. When compared with the complete transcriptome of *Octopus bimaculoides* (Albertin et al., 2015) our peptide dataset leads few hits, mainly with actin-like proteins (3 hits, exhibited by all the extract of the organic matrix of *Spirula spirula*) and H4 histones (9 hits in total, only with the ASM2 extract).

Such a paucity of hits with already studied cephalopods is intriguing but may be related to the uncertainty about the phylogenetic position of *Spirula* and to its unique evolutionary history, *i.e.,* its relative evolutionary remoteness from other members of this mollusc class: indeed, within Decabrachia, the placement of the order Spirulida has historically been contentious (reviewed within Lindgren and Anderson, 2018). On the one hand, early phylogenetic studies based on a small number of genes proposed that *Spirula* occupies either a basal position relative to other Decabrachia (Allcock et al., 2011; Warnke et al., 2003) or that a sister taxa relationship exists between the two families Spirulidae and Sepiidae (Strugnell et al., 2005). These views are congruent with morphological characters such as a well-developed shell with chambers and septa. On the other hand, recent studies, including those based on more

comprehensive molecular datasets, have recovered a sister taxon relationship between Spirulida and a clade containing Oegopsida (‘open eyed' squids) and Bathyteuthoidea (mesopelagic to bathopelagic squids) (Lindgren et al., 2012; Strugnell et al., 2017). Similarly, Tanner et al. (2017) reported a sister taxon relationship between spirulid and oegopsid squids. Although these authors did not include bathyteuthoid squids in their analyses, a close phylogenetic relationship between bathyteuthoids and oegopsids is well established (*e.g.* Lindgren et al., 2012; Strugnell et al., 2005). Recent divergence time estimates, incorporating molecular and fossil data, suggest that spirulids and oegospids diverged around 128 Ma (Tanner et al., 2017), which accords with appearance of stem group spirulids in the latest Cretaceous (~66-72 Ma) (Fuchs et al., 2012).

The absence of any similarity with Nautilus is not surprising, considering the deep divergence time between the subclasses Nautiloidea (Nautilus) and Coleoidea (octopus, cuttlefishes and squids) (~416 ± 60 Ma, Kröger et al., 2011). In addition, our study suggests that *Spirula* and *Sepia* have completely different molecular tools to construct their shell. This result is also not surprising. Sepiids and spirulids both possess an internal calcareous shell with a phragmocone and, as a result, they have extensive fossil records in comparison to other extant decabrachian lineages that have lost their shell (*i.e.*, oegopsids, myopsids). The phragmocone is a plesiomorphic character within cephalopods and, therefore, the fact that it is present within sepiids and spirulids does not indicate that these lineages are closely related. In support of this, several recent phylogenetic studies have recovered a sister taxon relationship between sepiids and all remaining decabrachians (including spirulids) (Strugnell et al., 2017; Tanner et al., 2017; Uribe and Zardoya, 2017) with divergence of these two clades estimated to have occurred in the mid Jurassic (Tanner et al., 2017) or in the early Cretaceous (Uribe and Zardoya, 2017). This lapse of time is long enough to allow the independent evolution of the organic matrix proteins: recent findings suggest indeed that several skeletal matrix proteins exhibit a fast evolution rate (Jackson et al., 2006; McDougall et al., 2013). However, a rapid evolution of matrix components is not incompatible with the occurrence of very conservative characters, such as the abundance of chitin or the overall morphological similarities (septate inner shell, for example). Given the modern phylogenetic view of coleoids, which brings Spirulida as a sister group of the non-calcifying oegopsids, it would be relevant to determine how the molecular functions of the proteins used for calcification within *Spirula* have evolved in this latter clade.

Acknowledgements

The authors thank 3P5 proteomic platform (University of Paris, Cochin Institute, INSERM U1016, CNRS UMR8104, Paris). They also thank both Emmanuel Fara and Arnaud Brayard (UMR 6282 Biogeosciences, Dijon) for their valuable comments and advices, and two anonymous reviewers who contributed to improve the manuscript. At last, F. M. and M. O. thank Yannicke Dauphin (UMR CNRS 7205 ISYEB, MNHN, Paris) for providing minor element data on the shell of *S. spirula*. The work of M. Oudot, P. Neige and F. Marin was supported mostly by annual recurrent CNRS funding of UMR Biogéosciences. Complementary funding included CNRS-INTERRVIE project 'BECOME FREE' (F. Marin, 2019). The contribution of I. Ben Shir and A. Schmidt was supported by the Israel Science Foundation grant 2001/17.

Authors contribution

M. O. conducted the biochemical/structural characterization under the guidance of F. M.; L. P. performed FTIR spectroscopy, C. B., proteomic analyses, I. B. S. and A. S, SS-NMR characterization. J. M. S., R. H., P. N. and A. L. provided samples. J. M. S. provided transcriptomic data. M. O., F. M., and P. N. wrote the manuscript with inputs from A. S., L. P and J. M. S.

Legends of figures

Figure 1: *Spirula spirula*. A, specimen from East Australian coast (Townsville, Queensland). B, specimen from Thailand. Note that the first specimen still possesses its initial chamber, while it is missing in the second one.

