



HAL
open science

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

► To cite this version:

Morgane Oudot, Pascal Neige, Ira Ben Shir, Asher Schmidt, Jan M. Strugnell, et al.. The shell matrix and microstructure of the Ram's Horn squid: molecular and structural characterization.. *Journal of Structural Biology*, 2020, 211 (1), pp.107507. 10.1016/j.jsb.2020.107507 . hal-02906412

HAL Id: hal-02906412

<https://hal.science/hal-02906412v1>

Submitted on 23 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

3
4 Morgane Oudot¹, Pascal Neige¹, Ira Ben Shir², Asher Schmidt², Jan M. Strugnell^{3,4}, Laurent
5 Plasseraud⁵, Cédric Broussard⁶, René Hoffmann⁷, Alexander Lukeneder⁸ and Frédéric Marin^{1*}

6
7 Authors affiliations:

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

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

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

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

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

18 ⁶ 3P5 Proteomic Platform, Cochin Institute, 22 Rue Méchain, Paris, France. 3P5
19 Proteomic Platform, University of Paris, Cochin Institute, INSERM, U1016, CNRS,
20 UMR8104, F-75014 PARIS, France

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

23 ⁸ Natural History Museum Vienna, Burgring 7, 1010 Vienna, Austria.

24
25
26 * To whom correspondence should be sent.

27

28 **Abstract**

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

51

52

53 **Introduction**

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

60 epithelium that delineates the mantle, the ciliated organ that coats the inner surface of the shell
61 (Checa, 2018; Simkiss and Wilbur, 1989). The shell fabrication requires specific cellular and
62 molecular machinery: the mineral deposition is indeed controlled by an extracellular organic
63 matrix, which is subsequently occluded. Since its discovery as a key-player in regulating shell
64 biomineralization, the organic matrix has been the subject of a large number of biochemical
65 analyses focused on its protein moieties (Crenshaw, 1972; Hare, 1963; Marin and Luquet, 2004;
66 Rusenko et al., 1991; Sarashina et al., 2006; Weiner, 1983). Most of these studies concentrated
67 their efforts on bivalves and gastropods, the two modern largest mollusc classes. Far less studied
68 are the cephalopods. They represent however an important class, with about 800 living and
69 more than 10 000 fossil species (<http://cephbase.eol.org/>; Nishiguchi and Mapes, 2008). The
70 fossil record indicates that the first indisputable cephalopods appeared in the Middle Cambrian
71 (Kröger et al., 2011) but they may have been present as early as the Lower Cambrian with
72 *Tannuella elinorae* (~530 Ma, Brock and Paterson, 2004; Kröger et al., 2011; Vinther, 2015).
73 Since then, cephalopods experienced a complex evolutionary history with important
74 Ordovician, Devonian-Carboniferous, Triassic and Cenozoic radiations and major crises,
75 including the end-Permian and end-Cretaceous mass extinctions (Brayard et al., 2009; Murray,
76 1985; Nishiguchi and Mapes, 2008). From a "biomineralization" viewpoint, cephalopods have
77 a singular evolutionary trajectory (Klug et al., 2019), with a general trend to shell size reduction
78 and internalization (cuttlefish bone), a drastic mineralization decrease (the organic squid pen),
79 leading ultimately to complete loss of skeletal structures, in the most recently appeared
80 cephalopods, *i.e.*, octopuses. Although the clade is commonly recognized as monophyletic,
81 several uncertainties remain regarding the phylogenetic relationships among the major lineages
82 of extant coleoid cephalopods, including that of the internally shelled sepiids (cuttlefishes) and
83 *Spirula* (Lindgren and Anderson, 2018; Strugnell et al., 2017; Tanner et al., 2017). In addition,
84 the molecular mechanism of shell formation remains poorly known for these taxa.

85 In this general framework, we focused on the Ram's Horn squid *Spirula spirula* (Linné,
86 1758). This species is a small-sized mesopelagic cephalopod (Coleoidea, Decabrachia), rarely
87 observed in its natural environment. It is considered as a monospecific genus (Haring et al.,
88 2012; Neige and Warnke, 2010) and even a monospecific order, whose shell exhibits peculiar
89 characteristics. Some of them are reminiscent of the distantly related *Nautilus*, such as
90 planispiral coiling and a chambered shell where chambers are separated by concave septa and
91 communicate via a siphuncle, a primordial morphology (plesiomorphic character) within
92 cephalopods. However, the siphuncle of *S. spirula* is in proximal position, while that of *Nautilus*

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

97 Even though the biology of *S. spirula* remains largely elusive (Ohkouchi et al., 2013),
98 it is quite common to find shells on beaches of the Atlantic, Indian and West Pacific oceans
99 (Bruun, 1943, Clarke, 1970) and different studies have investigated their microstructure (Cuif
100 et al., 1983, Bandel and Boletzky v., 1979, Dauphin, 1976, 1977). Nevertheless, the shell
101 biochemistry has been explored only very superficially (Dauphin, 1996; Dauphin and Marin,
102 1995; Degens et al., 1967). The aim of this work is to determine the biochemical characteristics
103 of the shell matrix of *S. spirula*, with a view to understanding its multiple molecular functions
104 in biomineralization. The second objective of our study is to compare our data with that
105 obtained from better known representatives of the class Cephalopoda and, by extension, of the
106 phylum Mollusca, in order to draw macroevolutionary conclusions.

107

108

109 **Material and Methods**

110 *Shell sampling*

111 All *Spirula spirula* shell samples used in this study were collected on the beach. They
112 were provided by four of us (R. H., J. M. S., P. N. and A. L.). For subsequent analyses, we used
113 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

145 days (6 water changes) and lyophilized. In parallel, the precipitate (*i.e.* pellet) containing the
146 acid insoluble matrix (AIM) was rinsed via 5 cycles of resuspension in Milli-Q water -
147 centrifugation, each supernatant being added to the ASM. Finally, the AIM pellet was
148 lyophilized. All the lyophilisates were weighed on a precision balance (Sartorius Quintix35-
149 1S) three to five times, and the mean value was calculated. In total, three extractions from the
150 Canaries samples were performed, and two from the Brazil ones, leading each time to four
151 batches of extracts, both ASM (ASM1_b ASM2_b, ASM1_c ASM2_c) and AIM (AIM1_B AIM2_B,
152 AIM1_C AIM2_C).

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

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

163 *Solid State Nuclear Magnetic Resonance (SS-NMR)*

164 ¹³C (75.4 MHz) and ³¹P (121.85 MHz) Magic Angle Spinning (MAS) NMR
165 measurements were carried out on 300 MHz solid state NMR Bruker AVANCE III spectrometer
166 using a 4 mm triple-resonance MAS NMR probe with zirconia rotors. Samples were spun at
167 10,000 ± 2 Hz. Two excitation techniques were employed: DE - direct excitation, and CP -
168 Cross Polarization (CP), both using rotor matched chemical shift echo (Stejskal et al., 1977).
169 All experiments employed: 5.0 μs π/2, 10.0 μs π pulse widths, and an echo interval of 100 μs
170 identical to the rotor period T_R ; TPPM ¹H decoupling (Bennett et al., 1995) with rf field strength
171 of 100 kHz. The Hartmann-Hahn CP employed 50 kHz rf level for the X-channel (¹³C, ³¹P) and
172 50-70 kHz ramped rf level for the ¹H channel with 1 and 2 ms duration (contact time, ct) for
173 ¹³C and ³¹P respectively. Experiments were repeated using relaxation delays of 3 sec (¹³C)
174 acquiring 8k and 40k transients (Canaries and Brazil samples, respectively) and 3 sec for ³¹P
175 acquiring 2k transients. Complementary ¹³C DE echo MAS NMR measurements were carried

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

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

186

187 *Mono-dimensional PAGE*

188 The shell extracts were analyzed by conventional mono-dimensional denaturing SDS-
189 PAGE (Bio-Rad, Mini Protean III gels), on precast gradient gels (Mini-PROTEAN TGX Gel
190 4-15% acrylamide, 90 x 70 mm, Bio-Rad) or on homemade gels (12% polyacrylamide). Prior
191 to migration, the samples were treated as follows: lyophilisates (both ASM and AIM) were
192 dissolved in 2X Laemmli Sample Buffer (LSB) to a final matrix concentration of 5 µg/µL. Both
193 preparations were heat-denatured at 100 °C: 5 min. for the ASM and 10 min. for the AIM. Note
194 that the AIMS are not fully solubilized by the Laemmli treatment: only the soluble fractions,
195 referred as Laemmli-Soluble AIMS or LS-AIMS, can be visualized on gels. The preparations
196 were cooled down and shortly centrifuged before being applied on the top of the gel.
197 Subsequently, gels were stained by silver nitrate (Morrissey, 1981), carbocyanine, *i.e.*, "Stains-
198 all" (Campbell et al., 1983) and with Bio-Rad blue dye. Other staining methods, such as PAS
199 for glycoproteins and Alcian blue for polyanionic substances, were applied too, but did not
200 produce significant results.

201 *Protein-calcium interaction*

202 The effect of the *Spirula spirula* ASM on the growth of calcium carbonate crystals was
203 tested *in vitro*, as previously described (Albeck et al., 1993) with some modifications (Pavat et
204 al., 2012). In this assay, crystal growth is obtained by slow diffusion of NH₄HCO₃ vapor into
205 the CaCl₂ solution. To this end, a 10 mM solution of CaCl₂ solutions was prepared fresh and
206 filtered (5 µm pore size) and used for serial dilutions of ASM in the range [0.25-16 µg/well,

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

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

219 *Enzyme-Linked Lectin Assay (ELLA)*

220 Screening of *Spirula spirula* ASMs by Enzyme-Linked Lectin Assay (ELLA) was
221 performed using a set of 21 biotinylated lectins (Vector Labs, Kits I to III, ref. BK-1000, BK-
222 2000, BK-3000 respectively). The binding preferences and specificities of each lectin are
223 indicated in Kanold et al. (2015), which compiles data from different bibliographical sources.

224 The test was carried out in 96-well microplates (MaxiSorp, Nunc™, Rockslide,
225 Denmark): the *Spirula* ASM diluted in TBS buffer (200 ng/well) was incubated for 90 min at
226 37 °C, immediately followed by three washing steps with TBS/Tween20. Wells were blocked
227 with 1X concentrated Carbo-free blocking solution (Vector Labs, ref. SP-5040) for at least 30
228 min at 37 °C. The diluted lectins (in TBS/Tween20) were incubated into the microplate for 90
229 min at 37 °C, at the following dilution factors: 200 times for kits I and II, and 100 times for kit
230 III. After three washes with TBS/Tween20, the microplates were incubated with a diluted
231 solution of Avidin-Alkaline Phosphatase conjugate (1:70000, ref. Sigma A7294, St. Louis, MO,
232 USA) for 90 min at 37 °C, for the detection of bound biotinylated lectins. They were then
233 washed five times with TBS/Tween20 and revealed with a substrate solution containing
234 aqueous diethanolamine (9.7% vol/vol, pH adjusted to 9.8 with HCl) and phosphatase substrate
235 (0.5 mg/mL, pNPP Tablet, Sigma, St. Louis, MO, USA). The microplates were read at 405 nm
236 every 15 minutes (BioRad Model 680). A check of the background signal without ASM, lectin
237 or Avidin-AP showed no reaction with the substrate solution. This test was performed with
238 quadruplicates of each lectin and the results were normalized and translated in percentage of

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

241

242 *RNA extraction, transcriptome sequencing and assembly*

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

256

257 *Proteomic analysis of the shell extracts and subsequent in silico analysis*

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

270 For MS and MS/MS ORBITRAP, analyses were performed using an Ultimate 3000
271 Rapid Separation Liquid Chromatographic (RSLC) system (Thermo Fisher Scientific) online
272 with a Q-Exactive Plus hybrid quadrupole Orbitrap mass spectrometer (Thermo Fisher
273 Scientific). Briefly, peptides were dissolved in 10 μ L of 10 % ACN - 0.1 % trifluoroacetic acid
274 (TFA). Then, peptides were loaded and washed on a C₁₈ reverse phase pre-column (3 μ m
275 particle size, 100 Å pore size, 75 μ m i.d., 2 cm length). The loading buffer contained 98 % H₂O,
276 2 % ACN and 0.1 % TFA. Peptides were then separated on a C₁₈ reverse phase resin (2 μ m
277 particle size, 100 Å pore size, 75 μ m i.d., 25 cm length) with a 1 hour gradient from 99 % A
278 (0.1 % FA and 100 % H₂O) to 90 % B (80 % ACN, 0.085 % FA and 20 % H₂O).

