

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Shell palaeoproteomics: First application of peptide mass fingerprinting for the rapid identification of mollusc shells in archaeology

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1755182> since 2020-09-09T14:22:04Z

Published version:

DOI:10.1016/j.jprot.2020.103920

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

1 Shell palaeoproteomics: First application of peptide mass fingerprinting
2 for the rapid identification of mollusc shells in archaeology

3 List of authors:

4 Jorune Sakalauskaite^{1,2}, Frédéric Marin², Barbara Pergolizzi³, Beatrice Demarchi¹

6 Affiliations

7 ¹Department of Life Sciences and Systems Biology, University of Turin, Via Accademia
8 Albertina 13, 10123 Turin, Italy.

9
10 ²Biogeosciences, UMR CNRS 6282, University of Burgundy-Franche-Comté, 6 Boulevard
11 Gabriel, 21000 Dijon, France.

12
13 ³Department of Clinical and Biological Sciences, University of Turin, AOU S. Luigi, 10043
14 Orbassano TO, Italy.

16 Corresponding Authors:

17 **Jorune Sakalauskaite**, jorune.sakalauskaite@unito.it

18 Address 1: Department of Life Sciences and Systems Biology, University of Turin, Via
19 Accademia Albertina 13, 10123 Turin, Italy;

20 Address 2: Biogeosciences, UMR CNRS 6282, University of Burgundy-Franche-Comté, 6
21 Boulevard Gabriel, 21000 Dijon, France.

22
23 **Beatrice Demarchi**, beatrice.demarchi@unito.it

24 Department of Life Sciences and Systems Biology, University of Turin, Via Accademia
25 Albertina 13, 10123 Turin, Italy.

26 **Keywords**

27 Palaeoproteomics, mollusc shell, shellomics, MALDI-TOF mass spectrometry, peptide mass
28 fingerprint

29

30 Abstract

31 Molluscs were one of the most widely-used natural resources in the past, and their shells are
32 abundant among archaeological findings. However, our knowledge of the variety of shells that
33 were circulating in prehistoric times (and thus their socio-economic and cultural value) is
34 scarce due to the difficulty of achieving taxonomic determination of fragmented and/or worked
35 remains. This study aims to obtain molecular barcodes based on peptide mass fingerprints,
36 (PMFs) of intracrystalline proteins, in order to obtain shell identification. Palaeoproteomic
37 applications on shells are challenging, due to low concentration of molluscan proteins and an
38 incomplete understanding of their sequences. We explore different approaches for protein
39 extraction from small-size samples (<20 mg), followed by MALDI-TOF-MS analysis. The SP3
40 (single-pot, solid-phase) sample preparation method was found to be the most successful in
41 retrieving the intracrystalline protein fraction from seven molluscan shell taxa, which belong to
42 different phylogenetic groups, possess distinct microstructures and are relevant for
43 archaeology. Furthermore, all the shells analysed, including a 7000-year-old specimen of the
44 freshwater bivalve *Pseudunio*, yielded good-quality distinctive spectra, demonstrating that
45 PMFs can be used for shell taxon determination. Our work suggests good potential for large-
46 scale screening of archaeological molluscan remains.

47

48 APPROXIMATE LOCATION OF GRAPHICAL ABSTRACT

49 1. Introduction

50 Molluscs have been an important natural resource throughout human history; they were
51 exploited as a foodstuff and their shells were perforated and, presumably, worn as ornaments
52 by both early modern humans [e.g. 1–3] and Neanderthals [e.g. 4]. The tradition of “shell
53 jewelry” continued throughout the Middle and Upper Palaeolithic and further expanded during
54 the Neolithic, when shells were extensively used as a raw material, fashioned into pendants,
55 bracelets and beads of a variety of shapes and types [5]. While research into Palaeolithic
56 ornaments has been especially fruitful, the same cannot be said for later prehistory and
57 historical times [6], despite the growing number of studies at the regional and supra-regional
58 scale, particularly for the European Neolithic [e.g. 7,8]. One line of enquiry concerns the
59 discovery of the diversity of shells used as raw materials by prehistoric societies, and, above
60 all, the reasons behind their choice: was the selection of certain species based on their
61 prestige, material qualities or socio-cultural significance [6,9,10]? Answering these questions
62 could help us to better understand shifting cultural and biological boundaries in the past, to
63 track people’s interactions, migrations and mobility, as well as to reconstruct their strategies
64 for adapting to new environments [8,11–13].

65

66 Archaeological shell artefacts, particularly ornaments or tools, are often found heavily
67 fragmented, worked and/or degraded and thus taxonomic identification becomes problematic,
68 if not impossible [14]. This is because most (if not all) morphological features, such as outer
69 surface ornamentations, are usually absent. The microstructural/mineralogical characteristics
70 of the material can only give broad information on the shell type used [15,16], because the
71 most commonly encountered microstructure types, such as nacreous, prismatic and crossed-
72 lamellar, are found across many different mollusc families, chiefly among bivalves and
73 gastropods [17]. The development of different biomolecular tools has advanced research into

74 the origin of small, old and fragmented biological remains from archaeological and
75 paleontological contexts. In particular, in the past decade, ancient protein research
76 (palaeoproteomics) has been extremely successful with respect to collagen-based and
77 keratin-based substrates [18–24], while newer applications include the characterisation of
78 more complex mineralised proteomes, such as those of dental calculus, dental enamel and
79 avian eggshell [16,25–29]. Peptide mass fingerprinting (PMF) by MALDI-TOF has been
80 particularly useful, allowing rapid large-scale screening of artefacts for species identification
81 [30]. However, mollusc shells, and more generally all invertebrate organisms, are still
82 underrepresented in palaeoproteomic studies.

83 Mollusc shells are organo-mineral nanocomposites of calcium carbonate (calcite and/or
84 aragonite) and a small organic fraction (~0.01-2%) comprising a mixture of proteins,
85 saccharides, lipids and pigments [31]. There are two key features that make shells a very
86 interesting system for ancient protein studies. Firstly, shell protein sequences can vary
87 considerably across taxa [32–34], which is useful when attempting to determine specific
88 molecular “barcodes”. Secondly, shells retain a small fraction of their proteins occluded in the
89 mineral crystals; these are known as “intracrystalline proteins” and may represent a so-called
90 ‘closed system’, remaining inaccessible by environmental contamination and protected from
91 rapid *in situ* degradation processes (i.e., diagenesis) over archaeological/geological
92 timescales [35–40].

93

94 However, shells are also a challenging substrate for biomolecular studies, partially explaining
95 why “palaeoshellomics” is only just beginning to catch up. The main challenges are due to the
96 low abundance of the intracrystalline proteins and our limited knowledge of shell protein
97 sequences. Firstly, the intracrystalline shell protein fraction, which is typically isolated by a
98 strong bleaching step, represents around 0.001-0.01% of the total shell mass [37]. Nacreous
99 shells such as the freshwater mother-of-pearl mussels (e.g. *Unio* and *Margaritifera*), pearl
100 oysters (*Pinctada*), abalone shells (*Haliotis*) and many others, have a relatively organic-rich
101 framework. These structures are dominated by *intercrystalline* organics, which can constitute
102 up to 1-2% by weight of the total shell [37,41]. In contrast, for crossed-lamellar shells, e.g.
103 *Spondylus*, *Glycymeris*, *Cardiidae*, *Strombus gigas* and some foliated shells, e.g. *Pecten*,
104 *Crassostrea*, the shell matrix content may be as low as ~0.004% by weight [42–45]. Therefore,
105 considering that sample size is usually a limiting factor for the application of palaeoproteomics
106 to unique archaeological artefacts, sample preparation protocols commonly employed on
107 small samples (typically, <20 mg) of other biomineralised tissues, e.g. eggshell [26,46,47],
108 may not be adequate for most mollusc shell substrates, and will need to be revised.

109 Secondly, there is a great diversity of mollusc shell proteins, most of which are currently not
110 fully characterised and thus remain largely unknown [48]. The peculiarity of shell proteins is
111 attested by the fact that they neither carry a simple phylogenetic signal, nor are associated to
112 specific microstructural features [33,49,50]. One of their most prominent characteristics is the
113 presence of repetitive low complexity domains (RLCDs) [34,50], which are made of blocks of
114 several to tens of poly-Ala, poly-Gly and poly-Ser. Such domains are difficult to cleave with
115 routinely used proteases (e.g. trypsin), thus are often “missed” in proteomic analyses. The
116 presence of different post-translational modifications (such as glycosylation, phosphorylation)
117 may also hinder the detection/characterisation by mass spectrometric analyses [48,51].