Figure 2: Shell microstructures of Spirula spirula, in the insertion zone of a septum into the shell wall. A) General view, equatorial section. B) General view of the septum insertion in the outer wall. Note the beveled shape of the whole structure. C) Zoom on the very beginning of the septum into the shell outer wall, showing its organization into an internal unit, with an irregular semi-prismatic layer (s.irr. p). D) Zoom on the external unit of the shell wall, showing its tuberculated and regular semi-prismatic structure (t-r.p). E) Zoom on the lamello-fibrillar structure of the septum. F) Zoom on the contact zone between the septum and the internal shell unit, at the very beginning of the insertion, showing the transformation from the irregular semi-prismatic structure into the lamello-fibrillar one. G) Zoom on the supraseptal annular ridge, observed on the aboral side, dorsal corner, of the septum. In spite of a thin delimitation between the annular ridge and the internal unit, one notices a microstructural continuity.

Figure 3: FT-IR spectra of the shell powder of *Spirula spirula*, of commercial α-chitin and of the different organic fraction of the shell matrix (AIM1, AIM2, ASM1, ASM2) from both localities (i.e. Canaries, C, and Brazil, B). The “domains” of the characteristic absorption bands of protein and saccharidic moieties are highlighted, respectively, in pink and yellow. The values of the main absorption peaks are indicated under each of them, and colored in pink when identical to those of the commercial chitin.

Figure 4: 75.4 MHz ^{13}C CPMAS spectra of acid insoluble extracts (AIM1) of *Spirula spirula* shells from the Canaries and Brazil, and of α-chitin standard (shrimp) showing that the organic content of AIM1 consists primarily of chitin (the dashed lines denote the characteristic peaks of chitin).

Fig. 5: 121.85 MHz ^{31}P CP MAS NMR spectra of the acid insoluble extract (AIM1) of *Spirula spirula* shells from the Canaries and Brazil. The peaks are centered at -0.2 ppm and are

of similar normalized intensities (accounting for sample weights of 16.8 and 6.0 mg, respectively). Each spectrum was obtained acquiring 2k transients.

Figure 6: SDS-Page gel electrophoresis on *Spirula spirula*. A) Silver-staining; B) Bio-rad blue staining; C) Stains-all. PM: standard molecular weight (Fermentas).

Figure 7: *In-vitro* crystallization assay performed on both soluble matrix extracts of *Spirula spirula* (Canaries Islands) ASM1 (top) and ASM2 (bottom). The black arrows indicate the increasing matrix concentration within each well, from 0 (blank) to 16 µg/well (200 µL per well).

Table 1: Summary of the quantity and proportion of each organic shell matrix fraction (ASM, AIM) after on (batch 1) or two (batch 2) bleaching treatments. Separate data were obtained for samples of two different geographical sources, Canary Islands and Brazil.

Table 2: Summary of the reactivity pattern of lectins on both soluble fractions (ASM1 and ASM2) for both localities. 'XXXX' indicates a reactivity that exceeds 60%, 'XXX', a reactivity comprised between 40 and 60%, 'XX' between 20-40%, 'X', between 10-20% and '-' for a reactivity below 10%.

Table 3: Overview of the main proteomic results by MASCOT. A) Summary of the identified hits per extract. B) Cross-comparison of shared hits in the 4 extracts. C) Summary of the identified peptides per extract and occurrence of the most represented AA residues (sorted by decreasing order and colored according to their chemical properties: Blue: basic; red: acidic; orange; hydroxylated; green: hydrophobic)

Table 4: List of the peptides identified by MASCOT within each organic fraction of the shell matrix of Spirula spirula. For each fraction, the numbers in parentheses correspond to the number of peptides (number of queries) detected in total.

Table 5: Summary of the main proteomic hits obtained by MASCOT and their presence/absence (full box/empty box) within each fraction of the organic matrix. The hits mentioned here only represent those shared by, at least, two fractions; the list of the hits that are unique to each extract is available in the supplementary data.

Table 6: Summary of the protein hits identified by MASCOT against the protein-translated transcriptome of *Spirula spirula*. Each of the four fraction was analysed separately. The left column indicates the proteins identified, from the most (top) to the least (bottom) one represented. For each protein, the central colum indicates its occurrence in each fraction (**X**), the peptidic coverage of the sequence (in %) and the three successive numbers, its length (number of AA residues), its molecular weight (in kDa) and its isoelectric point, respectively. Five sequences exhibit an additional number in parentheses, indicating the full length of the sequence with it signal peptide. The central column lists the functional domains, identified by CD-search at NCBI, and some motifs, identified manually and currently found in skeletal matrix proteins. The numbers in parentheses indicate the starting and final AA residue of the conserved domains, and their accession number. Finally, the right column lists the putative function of the identified protein, in relation to shell formation. In particular, several proteins that are primarily cytoplasmic or nuclear (like histones) can also be found in extracellular environment. For these proteins, we tentatively list their function(s) in the context of extracellular biomineralization.

Supp. Fig. 1: Complete FT-IR spectra (extended to 4000 cm^{-1}) of, the shell powder of *Spirula spirula*, of commercial α-chitin and of the different organic fraction of the shell matrix.