279 The mass spectrometer acquired data throughout the elution process and operated in a
280 data dependent scheme with full MS scans acquired with the Orbitrap, followed by up to 10
281 MS/MS HCD spectra. Mass spectrometer settings were full MS (AGC: 3x10⁶, resolution:
282 7x10⁴, m/z range 400-2000, maximum ion injection time: 100 ms); MS/MS (AGC: 1x10⁵,
283 maximum injection time: 100 ms, isolation windows: 4m/z Da, dynamic exclusion time setting:
284 15 s). The fragmentation was permitted for precursors with a charge state of 2, 3, 4 and up. For
285 the spectral processing, the software used to generate .mgf files is Proteome discoverer 1.4.

286 Database searches were carried out using Mascot version 2.5 (MatrixScience, London,
287 UK) on 'Other Metazoa' proteins (2,282,777 sequences) from NCBI Inr databank containing
288 67,337,701 sequences; 24,122,812,982 residues (April 2018) (<http://www.ncbi.nlm.nih.gov/>).
289 A second search was performed with the 16405 sequences of the protein-translated
290 transcriptome built from the mantle tissue of *Spirula spirula*. In each case, the enzyme
291 specificity was trypsin's and up to one missed tryptic cleavage was tolerated. The precursor
292 mass tolerance was set to 4 ppm and the fragment mass tolerance to 20 mmu for Q-Exactive
293 data. Carbamidomethylation of cysteins and oxidation of methionines were set as variable
294 modifications. Hits with human keratines were ignored.

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

303

304

Results

305

Shell microstructures

306

The different shell microstructures identified in the *S. spirula* are shown in Figure 2.

307

SEM observations confirmed that the shell wall is constituted of two layers: inner and outer.

308

However, we noticed that the internal one is irregular semi-prismatic (Fig. 2C), while the

309

external one appears as tuberculated and regular semi-prismatic (Fig. 2D), with well-developed

310

spicular crystals. Between these two layers, we can observe a thin blackish layer. Moreover,

311

the septa show a lamello-fibrillar pattern (Fig. 2E), which differs from a nacreous

312

microstructure *sensu stricto*. We also noticed that the septa progressively insert into the shell

313

wall in the form of a bevel. An observation made at higher magnification revealed that the

314

internal layer progressively changes into the septa (Fig. 2E) with very gradual transition from

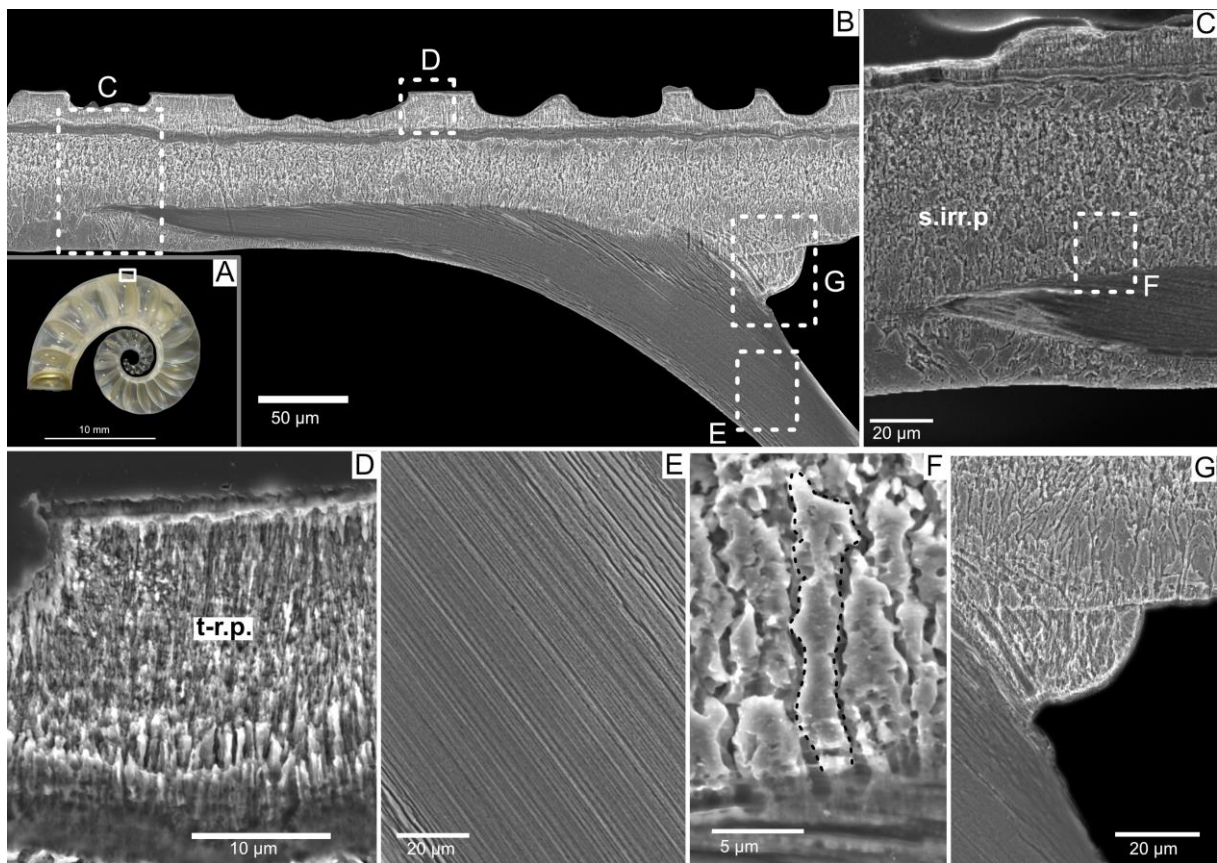


Figure 2: Shell microstructures of *Spirula spirula*, in the insertion zone of a septum into the shell wall. A) General view, equatorial section. B) General view of the septum insertion in the outer wall. Note the beveled shape of the whole structure. C) Zoom on the very beginning of the septum into 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 suprasedal 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.

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

317

318 *Quantification of the organic matrix in the shell*

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

329

330 *FT-IR spectroscopy*

331 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.

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

334 the comparison, the spectra exclusively
 335 focus on the range of 2000-400 cm^{-1} (full
 336 FT-IR data are depicted in the
 337 supplementary Figure 1). Firstly, we
 338 verified that the mineralogy of the samples
 339 was fully aragonitic - characterized in
 340 particular by the 700-713 cm^{-1} double peak
 341 (ν_4 internal vibration modes of CO_3^{2-})
 342 (Balmain et al., 1999) - by checking
 343 different zones of the shell (initial
 344 chamber, one random intermediate
 345 chamber, last preserved chamber). For
 346 clarity, only the spectrum acquired from
 347 the powder of the last chamber is shown in
 348 Fig. 3 (top spectrum-shell powder). In the
 349 2500-4000 cm^{-1} range (not shown in Fig.
 350 3), the shell matrix samples exhibit one
 351 broad band located at 3271 cm^{-1} , which is
 352 usually attributed to the amide A group
 353 [$\nu(\text{N-H})$] (Kong and Yu, 2007). In
 354 addition, one low amplitude enlarged band
 355 centered at 2932 cm^{-1} can be observed, and
 356 usually assigned to the $\nu(\text{C-H})$ stretching
 357 vibrations.

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

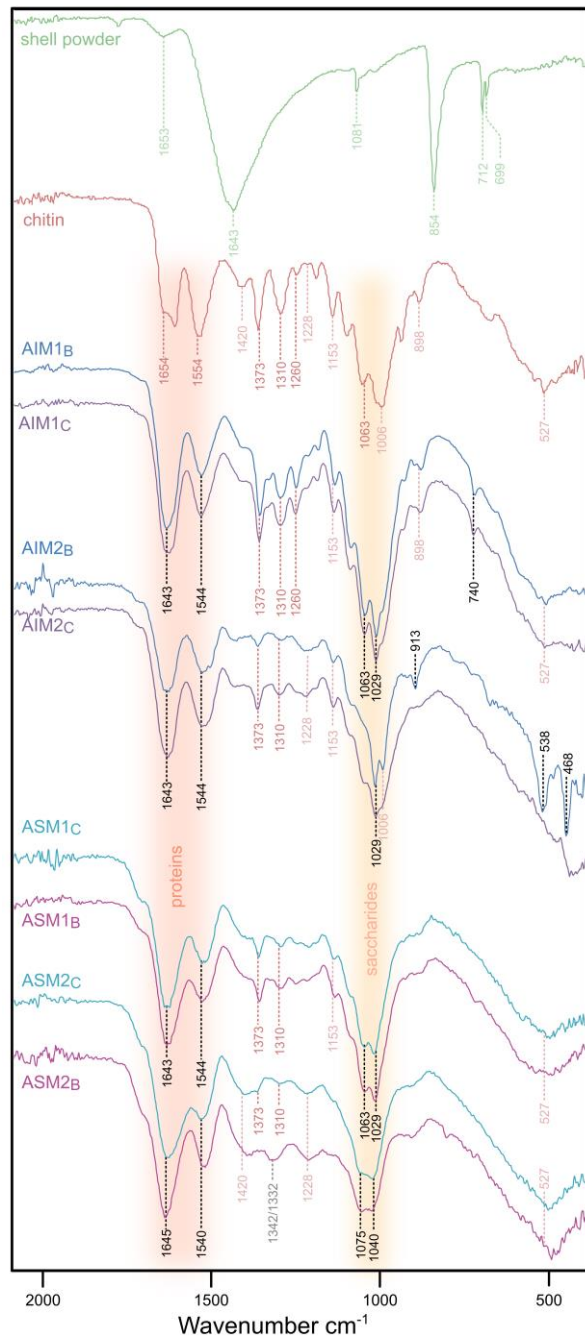


Figure 3: FT-IR spectra of the shell powder of *Spirula spirula*, of commercial α -chitin and of the different organic fraction of the shell matrix (AIM1, AIM2, ASM1, ASM2) from both localities (i.e. Canaries and Brazil). The “domains” of the characteristic absorption bands of protein and saccharidic 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.

367 fractions are also characterized by a high density of absorption bands between 1500 and 1000
368 cm^{-1} , very similar to those observed for commercial α -chitin. The most significant are located
369 at 1420, 1373, 1310 and 1260 cm^{-1} , and can be attributed to $\delta(\text{CH}_2)$, $\delta_s(\text{CH}_3)$, $\delta(\text{C-H})$ and amide
370 I bands, respectively. Moreover, absorptions of stretching vibrations of saccharidic moieties are
371 clearly revealed by the presence of intense absorptions at 1153, 1063 and 1006 cm^{-1}
372 corresponding to $\nu_s(\text{C-O-C})$, $\nu_{as}(\text{C-O-C})$, and $\nu(\text{C-O})$, respectively (Kaya et al., 2014). The
373 spectra from both origins – Canary Islands and Brazil – are very similar. It is interesting to note
374 that the AIM2_B spectrum exhibits additional bands, sharp and of medium intensities, between
375 1000 and 400 cm^{-1} (located at 913, 538, 468 cm^{-1}). To the best of our knowledge, and based on
376 the existing bibliography, these bands suggest the presence of phosphate groups (PO_4^{3-}) (Radev
377 et al., 2012). Thus, the band at 913 cm^{-1} can be assigned to the stretching P–O vibrations and
378 those at 538 and 468 cm^{-1} can be related to the bending O–P–O vibrations of phosphate groups
379 (Jastrzbski et al., 2011). At this stage of investigations, the reason of this variance from other
380 samples remains still undetermined. However, in recent studies on other shells of marine
381 animals, such as *Pinctada margaritifera* (Schmitt et al., 2018) and *Porites australiensis*
382 (Takeuchi et al., 2018), we already made the same spectroscopic observations on AIM fractions
383 leading to comparable suppositions about the presence of PO_4^{3-} groups.

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

389

390 *Solid State Nuclear Magnetic Resonance*

391 The ^{13}C CP MAS NMR spectra of the two *Spirula spirula* extracts (AIM1) and of α -
392 chitin standard are shown in Figure 4. The spectra from both origins – Canary Islands and Brazil
393 – are closely identical, and are dominated by the characteristic chitin peaks. The latter are
394 superimposed on a much weaker set of partially resolved peaks between 20 and 55 ppm, which
395 are consistent with small residual proteinaceous content. The merged C_3 and C_5 peaks, the
396 broadened methyl peak, and the chemical shifts of C_1 and C_4 peaks (103.9 and 83.9 ppm) in the
397 ^{13}C CP MAS spectra of the *S. spirula* samples are consistent with β -chitin (Jang et al., 2004;
398 Kaya et al., 2017).