118

119 This work therefore aimed to develop a simple proteomic approach based on MALDI-TOF
120 mass spectrometry, in order to obtain molecular barcodes for the taxonomic identification of
121 archaeological shell artefacts. The main objectives were:

- 122 1) Method development: to test different preparation protocols, which could be used for
123 small-size shell samples (<20 mg).
124 2) Application: to explore the viability of generating peptide mass fingerprints (PMFs) for
125 the intracrystalline proteins of different mollusc shell taxa.
126

127 To achieve the first objective, three specimens of *Unio pictorum* (Bivalvia, Unionida, Unionidae
128 - freshwater), *Ostrea edulis* (Bivalvia, Ostreida, Ostreidae - marine) and *Spondylus*
129 *gaederopus* (Bivalvia, Pectinida, Spondylidae - marine) were used in order to develop a
130 suitable method for shell protein extraction and characterisation. These three species are
131 important for archaeological research in the Mediterranean basin and in central-northern
132 Europe. In addition, they represent three different microstructures: aragonitic nacre, calcitic
133 foliated and aragonitic crossed-lamellar. They were also selected on the basis of their bulk
134 amino acid composition: in the dataset reported by ref [14] (their Figure 3) *Unio* sp. could easily
135 be distinguished from other taxa, while both *Spondylus* and *Ostrea* yielded a more uncertain
136 classification. Our hypothesis was that, by retrieving the peptide mass fingerprints (PMFs) of
137 the intracrystalline shell proteins, more secure taxonomic identification could be achieved,
138 especially as new data on *Spondylus* showed that this mollusc shell has a very distinct protein
139 makeup [45]. This encompasses the second objective of this work, i.e. to apply the optimised
140 preparation method to a further set of shells: *Pecten maximus* (Bivalvia, Pectinida, Pectinidae
141 - marine), *Patella vulgata* (Gastropoda, Patellidae - marine), *Phorcus turbinatus* (Gastropoda,
142 Trochida, Trochidae - marine) and a 7000-year-old archaeological specimen, *Pseudunio*
143 *auricularius* (Bivalvia, Unionida, Margaritiferidae - freshwater) [52,53]. These taxa were
144 selected as they had been previously tested for their ability to preserve a fraction of
145 intracrystalline proteins, which is stable over archaeological and geological timescales
146 [16,39,40,44,54].

147 2. Material and Methods

148 2.1. Samples

149 2.1.1. Method development

150 Three bivalve species were studied in order to optimise a suitable method for shell protein
151 analysis, testing different bleaching and protein extraction techniques in small-size samples
152 (<20 mg):

- 153 ● 1) *Spondylus gaederopus* is a Mediterranean bivalve, which belongs to a small family,
154 Spondylidae (order Pectinida), and has a complex microstructure, composed of
155 aragonitic crossed-lamellar and prismatic layers and an upper calcitic foliated layer.
156 The shell was purchased from Conchology, Inc [55]; it had been collected alive by
157 diving to a depth of 15 m in the area of Saronikos, Greece, in 2010 (as indicated by
158 the vendors);
- 159 ● 2) *Unio pictorum* is a freshwater bivalve and belongs to the family Unionidae. Its shell
160 is completely aragonitic, comprising nacreous and prismatic layers. The shell used in
161 this study was collected in a stream close to Izeure (Burgundy, France) by one of the
162 authors (F.M.);
- 163 ● 3) *Ostrea edulis* is a marine bivalve, commonly known as the European flat oyster, and
164 belong to the family Ostreidae. The shell is foliated calcitic with the presence of
165 discontinuous chalky lenses. The specimen was collected in northern Jutland

166 (Denmark), and obtained from the personal collection of collaborator Søren H.
167 Andersen [16].
168

169 2.1.2 Method application

170 Three modern shells (specimens from the reference collection of one of the authors, B.D.) and
171 one archaeological shell were studied to evaluate the optimal method for intracrystalline
172 protein extraction and analysis by peptide mass fingerprinting (PMF).

- 173 • *Patella vulgata* is a marine gastropod with calcite and aragonite layered in several
174 different microstructures (prismatic, foliated, crossed-lamellar). The intracrystalline
175 shell proteins display a closed-system behaviour [39];
- 176 • *Phorcus turbinatus* is a marine gastropod, mainly nacreous (aragonitic) with a thin
177 upper calcitic layer (prismatic and foliated). The intracrystalline protein fraction was
178 observed to behave as a closed system [40,54];
- 179 • *Pecten maximus* is a marine bivalve, commonly known as the great scallop. The shell
180 is composed mainly of foliated calcite. *Pecten* shell also retains a small intracrystalline
181 protein fraction that was found to behave as a closed system [44];
- 182 • *Pseudunio auricularius* is a freshwater bivalve with a fully aragonitic shell, comprising
183 nacreous and prismatic layers. This specimen comes from the Neolithic site of Isorella
184 in the Po Plain, Italy, dated to 5226–5023 cal BCE [16,52,53].

185 2.2 Analytical procedure: Method development

186 All of the shell samples were already available as fine-grained powders (particle size: 200-500
187 µm) as they had been used for previous studies [14,16,39,44,45]. The powders represent the
188 bulk fraction of the shell, i.e. where all (or most) microstructural layers are represented.

189 2.2.1 Bleaching

190 Bleaching is a vigorous cleaning approach which involves the use of sodium hypochlorite
191 (NaOCl) in order to remove surface contamination and/or weakly bound intercrystalline
192 organics from targeted samples. It is used routinely to treat biomineralised tissues and
193 organisms, such as eggshell, mollusc shell or coral, before ancient protein analysis.

194 Developed originally for amino acid racemization geochronology [36,37] it is also used in
195 shell proteomics in order to reduce the pool of analysed proteins and isolate those that are
196 truly associated to the mineral phase [56].

197 *Unio*, *Spondylus* and *Ostrea* shell powders were carefully weighted and placed in clean
198 eppendorf vials. Twelve 20 mg samples were prepared for each shell (Figure 1) so that three
199 different bleaching exposures could be tested on four 20 mg aliquots:

- 200 1) Mild bleaching: 1 mL of NaOCl (diluted to an approximate concentration of 1.0-1.5%)
201 was added and powders were left to soak for 4 hours; this type of bleaching was
202 selected as it was used in a previous palaeoshellomics study [16];
- 203 2) Intermediate bleaching: 1 mL of NaOCl (diluted to an approximate concentration of
204 1.0-1.5%) was added and powders were left to soak for 24 hours; this type of bleaching
205 was selected as an intermediate step between the “mild” and “strong”;
- 206 3) Strong bleaching: 1 mL of NaOCl (concentrated, 10-15%) was added and the
207 powders soaked for 48 hours - this step is typically used to isolate the intracrystalline
208 fraction of proteins in mollusc shells [37]. In this paper, for convenience, we refer to
209 this 48-hr-bleached fraction as “intracrystalline”. However, we note that a series of

210 experiments should be performed for each of the shells separately in order to verify
211 the optimal bleaching times and to test the closed-system behaviour [see e.g. 57].

212

213 After bleaching, all of the samples were thoroughly rinsed with ultrapure water (5 times) and
214 air-dried.

215 2.2.2 Demineralisation

216 The bleached powders of each shell were divided into two subsets in order to test two
217 demineralisation approaches (Table 1):

- 218 1) Acetic acid: the first set was demineralised with cold acetic acid (10% v/v) adding 100
219 μL every hour, thoroughly mixing, to a final volume of 300 μL (in the case of *Spondylus*,
220 which was not fully demineralised, an additional 30 μL aliquot was added to obtain
221 complete demineralisation);
- 222 2) EDTA: the second set was demineralised with a 0.5 M EDTA solution (Sigma-Aldrich,
223 E7889, pH 8, $\sim 0.5\text{M}$) by adding 500 μL to each of the powdered samples and
224 thoroughly mixing with a vortex for ~ 4 hours.

225

226 All of the extracts were kept at 4 °C until the protein purification step was carried out.

227 2.2.3 Protein purification and processing

228 All of the demineralised shell samples were again divided into two subsets and two separate
229 desalting/protein purification approaches were applied: filter aided sample preparation (FASP)
230 and single-pot, solid-phase sample preparation (SP3) (Table 1).

