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

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

List of authors:

Jorune Sakalauskaite[1,2], Frédéric Marin[2], Barbara Pergolizzi[3], Beatrice Demarchi[1]

Affiliations

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

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

[3]Department of Clinical and Biological Sciences, University of Turin, AOU S. Luigi, 10043 Orbassano TO, Italy.

Corresponding Authors:

Jorune Sakalauskaite, jorune.sakalauskaite@unito.it
Address 1: Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Turin, Italy;
Address 2: Biogeosciences, UMR CNRS 6282, University of Burgundy-Franche-Comté, 6 Boulevard Gabriel, 21000 Dijon, France.

Beatrice Demarchi, beatrice.demarchi@unito.it
Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Turin, Italy.

Keywords

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

Abstract

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

APPROXIMATE LOCATION OF GRAPHICAL ABSTRACT

1. Introduction

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

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

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

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

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

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

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

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

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

2. Material and Methods

2.1. Samples

2.1.1. Method development

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

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

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

2.1.2 Method application

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

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

2.2 Analytical procedure: Method development

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

2.2.1 Bleaching

Bleaching is a vigorous cleaning approach which involves the use of sodium hypochlorite (NaOCl) in order to remove surface contamination and/or weakly bound intercrystalline organics from targeted samples. It is used routinely to treat biomineralised tissues and organisms, such as eggshell, mollusc shell or coral, before ancient protein analysis. Developed originally for amino acid racemization geochronology [36,37] it is also used in shell proteomics in order to reduce the pool of analysed proteins and isolate those that are truly associated to the mineral phase [56].

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

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

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

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

2.2.2 Demineralisation

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

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

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

2.2.3 Protein purification and processing

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

FASP extraction

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

SP3 extraction

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

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

APPROXIMATE LOCATION OF TABLE 1

2.3 MALDI-MS Analysis

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

2.4 PMF library preparation

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

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

2.5 Method application

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

3. Results and Discussion

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

3.1 Method Development

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

APPROXIMATE LOCATION OF FIGURE 1

3.1.1 Intracrystalline shell proteins and effect of bleaching

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

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

The results showed that intracrystalline PMFs can be obtained from 20 mg shell samples, regardless of their different microstructures (nacreous, crossed-lamellar, foliated) and their variable organic content: even the most organic-poor microstructure (i.e. crossed-lamellar in *Spondylus*) retained a sufficient fraction of Ic proteins.
Furthermore, we did not observe any simple correlation between bleaching time/NaOCl concentration and the number of potential marker peptides - i.e. shorter bleaching times do not imply better MALDI-TOF spectra and, *vice versa*, harsh bleaching treatments do not necessarily mean that protein concentrations will be too low for proteomics. This is interesting as many “shellomics” studies encourage bleaching as a cleaning pretreatment [56], but generally avoid higher concentrations of NaOCl and longer exposure times, presuming that shell proteins would be fully hydrolyzed. In the case of *Unio*, no compelling difference was observed between the spectra of the (inter+intra)crystalline fraction (“mild” bleaching) and the intracrystalline (Ic) fraction (“strong” bleaching). Remarkably, for *Spondylus*, the intensity and number of potential marker peaks is considerably higher in the intracrystalline fraction compared to the spectra obtained after just 4 hours of bleaching, for which the PMFs were barely detectable. This effect has also been observed in other shells [60], including a study of the *Spondylus* proteome by tandem mass spectrometry [45], and may be due to the difficulty of breaking down complex networks of proteins with other shell matrix macromolecules, such as chitin. It is likely that the presence of glycosylated proteins, lipoproteins, phospholipids, or proteins with repetitive low complexity domains (RLCD), could influence signal detection (or suppress it completely) [51,61]. This would explain why a strong oxidative treatment, which removes a large quantity of these macromolecules, may be advantageous in shell protein analyses. In addition, the intracrystalline proteins have more acidic domains, which bind to the mineral [25], and thus are preferentially ionised, therefore their detection is favoured when analysed by MALDI-TOF mass spectrometry.