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

419 Both *S. spirula* AIM1 extracts show a
 420 broad -0.2 ppm phosphate peak in their ^{31}P CP
 421 MAS spectra (Figure 5), validating the above IR
 422 findings for the AIM2 samples. $^{13}\text{C}\{^{31}\text{P}\}$ CP
 423 MAS REDOR measurements applied to AIM1c
 424 showed no REDOR effect (not shown) therefore
 425 ruling out phosphorylation of any of the chitin
 426 carbons. In the absence of such measurable
 427 proximity and in view of the resilience of these P-
 428 species under the extraction procedures (AIM1
 429 and the more aggressive of AIM2), we conclude
 430 that the observed P-species are inseparable from
 431 the proteinaceous content. We have previously reported similar ^{31}P peak for decalcified

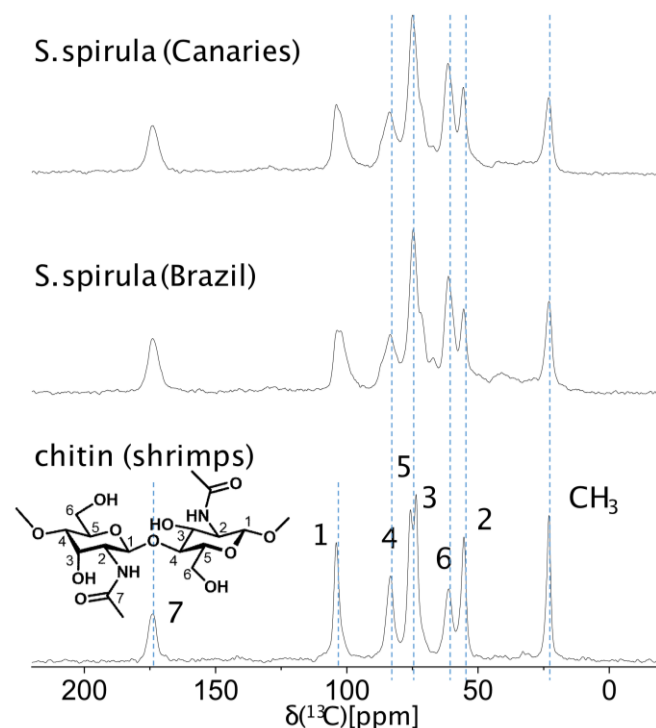


Figure 4: 75.4 MHz ^{13}C CPMAS spectra of acid insoluble extracts (AIM1) of *Spirula spirula* shells from the Canaries and Brazil, and of α -chitin standard (shrimp) showing that the organic content of AIM1 consists primarily of chitin (the dashed lines denote the characteristic peaks of chitin).

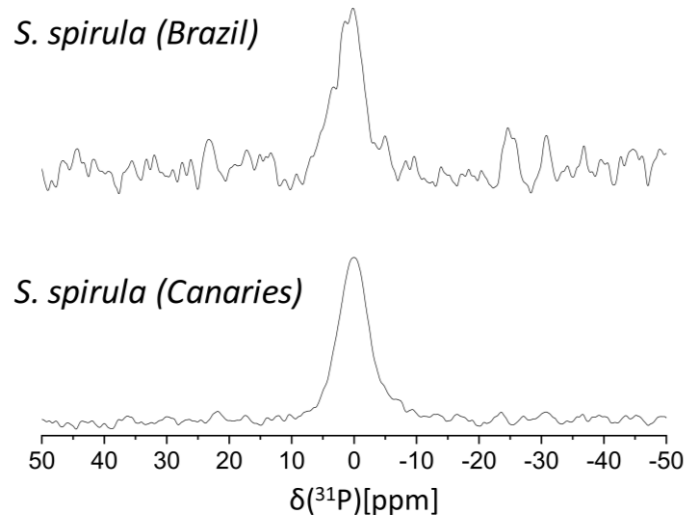


Figure 5: 121.85 MHz ^{31}P CP MAS NMR spectra of the acid insoluble extract (AIM1) of *Spirula spirula* shells from the Canaries and Brazil. The peaks are centered at -0.2 ppm and are of similar normalized intensities (accounting for sample weights of 16.8 and 6.0 mg, respectively). Each spectrum was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

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

434 *SDS-PAGE of the shell matrix extracts*

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

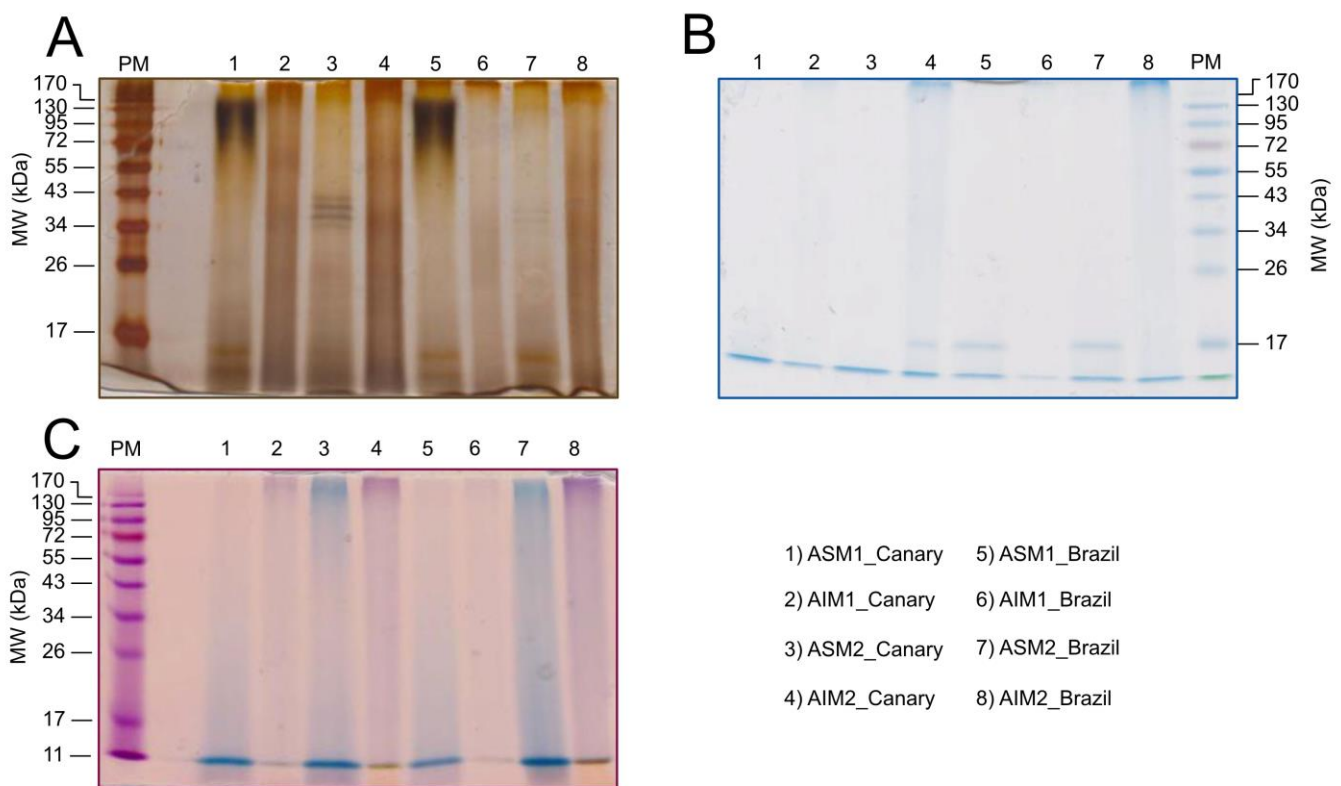


Figure 6: SDS-Page gel electrophoresis on *Spirula spirula*. A) Silver-staining; B) Bio-rad blue staining; C) Stains-all. PM: standard molecular weight (Fermentas).

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

457 *In vitro crystallization assay*

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

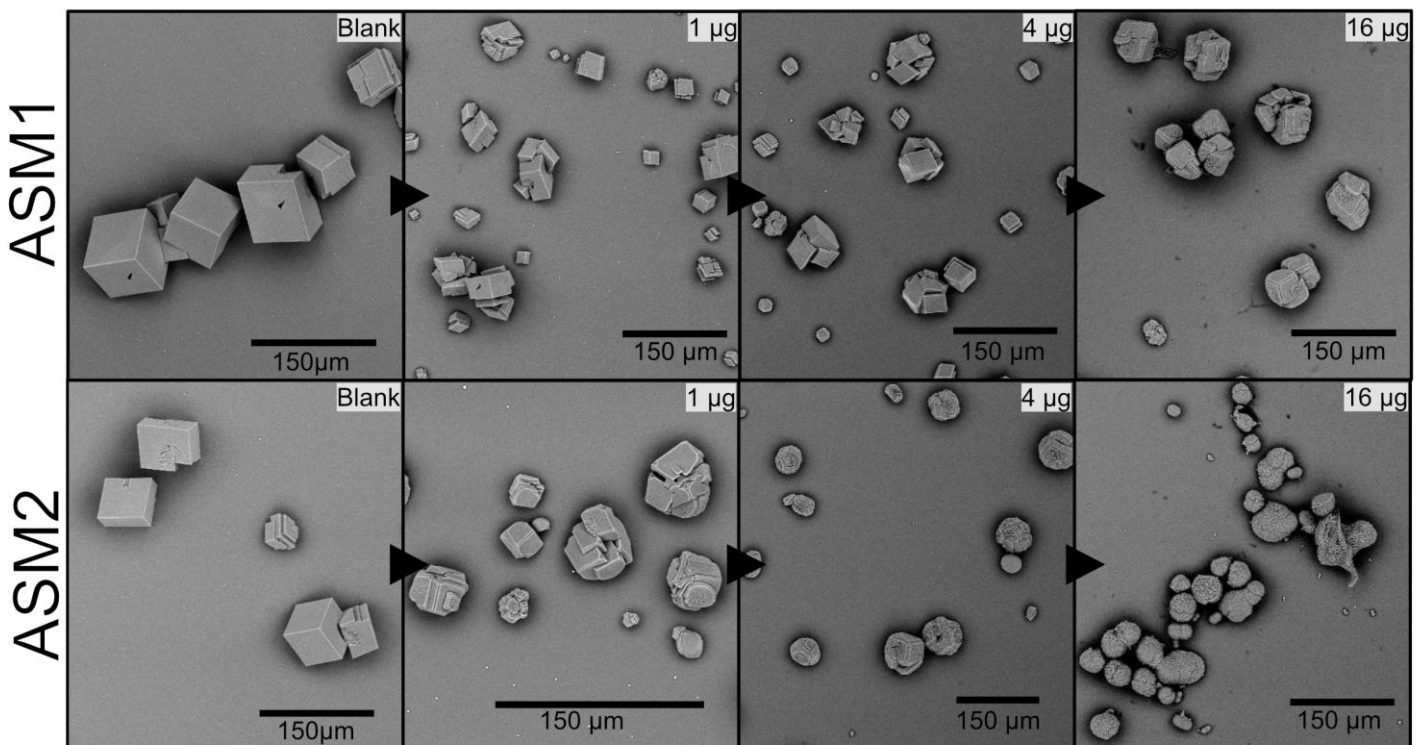


Figure 7: *In-vitro* crystallization assay performed on both soluble matrix extracts of *Spirula spirula*, ASM1 (top) and ASM2 (bottom). The black arrows indicate the increasing matrix concentration within each well, from 0 (blank) to 16 $\mu\text{g}/\text{well}$ (200 μL per well).

463 progressively visible, with the formation of grooves on the polycrystalline aggregates edges. At
 464 16 µg/well, the effect is more pronounced with the formation of rounded angles. No signs of
 465 crystallization inhibition - characterized by a drastic decrease of the size of polycrystalline
 466 aggregates - are recorded with ASM1.