231

232 **FASP extraction**

233 The extracts were concentrated using PALL Nanosep centrifugal devices (3kDa, 0.5 mL). For
234 the acetic acid extracts, which resulted in a mixture of acid soluble and acid insoluble matrices
235 (ASM and AIM), these were mixed and loaded to the same centrifugal device to minimise loss
236 due to separate washes. The EDTA extracts were solubilised and homogeneous. The
237 solutions were loaded onto spin filter columns and the samples were concentrated and
238 desalted washing five times with HPLC-grade water (0.5 mL, centrifuging at 11000 rpm, room
239 temperature), before exchanging to buffer (50 mM ammonium bicarbonate, pH 7.5-8). The
240 extracts were reduced using 1M DL-dithiothreitol (Sigma, Canada) for 1 hr at 65 °C, alkylated
241 with 0.5M iodoacetamide (Sigma, USA) for 45 min at room temperature in the dark and
242 digested with trypsin (0.5 μg , Promega, V5111, proteomics grade) overnight. Digestion was
243 stopped with 10% TFA (to a final TFA concentration of 0.1%), samples were purified using
244 C18 solid-phase extraction tips (Pierce zip-tip; Thermo-Fisher) and evaporated to dryness.

245

246 **SP3 extraction**

247 The samples were processed as described in a previous study [58]. For the EDTA extracts,
248 reduction and alkylation were performed before processing with the SP3 beads, and for the
249 acetic acid extracts it was performed after SP3 extraction and buffer exchange. 8 μL of Sera-
250 Mag SpeedBeads (1:1 mixture of hydrophobic and hydrophilic) were added to each of the
251 extracts. To induce binding, 100% EtOH (HPLC-grade) was added to a final EtOH
252 concentration of 50% and incubated at 24°C for 5 min at ~ 1000 rpm. The tubes were then
253 placed on a magnetic rack for separation, the supernatant removed and discarded. The
254 proteins bound to the beads were cleaned with 80% EtOH (3x), exchanged to buffer (50 mM

255 ammonium bicarbonate, pH 7.5-8) and the mixture sonicated for 30 sec. After this step, for
256 the EDTA extracts, enzymatic digestion was carried out directly, while for the acidic extracts,
257 reduction and alkylation were performed first. Trypsin was added (0.5 µg, Promega,
258 proteomics grade) for overnight digestion at 37°C and light shaking was applied (~1000 rpm).
259 Afterwards, the extracts were centrifuged for 1 min, placed on a magnetic rack, the
260 supernatants containing the digested peptides were transferred to separate tubes, acidified
261 with 10% TFA (to a final TFA concentration of 0.1%) and the samples purified using C18 solid-
262 phase extraction tips. Eluted peptides were evaporated to dryness.
263 Table 1 shows the full list of the twelve different treatments tested for each of the three shells
264 (a total of 36 samples were analysed). Additionally, four blank samples were included in the
265 study (AcOH vs EDTA; FASP vs SP3).

266

267 APPROXIMATE LOCATION OF TABLE 1

268 2.3 MALDI-MS Analysis

269 The samples were resuspended in 10 µL TFA solution (0.1%) and 0.7 µL aliquots were
270 mixed with 0.7 µL of α -cyano-4-hydroxycinnamic acid matrix solution (1%, prepared in 50%
271 acetonitrile/ 0.1% trifluoroacetic acid (v/v/v)) directly on a MBT Biotarget 96 MALDI plate. All
272 the samples were analysed on a bench-top Microflex LRF MALDI-TOF mass spectrometer
273 (Bruker Daltonics, Germany). Samples were analysed in reflector mode, using the following
274 parameter settings: ion source 1 18.96 kV; ion source 2 16.02 kV; lens voltage 9.05 kV,
275 reflector 20.01 kV, laser power 22–28%. Shell proteomes of *Ostrea edulis* and *Pseudunio*
276 *auricularius* were analysed with higher laser power (28%) than the rest of the shells (22%).
277 The spectrum collected for each sample resulted from the sum of 1000 laser shots. Mass
278 range was 800–4000 m/z and peptide masses below 650 Da were suppressed. The peptide
279 calibration standard (#8206195, Bruker Daltonics, Germany), a mixture of seven peptides
280 (Angiotensin II m/z = 1046.541, Angiotensin I m/z = 1296.685, Substance_P m/z =
281 1347.735, Bombesin m/z = 1619.822, ACTH (1–17 clip) m/z = 2093.086, ACTH (18–39 clip)
282 m/z = 2465.198 and Somatostatin m/z = 3147.471) was used for external mass calibration to
283 maximise mass accuracy. The spectra were exported as text files and further processed
284 using mMass, an open access mass spectrometry interpretation tool [59]. Two spectra were
285 obtained and averaged for each sample. All of the resulting spectra were processed by
286 performing baseline correction (precision: 100%, relative offset: 10-30%) and by smoothing
287 (Savitzky-Golay method, with a window size of 0.3 m/z , 1.5 cycles). Peak picking was
288 performed selecting an S/N threshold \geq 6, picking height of 100% and deisotoping using
289 standard mMass parameters. Internal mass calibration was carried out using trypsin, keratin
290 and matrix m/z values (reported in SI.1). All the spectra are reported in SI.1.

291 2.4 PMF library preparation

292 For marker peaks identification, samples were extracted in duplicate using the Ic_EDTA_SP3
293 method. Any m/z values corresponding to common laboratory contaminants (i.e. keratin,
294 trypsin, α -cyano MALDI matrix) were excluded from data interpretation (mass tolerance for
295 peak matching: 0.1 Da). Furthermore, in order to ensure that all possible contaminants were
296 taken into account, m/z values identified in blank samples (four samples prepared with the
297 AcOH/EDTA and FASP/SP3 methods) were added to the common contaminants peak list,
298 which was then used to exclude these values from samples PMFs. Finally, each shell
299 spectrum was checked manually for additional recurring peaks, i.e. if the same peak was

300 observed in most shell spectra, it was not taken into account; these might be genuine shell
301 peptides and not contamination, but their occurrence across taxa would prevent their use as
302 “markers”. We note that shell protein sequences can be very different in phylogenetically
303 distant taxa, therefore there is a high probability that recurring *m/z* values do not represent the
304 same peptide but different, isobaric, sequences. The full list of identified contaminant peaks is
305 presented in supplementary file SI.2.

306 2.5 Method application

307 For the four shells included in this part of the study (*Phorcus*, *Patella*, *Pecten* and
308 archaeological *Pseudunio*), the intracrystalline shell proteins were extracted using the optimal
309 method, noted as Ic_EDTA_SP3 (Table 1, method no. 12). In brief, intracrystalline proteins
310 were isolated after 48 hrs of bleaching with concentrated NaOCl (10-15%). Powders were
311 demineralised using EDTA and proteins were extracted and purified using the SP3 method.
312 Enzymatic digestion, peptide desalting and MS analyses were carried out as detailed in
313 section 2.2.3 and 2.3.

314 3. Results and Discussion

315 The first part of the results and discussion section aims to assess the most suitable approach
316 for extracting shell proteins for peptide mass fingerprint (PMF) characterisation. In the second
317 part we show the applicability of the optimised method to a wider variety of molluscan taxa,
318 including an archaeological specimen. Finally, the PMFs for the different shell taxa are
319 presented.

320 3.1 Method Development

321 For shell proteomics by MALDI-TOF, different bleaching, demineralisation and protein
322 purification steps were evaluated on three molluscan taxa (Figure 1, Table 1). We note that
323 the conditions needed to isolate the intracrystalline protein fraction in shells may differ for each
324 species and should be tested individually. However, as the future scope of this project is to
325 create a large library of “intracrystalline PMFs”, in this work, we give an “operational” definition
326 of the “intracrystalline” proteins as the fraction which can be isolated via a 48-hr bleaching
327 step using concentrated NaOCl (~12 %), which is effective for all shell taxa tested thus far,
328 regardless of their age (modern vs fossil) or provenance [14, 36-40, 45, 57].