Supp. Fig. 2: Additional SDS-Page gels after carbocyanine (Stains-all) staining, showing two more discrete bands on the ASMs, one at 17kDa and another one at very high molecular weight.

Supp. Table 1: List of the proteomic hits identified in one fraction only on the four tested. MASCOT was used against the "other metazoans" dataset.

Bibliography

Accinni, L., Natali, P.G., Silvestrini, M., De Martino, C., 1983. Actin in the extracellular matrix of smooth muscle cells. An immunoelectron microscopic study. Conn. TIss. Res. 11, 69-78.

Agbaje, O.B.A., Ben Shir, I., Zax, D.B., Schmidt, A., Jacob, D.E., 2018. Biomacromolecules within bivalve shells: Is chitin abundant? Acta Biomater. 80, 176–187. https://doi.org/10.1016/j.actbio.2018.09.009

Albeck, S., Aizenberg, J., Addadi, L., Weiner, S., 1993. Interactions of various skeletal intracrystalline components with calcite crystals. J. Am. Chem. Soc. 115, 11691–11697. https://doi.org/10.1021/ja00078a005

Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., Rokhsar, D.S., 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature 524, 220–224. https://doi.org/10.1038/nature14668

Allcock, A.L., Cooke, I.R., Strugnell, J.M., 2011. What can the mitochondrial genome reveal about higher-level phylogeny of the molluscan class Cephalopoda? Zool. J. Linn. Soc. 161, 573–586. https://doi.org/10.1111/j.1096-3642.2010.00656.x

Appellöf, A., 1893. Die schalen von *Sepia*, *Spirula* und *Nautilus* studien über den Bau und das Wachstum. K. Sven. VetenskAkad. Handl., Stock. 25, 1–106.

Arivalagan, J., Yarra, T., Marie, B., Sleight, V.A., Duvernois-Berthet, E., Clark, M.S., Marie, A., Berland, S., 2016. Insights from the shell proteome: Biomineralization to adaptation. Mol. Biol. Evol. 34, 66–77. https://doi.org/10.1093/molbev/msw219

Balmain, J., Hannoyer, B., Lopez, E., 1999. Fourier transform infrared spectroscopy (FTIR) and X-ray diffraction analyses and organic matrix during heating of mother of pearl (nacre) from the shell of the mollusk *Pinctada maxima*. J. Biomed. Mater. Res. 48, 749–754.

Bandel, K., Boletzky, S., 1979. A comparative study of the structure, development and morphological relationships of chambered cephalopod shells. Veliger 21, 313–354.

Barskov, S., 1973. Microstructure of the skeletal layers of *Sepia* and *Spirula* compared with the shell layers of other mollusks. J. Paleontol. 3, 285–294.

Bennett, A.E., Rienstra, C.M., Auger, M., Lakshmi, K. V, Griffin, R.G., 1995. Heteronuclear decoupling in rotating solids. J. Chem. Phys. 103, 6951–6958. https://doi.org/10.1063/1.470372

Bøggild, O.B., 1930. The shell structure of the mollusks. K. Dan. Vidensk. Selsk. Skr. Naturvidensk. Math. Afd. 9, 231–326.

Bouchet, P., Bary, S., Héros, V., Marani, G., 2016. How many species of molluscs are there in the world's oceans, and who is going to describe them? Mémoires du Muséum Natl. d'Histoire Nat. 208, 9–24.

Brauer, M., Sykes, B.D., 1984. Phosphorus-31 nuclear magnetic resonance studies of phosphorylated proteins. Methods Enzymol. 107, 36–81. https://doi.org/10.1016/0076-6879(84)07005-1

Brayard, A., Escarguel, G., Bucher, H., Monnet, C., Brühwiler, T., Goudemand, N., Galfetti, T., Guex, J., 2009. Good genes and good luck: Ammonoid diversity and the end-permian mass extinction. Science 325, 1118–1121. https://doi.org/10.1126/science.1174638

Brock, G., Paterson, J., 2004. A new species of *Tannuella* (Helcionellida, Mollusca) from the early Cambrian of South Australia. Assoc. Australas. Palaeontol. Mem. 30, 133–143.

Bruun, A., 1943. The biology of *Spirula spirula* (L.). Dana Rep. 24, 1–46.

Čadež, V., Škapin, S.D., Leonardi, A., Križaj, I., Kazazić, S., Salopek-Sondi, B., Sondi, I., 2017. Formation and morphogenesis of a cuttlebone's aragonite biomineral structures for the common cuttlefish (*Sepia officinalis*) on the nanoscale: Revisited. J. Colloid Interface Sci. 508, 95–104. https://doi.org/10.1016/j.jcis.2017.08.028

Calvo-Iglesias, J., Pérez-Estévez, D., González-Fernández, Á., 2017. MSP22.8 is a protease inhibitor-like protein involved in shell mineralization in the edible mussel *Mytilus galloprovincialis*. FEBS Open Bio 7, 1539–1556. https://doi.org/10.1002/2211-5463.12286

Campbell, K., MacLennan, D., Jorgensen, A.O., 1983. Staining of the Ca^{2+}-binding proteins, calsequestrin, calmodulin, troponin C, and S-100, with the cationic carbocyanine dye "Stains-all." J. Biol. Chem. 258, 11267–11273.