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

471

472 *Enzyme-Linked Lectin Assay (ELLA)*

473 All screening experiments
 474 performed with a large set of commercial
 475 lectins are summarized in Table 2. We
 476 analyzed only ASMs: the insolubility of
 477 AIMs in standard buffers precluded the
 478 use of ELLA. Firstly, we do not detect
 479 significant variations of lectin reactivity
 480 between Canaries Island extracts and that
 481 of Brazil: the few low amplitude
 482 variations concern a couple of lectins that
 483 either do not react or react very weakly
 484 (10-20%), such as LCA or GSL-1. To
 485 summarize, we obtain three classes of
 486 reactivity: high, intermediate and low. In
 487 the first class, one finds LEL, WGA, STL
 488 (only for ASMs1) and Jacalin. The three
 489 first lectins bind oligomers (dimers and
 490 trimers in the case of WGA) of N-
 491 acetylglucosamine and are usually considered as chitin-binding lectins. Jacalin binds
 492 preferentially D-galactose or oligomers that terminate with galactose residues, as well as N-
 493 acetylgalactosamine-D-galactose dimers that very often correspond to the first glycosylation
 494 steps of Ser/Thr residues in O-glycosylated proteins. The second category of reactivity
 495 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%.

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

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

510 *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)

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

AIM1 (13959)	AGFAGDDAPR ; GYSFTTAAER ; EISALAPATMK ; QEYDESGPSIVHR ; SYELPDGQVITIGNER ; VAPEEHPVLLTEAPLNPK ; DLYANTVLSGGSTMFPGIADR ; TTGIVMDSGDGVTHTVPIYEGYALPHAILR ; EITALAPSTMK ; VAIVITDGR ; DYAIIVDER ; STELLIR ; ISGLIYEETR ; TITLEVEPSDTIENVK ; DEVLGGILR ; LILPGELAK ; GDGTGGLSIYGDR ; IDVADLLR ; SPNQNVQAAAAGALR ; VAVDTSYR ; SPIDLVR ; IDLLIPR ; ITIDLFEK ; IVMSQGTHLYFDHPYEPDPEER ; VSIVITDGR ; LQSDIKPQMKR ; TPTDVPVNLSCGR ; QLVQDSLDSL ; ANILSILEK ; KLVLLNYER ; YSPVDPIR ; QSIESVSPR ; LGIYEDFGTK ; IVIGLFGK ; IAFLVLK ; DSTLIMQLLR ; ELQDLNER ; LFDQAFGLPR ; DALEELDADR ; KILSIVER ; EGIEAIEDDR ; LDILPLK ; DLVDELADAIR ; QLSGQQYR ; LDTELNTYR ; VLSIILEK ; QLLEEIANLQK ; NIIDLNTEMK ; LMASFER ; LLLTAAISADK ; KINPPYYR ; ATSQTLDDENDITFR ; ATGLLPQPK ; VATVPLSR ; TGVSEGEDTKR ; TTGIVMDSGDGVTHTVPIYEGYALPHAILR ; SPLDLVR ; VVMTQASHLYFDHPYEPDPEER ; ASLVLDTGR ; LDLLPLR ; IGGIGTVPVGR ; IADADFR ; IASYLEK ; LGTALPFVR ; DYLGIVDAR ; EGELSDLSR ; KLLSDILR ; LDILPIK ; AQIEIIR ; LGDVEILR ; IGSLSLPR ; LLLPGELAK ; RIAEDSLYFR ; ASCSERCVAR ; VATVPSLR ; LAVVITDGR ; EIIDLVLDR ; SFFVAR ; GDFQLPALR ; QEGAKDLLLL ; ATNLSNFR ; FASFLDK
ASM1 (11712)	AGFAGDDAPR ; GYSFTTAAER ; SYELPDGQVITIGNER ; VAPEEHPVLLTEAPLNPK ; VAIVITDGR ; VATVSLPR ; DYAIIVDER ; LVALFDLR ; FQNALLVR ; EDQTEYLEER ; QSIESVSPR ; IVIGLFGK ; LPVLETER ; FDCLLSR ; KLVLLNYER ; QLSGQQYR ; LLLTAAISADK ; IAFLVLK ; EPSPQSPQLHR ; VAVDTSYR
AIM2 (13977)	AGFAGDDAPR ; GYSFTTAAER ; EITALAPSTMK ; QEYDESGPSIVHR ; SYELPDGQVITIGNER ; VAPEEHPVLLTEAPLNPK ; DLYANTVLSGGSTMFPGIADR ; EISALAPATMK ; TTGIVMDSGDGVTHTVPIYEGYALPHAILR ; AVFSPVGR ; DYAIIVDER ; NFDGLDLWEYPANR ; VAIVITDGR ; IVMSQGTHLYFDHPYEPDPEER ; IALVITDGR ; GDGTGGLSIYGDR ; IDVADLLR ; QLFQDVVK ; IGFPWSEIR ; APDFVYAPR ; FGGTLADCI ; STELLIR ; LILPGELAK ; ISGLIYEETR ; EGIEAIEDDR ; IAIVITDGR ; IIIVDSLGR ; GWLEPLDR ; IDLLIPR ; ITQSNAILR ; VSIVITDGR ; LVALFDLR ; TITLEVEPSDTIENVK ; SETLFLR ; DSTLIMQLLR ; QFGSSIAR ; IPVLLNYMR ; ASDGVPQVIVVLTDR ; IATFVLR ; QSIESVSPR ; LDDVHLR ; YSPVDPIR ; IAFLVLK ; IVIGLFGK ; EPSPQSPQLHR ; VAVDTSYR ; LSSPTALSDR ; HALSDAAVR ; VESVASLKER ; VPGVVTFR ; TPTDVPVNLSCGR ; DLVDELADAIR ; SPIDLVR ; LIGWGQIR ; ILITGGAGFVGSMLVDR ; ;
ASM2 (13218)	FESNFNTQATNR ; NTDGSTDYGILQINSR ; DYAIIVDER ; LLLPGELAK ; VFLENVIR ; ISGLIYEETR ; VAIVITDGR ; STELLIR ; LILPGELAK ; AGFAGDDAPR ; GYSFTTAAER ; EITALAPSTMK ; QEYDESGPSIVHR ; VAPEEHPVLLTEAPLNPK ; VSIVITDGR ; LPLQDVYK ; YVVTIIDAPGHR ; TITLEVEPSDTIENVK ; EGIEAIEDDR ; EDQTEYLEER ; EISALAPATMK ; LTAFLVR ; TGVSVSSLR ; YSPVDPIR ; ILVPQLDR ; EPSPQSPQLHR ; LSSPTALSDR ; IAFLVLK ; INHDLR ; GIVHTDLR ; ASDGVPQVIVVLTDR ; QAAIVNGTFR ; IHSDCANQQVYR ; IVIGLFGK ; VAVDTSYR ; KLVLLNYER ; TANLETSLR ; TLINVEDTR ; 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.

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

533 All the peptides identified per extract are listed in Table 4. The shortest ones comprise
 534 7 residues while the longest, 30. One can notice some variation on the identified number of
 535 peptides, depending on the bleaching. We observe that there is a slight decrease of the number
 536 of peptides from AIM1 to AIM2, and a significant increase from ASM1 to ASM2.

537 All the hits identified in 2, 3 or 4 extracts are summarized in Table 5. Of particular
 538 relevance are the hits with specific enzymes that are putatively involved in biomineralization,
 539 such as "putative tyrosinase-like protein Tyr-3", a "probable chitinase 2", and a "serine protease
 540 K12H4.7 precursor". Other candidates of interest are proteins that exhibit calcium-binding
 541 properties (CBR-MUP-4 protein, transmembrane cell adhesion receptor MUA-3 precursor, and
 542 a hypothetical protein HELRODRAFT_161577) and an immunoglobulin protein (IgGfC-
 543 binding protein). Five identified proteins are thus found into the four matrices extracts: CBR-
 544 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	<i>Caenorhabditis briggsae</i> - nematode				
gi 405964315	Putative tyrosinase-like protein tyr-3	<i>Crassostrea gigas</i> - pacific oyster				
gi 405970334	IgGfC-binding protein	<i>Crassostrea gigas</i> - pacific oyster				
gi 268570024	C. briggsae CBR-CYN-5 protein	<i>Caenorhabditis briggsae</i> - nematode				
gi 158297124	AGAP008053-PA, partial	<i>Anopheles gambiae</i> str. PEST - african malaria mosquito				
gi 284999807	beta actin	<i>Idiosepius paradoxus</i> - nothern pygmy squid				
gi 323230	polyprotein [Bovine viral diarrhea virus 1-Osloss]					
gi 510849995	core histone H2A/H2B/H3/H4	<i>Ancylostoma ceylanicum</i> - nematode				
gi 156359527	predicted protein	<i>Nematostella vectensis</i> - starlet sea anemone				
gi 170578661	Transmembrane cell adhesion receptor mua-3 precursor, partial	<i>Brugia malayi</i> - filarial nematode worm				
gi 675868340	hypothetical protein HELRODRAFT_161577	<i>Helobdella robusta</i> - californian leech				
gi 158300608	AGAP012043-PA, partial	<i>Anopheles gambiae</i> str. PEST - african malaria mosquito				
gi 768189576	hypothetical protein DICVIV_07943, partial	<i>Dictyocaulus viviparus</i> - bovine lungworm				
gi 669305641	hypothetical protein M513_09906	<i>Trichuris suis</i> - pig whipworm				
gi 163637123	ribosomal protein S3, partial	<i>Crassostrea gigas</i> - pacific oyster				
gi 223016075	actin	<i>Octopus vulgaris</i> - common octopus				
gi 524896874	PREDICTED: peptidyl-prolyl cis-trans isomerase B-like	<i>Aplysia californica</i> - california sea hare				
gi 170586872	hypothetical protein Bm1_33780	<i>Brugia malayi</i> - filarial nematode worm				
gi 33321997	gamma-glutamyl phosphate reductase, partial	<i>Lactobacillus delbrueckii</i> subsp. <i>Lactis</i> - bacteria				
gi 676489609	hypothetical protein LOTGIDRAFT_236342	<i>Lottia gigantea</i> - owl limpet				
gi 748378928	homeobox domain protein, partial	<i>Ancylostoma duodenale</i> - nematode				
gi 1360640	14-3-3 zeta protein	<i>Xenopus laevis</i> - African clawed frog				
gi 156544592	PREDICTED: alpha-L-fucosidase-like	<i>Nasonia vitripennis</i> - jewel wasp				
gi 193591718	PREDICTED: signal peptide, CUB and EGF-like domain-containing protein 1	<i>Acyrtosiphon pisum</i> - pea aphid				
gi 542607731	carbomoylphosphate synthase, partial	<i>Culicoides immaculatus</i> - diptera				
gi 307203534	Probable chitinase 2	<i>Harpegnathos saltator</i> - Jerdon's jumping ant				
gi 358337009	monocarboxylate transporter 14, partial	<i>Clonorchis sinensis</i> - chinese liver fluke				
gi 585686849	PREDICTED: collagen alpha-3(VI) chain isoform X1	<i>Elephantulus edwardii</i> - cape elephant shrew				
gi 225718928	serine protease K12H4.7 precursor	<i>Caligus clemensi</i> - sea louse				
gi 8272602	82 kDa heat shock protein 3, partial	<i>Philodina roseola</i> - 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

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

552 The results of the search against the *S. spirula* transcriptome brought additional results
553 that are summarized in Table 6. In total, we obtained 39 hits: 27 with AIM1, 31 with AIM2, 10
554 with ASM1 and 12 with ASM2. Among them, 22 are shared: 6 between the 4 extracts, 7
555 between 3 of them and 9 between 2. In the shared hits, the percentage of coverage varies from
556 almost 60 to 1.4%. Furthermore, 17 hits are unique to one of the four extracts: 6 to AIM1, 9 to
557 AIM2, 1 to ASM1 and 1 to ASM2. It is interesting to note that the two AIMs generate more hits
558 than the ASMs, and the second extracts (AIM2, ASM2), more than the first ones (AIM1,
559 ASM1). Among the 6 hits shared by the four extracts, 4 of them are invariably the ones with
560 the highest peptidic coverage. They include a true transferrin, a serine protease inhibitor
561 (elastase-inhibitor like), a matrilin, and actin. The two other "shared by four" hits are an
562 uncharacterized protein with a heme-binding domain and a peptidyl prolyl cis trans isomerase.
563 Among the 7 hits shared by 3 extracts, four of them exhibit a peptidic coverage higher than
564 10% of their sequence. They include α -2 macroglobulin, collagen (N-terminus), a
565 protein/nucleic acid deglycase and a histone. The three other proteins are one ubiquitin and two
566 histones. The proteins shared by two extracts are represented by a calcium-binding protein, a
567 peptidyl-prolyl cis-trans isomerase, a chitinase/sugar-binding protein, a GLIPR1 protein, a
568 transferrin fragment, a collagen N-terminus, and a lamin. The other 17 proteins found in one
569 extract include, among others, a Cys-rich secretory protein, a galectin3 binding-protein, a
570 calcium-binding protein and four enzymes involved in saccharidic moieties modifications
571 (heparan sulfate glucosamine sulfotransferase, N-acetyl-galactosaminidase, chitin deacetylase,
572 UDP-glucuronic acid decarboxylase)