329

330 **APPROXIMATE LOCATION OF FIGURE 1**

331

332 3.1.1 Intracrystalline shell proteins and effect of bleaching

333 Mollusc shell proteins were successfully isolated, extracted and characterised by MALDI-TOF-
334 MS from all three samples - *Unio*, *Spondylus* and *Ostrea* (Figure 2). Comparing the PMFs of
335 the intracrystalline shell protein fraction (Figure 2a, c, e; spectra in red) and the fraction
336 obtained via “mild” bleaching (spectra in blue), we note that the two spectra are very similar
337 only for *Unio* (Figure 2a), while for *Spondylus* and *Ostrea* (Figure 2c, e), the spectra of the
338 fraction obtained after “mild” bleaching were of lower quality. This was particularly evident for
339 *Spondylus* (Figure 2c), for which the PMF of the Ic fraction was significantly better than that
340 obtained from both the 4-hr (Figure 2c, in blue) and the 24-hr bleached samples (SI.1).
341 Therefore, the “strong bleaching” step is preferable for the isolation and characterisation of

342 shell proteins by MALDI-TOF-MS. Importantly, *Unio*, *Spondylus* and *Ostrea* yielded individual
343 intracrystalline PMFs (Figure 2a, c, e), and we find that most of the potential marker peaks for
344 these shells appear in the 1000-2000 *m/z* range. The corresponding bulk amino acid
345 compositions are presented as pie charts next to the spectra (Figure 2b, d, f) and clearly show
346 that the differences in PMFs are far more evident than the differences in relative amino acid
347 composition.

348
349 The results showed that intracrystalline PMFs can be obtained from 20 mg shell samples,
350 regardless of their different microstructures (nacreous, crossed-lamellar, foliated) and their
351 variable organic content: even the most organic-poor microstructure (i.e. crossed-lamellar in
352 *Spondylus*) retained a sufficient fraction of Ic proteins.

353 Furthermore, we did not observe any simple correlation between bleaching time/NaOCl
354 concentration and the number of potential marker peptides - i.e. shorter bleaching times do
355 not imply better MALDI-TOF spectra and, *vice versa*, harsh bleaching treatments do not
356 necessarily mean that protein concentrations will be too low for proteomics. This is interesting
357 as many “shellomics” studies encourage bleaching as a cleaning pretreatment [56], but
358 generally avoid higher concentrations of NaOCl and longer exposure times, presuming that
359 shell proteins would be fully hydrolyzed. In the case of *Unio*, no compelling difference was
360 observed between the spectra of the (inter+intra)crystalline fraction (“mild” bleaching) and the
361 intracrystalline (Ic) fraction (“strong” bleaching). Remarkably, for *Spondylus*, the intensity and
362 number of potential marker peaks is considerably higher in the intracrystalline fraction
363 compared to the spectra obtained after just 4 hours of bleaching, for which the PMFs were
364 barely detectable. This effect has also been observed in other shells [60], including a study of
365 the *Spondylus* proteome by tandem mass spectrometry [45], and may be due to the difficulty
366 of breaking down complex networks of proteins with other shell matrix macromolecules, such
367 as chitin. It is likely that the presence of glycosylated proteins, lipoproteins, phospholipids, or
368 proteins with repetitive low complexity domains (RLCD), could influence signal detection (or
369 suppress it completely) [51,61]. This would explain why a strong oxidative treatment, which
370 removes a large quantity of these macromolecules, may be advantageous in shell protein
371 analyses. In addition, the intracrystalline proteins have more acidic domains, which bind to the
372 mineral [25], and thus are preferentially ionised, therefore their detection is favoured when
373 analysed by MALDI-TOF mass spectrometry.

374

375 **APPROXIMATE LOCATION OF FIGURE 2**

376 3.1.2 Extraction and purification

377 The SP3 method for shell protein isolation and purification was found to be more effective than
378 FASP. The intracrystalline PMFs obtained by SP3 were of better quality and displayed a higher
379 number of marker peaks for *Unio*, *Ostrea* and *Spondylus* (Figure 3a-c, spectra in red). On the
380 contrary, in the FASP PMFs, the relative proportion between marker peaks and the
381 trypsin/keratin peaks (common laboratory contaminants) was severely skewed towards the
382 latter (Figure 3a-c, green spectra). This is probably due to the fact that the (minimal) loss of
383 proteins which occurs during ultrafiltration is especially noticeable for protein-poor samples,
384 with enzymes and common contaminants thus being over-represented in the resulting
385 spectrum. The SP3 extraction is therefore better suited to shell samples [62].

386 In general, the issue of protein concentration will principally affect MALDI-TOF analyses of
387 proteins from crossed-lamellar and foliated microstructures; for example, a 20-mg *Spondylus*
388 sample may contain as low as ~200 ng of intracrystalline proteins. For nacreous shells,

389 which are generally more organic-rich, we can speculate that 10-15 mg samples should be
390 sufficient for obtaining good-quality PMFs. Obviously, diagenesis will inevitably impact on the
391 limit of detection.

392 There was no significant difference between spectra of samples demineralised with EDTA or
393 acetic acid (Figure 3d), but we note that it was much easier to handle the EDTA extracts
394 because 1) demineralisation with EDTA is less vigorous and 2) EDTA yields fully
395 demineralised extracts, while the acidic decalcification results in two fractions - the acid soluble
396 (ASM) and acid insoluble (AIM) matrices.

397

398 **APPROXIMATE LOCATION OF FIGURE 3**

399

400 The SP3 method, which had been developed for low-concentration samples [58] showed very
401 good results for shell proteins, and it appeared to be time and cost-effective (for such small-
402 size samples, and assuming similar cost for consumables, the SP3 method is ~25 times
403 cheaper than FASP). To our knowledge this is the first application of SP3 extraction for
404 “shellomics”, and it is not yet routinely employed in palaeoproteomics [63]. The efficiency of
405 SP3 was especially visible for “protein-poor” shells such as *Spondylus*. SP3 enabled us to
406 obtain good-quality spectra of the Ic fraction, whereas the same samples extracted by FASP
407 did not show any peptide markers at all. Therefore, SP3 outperforms FASP, a method used
408 in previous studies on molluscan shells and archaeological substrates.

409

410 Considering the results of all the tests conducted here, we conclude that the optimal method
411 for shell protein analyses is Ic_EDTA_SP3 (Table 1, method no.12), which consists of three
412 steps:

- 413 1) isolation of the intracrystalline protein fraction by bleaching the shell powder for 48 hours
414 using concentrated NaOCl (10-15%),
- 415 2) demineralisation of the shell powder using EDTA (0.5 M)
- 416 3) protein purification by single-pot, solid-phase sample preparation (SP3).

417 **3.2. The application of “palaeoshellomics”: shell PMFs**

418 The extraction approach Ic_EDTA_SP3 was tested on a set of different shells, in order to
419 validate the method. The set included a marine bivalve shell (the scallop *Pecten maximus*),
420 two gastropods (*Patella vulgata* and *Phorcus turbinatus*) and an archaeological freshwater
421 mussel, *Pseudunio auricularius*. Protein extraction was successful for all shells, including the
422 archaeological *Pseudunio*. Figure 4 shows the PMFs obtained; as noted for *Spondylus*, *Ostrea*
423 and *Unio* (Figure 2), most of the marker peptides were observed in the 1000-2000 *m/z* range
424 (Figure 4a-d, markers represented by asterisks).

425

426 **APPROXIMATE LOCATION OF FIGURE 4**

427

428 Table 2 summarises the peaks that were found to be taxon-specific in this pilot study, i.e. did
429 not pertain to any of the identified laboratory contaminants (see section 2.4 for more details)
430 and did not occur in any of the other species tested (except for *Unio* and *Pseudunio*, which
431 are phylogenetically close and have similar proteomes, see discussion below). Excluding *m/z*
432 values which may represent genuine shell peptides but which recur in different taxa is a
433 cautious approach, but in the absence of sequence information we are unable to evaluate if
434 these *m/z* values represent identical peptides or different peptides with the same mass, and

435 therefore assess their phylogenetic significance. We hope to revise this information in the
436 future. Nonetheless, the unique peptides were sufficient to discriminate between taxa. The
437 two gastropod shells, *Patella* and *Phorcus*, yielded very distinctive PMFs, with 24 and 18
438 markers identified respectively. Among the marine bivalves, 6 markers were identified for
439 *Pecten*, 13 markers for *Spondylus* and 15 markers for *Ostrea*. The freshwater mother-of-pearl
440 mussel *Unio* yielded 10 individual markers and 14 peptide markers were identified from the
441 archaeological *Pseudunio* (Table 2). Overall, there is noticeable variation in the number of
442 markers identified per taxon; this may imply that *Pecten* (6 markers) may be more difficult to
443 identify than *Patella* (24 markers) in the archaeological record, as diagenesis is expected to
444 cause the disappearance of some of these markers over time. We are currently conducting
445 artificial diagenesis experiments on *Spondylus* intracrystalline proteins and preliminary data
446 show the persistence of eight (out of thirteen) peptide markers after 96 hours continuous
447 heating at 80 °C and of three after 4800 hours (Sakalauskaite et al., unpublished data).