Checa, A.G., 2018. Physical and biological determinants of the fabrication of molluscan shell microstructures. Front. Mar. Sci. 5, 353. https://doi.org/10.3389/fmars.2018.00353

Clarke, M.R., 1970. Growth and development of *Spirula spirula*. J. Mar. Biol. Assoc. United Kingdom 50, 53–64. https://doi.org/10.1017/S002531540000059X

Crenshaw, M.A., 1972. The soluble matrix from *Mercenaria mercenaria* shell. Biomineralization 6, 6–11.

Cuif, J., Dauphin, Y., Denis, A., Gaspard, D., Keller, J., 1983. Etude des caractéristiques de la phase minérale dans les structures prismatiques du test de quelques mollusques. Bull. du Muséum Natl. d'histoire Nat. Sect. A, Zool. Biol. Ecologie Anim. 5, 679–717.

Dauphin, Y., 1996. The organic matrix of coleoid cephalopod shells: molecular weights and isoelectric properties of the soluble matrix in relation to biomineralization processes. Mar. Biol. 125, 525–529. https://doi.org/10.1007/BF00353265

Dauphin, Y., 1977. Microstructure et flottabilité chez la spirule (Cephalopoda). Comptes rendus Hebd. des séances l'Académie des Sci. Paris D 284, 2483–2485.

Dauphin, Y., 1976. Microstructure des coquilles de céphalopodes. I. *Spirula spirula* L. (Dibranchiata, Decapoda). Bull. du muséum natl. d'histoire Nat. 3ème série, n° 382, Sci. de la Terre 54, 197–238.

Dauphin, Y., Marin, F., 1995. The compositional analysis of recent cephalopod shell carbohydrates by Fourier transform infrared spectrometry and high performance anion

exchange-pulsed amperometric detection. Experientia 51, 278–283. https://doi.org/10.1007/BF01931112

Degens, E.T., Spencer, D.W., Parker, R.H., 1967. Paleobiochemistry of molluscan shell proteins. Comp. Biochem. Physiol., 20, 553–579. doi:10.1016/0010-406x(67)90269-1

Doguzhaeva, L., 1996. Two early Cretaceous spirulid coleoids of the north-western Caucasus: their shell ultrastructure and evolutionary implications. Palaeontology 39, 681–707.

Erben, H.K., 1972. Uber die Bildung und das Wachstum von Perlmutt. Biomineralization 4, 15–46.

Feng, D., Li, Q., Yu, H., Kong, L., Du, S., 2017. Identification of conserved proteins from diverse shell matrix proteome in *Crassostrea gigas*: characterization of genetic bases regulating shell formation. Sci. Rep. 7, 45754. https://doi.org/ 10.1038/srep45754

Fresquet, M., Jowitt, T.A., Stephen, L.A., Ylöstalo, J., Briggs, M.D., 2010. Structural and functional investigations of Matrilin-1 A-domains reveal insights into their role in cartilage ECM assembly. J. Biol. Chem. 285, 34048—34061. https://doi.org/10.1074/jbc.M110.154443

Fuchs, D., Iba, Y., Ifrim, C., Nishimura, T., Kennedy, W.J., Keupp, H., Stinnesbeck, W., Tanabe, K., 2013. Longibelus gen. nov., a new Cretaceous coleoid genus linking Belemnoidea and early Decabrachia. Palaeontology 56, 1081–1106. https://doi.org/10.1111/pala.12036

Fuchs, D., Keupp, H., Trask, P., Tanabe, K., 2012. Taxonomy, morphology and phylogeny of Late Cretaceous spirulid coleoids (Cephalopoda) from Greenland and Canada. Palaeontology 55, 285–303. https://doi.org/10.1111/j.1475-4983.2011.01125.x

Giansanti, F., Leboffe, L., Pitari, G., Ippoliti, R., Antonini, G., 2012. Physiological roles of ovotransferrin. Biochim. Biophys. Acta 1820, 218—225. https://doi.org/10.1016/j.bbagen.2011.08.004

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883

Harada, N., Iijima, S., Kobayashi, K., Yoshida, T., Brown, W.R., Hibi, T., Oshima, A., Morikawa, M., 1997. Human IgGFc binding protein (FcγBP) in colonic epithelial cells exhibits mucin-like structure. J. Biol. Chem. 272, 15232–15241. https://doi.org/10.1074/jbc.272.24.15232

Hare, P.E., 1963. Amino Acids in the Proteins from Aragonite and Calcite in the Shells of *Mytilus californianus*. Science 139, 216-217.

Haring, E., Kruckenhauser, L., Lukeneder, A., 2012. New DNA sequence data on the enigmatic *Spirula spirula* (Linnaeus, 1758)(Decabrachia, suborder Spirulina). Ann. Naturhist. Mus. Wien. Ser. B Bot. Zool. 113, 37–48.