APPROXIMATE LOCATION OF FIGURE 2

3.1.2 Extraction and purification

The SP3 method for shell protein isolation and purification was found to be more effective than FASP. The intracrystalline PMFs obtained by SP3 were of better quality and displayed a higher number of marker peaks for *Unio*, *Ostrea* and *Spondylus* (Figure 3a-c, spectra in red). On the contrary, in the FASP PMFs, the relative proportion between marker peaks and the trypsin/keratin peaks (common laboratory contaminants) was severely skewed towards the latter (Figure 3a-c, green spectra). This is probably due to the fact that the (minimal) loss of proteins which occurs during ultrafiltration is especially noticeable for protein-poor samples, with enzymes and common contaminants thus being over-represented in the resulting spectrum. The SP3 extraction is therefore better suited to shell samples [62].
In general, the issue of protein concentration will principally affect MALDI-TOF analyses of proteins from crossed-lamellar and foliated microstructures; for example, a 20-mg *Spondylus* sample may contain as low as ~200 ng of intracrystalline proteins. For nacroprismatic shells,

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

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

APPROXIMATE LOCATION OF FIGURE 3

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

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

1) isolation of the intracrystalline protein fraction by bleaching the shell powder for 48 hours using concentrated NaOCl (10-15%),

2) demineralisation of the shell powder using EDTA (0.5 M)

3) protein purification by single-pot, solid-phase sample preparation (SP3).

3.2. The application of "palaeoshellomics": shell PMFs

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

APPROXIMATE LOCATION OF FIGURE 4

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

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

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

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

We highlight that:

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

APPROXIMATE LOCATION OF TABLE 2

4. Conclusions

In this work we find that:

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

- 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;
- 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.

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

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

Acknowledgements

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

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

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

References

[1] F. d’Errico, C. Henshilwood, M. Vanhaeren, K. van Niekerk, *Nassarius kraussianus* shell beads from Blombos Cave: evidence for symbolic behaviour in the Middle Stone Age, J. Hum. Evol. 48 (2005) 3–24.

[2] M. Vanhaeren, F. d’Errico, C. Stringer, S.L. James, J.A. Todd, H.K. Mienis, Middle Paleolithic shell beads in Israel and Algeria, Science. 312 (2006) 1785–1788.

[3] A. Bouzouggar, N. Barton, M. Vanhaeren, F. d’Errico, S. Collcutt, T. Higham, E. Hodge, S. Parfitt, E. Rhodes, J.-L. Schwenninger, Others, 82,000-year-old shell beads from North Africa and implications for the origins of modern human behavior, Proceedings of the National Academy of Sciences. 104 (2007) 9964–9969.

[4] D.L. Hoffmann, D.E. Angelucci, V. Villaverde, J. Zapata, J. Zilhão, Symbolic use of marine shells and mineral pigments by Iberian Neandertals 115,000 years ago, Sci Adv. 4 (2018) eaar5255.

[5] H.C. Beck, Classification and Nomenclature of Beads and Pendants, Archaeologia. 77 (1928) 1–76.

[6] E.L. Baysal, Personal Ornaments in Prehistory: An exploration of body augmentation from the Palaeolithic to the Early Bronze Age, Oxbow Books, 2019.

[7] S. Rigaud, F. d’Errico, M. Vanhaeren, Ornaments reveal resistance of North European cultures to the spread of farming, PLoS One. 10 (2015) e0121166.

[8] S. Rigaud, C. Manen, I. García-Martínez de Lagrán, Symbols in motion: Flexible cultural boundaries and the fast spread of the Neolithic in the western Mediterranean, PLoS One. 13 (2018) e0196488.

[9] A.C. Paulsen, The Thorny Oyster and the Voice of God: *Spondylus* and *Strombus* in Andean Prehistory, Am. Antiq. 39 (1974) 597–607.