573

Protein hit * indicates complete sequences	Shell organic matrix extract				Functional domains and motifs (starting AA - final AA, domain)	Putative function in biomineralization
	AIM1	AIM2	ASM1	ASM2		
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 <i>Octopus</i>)	% 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 <i>Octopus</i>
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, pfam00333) 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, pfam00333) 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-αχετυλγλεξοσαμινιδάση*	% cover. AA / MW /	X 2.6 860 / 98.40 / 5.38	X 2.6	Carbohydrate-binding (13-160, pfam03173) Glycosylhydrolase 20 (198-322, pfam00021)
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, pfam00021)
Uncharacterized / Cys-rich secretory protein	% cover. AA / MW /	X 10.3 106 / 12.28 / 8.06		Cap: Cys-rich secretory protein (42-103, cl00133)
α-1-Τυβουλιν (N term)	% cover. AA / MW /	X 5.9 152 / 16.60 / 5.46		Tubulin α-χηαιν (whole seq., cl30502)
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 (82-225, pfam00112)
Heparan sulfate glucosamine sulfotransferase	% cover. AA / MW /	X 2.7 261 / 30.23 / 9.74		Sulfotransferase (43-260, pfam00685)
Alpha-N-acetyl-galactosaminidase-like*	% cover. AA / MW /	X 2.6 389 (411) / 43.97 / 5.40		α-γαλαχτοσιδάση A (30-309, pfam16499) α-Γαλαχτοσιδάση A (312-396, pfam16499)
Elongation factor 1-a*	% cover. AA / MW /	X 2.4 463 / 51.11 / 9.03		Elongation factor 1-a (1-454, PTZ00141)
Methylmalonic aciduria & homocystinuria type C	% cover. AA / MW /	X 12.8 242 / 28.15 / 9.43		MMACH type C (7-225, pfam16690)
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, pfam00021)
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
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)
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)
Moesin / Vilin 2 / Radixin *	% cover. AA / MW /	X 4.6 583 / 68.82 / 5.76		FERM (203-299, cd13194) ERM (7-209, smart00295)
UDP-glucuronic acid decarboxylase 1	% cover. AA / MW /	X 4.1 417 / 47.79 / 9.42		DUF3573 (24-91, cl13524) UGD (95-398, cd05230)
Arginine kinase*	% cover. AA / MW /	X 2.4 409 / 46.24 / 8.57		Arginine kinase (68-409, cd07932)
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)
Histone H2A*	% cover. AA / MW /	X 7.0 128 / 13.42 / 10.58		Histone H2A (6-121, cl30550)
Elongation factor 1-a (fragment)	% cover. AA / MW /	X 7.5 107 / 12.09 / 7.99		Elongation factor 1-a (whole seq., cl33168)

574

575

Discussion

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

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

587 We also observe the lamello-fibrillar microstructure of the septa, usually considered as
588 a nacreous layer (Barskov, 1973; Bøggild, 1930; Dauphin, 1976; Mutvei, 1964).
589 Macroscopically, the septa exhibit indeed iridescence properties. However, we feel that the
590 "nacreous" terminology is misleading and we follow the assertion of Erben (1972) who claimed
591 that the nacre of *Spirula* and other coleoids is rather a 'false nacre' and recommended the use of
592 "lamello-fibrillar" terminology. This expression is entirely justified in view of a very recently
593 published work (Lemanis *et al.*, 2020), which precisely visualizes the substructures of the
594 lamello-fibrillar layer, and describes it as the stacking of "thin lamellae that are themselves
595 composed of aragonitic rods that tend to be parallel within a lamella." The lamello-fibrillar
596 terminology was also used by Fuchs *et al.* (2012) to describe the septal microstructure of
597 ancestral spirulid coleoids. Two patterns are of particular importance: the insertion of the septa
598 in the shell walls, and the presence of what resembles cameral deposit at the back of the
599 insertion zone. The first one, although mentioned in earlier works (Bandel and Boletzky, 1979;
600 Dauphin, 1976; Hoffmann *et al.*, 2018), was neither fully described, nor schematized (Bandel
601 and Boletzky, 1979). We confirm that the septa are sandwiched as a bevel in the internal wall.
602 Such structures were also observed among the fossil relatives of *Spirula* (Doguzhaeva, 1996)
603 and defined as "mural flap" (Fuchs *et al.*, 2013): this character state may be an apomorphy of
604 the Spirulida order. Given the thinness of the *S. spirula* shell, one can assume that a "sandwich"
605 insertion considerably reinforces the overall shell mechanical properties and was consequently
606 selected by evolution. However, Lemanis *et al.* (2020) questioned such hypotheses,
607 demonstrating that mural flaps have "no notable structural function". The second pattern -
608 deposits in the dorsal septal corner and mentioned by Appelöf (1893), Bøggild (1930) and
609 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

644 ~80% of the insoluble fraction. This percentage is higher than that indicated in earlier works,
645 which found a protein/chitin ratio around two (Degens et al., 1967; Hunt and Nixon, 1981).
646 However, this latter value is certainly overestimated (in favor of proteins) because these authors
647 employed harsh hydrolytic conditions that have likely degraded a part of the saccharide
648 moieties. In our case, the abundance of chitin identified by SS-NMR confirms earlier findings
649 based on monosaccharide analysis, which showed that both ASM and AIM are composed
650 predominantly of the deacetylated form - due to hydrolytic conditions - of N-acetyl-
651 glucosamine, the monomer of chitin (Dauphin and Marin, 1995). Such a composition where
652 sugars predominate is rather unusual for matrices associated with mollusc shells: most of the
653 works published so far on this type of materials tend to show that the saccharidic moieties are
654 often much less abundant than the protein ones (Agbaje et al., 2018; Immel et al., 2016).
655 Similarities of ASMs profiles with chitin, seen by FTIR, are also in agreement with the ELLA
656 test, since the main chitin-binding lectins such as WGA, STL and LEL give the highest
657 reactivity. The presence of chitin in ASMs may appear odd as this polymer is supposed to be
658 particularly insoluble. This suggests different potential explanations that do not mutually
659 exclude each other: 1. Soluble non-crossed-linked chitin may exist *per se* in the ASM, which
660 would be rather unusual. 2. N-acetylglucosamine oligomers (that "mimic" chitin and are
661 consequently targeted by chitin-binding lectins) are bound to soluble protein core. 3. From our
662 discussion above, the chitin framework occupies an intercrystalline position in the biominerals;
663 consequently, this framework is partly cleaved chemically in a random manner by sodium
664 hypochlorite during the bleaching treatment, and then, released in solution during
665 decalcification. Then, the cleaved fragments remain large enough not to be eliminated by
666 subsequent ultrafiltration. It is interesting to notice that if chitin represents about 80% of the
667 AIM, and that AIM itself accounts for about 70% of the whole matrix, this means that chitin
668 represents between 50 and 60% of the matrix: in other words, this means that the matrix is more
669 chitinous than proteinaceous. This character brings it close to the cuttlefish matrix (Stegemann,
670 1963), and consequently, to the Sepiidae.

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

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

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

693 Other technical biases, not discussed here, may hamper the interpretation of the
694 proteomic results (see Marin et al., 2016 for more details). Interestingly, the number of
695 identified peptides significantly increased between ASM1 and ASM2. Similarly to what likely
696 happens to chitin during the second bleaching treatment, we suggest that insoluble proteins are
697 randomly cleaved by sodium hypochlorite and consequently, partly solubilized, becoming
698 components of the ASM2 fraction. In total, the overall peptides lead to the identification of 91
699 protein hits, which constitutes a low score, given the number of peptides. In any event, the
700 peptides recognized several interesting hits, namely specific enzymes, such as tyrosinase,
701 chitinase and serine protease, which constitute a part of the molecular toolkit frequently found
702 in shell matrices (Marie et al., 2012; Marin et al., 2014). Tyrosinase is thought to be
703 biomineralization-linked either by being involved in the melanin biosynthesis or in the
704 periostracum formation (Arivalagan et al., 2016; Nagai et al., 2007; Zhang et al., 2006), while
705 chitinase and serine protease are associated to organic matrix remodelling (Marin et al., 2016;
706 Yonezawa et al., 2016). We also identified a set of sequences that can be related to non-
707 molluscan calcium-binding proteins, with either EF-hand, or EGF-like domains, both belonging
708 to the "high-affinity - low capacity" category. Among the unexpected hits, one finds actin, the
709 main cytoskeletal protein, and histone, a nuclear protein. The actin identified here is identical
710 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

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

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

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

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

779 development in murine model (Pandya et al., 2017). Interestingly, in the shell matrix of *S.*
780 *spirula*, beside the full-length member, we detected a shortened version of peptidyl prolyl cis-
781 trans isomerase B, with a reasonably good 17% peptidic coverage, and corresponding to 60%
782 of the C-terminus of the first sequence. This fragment is not identical but a variant that exhibits
783 75% identity (and 89% similarity) with the complete form. Curiously, while the latter is present
784 in the four extracts, the short variant has been detected only in the two AIMs, which suggests
785 that it is cross-linked with other matrix components.

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

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

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

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

826 To summarize, the shell matrix of *Spirula spirula*, as seen by the MASCOT analysis of
827 the protein-translated transcriptome of *Spirula spirula*, contains numerous molecular actors,
828 including a minimum of 11 true enzymes and 28 proteins of very different functions. All these
829 39 members can be categorized as follows: extracellular matrix proteins (matrilin, collagen-
830 like) together with ECM-binding partners and ECM-interacting enzymes (like cis-trans
831 isomerase) required for ECM protein folding and activation; a set of enzymes that regulate,
832 modify and remodel the saccharidic moieties (chitin deacetylase, N-acetylgalactosaminidase...)
833 together with saccharide-binding proteins that are not enzymes; a set of cytoskeletal proteins
834 (actin, lamin, tubulin) that function in extracellular environment together with their binding
835 partners (elongation factor, moesin). The system includes also proteins involved in cell
836 signaling (14-3-3 protein zeta, ubiquitin, protein with MMACH domain...) and calcium-binding
837 proteins (hippocalcin-like, calmodulin-like). The system possesses also its protective functions
838 at molecular and cellular levels, respectively: at molecular level, these are protease inhibitors
839 on the one hand, and putative bactericidal factors (histones?), immunity-related proteins and
840 possibly proteins that bind iron, on the other hand. Finally, the system contains a set of proteins
841 of totally unknown functions, as exemplified by the unhealthy ribosome biogenesis protein 2
842 and GLIPR1 protein 1.

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

847 *spirula*, we identified three CA sequences, which possess tryptic cleavage sites (data not shown)
848 of which no peptide is present in the matrix proteomic results. This is a strong indication that
849 CA is not incorporated as a shell matrix protein in the Ram's Horn model. This conclusion
850 confirms earlier finding on the shell matrix of the abalone *Haliotis tuberculata* (Le Roy et al.,
851 2014) and shows that CA, depending on the mollusc model is or is not a shell matrix protein.
852 The other absent players of the list are proteins that contain low complexity domains or
853 repetitive low complexity domains. In calcium carbonate biomineralization, these domains can
854 be of several types, the most known being the acidic, aspartic acid-rich ones, and the
855 hydrophobic "alanine/glycine-rich" domains (Marin et al., 2016). We admit that their
856 underrepresentation in the protein list may be due to technical bias, such domains having few
857 or no trypsin cleavage sites. Clearly, the absence of peptides exhibiting low complexity
858 sequences does not mean that the corresponding proteins are absent from the shell forming-
859 machinery: a separate analysis of the proteome-translated transcriptome (not shown) reveals
860 putative candidates, which - for an unknown reason - are not incorporated to the matrix.