448
449 Some interesting observations can be made with regard to the similarities (or lack thereof) of
450 species that are phylogenetically related. For example, the remarkable difference between
451 *Spondylus* and *Pecten* PMFs supports a recent study showing that spondylids may have
452 followed a distinct evolutionary pathway from the other pectinoid molluscs [45]. Furthermore,
453 we identified one marker peak (m/z 1570.8), that likely corresponds to a peptide shared by the
454 two Unionida shells (freshwater bivalves) *Unio pictorum* (family Unionidae) and *Pseudunio*
455 *auricularius* (family Margaritiferidae). We suggest that the peptide at m/z 1570.8 (Table 2)
456 belongs to protein Hic74 [62], which was found to be the dominant protein in unionoid shells
457 [16]. The peak can be assigned to peptide sequence EAD(-18.01)DLALLSLLFGGR and it was
458 previously identified by LC-MS/MS analyses.

459
460 In summary, distinct PMFs can be obtained for intracrystalline shell proteins for different taxa
461 (Figure 4). Our suggested extraction approach was effective on 20 mg bleached shell samples
462 of both bivalves (freshwater and marine) and gastropods, regardless of their microstructure.
463 The method was also successful in extracting and characterising proteins from an
464 archaeological sample of *Pseudunio auricularius*.

465 We highlight that:

- 466 1) shells with the same microstructure, e.g. nacre (*Unio/Pseudunio*), foliated
467 (*Pecten/Ostrea*), crossed lamellar (*Patella/Spondylus*), yielded distinct PMFs;
- 468 2) species that belong to the same order (e.g. Pectinida: *Pecten/Spondylus*; Unionida:
469 *Unio/Pseudunio*) display different PMFs (with the exception of one marker peak that is
470 likely shared by *Pseudunio* and *Unio*).

471 472 **APPROXIMATE LOCATION OF TABLE 2**

473 **4. Conclusions**

474 In this work we find that:

- 475 • The most suitable method for mollusc shell protein extraction from small-size samples
476 includes a strong bleaching step (12% NaOCl for 48 hrs), followed by EDTA
477 demineralisation and SP3 extraction. This is the first application of the SP3 method for
478 “shellomic” studies.

- 479
- Using this method, the intracrystalline shell proteins can be successfully extracted and analysed by MALDI-TOF-MS and they yield unique PMFs, which enable us to discriminate between different shell taxa;
- 481
- The optimised method was employed to study several modern shells and one archaeological specimen, showing the effectiveness of this approach, regardless of species, microstructure or age of the samples.
- 482
- 483
- 484
- 485

486 Overall, this preliminary work strongly indicates that different molecular barcodes based on
487 PMFs of intracrystalline shell proteins can be obtained from small-size samples and used for
488 taxonomic identification of shells. Importantly, the method was found to be effective on a sub-
489 fossil shell, suggesting excellent potential for archaeological applications. We also highlight
490 current challenges facing “palaeoshellomics”. First of all, many shell species, including those
491 that were widely exploited in the past, lack reference sequences at genomic or transcriptomic
492 level. In the future we will build a larger reference dataset of molluscan shell PMFs and test
493 the intra-specific variability by analysing a higher number of specimens per taxon. We also
494 hope to link PMFs to sequence data and thus be able to test patterns of phylogenetic
495 relatedness more rigorously. The second challenge concerns our poor understanding of
496 peptide bond stability over archaeological timescales; however, artificial diagenesis
497 experiments show that intracrystalline shell proteins yield identifiable PMFs even after
498 prolonged heating. Moreover, previous studies on Neolithic shell ornaments had already
499 demonstrated excellent protein sequence recovery [16].

500

501 Fast and reliable molecular identification of shells from archaeological sites could represent
502 an important contribution to archaeological, palaeoenvironmental and geoarchaeological
503 research. Given the challenges above, we are focusing our investigation on taxa that are
504 especially relevant for the study of the past, such as molluscs exploited as a food resource
505 (e.g. oysters, mussels) or as raw materials for making tools and ornaments (e.g. pearl
506 mussels, *Spondylus*, *Glycymeris*). As an example, in this study we report that *Spondylus*
507 displays a set of unique markers. This is archaeologically significant, because *Spondylus* was
508 one of the most important and prestigious shells in prehistory, with numerous archaeological
509 finds from both Neolithic Europe and pre-Columbian South America [65–68]. However, the
510 majority of presumed *Spondylus* ornaments are poorly preserved and morphologically
511 undiagnostic, therefore our work will allow archaeologists to gain a deeper insight into the
512 circulating “shell economy” of prehistoric times.

513

514 Acknowledgements

515 The authors would like to thank Søren H. Andersen (Moesgaard Museum), Elisabetta Starnini
516 (University of Pisa) and Alberto Girod (Italian Malacological Society) for originally providing
517 the *Ostrea* and *Pseudunio* samples used in this study, as well as Emmanuel Fara (University
518 of Burgundy-Franche-Comté) and Matthew Collins (Universities of Copenhagen and
519 Cambridge) for valuable insights and discussions. BD is grateful to Kirsty Penkman, Jane
520 Thomas-Oates and Julie Wilson (University of York) for support.

521

522 The authors are grateful to two anonymous reviewers and to Frido Welker for their insightful
523 and valuable comments, which have improved the manuscript.

524

525 JS, FM and BD are supported by the PHC Galilée programme, Italo-French University
526 (UIF/UFI) (project G18-464/39612SB) and JS acknowledges the support of the Campus
527 France fund obtained through the program “Eiffel”. BD is funded by the “Giovani Ricercatori -
528 Rita Levi Montalcini” Programme (MIUR; Ministero dell'Istruzione dell'Università e della
529 Ricerca).

530 References

- 531 [1] F. d'Errico, C. Henshilwood, M. Vanhaeren, K. van Niekerk, *Nassarius kraussianus* shell
532 beads from Blombos Cave: evidence for symbolic behaviour in the Middle Stone Age, *J.*
533 *Hum. Evol.* 48 (2005) 3–24.
- 534 [2] M. Vanhaeren, F. d'Errico, C. Stringer, S.L. James, J.A. Todd, H.K. Mienis, Middle
535 Paleolithic shell beads in Israel and Algeria, *Science*. 312 (2006) 1785–1788.
- 536 [3] A. Bouzouggar, N. Barton, M. Vanhaeren, F. d'Errico, S. Collcutt, T. Higham, E. Hodge,
537 S. Parfitt, E. Rhodes, J.-L. Schwenninger, Others, 82,000-year-old shell beads from North
538 Africa and implications for the origins of modern human behavior, *Proceedings of the*
539 *National Academy of Sciences*. 104 (2007) 9964–9969.
- 540 [4] D.L. Hoffmann, D.E. Angelucci, V. Villaverde, J. Zapata, J. Zilhão, Symbolic use of marine
541 shells and mineral pigments by Iberian Neandertals 115,000 years ago, *Sci Adv.* 4 (2018)
542 eaar5255.
- 543 [5] H.C. Beck, Classification and Nomenclature of Beads and Pendants, *Archaeologia*. 77
544 (1928) 1–76.
- 545 [6] E.L. Baysal, *Personal Ornaments in Prehistory: An exploration of body augmentation from*
546 *the Palaeolithic to the Early Bronze Age*, Oxbow Books, 2019.
- 547 [7] S. Rigaud, F. d'Errico, M. Vanhaeren, Ornaments reveal resistance of North European
548 cultures to the spread of farming, *PLoS One*. 10 (2015) e0121166.
- 549 [8] S. Rigaud, C. Manen, I. García-Martínez de Lagrán, Symbols in motion: Flexible cultural
550 boundaries and the fast spread of the Neolithic in the western Mediterranean, *PLoS One*.
551 13 (2018) e0196488.
- 552 [9] A.C. Paulsen, The Thorny Oyster and the Voice of God: *Spondylus* and *Strombus* in
553 Andean Prehistory, *Am. Antiq.* 39 (1974) 597–607.
- 554 [10] M.B.D. Trubitt, The Production and Exchange of Marine Shell Prestige Goods, *Journal of*
555 *Archaeological Research*. 11 (2003) 243–277.
- 556 [11] D.E. Bar-Yosef Mayer, Shell ornaments and artifacts in Neolithic Cyprus and correlations
557 with other Mediterranean regions, *Quat. Int.* 464 (2018) 206–215.
- 558 [12] S.L. Kuhn, M.C. Stiner, D.S. Reese, E. Güleç, Ornaments of the earliest Upper Paleolithic:
559 new insights from the Levant, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 7641–7646.
- 560 [13] M. Vanhaeren, F. d'Errico, Aurignacian ethno-linguistic geography of Europe revealed by
561 personal ornaments, *J. Archaeol. Sci.* 33 (2006) 1105–1128.
- 562 [14] B. Demarchi, S. O'Connor, A. de Lima Ponzoni, R. de Almeida Rocha Ponzoni, A.
563 Sheridan, K. Penkman, Y. Hancock, J. Wilson, An integrated approach to the taxonomic
564 identification of prehistoric shell ornaments, *PLoS One*. 9 (2014) e99839.
- 565 [15] *Spondylus gaederopus*: applicazione sperimentale dell'analisi microCT per la
566 determinazione del genere, in: E. Borgna, P.C. Guida, S. Corazza (Eds.), *Preistoria E*
567 *Protostoria Del “Caput Adriae,”* Istituto italiano di preistoria e protostoria, Firenze, 2018:
568 pp. 265–278.
- 569 [16] J. Sakalauskaite, S.H. Andersen, P. Biagi, M.A. Borrello, T. Cocquerez, A.C. Colonese,
570 F. Dal Bello, A. Girod, M. Heumüller, H. Koon, G. Mandili, C. Medana, K.E. Penkman, L.