[10] M.B.D. Trubitt, The Production and Exchange of Marine Shell Prestige Goods, Journal of Archaeological Research. 11 (2003) 243–277.

[11] D.E. Bar-Yosef Mayer, Shell ornaments and artifacts in Neolithic Cyprus and correlations with other Mediterranean regions, Quat. Int. 464 (2018) 206–215.

[12] S.L. Kuhn, M.C. Stiner, D.S. Reese, E. Güleç, Ornaments of the earliest Upper Paleolithic: new insights from the Levant, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 7641–7646.

[13] M. Vanhaeren, F. d’Errico, Aurignacian ethno-linguistic geography of Europe revealed by personal ornaments, J. Archaeol. Sci. 33 (2006) 1105–1128.

[14] B. Demarchi, S. O’Connor, A. de Lima Ponzoni, R. de Almeida Rocha Ponzoni, A. Sheridan, K. Penkman, Y. Hancock, J. Wilson, An integrated approach to the taxonomic identification of prehistoric shell ornaments, PLoS One. 9 (2014) e99839.

[15] *Spondylus gaederopus*: applicazione sperimentale dell’analisi microCT per la determinazione del genere, in: E. Borgna, P.C. Guida, S. Corazza (Eds.), Preistoria E Protostoria Del “Caput Adriae,” Istituto italiano di preistoria e protostoria, Firenze, 2018: pp. 265–278.

[16] J. Sakalauskaite, S.H. Andersen, P. Biagi, M.A. Borrello, T. Cocquerez, A.C. Colonese, F. Dal Bello, A. Girod, M. Heumüller, H. Koon, G. Mandili, C. Medana, K.E. Penkman, L.

Plasseraud, H. Schlichtherle, S. Taylor, C. Tokarski, J. Thomas, J. Wilson, F. Marin, B. Demarchi, “Palaeoshellomics” reveals the use of freshwater mother-of-pearl in prehistory, Elife. 8 (2019). https://doi.org/10.7554/eLife.45644.

[17] J.G. Carter, ed., Skeletal Biomineralization: Pattern, Processes, and Evolutionary Trends, Amer Geophysical Union, 1990.

[18] M. Buckley, M.J. Collins, Collagen survival and its use for species identification in Holocene-lower Pleistocene bone fragments from British archaeological and paleontological sites, Antiqua. 1 (2011) 1.

[19] K.K. Richter, J. Wilson, A.K.G. Jones, M. Buckley, N. van Doorn, M.J. Collins, Fish’n chips: ZooMS peptide mass fingerprinting in a 96 well plate format to identify fish bone fragments, J. Archaeol. Sci. 38 (2011) 1502–1510.

[20] C. Solazzo, M. Wadsley, J.M. Dyer, S. Clerens, M.J. Collins, J. Plowman, Characterisation of novel α-keratin peptide markers for species identification in keratinous tissues using mass spectrometry, Rapid Commun. Mass Spectrom. 27 (2013) 2685–2698.

[21] F. Welker, M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, S.T. Turvey, M. Reguero, J.N. Gelfo, A. Kramarz, J. Burger, J. Thomas-Oates, D.A. Ashford, P.D. Ashton, K. Rowsell, D.M. Porter, B. Kessler, R. Fischer, C. Baessmann, S. Kaspar, J.V. Olsen, P. Kiley, J.A. Elliott, C.D. Kelstrup, V. Mullin, M. Hofreiter, E. Willerslev, J.-J. Hublin, L. Orlando, I. Barnes, R.D.E. MacPhee, Ancient proteins resolve the evolutionary history of Darwin’s South American ungulates, Nature. 522 (2015) 81–84.

[22] C. Solazzo, W. Fitzhugh, S. Kaplan, C. Potter, J.M. Dyer, Molecular markers in keratins from Mysticeti whales for species identification of baleen in museum and archaeological collections, PLoS One. 12 (2017) e0183053.