861 From an evolutionary viewpoint, it is puzzling that most of the peptides obtained do not
862 exhibit any similarity with already known molluscan shell proteins. Among the few hits with
863 molluscan proteins, the majority belongs to a bivalve, the Pacific oyster *Crassostrea gigas* (5
864 hits in total), while there are only two hits with a cephalopod, the common octopus *Octopus*
865 *vulgaris*. A comparison of our peptide list with already published data on *Sepia* (Čadež et al.,
866 2017; Le Pabic et al., 2017) and *Nautilus* (Marie et al., 2011, 2009) did not reveal any similarity.
867 When compared with the complete transcriptome of *Octopus bimaculoides* (Albertin et al.,
868 2015) our peptide dataset leads few hits, mainly with actin-like proteins (3 hits, exhibited by all
869 the extract of the organic matrix of *Spirula spirula*) and H4 histones (9 hits in total, only with
870 the ASM2 extract).

871 Such a paucity of hits with already studied cephalopods is intriguing but may be related
872 to the uncertainty about the phylogenetic position of *Spirula* and to its unique evolutionary
873 history, *i.e.*, its relative evolutionary remoteness from other members of this mollusc class:
874 indeed, within Decabrachia, the placement of the order Spirulida has historically been
875 contentious (reviewed within Lindgren and Anderson, 2018). On the one hand, early
876 phylogenetic studies based on a small number of genes proposed that *Spirula* occupies either a
877 basal position relative to other Decabrachia (Allcock et al., 2011; Warnke et al., 2003) or that a
878 sister taxa relationship exists between the two families Spirulidae and Sepiidae (Strugnell et al.,
879 2005). These views are congruent with morphological characters such as a well-developed shell
880 with chambers and septa. On the other hand, recent studies, including those based on more

881 comprehensive molecular datasets, have recovered a sister taxon relationship between Spirulida
882 and a clade containing Oegopsida ('open eyed' squids) and Bathyteuthoidea (mesopelagic to
883 bathopelagic squids) (Lindgren et al., 2012; Strugnell et al., 2017). Similarly, Tanner et al.
884 (2017) reported a sister taxon relationship between spirulid and oegopsid squids. Although
885 these authors did not include bathyteuthoid squids in their analyses, a close
886 phylogenetic relationship between bathyteuthoids and oegopsids is well established (e.g.
887 Lindgren et al., 2012; Strugnell et al., 2005). Recent divergence time estimates, incorporating
888 molecular and fossil data, suggest that spirulids and oegopsids diverged around 128 Ma (Tanner
889 et al., 2017), which accords with appearance of stem group spirulids in the latest Cretaceous
890 (~66-72 Ma) (Fuchs et al., 2012).

891 The absence of any similarity with Nautilus is not surprising, considering the deep
892 divergence time between the subclasses Nautiloidea (Nautilus) and Coleoidea (octopus,
893 cuttlefishes and squids) ($\sim 416 \pm 60$ Ma, Kröger et al., 2011). In addition, our study suggests
894 that *Spirula* and *Sepia* have completely different molecular tools to construct their shell. This
895 result is also not surprising. Sepiids and spirulids both possess an internal calcareous shell with
896 a phragmocone and, as a result, they have extensive fossil records in comparison to other extant
897 decabrachian lineages that have lost their shell (i.e., oegopsids, myopsids). The phragmocone
898 is a plesiomorphic character within cephalopods and, therefore, the fact that it is present within
899 sepiids and spirulids does not indicate that these lineages are closely related. In support of this,
900 several recent phylogenetic studies have recovered a sister taxon relationship between sepiids
901 and all remaining decabrachians (including spirulids) (Strugnell et al., 2017; Tanner et al.,
902 2017; Uribe and Zardoya, 2017) with divergence of these two clades estimated to have occurred
903 in the mid Jurassic (Tanner et al., 2017) or in the early Cretaceous (Uribe and Zardoya, 2017).
904 This lapse of time is long enough to allow the independent evolution of the organic matrix
905 proteins: recent findings suggest indeed that several skeletal matrix proteins exhibit a fast
906 evolution rate (Jackson et al., 2006; McDougall et al., 2013). However, a rapid evolution of
907 matrix components is not incompatible with the occurrence of very conservative characters,
908 such as the abundance of chitin or the overall morphological similarities (septate inner shell,
909 for example). Given the modern phylogenetic view of coleoids, which brings Spirulida as a
910 sister group of the non-calcifying oegopsids, it would be relevant to determine how the
911 molecular functions of the proteins used for calcification within *Spirula* have evolved in this
912 latter clade.

913

914 Acknowledgements

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

925

926 Authors contribution

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

932

933

934

Legends of figures

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

938

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

950

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

957

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

962

963

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

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

968

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

971

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

976

977

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

981

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

986

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

992

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

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

1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027

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 column 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 its 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

- 1029 Accinni, L., Natali, P.G., Silvestrini, M., De Martino, C., 1983. Actin in the extracellular matrix
1030 of smooth muscle cells. An immunoelectron microscopic study. *Conn. Tiss. Res.* 11, 69-
1031 78.
- 1032 Agbaje, O.B.A., Ben Shir, I., Zax, D.B., Schmidt, A., Jacob, D.E., 2018. Biomacromolecules
1033 within bivalve shells: Is chitin abundant? *Acta Biomater.* 80, 176–187.
1034 <https://doi.org/10.1016/j.actbio.2018.09.009>
- 1035 Albeck, S., Aizenberg, J., Addadi, L., Weiner, S., 1993. Interactions of various skeletal
1036 intracrystalline components with calcite crystals. *J. Am. Chem. Soc.* 115, 11691–11697.
1037 <https://doi.org/10.1021/ja00078a005>
- 1038 Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E.,
1039 Brenner, S., Ragsdale, C.W., Rokhsar, D.S., 2015. The octopus genome and the evolution
1040 of cephalopod neural and morphological novelties. *Nature* 524, 220–224.
1041 <https://doi.org/10.1038/nature14668>
- 1042 Allcock, A.L., Cooke, I.R., Strugnell, J.M., 2011. What can the mitochondrial genome reveal
1043 about higher-level phylogeny of the molluscan class Cephalopoda? *Zool. J. Linn. Soc.*
1044 161, 573–586. <https://doi.org/10.1111/j.1096-3642.2010.00656.x>
- 1045 Appellöf, A., 1893. Die schalen von *Sepia*, *Spirula* und *Nautilus* studien über den Bau und das
1046 Wachstum. *K. Sven. VetenskAkad. Handl.*, Stock. 25, 1–106.
- 1047 Arivalagan, J., Yarra, T., Marie, B., Sleight, V.A., Duvernois-Berthet, E., Clark, M.S., Marie,
1048 A., Berland, S., 2016. Insights from the shell proteome: Biomineralization to adaptation.
1049 *Mol. Biol. Evol.* 34, 66–77. <https://doi.org/10.1093/molbev/msw219>
- 1050 Balmain, J., Hannoyer, B., Lopez, E., 1999. Fourier transform infrared spectroscopy (FTIR)
1051 and X-ray diffraction analyses and organic matrix during heating of mother of pearl (nacre)
1052 from the shell of the mollusk *Pinctada maxima*. *J. Biomed. Mater. Res.* 48, 749–754.
- 1053 Bandel, K., Boletzky, S., 1979. A comparative study of the structure, development and
1054 morphological relationships of chambered cephalopod shells. *Veliger* 21, 313–354.
- 1055 Barskov, S., 1973. Microstructure of the skeletal layers of *Sepia* and *Spirula* compared with the
1056 shell layers of other mollusks. *J. Paleontol.* 3, 285–294.
- 1057 Bennett, A.E., Rienstra, C.M., Auger, M., Lakshmi, K. V, Griffin, R.G., 1995. Heteronuclear
1058 decoupling in rotating solids. *J. Chem. Phys.* 103, 6951–6958.
1059 <https://doi.org/10.1063/1.470372>
- 1060 Bøggild, O.B., 1930. The shell structure of the mollusks. *K. Dan. Vidensk. Selsk. Skr.*
1061 *Naturvidensk. Math. Afd.* 9, 231–326.
- 1062 Bouchet, P., Bary, S., Héros, V., Marani, G., 2016. How many species of molluscs are there in
1063 the world's oceans, and who is going to describe them? *Mémoires du Muséum Natl.*
1064 *d'Histoire Nat.* 208, 9–24.

- 1065 Brauer, M., Sykes, B.D., 1984. Phosphorus-31 nuclear magnetic resonance studies of
1066 phosphorylated proteins. *Methods Enzymol.* 107, 36–81. [https://doi.org/10.1016/0076-](https://doi.org/10.1016/0076-6879(84)07005-1)
1067 [6879\(84\)07005-1](https://doi.org/10.1016/0076-6879(84)07005-1)
- 1068 Brayard, A., Escarguel, G., Bucher, H., Monnet, C., Brühwiler, T., Goudemand, N., Galfetti,
1069 T., Guex, J., 2009. Good genes and good luck: Ammonoid diversity and the end-permian
1070 mass extinction. *Science* 325, 1118–1121. <https://doi.org/10.1126/science.1174638>
- 1071 Brock, G., Paterson, J., 2004. A new species of *Tannuella* (Helcionellida, Mollusca) from the
1072 early Cambrian of South Australia. *Assoc. Australas. Palaeontol. Mem.* 30, 133–143.
- 1073 Bruun, A., 1943. The biology of *Spirula spirula* (L.). *Dana Rep.* 24, 1–46.
- 1074 Čadež, V., Škapin, S.D., Leonardi, A., Križaj, I., Kazazić, S., Salopek-Sondi, B., Sondi, I.,
1075 2017. Formation and morphogenesis of a cuttlebone's aragonite biomineral structures for
1076 the common cuttlefish (*Sepia officinalis*) on the nanoscale: Revisited. *J. Colloid Interface*
1077 *Sci.* 508, 95–104. <https://doi.org/10.1016/j.jcis.2017.08.028>
- 1078 Calvo-Iglesias, J., Pérez-Estévez, D., González-Fernández, Á., 2017. MSP22.8 is a protease
1079 inhibitor-like protein involved in shell mineralization in the edible mussel *Mytilus*
1080 *galloprovincialis*. *FEBS Open Bio* 7, 1539–1556. [https://doi.org/10.1002/2211-](https://doi.org/10.1002/2211-5463.12286)
1081 [5463.12286](https://doi.org/10.1002/2211-5463.12286)
- 1082 Campbell, K., MacLennan, D., Jorgensen, A.O., 1983. Staining of the Ca²⁺-binding proteins,
1083 calsequestrin, calmodulin, troponin C, and S-100, with the cationic carbocyanine dye
1084 “Stains-all.” *J. Biol. Chem.* 258, 11267–11273.
- 1085 Checa, A.G., 2018. Physical and biological determinants of the fabrication of molluscan shell
1086 microstructures. *Front. Mar. Sci.* 5, 353. <https://doi.org/10.3389/fmars.2018.00353>
- 1087 Clarke, M.R., 1970. Growth and development of *Spirula spirula*. *J. Mar. Biol. Assoc. United*
1088 *Kingdom* 50, 53–64. <https://doi.org/10.1017/S002531540000059X>
- 1089 Crenshaw, M.A., 1972. The soluble matrix from *Mercenaria mercenaria* shell.
1090 *Biomineralization* 6, 6–11.
- 1091 Cuif, J., Dauphin, Y., Denis, A., Gaspard, D., Keller, J., 1983. Etude des caractéristiques de la
1092 phase minérale dans les structures prismatiques du test de quelques mollusques. *Bull. du*
1093 *Muséum Natl. d'histoire Nat. Sect. A, Zool. Biol. Ecologie Anim.* 5, 679–717.
- 1094 Dauphin, Y., 1996. The organic matrix of coleoid cephalopod shells: molecular weights and
1095 isoelectric properties of the soluble matrix in relation to biomineralization processes. *Mar.*
1096 *Biol.* 125, 525–529. <https://doi.org/10.1007/BF00353265>
- 1097 Dauphin, Y., 1977. Microstructure et flottabilité chez la spirule (Cephalopoda). *Comptes rendus*
1098 *Hebd. des séances l'Académie des Sci. Paris D* 284, 2483–2485.
- 1099 Dauphin, Y., 1976. Microstructure des coquilles de céphalopodes. I. *Spirula spirula* L.
1100 (Dibranchiata, Decapoda). *Bull. du muséum natl. d'histoire Nat.* 3ème série, n° 382, *Sci.*
1101 *de la Terre* 54, 197–238.
- 1102 Dauphin, Y., Marin, F., 1995. The compositional analysis of recent cephalopod shell
1103 carbohydrates by Fourier transform infrared spectrometry and high performance anion