Hoffmann, R., Lemanis, R.E., Wulff, L., Zachow, S., Lukeneder, A., Klug, C., Keupp, H., 2018. Traumatic events in the life of the deep-sea cephalopod mollusc, the coleoid *Spirula spirula*. Deep. Res. Part I Oceanogr. Res. Pap. 142, 127–144. https://doi.org/10.1016/j.dsr.2018.10.007

Holliday, L.S., De Faria, L.P., Rody, W.J., 2020. Actin and actin-associated proteins in extracellular vesicles shed by osteoclasts. Int. J. Mol. Sci., 21, art. numb. 158. https://doi.org/10.3390/ijms21010158

Hunt, S., Nixon, M., 1981. A comparative study of protein composition in the chitin-protein complexes of the beak, pen, sucker disc, radula and oesophageal cuticle of cephalopods. Comp. Biochem. Physiol., 68 B, 535-546. https://doi.org/10.1016/0305-0491(81)90071-7

Immel, F., Broussard, C., Catherinet, B., Plasseraud, L., Alcaraz, G., Bundeleva, I., Marin, F., 2016. The shell of the invasive bivalve species Dreissena polymorpha: biochemical, elemental and textural investigations. PLoS One 11, 1–28. https://doi.org/10.1371/journal.pone.0154264

Jackson, D.J., McDougall, C., Green, K., Simpson, F., Wörheide, G., Degnan, B.M., 2006. A rapidly evolving secretome builds and patterns a sea shell. BMC Biol. 4, 40. https://doi.org/10.1186/1741-7007-4-40

Jang, M.-K., Kong, B.-G., Jeong, Y.-I., Lee, C.H., Nah, J.-W., 2004. Physicochemical characterization of α-chitin, β-chitin, and γ-chitin separated from natural resources. J. Polym. Sci. Part A Polym. Chem. 42, 3423–3432. https://doi.org/10.1002/pola.20176

Jastrzbski, W., Sitarz, M., Rokita, M., Bułat, K., 2011. Infrared spectroscopy of different phosphates structures. Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 79, 722–727. https://doi.org/10.1016/j.saa.2010.08.044

Kanold, J., Guichard, N., Immel, F., Plasseraud, L., Corneillat, M.,, Alcaraz, G., Brümmer, F., Marin, F., 2015. Spine and test skeletal matrices of the Mediterranean sea urchin *Arbacia lixula* - a comparative characterization of their sugar signature. FEBS J. 282, 1891-1905. https://doi.org/10.1111/febs.13242

Kaya, M., Baran, T., Mentes, A., Asaroglu, M., Sezen, G., Tozak, K.O., 2014. Extraction and Characterization of α-Chitin and Chitosan from Six Different Aquatic Invertebrates. Food Biophys. 9, 145–157.

Kaya, M., Mujtaba, M., Ehrlich, H., Salaberria, A.M., Baran, T., Amemiya, C.T., Galli, R., Akyuz, L., Sargin, I., Labidi, J., 2017. On chemistry of γ-chitin. Carbohydr. Polym. 176, 177–186. https://doi.org/10.1016/j.carbpol.2017.08.076

Klatt, R.A., Nitsche, P., Kobbe, B., Mörgelin, M., Paulsson, M., Wagener, R., 2000. Molecular structure and tissue distribution of matrilin-3, a filament-forming extracellular matrix protein expressed during skeletal development. J. Biol. Chem. 275, 3999–4006.

Klug, C., Landman, N.H., Fuchs, D., Mapes, R.H., Pohle, A., Guériau, P., Reguer, S., Hoffmann, R., 2019. Anatomy and evolution of the first Coleoidea in the Carboniferous. Commun. Biol. 2, art. number 280. https://doi.org/10.1038/s42003-019-0523-2

Kong, J., Yu, S., 2007. Fourier transform infrared spectroscopic analysis of protein secondary structures. Acta Biochim. Biophys. Sin. 39, 549–559.

Kröger, B., Vinther, J., Fuchs, D., 2011. Cephalopod origin and evolution: a congruent picture emerging from fossils, development and molecules: Extant cephalopods are younger than previously realised and were under major selection to become agile, shell-less predators. BioEssays 33, 602–613. https://doi.org/10.1002/bies.201100001

Lemanis, R., Stier, D., Zlotnikov, I., Zaslansky, P., Fuchs, D., 2020. The role of mural mechanics on cephalopod palaeoecology. J. R. Soc. Interface 17, 20200009. http://dx.doi.org/10.1098/rsif.2020.0009

Le Pabic, C., Marie, A., Marie, B., Percot, A., Bonnaud-Ponticelli, L., Lopez, P.J., Luquet, G., 2017. First proteomic analyses of the dorsal and ventral parts of the *Sepia officinalis* cuttlebone. J. Proteomics 150, 63–73. https://doi.org/10.1016/j.jprot.2016.08.015

Le Roy, N., Jackson, D.J., Marie, B., Ramos-Silva, P., Marin, F., 2014. The evolution of metazoan alpha-carbonic anhydrases and their roles in $CaCO_3$ biomineralization. Front. Zool. 11, art. numb. 75. https://doi.org/10.1186/s12983-014-0075-8.