- 571 Plasseraud, H. Schlichtherle, S. Taylor, C. Tokarski, J. Thomas, J. Wilson, F. Marin, B.
572 Demarchi, "Palaeoshellomics" reveals the use of freshwater mother-of-pearl in prehistory,
573 *Elife*. 8 (2019). <https://doi.org/10.7554/eLife.45644>.
- [17] J.G. Carter, ed., *Skeletal Biomineralization: Pattern, Processes, and Evolutionary Trends*,
574 Amer Geophysical Union, 1990.
575
- [18] M. Buckley, M.J. Collins, Collagen survival and its use for species identification in
576 Holocene-lower Pleistocene bone fragments from British archaeological and
577 paleontological sites, *Antiqua*. 1 (2011) 1.
578
- [19] K.K. Richter, J. Wilson, A.K.G. Jones, M. Buckley, N. van Doorn, M.J. Collins, Fish'n
579 chips: ZooMS peptide mass fingerprinting in a 96 well plate format to identify fish bone
580 fragments, *J. Archaeol. Sci.* 38 (2011) 1502–1510.
581
- [20] C. Solazzo, M. Wadsley, J.M. Dyer, S. Clerens, M.J. Collins, J. Plowman,
582 Characterisation of novel α -keratin peptide markers for species identification in keratinous
583 tissues using mass spectrometry, *Rapid Commun. Mass Spectrom.* 27 (2013) 2685–
584 2698.
585
- [21] F. Welker, M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, S.T. Turvey,
586 M. Reguero, J.N. Gelfo, A. Kramarz, J. Burger, J. Thomas-Oates, D.A. Ashford, P.D.
587 Ashton, K. Rowsell, D.M. Porter, B. Kessler, R. Fischer, C. Baessmann, S. Kaspar, J.V.
588 Olsen, P. Kiley, J.A. Elliott, C.D. Kelstrup, V. Mullin, M. Hofreiter, E. Willerslev, J.-J.
589 Hublin, L. Orlando, I. Barnes, R.D.E. MacPhee, Ancient proteins resolve the evolutionary
590 history of Darwin's South American ungulates, *Nature*. 522 (2015) 81–84.
591
- [22] C. Solazzo, W. Fitzhugh, S. Kaplan, C. Potter, J.M. Dyer, Molecular markers in keratins
592 from Mysticeti whales for species identification of baleen in museum and archaeological
593 collections, *PLoS One*. 12 (2017) e0183053.
594
- [23] R. Sawafuji, E. Cappellini, T. Nagaoka, A.K. Fotakis, R.R. Jersie-Christensen, J.V. Olsen,
595 K. Hirata, S. Ueda, Proteomic profiling of archaeological human bone, *R Soc Open Sci.*
596 4 (2017) 161004.
597
- [24] F. Welker, Palaeoproteomics for human evolution studies, *Quat. Sci. Rev.* 190 (2018)
598 137–147.
599
- [25] B. Demarchi, S. Hall, T. Roncal-Herrero, C.L. Freeman, J. Woolley, M.K. Crisp, J. Wilson,
600 A. Fotakis, R. Fischer, B.M. Kessler, R. Rakownikow Jersie-Christensen, J.V. Olsen, J.
601 Haile, J. Thomas, C.W. Marean, J. Parkington, S. Presslee, J. Lee-Thorp, P. Ditchfield,
602 J.F. Hamilton, M.W. Ward, C.M. Wang, M.D. Shaw, T. Harrison, M. Domínguez-Rodrigo,
603 R.D.E. MacPhee, A. Kwekason, M. Ecker, L. Kolska Horwitz, M. Chazan, R. Kröger, J.
604 Thomas-Oates, J.H. Harding, E. Cappellini, K. Penkman, M.J. Collins, Protein sequences
605 bound to mineral surfaces persist into deep time, *Elife*. 5 (2016).
606 <https://doi.org/10.7554/eLife.17092>.
607
- [26] S. Presslee, J. Wilson, J. Woolley, J. Best, D. Russell, A. Radini, R. Fischer, B. Kessler,
608 R. Boano, M. Collins, B. Demarchi, The identification of archaeological eggshell using
609 peptide markers, *STAR: Science & Technology of Archaeological Research*. 3 (2017) 89–
610 99.
611
- [27] M. Mackie, J. Hendy, A.D. Lowe, A. Sperduti, M. Holst, M.J. Collins, C.F. Speller,
612 Preservation of the metaproteome: variability of protein preservation in ancient dental
613 calculus, *Sci Technol Archaeol Res.* 3 (2017) 74–86.
614
- [28] E. Cappellini, F. Welker, L. Pandolfi, J. Ramos-Madrigal, D. Samodova, P.L. Rüter, A.K.
615 Fotakis, D. Lyon, J. Víctor Moreno-Mayar, M. Bukhsianidze, R.R. Jersie-Christensen, M.
616 Mackie, A. Ginolhac, R. Ferring, M. Tappen, E. Palkopoulou, M.R. Dickinson, T.W.
617 Stafford, Y.L. Chan, A. Götherström, Senthilvel K S, P.D. Heintzman, J.D. Kapp, I.
618