- 1104 exchange-pulsed amperometric detection. *Experientia* 51, 278–283.
1105 <https://doi.org/10.1007/BF01931112>
- 1106 Degens, E.T., Spencer, D.W., Parker, R.H., 1967. Paleobiochemistry of molluscan shell
1107 proteins. *Comp. Biochem. Physiol.*, 20, 553–579. doi:10.1016/0010-406x(67)90269-1
- 1108 Doguzhaeva, L., 1996. Two early Cretaceous spirulid coleoids of the north-western Caucasus:
1109 their shell ultrastructure and evolutionary implications. *Palaeontology* 39, 681–707.
- 1110 Erben, H.K., 1972. Über die Bildung und das Wachstum von Perlmutter. *Biomaterialization* 4,
1111 15–46.
- 1112 Feng, D., Li, Q., Yu, H., Kong, L., Du, S., 2017. Identification of conserved proteins from
1113 diverse shell matrix proteome in *Crassostrea gigas*: characterization of genetic bases
1114 regulating shell formation. *Sci. Rep.* 7, 45754. <https://doi.org/10.1038/srep45754>
- 1115 Fresquet, M., Jowitt, T.A., Stephen, L.A., Ylöstalo, J., Briggs, M.D., 2010. Structural and
1116 functional investigations of Matrilin-1 A-domains reveal insights into their role in cartilage
1117 ECM assembly. *J. Biol. Chem.* 285, 34048–34061.
1118 <https://doi.org/10.1074/jbc.M110.154443>
- 1119 Fuchs, D., Iba, Y., Ifrim, C., Nishimura, T., Kennedy, W.J., Keupp, H., Stinnesbeck, W.,
1120 Tanabe, K., 2013. Longibelus gen. nov., a new Cretaceous coleoid genus linking
1121 Belemnoidea and early Decabrachia. *Palaeontology* 56, 1081–1106.
1122 <https://doi.org/10.1111/pala.12036>
- 1123 Fuchs, D., Keupp, H., Trask, P., Tanabe, K., 2012. Taxonomy, morphology and phylogeny of
1124 Late Cretaceous spirulid coleoids (Cephalopoda) from Greenland and Canada.
1125 *Palaeontology* 55, 285–303. <https://doi.org/10.1111/j.1475-4983.2011.01125.x>
- 1126 Giansanti, F., Leboffe, L., Pitari, G., Ippoliti, R., Antonini, G., 2012. Physiological roles of
1127 ovotransferrin. *Biochim. Biophys. Acta* 1820, 218–225.
1128 <https://doi.org/10.1016/j.bbagen.2011.08.004>
- 1129 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X.,
1130 Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A.,
1131 Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N.,
1132 Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a
1133 reference genome. *Nat. Biotechnol.* 29, 644–652. <https://doi.org/10.1038/nbt.1883>
- 1134 Harada, N., Iijima, S., Kobayashi, K., Yoshida, T., Brown, W.R., Hibi, T., Oshima, A.,
1135 Morikawa, M., 1997. Human IgG Fc binding protein (FcγBP) in colonic epithelial cells
1136 exhibits mucin-like structure. *J. Biol. Chem.* 272, 15232–15241.
1137 <https://doi.org/10.1074/jbc.272.24.15232>
- 1138 Hare, P.E., 1963. Amino Acids in the Proteins from Aragonite and Calcite in the Shells of
1139 *Mytilus californianus*. *Science* 139, 216–217.
- 1140 Haring, E., Kruckenhauser, L., Lukeneder, A., 2012. New DNA sequence data on the enigmatic
1141 *Spirula spirula* (Linnaeus, 1758)(Decabrachia, suborder Spirulina). *Ann. Naturhist. Mus.*
1142 *Wien. Ser. B Bot. Zool.* 113, 37–48.

- 1143 Hoffmann, R., Lemanis, R.E., Wulff, L., Zachow, S., Lukeneder, A., Klug, C., Keupp, H., 2018.
1144 Traumatic events in the life of the deep-sea cephalopod mollusc, the coleoid *Spirula*
1145 *spirula*. Deep. Res. Part I Oceanogr. Res. Pap. 142, 127–144.
1146 <https://doi.org/10.1016/j.dsr.2018.10.007>
- 1147 Holliday, L.S., De Faria, L.P., Rody, W.J., 2020. Actin and actin-associated proteins in
1148 extracellular vesicles shed by osteoclasts. Int. J. Mol. Sci., 21, art. numb. 158.
1149 <https://doi.org/10.3390/ijms21010158>
- 1150 Hunt, S., Nixon, M., 1981. A comparative study of protein composition in the chitin-protein
1151 complexes of the beak, pen, sucker disc, radula and oesophageal cuticle of cephalopods.
1152 Comp. Biochem. Physiol., 68 B, 535-546. [https://doi.org/10.1016/0305-0491\(81\)90071-7](https://doi.org/10.1016/0305-0491(81)90071-7)
- 1153 Immel, F., Broussard, C., Catherinet, B., Plasseraud, L., Alcaraz, G., Bundeleva, I., Marin, F.,
1154 2016. The shell of the invasive bivalve species *Dreissena polymorpha*: biochemical,
1155 elemental and textural investigations. PLoS One 11, 1–28.
1156 <https://doi.org/10.1371/journal.pone.0154264>
- 1157 Jackson, D.J., McDougall, C., Green, K., Simpson, F., Wörheide, G., Degnan, B.M., 2006. A
1158 rapidly evolving secretome builds and patterns a sea shell. BMC Biol. 4, 40.
1159 <https://doi.org/10.1186/1741-7007-4-40>
- 1160 Jang, M.-K., Kong, B.-G., Jeong, Y.-I., Lee, C.H., Nah, J.-W., 2004. Physicochemical
1161 characterization of α -chitin, β -chitin, and γ -chitin separated from natural resources. J.
1162 Polym. Sci. Part A Polym. Chem. 42, 3423–3432. <https://doi.org/10.1002/pola.20176>
- 1163 Jastrzbski, W., Sitarz, M., Rokita, M., Bułat, K., 2011. Infrared spectroscopy of different
1164 phosphates structures. Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 79, 722–727.
1165 <https://doi.org/10.1016/j.saa.2010.08.044>
- 1166 Kanold, J., Guichard, N., Immel, F., Plasseraud, L., Corneillat, M., Alcaraz, G., Brümmer, F.,
1167 Marin, F., 2015. Spine and test skeletal matrices of the Mediterranean sea urchin *Arbacia*
1168 *lixula* - a comparative characterization of their sugar signature. FEBS J. 282, 1891-1905.
1169 <https://doi.org/10.1111/febs.13242>
- 1170 Kaya, M., Baran, T., Mentés, A., Asaroglu, M., Sezen, G., Tozak, K.O., 2014. Extraction and
1171 Characterization of α -Chitin and Chitosan from Six Different Aquatic Invertebrates. Food
1172 Biophys. 9, 145–157.
- 1173 Kaya, M., Mujtaba, M., Ehrlich, H., Salaberria, A.M., Baran, T., Amemiya, C.T., Galli, R.,
1174 Akyuz, L., Sargin, I., Labidi, J., 2017. On chemistry of γ -chitin. Carbohydr. Polym. 176,
1175 177–186. <https://doi.org/10.1016/j.carbpol.2017.08.076>
- 1176 Klatt, R.A., Nitsche, P., Kobbe, B., Mörgelin, M., Paulsson, M., Wagener, R., 2000. Molecular
1177 structure and tissue distribution of matrilin-3, a filament-forming extracellular matrix
1178 protein expressed during skeletal development. J. Biol. Chem. 275, 3999–4006.
- 1179 Klug, C., Landman, N.H., Fuchs, D., Mapes, R.H., Pohle, A., Guériau, P., Reguer, S.,
1180 Hoffmann, R., 2019. Anatomy and evolution of the first Coleoidea in the Carboniferous.
1181 Commun. Biol. 2, art. number 280. <https://doi.org/10.1038/s42003-019-0523-2>

- 1182 Kong, J., Yu, S., 2007. Fourier transform infrared spectroscopic analysis of protein secondary
1183 structures. *Acta Biochim. Biophys. Sin.* 39, 549–559.
- 1184 Kröger, B., Vinther, J., Fuchs, D., 2011. Cephalopod origin and evolution: a congruent picture
1185 emerging from fossils, development and molecules: Extant cephalopods are younger than
1186 previously realised and were under major selection to become agile, shell-less predators.
1187 *BioEssays* 33, 602–613. <https://doi.org/10.1002/bies.201100001>
- 1188 Lemanis, R., Stier, D., Zlotnikov, I., Zaslansky, P., Fuchs, D., 2020. The role of mural
1189 mechanics on cephalopod palaeoecology. *J. R. Soc. Interface* 17, 20200009.
1190 <http://dx.doi.org/10.1098/rsif.2020.0009>
- 1191 Le Pabic, C., Marie, A., Marie, B., Percot, A., Bonnaud-Ponticelli, L., Lopez, P.J., Luquet, G.,
1192 2017. First proteomic analyses of the dorsal and ventral parts of the *Sepia officinalis*
1193 cuttlebone. *J. Proteomics* 150, 63–73. <https://doi.org/10.1016/j.jprot.2016.08.015>
- 1194 Le Roy, N., Jackson, D.J., Marie, B., Ramos-Silva, P., Marin, F., 2014. The evolution of
1195 metazoan alpha-carbonic anhydrases and their roles in CaCO₃ biomineralization. *Front.*
1196 *Zool.* 11, art. numb. 75. <https://doi.org/10.1186/s12983-014-0075-8>.
- 1197 Lindgren, A.R., Anderson, F.E., 2018. Assessing the utility of transcriptome data for inferring
1198 phylogenetic relationships among coleoid cephalopods. *Mol. Phylogenet. Evol.* 118, 330–
1199 342. <https://doi.org/10.1016/j.ympev.2017.10.004>
- 1200 Lindgren, A.R., Pankey, M.S., Hochberg, F.G., Oakley, T.H., 2012. A multi-gene phylogeny
1201 of Cephalopoda supports convergent morphological evolution in association with multiple
1202 habitat shifts in the marine environment. *BMC Evol. Biol.* 12.
1203 <https://doi.org/10.1186/1471-2148-12-129>
- 1204 Linné, C. von, 1758. *Systema naturae per regna tria naturae :secundum classes, ordines, genera,*
1205 *species, cum characteribus, differentiis, synonymis, locis. Holmiae.*
- 1206 Lowenstam, H.A., 1981. Minerals formed by organisms. *Science* 211, 1126–1131.
1207 <https://doi.org/10.1126/science.7008198>
- 1208 Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M.,
1209 Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang,
1210 D., Zheng, C., Lanczycki, C.J., Marchler-Bauer, A., 2020. CDD/SPARCLE: the conserved
1211 domain database in 2020. *Nucleic Acids Res.* 48, D265–D268. [https://doi:](https://doi:10.1093/nar/gkz991)
1212 [10.1093/nar/gkz991](https://doi.org/10.1093/nar/gkz991)
- 1213 Marie, B., Joubert, C., Tayalé, A., Zanella-Cleón, I., Belliard, C., Piquemal, D., Cochenne-
1214 Laureau, N., Marin, F., Gueguen, Y., Montagnani, C., 2012. Different secretory repertoires
1215 control the biomineralization processes of prism and nacre deposition of the pearl oyster
1216 shell. *Proc. Natl. Acad. Sci. U. S. A.* 109, 20986–20991.
1217 <https://doi.org/10.1073/pnas.1210552109>
- 1218 Marie, B., Marin, F., Marie, A., Bédouet, L., Dubost, L., Alcaraz, G., Milet, C., Luquet, G.,
1219 2009. Evolution of nacre: biochemistry and proteomics of the shell organic matrix of the
1220 cephalopod *Nautilus macromphalus*. *ChemBioChem* 10, 1495–1506.
1221 <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-](https://doi.org/10.1111/j.1742-4658.2011.08129.x)
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](https://doi.org/10.1016/S0070-2153(07)80006-8)
- 1242 Masuda, F., Hirano, M., 1980. Chemoical composition of some modern 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](https://doi.org/10.1016/0003-2697(81)90783-1)
- 1253 Murray, J.W., 1985. Atlas of invertebrate microfossils. 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>