[23] R. Sawafuji, E. Cappellini, T. Nagaoka, A.K. Fotakis, R.R. Jersie-Christensen, J.V. Olsen, K. Hirata, S. Ueda, Proteomic profiling of archaeological human bone, R Soc Open Sci. 4 (2017) 161004.

[24] F. Welker, Palaeoproteomics for human evolution studies, Quat. Sci. Rev. 190 (2018) 137–147.

[25] B. Demarchi, S. Hall, T. Roncal-Herrero, C.L. Freeman, J. Woolley, M.K. Crisp, J. Wilson, A. Fotakis, R. Fischer, B.M. Kessler, R. Rakownikow Jersie-Christensen, J.V. Olsen, J. Haile, J. Thomas, C.W. Marean, J. Parkington, S. Presslee, J. Lee-Thorp, P. Ditchfield, J.F. Hamilton, M.W. Ward, C.M. Wang, M.D. Shaw, T. Harrison, M. Domínguez-Rodrigo, R.D.E. MacPhee, A. Kwekason, M. Ecker, L. Kolska Horwitz, M. Chazan, R. Kröger, J. Thomas-Oates, J.H. Harding, E. Cappellini, K. Penkman, M.J. Collins, Protein sequences bound to mineral surfaces persist into deep time, Elife. 5 (2016). https://doi.org/10.7554/eLife.17092.

[26] S. Presslee, J. Wilson, J. Woolley, J. Best, D. Russell, A. Radini, R. Fischer, B. Kessler, R. Boano, M. Collins, B. Demarchi, The identification of archaeological eggshell using peptide markers, STAR: Science & Technology of Archaeological Research. 3 (2017) 89–99.

[27] M. Mackie, J. Hendy, A.D. Lowe, A. Sperduti, M. Holst, M.J. Collins, C.F. Speller, Preservation of the metaproteome: variability of protein preservation in ancient dental calculus, Sci Technol Archaeol Res. 3 (2017) 74–86.

[28] E. Cappellini, F. Welker, L. Pandolfi, J. Ramos-Madrigal, D. Samodova, P.L. Rüther, A.K. Fotakis, D. Lyon, J. Víctor Moreno-Mayar, M. Bukhsianidze, R.R. Jersie-Christensen, M. Mackie, A. Ginolhac, R. Ferring, M. Tappen, E. Palkopoulou, M.R. Dickinson, T.W. Stafford, Y.L. Chan, A. Götherström, Senthilvel K S, P.D. Heintzman, J.D. Kapp, I.

Kirillova, Y. Moodley, J. Agusti, R.-D. Kahlke, G. Kiladze, B. Martínez-Navarro, S. Liu, M.S. Velasco, M.-H.S. Sinding, C.D. Kelstrup, M.E. Allentoft, L. Orlando, K. Penkman, B. Shapiro, L. Rook, L. Dalén, M.T.P. Gilbert, J.V. Olsen, D. Lordkipanidze, E. Willerslev, Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny, Nature. (2019) 1–5.

[29] F. Lugli, G. Di Rocco, A. Vazzana, F. Genovese, D. Pinetti, E. Cilli, M.C. Carile, S. Silvestrini, G. Gabanini, S. Arrighi, L. Buti, E. Bortolini, A. Cipriani, C. Figus, G. Marciani, G. Oxilia, M. Romandini, R. Sorrentino, M. Sola, S. Benazzi, Enamel peptides reveal the sex of the Late Antique “Lovers of Modena,” Sci. Rep. 9 (2019) 13130.

[30] M. Buckley, Zooarchaeology by Mass Spectrometry (ZooMS) Collagen Fingerprinting for the Species Identification of Archaeological Bone Fragments, in: C.M. Giovas, M.J. LeFebvre (Eds.), Zooarchaeology in Practice: Case Studies in Methodology and Interpretation in Archaeofaunal Analysis, Springer International Publishing, Cham, 2018: pp. 227–247.