- 619 Kirillova, Y. Moodley, J. Agusti, R.-D. Kahlke, G. Kiladze, B. Martínez-Navarro, S. Liu,
620 M.S. Velasco, M.-H.S. Sinding, C.D. Kelstrup, M.E. Allentoft, L. Orlando, K. Penkman, B.
621 Shapiro, L. Rook, L. Dalén, M.T.P. Gilbert, J.V. Olsen, D. Lordkipanidze, E. Willerslev,
622 Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny,
623 Nature. (2019) 1–5.
- 624 [29] F. Lugli, G. Di Rocco, A. Vazzana, F. Genovese, D. Pinetti, E. Cilli, M.C. Carile, S.
625 Silvestrini, G. Gabanini, S. Arrighi, L. Buti, E. Bortolini, A. Cipriani, C. Figus, G. Marciani,
626 G. Oxilia, M. Romandini, R. Sorrentino, M. Sola, S. Benazzi, Enamel peptides reveal the
627 sex of the Late Antique “Lovers of Modena,” Sci. Rep. 9 (2019) 13130.
- 628 [30] M. Buckley, Zooarchaeology by Mass Spectrometry (ZooMS) Collagen Fingerprinting for
629 the Species Identification of Archaeological Bone Fragments, in: C.M. Giovas, M.J.
630 LeFebvre (Eds.), Zooarchaeology in Practice: Case Studies in Methodology and
631 Interpretation in Archaeofaunal Analysis, Springer International Publishing, Cham, 2018:
632 pp. 227–247.
- 633 [31] F. Marin, G. Luquet, Molluscan shell proteins, C. R. Palevol. 3 (2004) 469–492.
- 634 [32] F. Marin, G. Luquet, B. Marie, D. Medakovic, Molluscan Shell Proteins: Primary Structure,
635 Origin, and Evolution, in: Current Topics in Developmental Biology, Academic Press,
636 2007: pp. 209–276.
- 637 [33] D.J. Jackson, C. McDougall, B. Woodcroft, P. Moase, R.A. Rose, M. Kube, R. Reinhardt,
638 D.S. Rokhsar, C. Montagnani, C. Joubert, D. Piquemal, B.M. Degnan, Parallel evolution
639 of nacre building gene sets in molluscs, Mol. Biol. Evol. 27 (2010) 591–608.
- 640 [34] K.M. Kocot, F. Aguilera, C. McDougall, D.J. Jackson, B.M. Degnan, Sea shell diversity
641 and rapidly evolving secretomes: insights into the evolution of biomineralization, Front.
642 Zool. 13 (2016) 23.
- 643 [35] K.M. Towe, G.R. Thompson, The structure of some bivalve shell carbonates prepared by
644 ion-beam thinning, Calcif. Tissue Res. 10 (1972) 38–48.
- 645 [36] G.A. Sykes, M.J. Collins, D.I. Walton, The significance of a geochemically isolated
646 intracrystalline organic fraction within biominerals, Org. Geochem. 23 (1995) 1059–1065.
- 647 [37] K.E.H. Penkman, D.S. Kaufman, D. Maddy, M.J. Collins, Closed-system behaviour of the
648 intra-crystalline fraction of amino acids in mollusc shells, Quat. Geochronol. 3 (2008) 2–
649 25.
- 650 [38] K.E.H. Penkman, R.C. Preece, D.R. Bridgland, D.H. Keen, T. Meijer, S.A. Parfitt, T.S.
651 White, M.J. Collins, A chronological framework for the British Quaternary based on
652 *Bithynia opercula*, Nature. 476 (2011) 446–449.
- 653 [39] B. Demarchi, K. Rogers, D.A. Fa, C.J. Finlayson, N. Milner, K.E.H. Penkman, Intra-
654 crystalline protein diagenesis (IcPD) in *Patella vulgata*. Part I: Isolation and testing of the
655 closed system, Quat. Geochronol. 16 (2013) 144–157.
- 656 [40] J.E. Ortiz, Y. Sánchez-Palencia, I. Gutiérrez-Zugasti, T. Torres, M. González-Morales,
657 Protein diagenesis in archaeological gastropod shells and the suitability of this material
658 for amino acid racemisation dating: *Phorcus lineatus* (da Costa, 1778), Quat. Geochronol.
659 46 (2018) 16–27.
- 660 [41] F. Marin, N. Le Roy, B. Marie, The formation and mineralization of mollusk shell, Front.
661 Biosci. 4 (2012) 1099–1125.
- 662 [42] A. Osuna-Mascaró, T. Cruz-Bustos, S. Benhamada, N. Guichard, B. Marie, L.
663 Plasseraud, M. Corneillat, G. Alcaraz, A. Checa, F. Marin, The shell organic matrix of the
664 crossed lamellar queen conch shell (*Strombus gigas*), Comp. Biochem. Physiol. B
665 Biochem. Mol. Biol. 168 (2014) 76–85.
- 666 [43] O.B.A. Agbaje, D.E. Thomas, J.G. Dominguez, B.V. McInerney, M.A. Kosnik, D.E. Jacob,

- 667 Biomacromolecules in bivalve shells with crossed lamellar architecture, *J. Mater. Sci.* 54
668 (2019) 4952–4969.
- 669 [44] F. Pierini, B. Demarchi, J. Turner, K. Penkman, *Pecten* as a new substrate for IcPD dating:
670 The quaternary raised beaches in the Gulf of Corinth, Greece, *Quat. Geochronol.* 31
671 (2016) 40–52.
- 672 [45] J. Sakalauskaite, L. Plasseraud, J. Thomas, M. Albéric, M. Thoury, J. Perrin, F. Jamme,
673 C. Broussard, B. Demarchi, F. Marin, The shell matrix of the European thorny oyster,
674 *Spondylus gaederopus*: microstructural and molecular characterization, *Journal of*
675 *Structural Biology.* (2020).
- 676 [46] T. Jonuks, E. Oras, J. Best, B. Demarchi, R. Mänd, S. Presslee, S. Vahur, Multi-method
677 Analysis of Avian Eggs as Grave Goods: Revealing Symbolism in Conversion Period
678 Burials at Kukruse, NE Estonia, *Environ. Archaeol.* 23 (2018) 109–122.
- 679 [47] B.M. Sichert, P. Rentzel, B. Demarchi, J. Best, A. Negri, S. Deschler-Erb, Incubated eggs
680 in a Roman burial? A preliminary investigation on how to distinguish between the effects
681 of incubation and taphonomy on avian eggshell from archaeological sites, *Journal of*
682 *Archaeological Science: Reports.* 26 (2019) 101845.
- 683 [48] F. Marin, B. Marie, S.B. Hamada, P. Ramos-Silva, N. Le Roy, N. Guichard, S.E. Wolf, C.
684 Montagnani, C. Joubert, D. Piquemal, D. Saulnier, Y. Gueguen, “Shellome”: Proteins
685 involved in mollusk shell biomineralization-diversity, functions, in: S. Watabe, K.
686 Maeyama, H. Nagasawa (Eds.), Terrapub Tokyo, 2013: p. 149:166.
- 687 [49] B. Marie, D.J. Jackson, P. Ramos-Silva, I. Zanella-Cléon, N. Guichard, F. Marin, The
688 shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-
689 specific novelties, *FEBS J.* 280 (2013) 214–232.
- 690 [50] F. Marin, I. Bundeleava, T. Takeuchi, F. Immel, D. Medakovic, Organic matrices in
691 metazoan calcium carbonate skeletons: Composition, functions, evolution, *J. Struct. Biol.*
692 196 (2016) 98–106.
- 693 [51] K. Mann, E. Edsinger, The *Lottia gigantea* shell matrix proteome: re-analysis including
694 MaxQuant iBAQ quantitation and phosphoproteome analysis, *Proteome Sci.* 12 (2014)
695 28.
- 696 [52] A. Girod, Lavorazioni su conchiglia di *Pseudunio auricularius* (Mollusca, Bivalvia) nel
697 Neolitico antico a Isorella (BS), Italia settentrionale, *Atti Della Società Italiana Di Scienze*
698 *Naturali E Del Museo Civico Di Storia Naturale Di Milano.* 151 (2010) 89–98.
- 699 [53] E. Starnini, P. Biagi, N. Mazzucco, The beginning of the Neolithic in the Po Plain (northern
700 Italy): Problems and perspectives, *Quat. Int.* 470 (2018) 301–317.
- 701 [54] M.D. Bosch, M.A. Mannino, A.L. Prendergast, T.C. O’Connell, B. Demarchi, S.M. Taylor,
702 L. Niven, J. van der Plicht, J.-J. Hublin, New chronology for Ksâr ’Akil (Lebanon) supports
703 Levantine route of modern human dispersal into Europe, *Proc. Natl. Acad. Sci. U. S. A.*
704 112 (2015) 7683–7688.
- 705 [55] Conchology, Inc, (n.d.). <https://www.conchology.be/?t=1> (accessed November 22, 2019).
- 706 [56] B. Marie, P. Ramos-Silva, F. Marin, A. Marie, Proteomics of CaCO₃ biomineral-
707 associated proteins: how to properly address their analysis, *Proteomics.* 13 (2013) 3109–
708 3116.
- 709 [57] B. Demarchi, *Amino Acids and Proteins in Fossil Biominerals: An Introduction for*
710 *Archaeologists and Palaeontologists*, John Wiley & Sons, 2020.
- 711 [58] C.S. Hughes, S. Moggridge, T. Müller, P.H. Sorensen, G.B. Morin, J. Krijgsveld, Single-
712 pot, solid-phase-enhanced sample preparation for proteomics experiments, *Nat. Protoc.*
713 14 (2019) 68–85.
- 714 [59] T.H.J. Niedermeyer, M. Strohal, mMass as a software tool for the annotation of cyclic