- 1262 Neige, P., Warnke, K., 2010. Just how many species of *Spirula* are there ? A morphometric
1263 approach, , in: Tanabe, K., Shigeta, Y., Sasaki, T. & Hirano, H. (Eds.), Cephalopods -
1264 Present and Past. Tokai University Press, Tokyo, pp. 77–84.
- 1265 Nishiguchi, M.K., Mapes, R.H., 2008. Cephalopoda, in: Ponder, W.F. & Lindberg, D.R. (Eds.),
1266 Phylogeny and Evolution of the Mollusca. University of California Press, Berkeley, CA,
1267 pp. 163–199.
- 1268 Nys, Y., Hincke, M.T., Arias, J.L., Garcia-Ruiz, J.M., Solomon, S.E., 1999. Avian eggshell
1269 mineralization. Avian Poult. Biol. Rev. 10, 143–166.
- 1270 Ohkouchi, N., Tsuda, R., Chikaraishi, Y., Tanabe, K., 2013. A preliminary estimate of the
1271 trophic position of the deep-water ram’s horn squid *Spirula spirula* based on the nitrogen
1272 isotopic composition of amino acids. Mar. Biol. 160, 773–779.
1273 <https://doi.org/10.1007/s00227-012-2132-1>
- 1274 Palmer, A.R., 1992. Calcification in marine molluscs: how costly is it? Proc. Natl. Acad. Sci.
1275 U. S. A. 89, 1379–1382. <https://doi.org/10.1073/pnas.89.4.1379>
- 1276 Palmer, A.R., 1983. Relative cost of producing skeletal organic matrix versus calcification:
1277 Evidence from marine gastropods. Mar. Biol. 75, 287–292.
1278 <https://doi.org/10.1007/BF00406014>
- 1279 Pandya, M., Liu, H., Dangaria, S.J., Zhu, W., Li, L.L., Pan, S., Abufarwa, M., Davis, R.G.,
1280 Guggenheim, S., Keiderling, T., Luan, X., Diekwisch, T.G.H., 2017. Integrative temporo-
1281 spatial, mineralogic, spectroscopic, and proteomic analysis of postnatal enamel
1282 development in teeth with limited growth. Front. Physiol. 8, 793.
1283 <https://doi.org/10.3389/fphys.2017.00793>
- 1284 Pavat, C., Zanella-Cléon, I., Becchi, M., Medakovic, D., Luquet, G., Guichard, N., Alcaraz, G.,
1285 Dommergues, J.L., Serpentine, A., Lebel, J.M., Marin, F., 2012. The shell matrix of the
1286 pulmonate land snail *Helix aspersa maxima*. Comp. Biochem. Physiol. - B Biochem. Mol.
1287 Biol. 161, 303–314. <https://doi.org/10.1016/j.cbpb.2011.12.003>
- 1288 Ponder, W.F., Lindberg, D.R., Ponder, J.M., 2019. Biology and Evolution of the Mollusca.
1289 CRC Press.
- 1290 Radev, L., Y. Mostafa, N., Michailova, I., M. M. Salvado, I., H. V. Fernandes, M., 2012. In
1291 Vitro bioactivity of collagen/calcium phosphate silicate composites, cross-linked with
1292 chondroitin sulfate. Int. J. Mater. Chem. 2, 1–9.
1293 <https://doi.org/10.5923/j.ijmc.20120201.01>
- 1294 Réhault-Godbert, S., Hervé-Grépinet, V., Gautron, J., Cabau, C., Nys, Y., Hincke, M., 2011.
1295 Molecules involved in chemical defence of the chicken egg, in: Nys, Y., Bain, M., van
1296 Immerseel, F. (Eds.), Improving the Safety and Quality of Eggs and Egg Products - Egg
1297 Chemistry, Production and Consumption. Woodhead Publishing Series in Food Science,
1298 Technology and Nutrition , pp. 183–208. <https://doi.org/10.1533/9780857093912.2.183>
- 1299 Rosenthal, A.K., Gohr, C.M., Ninomiya, J., Wakim, B.T., 2011. Proteomic analysis of articular
1300 cartilage vesicles from normal and osteoarthritic cartilage. Arthritis Rheum. 63, 401–411.
1301 <https://doi.org/10.1002/art.30120>

- 1302 Rusenko, K.W., Donachy, J.E., Wheeler, A.P., 1991. Purification and characterization of a shell
1303 matrix phosphoprotein from the american oyster, in: Sykes, C.S., Wheeler, A.P. (Eds.),
1304 Surface Reactive Peptides and Polymers - Discovery and Commercialization. ACS
1305 Symposium Series, 444. American Chemical Society, pp. 107–124.
1306 <https://doi.org/doi:10.1021/bk-1991-0444.ch008>
- 1307 Sarashina, I., Yamaguchi, H., Haga, T., Iijima, M., Chiba, S., Endo, K., 2006. Molecular
1308 evolution and functionally important structures of molluscan dermatopontin: implications
1309 for the origins of molluscan shell matrix proteins. *J. Mol. Evol.* 62, 307–318.
1310 <https://doi.org/10.1007/s00239-005-0095-2>
- 1311 Sathyan, N., Rosamma, P., Chaithanya, E.R., Anil Kumar, P.R., 2012. Identification and
1312 molecular characterization of Molluskin, a histone-H2A-derived antimicrobial peptide
1313 from molluscs. *ISRN Mol. Biol.* 2012, 219656. <https://doi.org/10.5402/2012/219656>
- 1314 Schmitt, N., Marin, F., Thomas, J., Plasseraud, L., Demoy-Schneider, M., 2018. Pearl grafting:
1315 tracking the biological origin of nuclei by straightforward immunological methods. *Aquac.*
1316 *Res.* 49, 692-700. <https://doi.org/10.1111/are.13499>
- 1317 Simkiss, K., Wilbur, K.M., 1989. Biomineralization. Cell biology and mineral deposition.
1318 Academic Press, San Diego. [https://doi.org/https://doi.org/10.1016/B978-0-08-092584-](https://doi.org/https://doi.org/10.1016/B978-0-08-092584-4.50001-9)
1319 [4.50001-9](https://doi.org/https://doi.org/10.1016/B978-0-08-092584-4.50001-9)
- 1320 Stegemann, H., 1963. Proteine (Conchagene) und chitin im stützgewebe von tintenfischen.
1321 *Hoppe. Seylers. Z. Physiol. Chem.* 331, 269–279.
1322 <https://doi.org/10.1515/bchm2.1963.331.1.269>
- 1323 Stejskal, E.O., Schaefer, J., Waugh, J.S., 1977. Magic-angle spinning and polarization transfer
1324 in proton-enhanced NMR. *Journal of Magnetic Resonance* 28, 105-112.
1325 [https://doi.org/10.1016/0022-2364\(77\)90260-8](https://doi.org/10.1016/0022-2364(77)90260-8)
- 1326 Strugnell, J., Norman, M., Jackson, J., Drummond, A.J., Cooper, A., 2005. Molecular
1327 phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) using a multigene approach;
1328 the effect of data partitioning on resolving phylogenies in a Bayesian framework. *Mol.*
1329 *Phylogenet. Evol.* 37, 426–441. <https://doi.org/10.1016/j.ympev.2005.03.020>
- 1330 Strugnell, J.M., Hall, N.E., Vecchione, M., Fuchs, D., Allcock, A.L., 2017. Whole
1331 mitochondrial genome of the Ram’s Horn Squid shines light on the phylogenetic position
1332 of the monotypic order Spirulida (Haeckel, 1896). *Mol. Phylogenet. Evol.* 109, 296–301.
1333 <https://doi.org/10.1016/j.ympev.2017.01.011>
- 1334 Suzuki, M., Saruwatari, K., Kogure, T., Yamamoto, Y., Nishimura, T., Kato, T., Nagasawa, H.,
1335 2009. An acidic matrix protein, Pif, is a key macromolecule for nacre formation. *Science*
1336 325, 1388-1390. <https://doi.org/10.1126/science.1173793>
- 1337 Sviben, S., Gal, A., Hood, M.A., Bertinetti, L., Politi, Y., Bennet, M., Krishnamoorthy, P.,
1338 Schertel, A., Wirth, R., Sorrentino, A., Pereiro, E., Faivre, D., Scheffel, A., 2016. A
1339 vacuole-like compartment concentrates a disordered calcium phase in a key
1340 coccolithophorid alga. *Nat. Commun.* 7, 11228. <https://doi.org/10.1038/ncomms11228>
- 1341 Takeuchi, T., Plasseraud, L., Ziegler-Devin, I., Brosse, N., Shinzato, C., Satoh, N., Marin, F.,
1342 2018. Biochemical characterization of the skeletal matrix of the massive coral, *Porites*

- 1343 *australiensis* – The saccharide moieties and their localization. J. Struct. Biol. 203, 219–
1344 229. <https://doi.org/10.1016/j.jsb.2018.05.011>
- 1345 Tanner, A.R., Fuchs, D., Winkelmann, I.E., Ribeiro, M., Gilbert, M.T.P., Pankey, M.S., Kocot,
1346 K.M., Halanych, K.M., Oakley, T.H., Fonseca, R.R., Pisani, D., Vinther, J., Vinther, J.,
1347 2017. Molecular clocks indicate turnover and diversification of modern coleoid
1348 cephalopods during the Mesozoic Marine Revolution. Proc. R. Soc. B 284, 20162818.
- 1349 Tyszkka, J., Bickmeyer, U., Raitzsch, M., Bijma, J., Kaczmarek, K., Mewes, A., Topa, P., Janse,
1350 M., 2019. Form and function of F-actin during biomineralization revealed from live
1351 experiments on foraminifera. Proc. Natl. Acad. Sci. USA, 116, 4111-4116.
1352 <https://doi.org/10.1073/pnas.1810394116>
- 1353 Uribe, J.E., Zardoya, R., 2017. Revisiting the phylogeny of Cephalopoda using complete
1354 mitochondrial genomes. J. Molluscan Stud. 83, 133–144.
1355 <https://doi.org/10.1093/mollus/eyw052>
- 1356 Vinther, J., 2015. The origins of molluscs. Palaeontology 58, 19–34.
1357 <https://doi.org/10.1111/pala.12140>
- 1358 Warnke, K., Plötner, J., Santana, J.I., Rueda, M.J., Llinas, O., 2003. Reflections on the
1359 phylogenetic position of *Spirula* (Cephalopoda): preliminary evidence from the 18S
1360 ribosomal RNA gene. Berliner Paläobiol. Abh. 3, 253–260.
- 1361 Weigle, J., Franz-Odenaal, T.A., Hilbig, R., 2016. Not all inner ears are the same: otolith
1362 matrix proteins in the inner ear of sub-adult cichlid fish, *Oreochromis Mossambicus*,
1363 reveal insights into the biomineralization process. Anat. Rec. 299, 234–245.
1364 <https://doi.org/10.1002/ar.23289>
- 1365 Weiner, S., 1983. Mollusk shell formation: isolation of two organic matrix proteins associated
1366 with calcite deposition in the bivalve *Mytilus californianus*. Biochemistry 22, 4139–4145.
1367 <https://doi.org/10.1021/bi00286a023>
- 1368 Weiss, I.M., Schönitzer, V., Eichner, N., Sumper, M., 2006. The chitin synthase involved in
1369 marine bivalve mollusk shell formation contains a myosin domain. FEBS Lett. 580, 1846–
1370 1852. <https://doi.org/10.1016/j.febslet.2006.02.044>
- 1371 Whittaker, C.A., Hynes, R.O., 2002. Distribution and evolution of von Willebrand: integrin a
1372 domains: widely dispersed adhesion and elsewhere. Mol. Biol. Cell 13, 3369-3387.
1373 <https://doi.org/10.1091/mbc.E02-05-0259>
- 1374 Yonezawa, M., Sakuda, S., Yoshimura, E., Suzuki, M., 2016. Molecular cloning and functional
1375 analysis of chitinases in the fresh water snail, *Lymnaea stagnalis*. J. Struct. Biol. 196, 107–
1376 118. <https://doi.org/10.1016/j.jsb.2016.02.021>
- 1377 Zhang, C., Xie, L., Huang, J., Chen, L., Zhang, R., 2006. A novel putative tyrosinase involved
1378 in periostracum formation from the pearl oyster (*Pinctada fucata*). Biochem. Biophys.
1379 Res. Commun. 342, 632–639. <https://doi.org/10.1016/j.bbrc.2006.01.182>
- 1380 Zhang, Y., Meng, Q., Jiang, T., Wang, H., Xie, L., Zhang, R., 2003. A novel ferritin subunit
1381 involved in shell formation from the pearl oyster (*Pinctada fucata*). Comp. Biochem.

1382
1383

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