Lindgren, A.R., Anderson, F.E., 2018. Assessing the utility of transcriptome data for inferring phylogenetic relationships among coleoid cephalopods. Mol. Phylogenet. Evol. 118, 330–342. https://doi.org/10.1016/j.ympev.2017.10.004

Lindgren, A.R., Pankey, M.S., Hochberg, F.G., Oakley, T.H., 2012. A multi-gene phylogeny of Cephalopoda supports convergent morphological evolution in association with multiple habitat shifts in the marine environment. BMC Evol. Biol. 12. https://doi.org/10.1186/1471-2148-12-129

Linné, C. von, 1758. Systema naturae per regna tria naturae :secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Holmiae.

Lowenstam, H.A., 1981. Minerals formed by organisms. Science 211, 1126-1131. https://doi.org/10.1126/science.7008198

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J., Marchler-Bauer, A., 2020. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 48, D265-D268. https://doi: 10.1093/nar/gkz991

Marie, B., Joubert, C., Tayalé, A., Zanella-Cleón, I., Belliard, C., Piquemal, D., Cochennec-Laureau, N., 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. U. S. A. 109, 20986–20991. https://doi.org/10.1073/pnas.1210552109

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, 1495–1506. https://doi.org/10.1002/cbic.200900009

1222 Marie, B., Zanella-Cléon, I., Corneillat, M., Becchi, M., Alcaraz, G., Plasseraud, L., Luquet,
1223 G., Marin, F., 2011. Nautilin-63, a novel acidic glycoprotein from the shell nacre of
1224 *Nautilus macromphalus*. FEBS J. 278, 2117–2130. https://doi.org/10.1111/j.1742-
1225 4658.2011.08129.x

1226 Marie, B., Zanella-Cléon, I., Le Roy;, N., Becchi, M., Luquet, G., Marin, F., 2010. Proteomic
1227 analysis of the acid-soluble nacre matrix of the bivalve *Unio pictorum*: detection of novel
1228 carbonic anhydrase and putative protease inhibitor proteins. ChemBioChem 11, 2138–
1229 2147. https://doi.org/10.1002/cbic.201000276

1230 Marin, F., Bundeleva, I., Takeuchi, T., Immel, F., Medakovic, D., 2016. Organic matrices in
1231 metazoan calcium carbonate skeletons: Composition, functions, evolution. J. Struct. Biol.
1232 196, 98–106. https://doi.org/10.1016/j.jsb.2016.04.006

1233 Marin, F., Le Roy, N., Marie, B., Ramos-Silva, P., Wolf, S., Benhamada, S., Guichard, N.,
1234 Immel, F., 2014. Synthesis of calcium carbonate biological materials: How many proteins
1235 are needed? Key Eng. Mater. 614, 52–61.
1236 https://doi.org/10.4028/www.scientific.net/KEM.614.52

1237 Marin, F., Luquet, G., 2004. Molluscan shell proteins. Comptes Rendus Palevol 3, 469–492.
1238 https://doi.org/10.1016/J.CRPV.2004.07.009

1239 Marin, F., Luquet, G., Marie, B., Medakovic, D., 2008. Molluscan shell proteins: primary
1240 structure, origin, and evolution. Curr. Top. Dev. Biol. 80, 209–276.
1241 https://doi.org/10.1016/S0070-2153(07)80006-8

1242 Masuda, F., Hirano, M., 1980. Chemoical composition of some moderm pelecypod shells. Sci.
1243 Rep. Inst. Geosci. Univ. Tsukuba, sec B 1, 163-177.

1244 McDougall, C., Aguilera, F., Degnan, B.M., 2013. Rapid evolution of pearl oyster shell matrix
1245 proteins with repetitive, low-complexity domains. J. R. Soc. Interface 10, 20130041.
1246 https://doi.org/10.1098/rsif.2013.0041

1247 Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., Worm, B., 2011. How many species are
1248 there on earth and in the ocean? PLoS Biol. 9, 1–8.
1249 https://doi.org/10.1371/journal.pbio.1001127

1250 Morrissey, J.H., 1981. Silver stain for proteins in polyacrylamide gels: a modified procedure
1251 with enhanced uniform sensitivity. Anal. Biochem. 117, 307–310.
1252 https://doi.org/10.1016/0003-2697(81)90783-1

1253 Murray, J.W., 1985. Atlas of invertebrate macrofossils. Longman, The Palaeontological
1254 Association, Harlow, England.

1255 Mutvei, H., 2018. Cameral deposits in Paleozoic cephalopods. GFF 140, 254–263.
1256 https://doi.org/10.1080/11035897.2018.1483966

1257 Mutvei, H., 1964. On the shells of *Nautilus* and *Spirula* with notes on the shell secretion in non-
1258 cephalopod molluscs. Ark. för Zool. 16, 221–278.

1259 Nagai, K., Yano, M., Morimoto, K., Miyamoto, H., 2007. Tyrosinase localization in mollusc
1260 shells. Comp. Biochem. Physiol. - B Biochem. Mol. Biol. 146, 207–214.
1261 https://doi.org/10.1016/j.cbpb.2006.10.105

Neige, P., Warnke, K., 2010. Just how many species of *Spirula* are there ? A morphometric approach, , in: Tanabe, K., Shigeta, Y., Sasaki, T. & Hirano, H. (Eds.), Cephalopods - Present and Past. Tokai University Press, Tokyo, pp. 77–84.