[31] F. Marin, G. Luquet, Molluscan shell proteins, C. R. Palevol. 3 (2004) 469–492.

[32] F. Marin, G. Luquet, B. Marie, D. Medakovic, Molluscan Shell Proteins: Primary Structure, Origin, and Evolution, in: Current Topics in Developmental Biology, Academic Press, 2007: pp. 209–276.

[33] D.J. Jackson, C. McDougall, B. Woodcroft, P. Moase, R.A. Rose, M. Kube, R. Reinhardt, D.S. Rokhsar, C. Montagnani, C. Joubert, D. Piquemal, B.M. Degnan, Parallel evolution of nacre building gene sets in molluscs, Mol. Biol. Evol. 27 (2010) 591–608.

[34] K.M. Kocot, F. Aguilera, C. McDougall, D.J. Jackson, B.M. Degnan, Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization, Front. Zool. 13 (2016) 23.

[35] K.M. Towe, G.R. Thompson, The structure of some bivalve shell carbonates prepared by ion-beam thinning, Calcif. Tissue Res. 10 (1972) 38–48.

[36] G.A. Sykes, M.J. Collins, D.I. Walton, The significance of a geochemically isolated intracrystalline organic fraction within biominerals, Org. Geochem. 23 (1995) 1059–1065.

[37] K.E.H. Penkman, D.S. Kaufman, D. Maddy, M.J. Collins, Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells, Quat. Geochronol. 3 (2008) 2–25.

[38] K.E.H. Penkman, R.C. Preece, D.R. Bridgland, D.H. Keen, T. Meijer, S.A. Parfitt, T.S. White, M.J. Collins, A chronological framework for the British Quaternary based on *Bithynia opercula*, Nature. 476 (2011) 446–449.

[39] B. Demarchi, K. Rogers, D.A. Fa, C.J. Finlayson, N. Milner, K.E.H. Penkman, Intra-crystalline protein diagenesis (IcPD) in *Patella vulgata*. Part I: Isolation and testing of the closed system, Quat. Geochronol. 16 (2013) 144–157.

[40] J.E. Ortiz, Y. Sánchez-Palencia, I. Gutiérrez-Zugasti, T. Torres, M. González-Morales, Protein diagenesis in archaeological gastropod shells and the suitability of this material for amino acid racemisation dating: *Phorcus lineatus* (da Costa, 1778), Quat. Geochronol. 46 (2018) 16–27.

[41] F. Marin, N. Le Roy, B. Marie, The formation and mineralization of mollusk shell, Front. Biosci. 4 (2012) 1099–1125.

[42] A. Osuna-Mascaró, T. Cruz-Bustos, S. Benhamada, N. Guichard, B. Marie, L. Plasseraud, M. Corneillat, G. Alcaraz, A. Checa, F. Marin, The shell organic matrix of the crossed lamellar queen conch shell (*Strombus gigas*), Comp. Biochem. Physiol. B Biochem. Mol. Biol. 168 (2014) 76–85.

[43] O.B.A. Agbaje, D.E. Thomas, J.G. Dominguez, B.V. Mclnerney, 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. Bundeleva, 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. Strohalm, mMass as a software tool for the annotation of cyclic

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

TABLES

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

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)		

Table 2. Peptide markers (*m/z* values) for *Unio pictorum*, *Spondylus gaederopus*, *Ostrea edulis*, *Phorcus turbinatus, Patella vulgata, Pecten maximus* and *Pseudunio auricularius* (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, Spondylidae	Gastropoda, Patellidae	Gastropoda, Trochida, Trochidae	Bivalvia, Unionida, Unionidae	Bivalvia, Unionida, Margaritiferidae
Species	*Ostrea edulis*	*Pecten maximus*	*Spondylus gaederopus*	*Patella vulgata*	*Phorcus turbinatus*	*Unio pictorum*	*Pseudunio auricularius*
Marker m/z 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

FIGURE LEGENDS

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

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

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

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