715 peptide tandem mass spectra, PLoS One. 7 (2012) e44913.
716 [60] K. Mann, E. Edsinger-Gonzales, M. Mann, In-depth proteomic analysis of a mollusc shell:
717 acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*, Proteome Sci. 10
718 (2012) 28.
719 [61] J.S. Evans, Aragonite-associated biomineralization proteins are disordered and contain
720 interactive motifs, Bioinformatics. 28 (2012) 3182–3185.
721 [62] M. Sielaff, J. Kuharev, T. Bohn, J. Hahlbrock, T. Bopp, S. Tenzer, U. Distler, Evaluation
722 of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram
723 Range, J. Proteome Res. 16 (2017) 4060–4072.
724 [63] T.P. Cleland, Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-
725 Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce
726 Humics, J. Proteome Res. 17 (2018) 3976–3983.
727 [64] X. Liu, C. Jin, L. Wu, S. Dong, S. Zeng, J. Li, Hic74, a novel alanine and glycine rich matrix
728 protein related to nacreous layer formation in the mollusc *Hyriopsis cumingii*, Aquaculture
729 and Fisheries. 2 (2017) 119–123.
730 [65] M.A. Borrello, R. Micheli, *Spondylus gaederopus* in prehistoric Italy: jewels from neolithic
731 and copper age sites, Spondylus in Prehistory. New Data and Approaches. Contributions
732 to the Archaeology of Shell Technologies. (2011) 25–37.
733 [66] B.P. Carter, P. Benjamin, *Spondylus* in South American Prehistory, in: Spondylus in
734 Prehistory: New Data and Approaches - Contributions to the Archaeology of Shell
735 Technologies, British Archaeological Reports - J & E Hedges Ltd, 2011.
736 [67] J. Chapman, B. Gaydarska, *Spondylus Gaederopus/Glycymeris* Exchange Networks in
737 the European Neolithic and Chalcolithic, in: C. Fowler, J. Harding, D. Hofmann (Eds.),
738 The Oxford Handbook of Neolithic Europe, Oxford University Press, 2015.
739 [68] A. Windler, The Use of *Spondylus gaederopus* during the Neolithic of Europe, Journal of
740 Open Archaeology Data. 7 (2019) 19.
741
742
743
744

745 TABLES

746 Table 1. List of shell protein extraction treatments tested in this study. FASP - filter aided sample
 747 preparation; SP3 - Solid-phase sample preparation; for demineralisation, 10% cold acetic acid
 748 (AcOH) and 0.5 M EDTA solutions were tested. "Mild" and "intermediate" bleaching steps (4 and
 749 24 hrs) were carried out using diluted NaOCl (1.0-1.5%), while the "strong" bleaching step (for a
 750 duration of 48 hrs) was carried out to isolate intracrystalline fraction (Ic) using concentrated
 751 NaOCl (10-15%).

752

No.	Method annotation	Bleaching (hrs)	Demineralisation	Protein purification
1	4hrs_AcOH_FASP	4	AcOH (10%)	FASP
2	24hrs_AcOH_FASP	24		
3	Ic_AcOH_FASP	48 (Ic)		
4	4hrs_EDTA_FASP	4	EDTA (0.5 M)	
5	24hrs_EDTA_FASP	24		
6	Ic_EDTA_FASP	48 (Ic)		
7	4hrs_AcOH_SP3	4	AcOH (10%)	SP3
8	24hrs_AcOH_SP3	24		
9	Ic_AcOH_SP3	48 (Ic)		
10	4hrs_EDTA_SP3	4	EDTA (0.5 M)	
11	24hrs_EDTA_SP3	24		
12	Ic_EDTA_SP3	48 (Ic)		

753

754 Table 2. Peptide markers (*m/z* values) for *Unio pictorum*, *Spondylus gaederopus*, *Ostrea edulis*,
 755 *Phorcus turbinatus*, *Patella vulgata*, *Pecten maximus* and *Pseudunio auricularius*
 756 (intracrystalline protein fraction). Values in bold indicate shared markers.

	Shells						
Age	Modern						Archaeological
Dominant microstructure	Foliated		Crossed-lamellar		Nacreous		
Mineralogy	Calcitic		Mostly aragonitic (thin upper layer - calcitic)	Calcitic and Aragonitic	Mostly aragonitic (thin calcitic prisms)	Aragonitic	
Taxonomy	Bivalvia, Ostreida, Ostreidae	Bivalvia, Pectinida, Pectinidae	Bivalvia, Pectinida, Spondylida	Gastropoda, Patellidae	Gastropoda, Trochida, Trochidae	Bivalvia, Unionida, Unionidae	Bivalvia, Unionida, Margaritiferidae
Species	<i>Ostrea edulis</i>	<i>Pecten maximus</i>	<i>Spondylus gaederopus</i>	<i>Patella vulgata</i>	<i>Phorcus turbinatus</i>	<i>Unio pictorum</i>	<i>Pseudunio auricularius</i>
Marker <i>m/z</i> values	1087.9	1095.5	1146.6	1001.5	1023.5	1049.5	1111.6
	1095.0	1134.6	1160.6	1096.6	1029.6	1080.6	1119.6

1109.2	1437.7	1258.7	1135.6	1053.5	1085.6	1164.7
1166.5	1681.8	1275.7	1192.6	1070.5	1113.4	1279.7
1168.5	2060.9	1279.6	1252.7	1123.6	1130.5	1300.7
1172.7	2100.0	1304.6	1268.8	1231.7	1154.5	1327.7
1182.7		1327.7	1290.8	1247.8	1268.5	1355.7
1281.4		1411.7	1332.8	1285.6	1570.8	1542.7
1311.5		1415.7	1353.8	1450.8	1764.8	1570.8
1387.9		1432.7	1361.8	1458.9	1805.8	1571.8
1480.1		1751.8	1445.8	1511.9		1699.8
1711.8		1823.9	1451.7	1552.8		1806.8
1770.6		1951.9	1472.8	1691.7		1892.9
1798.5			1584.9	1727.8		1975.9
1996.3			1585.9	1815.9		
			1601.8	1824.0		
			1783.9	1833.9		
			1799.9	1868.9		
			1874.9			
			1921.9			
			1941.9			
			1972.0			
			2094.0			
			2116.0			

757

758

759 **FIGURE LEGENDS**

760

761 Figure 1. Scheme showing the different approaches tested for shell protein extraction. “Mild”
762 and “Intermediate” bleaching steps (4 and 24 hrs) were performed using diluted ~1.0-1.5%
763 NaOCl; “strong” 48-hr bleaching was performed using concentrated ~10-15% NaOCl, which
764 isolates the “operational” intracrystalline fraction (Ic). Demineralisation was achieved using
765 10% acetic acid (AcOH) or EDTA (0.5M). Two different protein purification methods were
766 evaluated and compared: filter aided sample preparation (FASP) vs single-pot, solid-phase
767 sample preparation (SP3).

768

769 Figure 2. *Unio pictorum*, *Spondylus gaederopus* and *Ostrea edulis* (inter+intra)crystalline vs
770 intracrystalline (Ic) peptide mass fingerprints (PMFs) (a, c, e) and bulk amino acid (AA)
771 compositions corresponding to the Ic fraction (b, d, f) obtained from previously published work
772 [14]. Intracrystalline proteins (Ic) isolated by “strong bleaching” are shown in red and
773 (inter+intra)crystalline proteins, obtained via “mild” bleaching, are shown in blue. Asterisks
774 indicate the marker peptides for these shells.

775 Figure 3. Peptide mass fingerprints (PMFs) of the intracrystalline shell protein fraction
776 extracted from *Spondylus gaederopus*, *Unio pictorum* and *Ostrea edulis*. Spectra a-c show
777 PMFs of a) *Unio*, b) *Spondylus* and c) *Ostrea*, obtained by single-pot, solid-phase sample
778 preparation (SP3, in red) or filter aided sample preparation (FASP, in green). *Spondylus*
779 spectra in d) compare the demineralisation with EDTA (red) and AcOH (dark green). Asterisks
780 indicate the marker peptides identified for these shells.

781

782 Figure 4. Intracrystalline PMFs of the four different shell species that were used to validate the
783 protein extraction method (Ic_EDTA_SP3): a) *Phorcus turbinatus* (modern), b) *Patella vulgata*
784 (modern), c) *Pecten maximus* (modern), *Pseudunio auricularius* (Neolithic, 5226–5023 cal
785 BCE). Asterisks indicate the marker peptides identified for these shells.

786