Nishiguchi, M.K., Mapes, R.H., 2008. Cephalopoda, in: Ponder, W.F. & Lindberg, D.R. (Eds.), Phylogeny and Evolution of the Mollusca. University of California Press, Berkeley, CA, pp. 163–199.

Nys, Y., Hincke, M.T., Arias, J.L., Garcia-Ruiz, J.M., Solomon, S.E., 1999. Avian eggshell mineralization. Avian Poult. Biol. Rev. 10, 143–166.

Ohkouchi, N., Tsuda, R., Chikaraishi, Y., Tanabe, K., 2013. A preliminary estimate of the trophic position of the deep-water ram's horn squid *Spirula spirula* based on the nitrogen isotopic composition of amino acids. Mar. Biol. 160, 773–779. https://doi.org/10.1007/s00227-012-2132-1

Palmer, A.R., 1992. Calcification in marine molluscs: how costly is it? Proc. Natl. Acad. Sci. U. S. A. 89, 1379–1382. https://doi.org/10.1073/pnas.89.4.1379

Palmer, A.R., 1983. Relative cost of producing skeletal organic matrix versus calcification: Evidence from marine gastropods. Mar. Biol. 75, 287–292. https://doi.org/10.1007/BF00406014

Pandya, M., Liu, H., Dangaria, S.J., Zhu, W., Li, L.L., Pan, S., Abufarwa, M., Davis, R.G., Guggenheim, S., Keiderling, T., Luan, X., Diekwisch, T.G.H., 2017. Integrative temporo-spatial, mineralogic, spectroscopic, and proteomic analysis of postnatal enamel development in teeth with limited growth. Front. Physiol. 8, 793. https://doi.org/10.3389/fphys.2017.00793

Pavat, C., Zanella-Cléon, I., Becchi, M., Medakovic, D., Luquet, G., Guichard, N., Alcaraz, G., Dommergues, J.L., Serpentini, A., Lebel, J.M., Marin, F., 2012. The shell matrix of the pulmonate land snail *Helix aspersa maxima*. Comp. Biochem. Physiol. - B Biochem. Mol. Biol. 161, 303–314. https://doi.org/10.1016/j.cbpb.2011.12.003

Ponder, W.F., Lindberg, D.R., Ponder, J.M., 2019. Biology and Evolution of the Mollusca. CRC Press.

Radev, L., Y. Mostafa, N., Michailova, I., M. M. Salvado, I., H. V. Fernandes, M., 2012. In Vitro bioactivity of collagen/calcium phosphate silicate composites, cross-linked with chondroitin sulfate. Int. J. Mater. Chem. 2, 1–9. https://doi.org/10.5923/j.ijmc.20120201.01

Réhault-Godbert, S., Hervé-Grépinet, V., Gautron, J., Cabau, C., Nys, Y., Hincke, M., 2011. Molecules involved in chemical defence of the chicken egg, in: Nys, Y., Bain, M., van Immerseel, F. (Eds.), Improving the Safety and Quality of Eggs and Egg Products - Egg Chemistry, Production and Consumption. Woodhead Publishing Series in Food Science, Technology and Nutrition , pp. 183–208. https://doi.org/10.1533/9780857093912.2.183

Rosenthal, A.K., Gohr, C.M., Ninomiya, J., Wakim, B.T., 2011. Proteomic analysis of articular cartilage vesicles from normal and osteoarthritic cartilage. Arthritis Rheum. 63, 401-411. https://doi.org/10.1002/art.30120

Rusenko, K.W., Donachy, J.E., Wheeler, A.P., 1991. Purification and characterization of a shell matrix phosphoprotein from the american oyster, in: Sykes, C.S., Wheeler, A.P. (Eds.), Surface Reactive Peptides and Polymers - Discovery and Commercialization. ACS Symposium Series, 444. American Chemical Society, pp. 107–124. https://doi.org/doi:10.1021/bk-1991-0444.ch008

Sarashina, I., Yamaguchi, H., Haga, T., Iijima, M., Chiba, S., Endo, K., 2006. Molecular evolution and functionally important structures of molluscan dermatopontin: implications for the origins of molluscan shell matrix proteins. J. Mol. Evol. 62, 307–318. https://doi.org/10.1007/s00239-005-0095-2

Sathyan, N., Rosamma, P., Chaithanya, E.R., Anil Kumar, P.R., 2012. Identification and molecular characterization of Molluskin, a histone-H2A-derived antimicrobial peptide from molluscs. ISRN Mol. Biol. 2012, 219656. https://doi.org/10.5402/2012/219656

Schmitt, N., Marin, F., Thomas, J., Plasseraud, L., Demoy-Schneider, M., 2018. Pearl grafting: tracking the biological origin of nuclei by straightforward immunological methods. Aquac. Res. 49, 692-700. https://doi.org/10.1111/are.13499

Simkiss, K., Wilbur, K.M., 1989. Biomineralization. Cell biology and mineral deposition. Academic Press, San Diego. https://doi.org/https://doi.org/10.1016/B978-0-08-092584-4.50001-9

Stegemann, H., 1963. Proteine (Conchagene) und chitin im stützgewebe von tintenfischen. Hoppe. Seylers. Z. Physiol. Chem. 331, 269–279. https://doi.org/10.1515/bchm2.1963.331.1.269

Stejskal, E.O., Schaefer, J., Waugh, J.S., 1977. Magic-angle spinning and polarization transfer in proton-enhanced NMR. Journal of Magnetic Resonance 28, 105-112. https://doi.org/10.1016/0022-2364(77)90260-8

Strugnell, J., Norman, M., Jackson, J., Drummond, A.J., Cooper, A., 2005. Molecular phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) using a multigene approach; the effect of data partitioning on resolving phylogenies in a Bayesian framework. Mol. Phylogenet. Evol. 37, 426–441. https://doi.org/10.1016/j.ympev.2005.03.020

Strugnell, J.M., Hall, N.E., Vecchione, M., Fuchs, D., Allcock, A.L., 2017. Whole mitochondrial genome of the Ram's Horn Squid shines light on the phylogenetic position of the monotypic order Spirulida (Haeckel, 1896). Mol. Phylogenet. Evol. 109, 296–301. https://doi.org/10.1016/j.ympev.2017.01.011

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

Sviben, S., Gal, A., Hood, M.A., Bertinetti, L., Politi, Y., Bennet, M., Krishnamoorthy, P., Schertel, A., Wirth, R., Sorrentino, A., Pereiro, E., Faivre, D., Scheffel, A., 2016. A vacuole-like compartment concentrates a disordered calcium phase in a key coccolithophorid alga. Nat. Commun. 7, 11228. https://doi.org/10.1038/ncomms11228

Takeuchi, T., Plasseraud, L., Ziegler-Devin, I., Brosse, N., Shinzato, C., Satoh, N., Marin, F., 2018. Biochemical characterization of the skeletal matrix of the massive coral, *Porites*

australiensis – The saccharide moieties and their localization. J. Struct. Biol. 203, 219–229. https://doi.org/10.1016/j.jsb.2018.05.011

Tanner, A.R., Fuchs, D., Winkelmann, I.E., Ribeiro, M., Gilbert, M.T.P., Pankey, M.S., Kocot, K.M., Halanych, K.M., Oakley, T.H., Fonseca, R.R., Pisani, D., Vinther, J., Vinther, J., 2017. Molecular clocks indicate turnover and diversification of modern coleoid cephalopods during the Mesozoic Marine Revolution. Proc. R. Soc. B 284, 20162818.

Tyszka, J., Bickmeyer, U., Raitzsch, M., Bijma, J., Kaczmarek, K., Mewes, A., Topa, P., Janse, M., 2019. Form and function of F-actin during biomineralization revealed from live experiments on foraminifera. Proc. Natl. Acad. Sci. USA, 116, 4111-4116. https://doi.org/10.1073/pnas.1810394116

Uribe, J.E., Zardoya, R., 2017. Revisiting the phylogeny of Cephalopoda using complete mitochondrial genomes. J. Molluscan Stud. 83, 133–144. https://doi.org/10.1093/mollus/eyw052

Vinther, J., 2015. The origins of molluscs. Palaeontology 58, 19–34. https://doi.org/10.1111/pala.12140

Warnke, K., Plötner, J., Santana, J.I., Rueda, M.J., Llinas, O., 2003. Reflections on the phylogenetic position of *Spirula* (Cephalopoda): preliminary evidence from the 18S ribosomal RNA gene. Berliner Paläobiol. Abh. 3, 253–260.

Weigele, J., Franz-Odendaal, T.A., Hilbig, R., 2016. Not all inner ears are the same: otolith matrix proteins in the inner ear of sub-adult cichlid fish, *Oreochromis Mossambicus*, reveal insights into the biomineralization process. Anat. Rec. 299, 234–245. https://doi.org/10.1002/ar.23289

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. https://doi.org/10.1021/bi00286a023

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

Whittaker, C.A, Hynes, R.O., 2002. Distribution and evolution of von Willebrand:integrin a domains: widely dispersed adhesion and elsewhere. Mol. Biol. Cell 13, 3369-3387. https://doi.org/10.1091/mbc.E02-05-0259

Yonezawa, M., Sakuda, S., Yoshimura, E., Suzuki, M., 2016. Molecular cloning and functional analysis of chitinases in the fresh water snail, *Lymnaea stagnalis*. J. Struct. Biol. 196, 107–118. https://doi.org/10.1016/j.jsb.2016.02.021

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

Zhang, Y., Meng, Q., Jiang, T., Wang, H., Xie, L., Zhang, R., 2003. A novel ferritin subunit involved in shell formation from the pearl oyster (*Pinctada fucata*). Comp. Biochem.

1382 Physiol. Part B Biochem. Mol. Biol. 135, 43–54. https://doi.org/10.1016/S1096-
1383 4959(03)00050-2