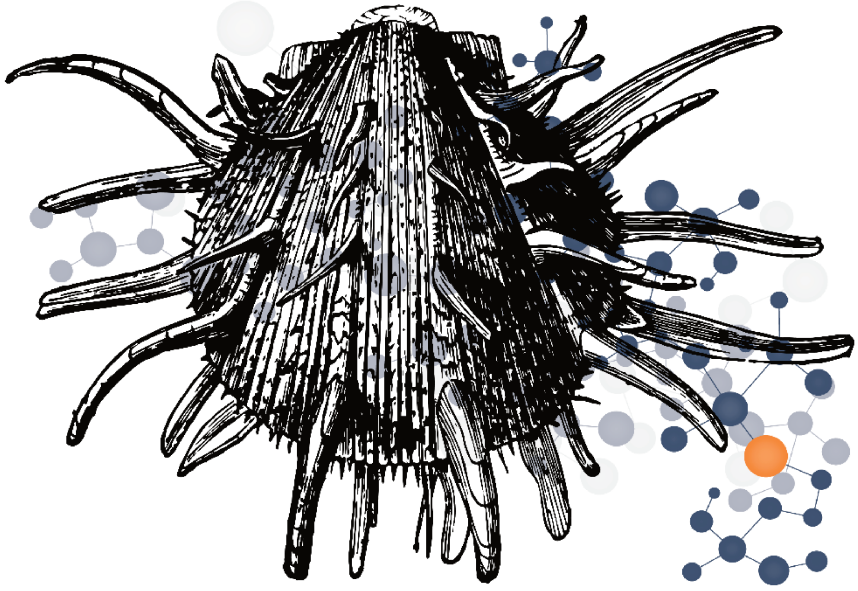
Università degli Studi di Torino
Department of Life Sciences and Systems Biology

University of Burgundy-Franche-Comté

PhD Programme in Biological Sciences and
Applied Biotechnologies & Palaeontology



Proteomic approaches for archaeology and cultural heritage: characterising ancient proteins preserved in mollusc shells

Jorune Sakalauskaite

Università degli Studi di Torino

Tesi di Dottorato di Ricerca in Scienze Biologiche
e Biotecnologie Applicate

PhD Thesis in Biological Sciences
and Applied Biotechnologies

Proteomic approaches for archaeology and cultural heritage: characterising ancient proteins preserved in mollusc shells

Jorune Sakalauskaite

Tutor: Beatrice Demarchi
Tutor from co-partnership institution: Frédéric Marin

XXXIII Cycle: 2017 – 2020

Université de Bourgogne-Franche-Comté

Université de Bourgogne
UMR CNRS 6282 Biogéosciences
Ecole Doctorale ES, Environnement - Santé

Thèse présentée par Mlle Jorune SAKALAUSKAITE
pour l'obtention du diplôme de Doctorat de 3ème cycle
de l'Université de Bourgogne-Franche-Comté

Approches protéomiques en archéologie et héritage culturel: caractérisation de protéines anciennes préservées dans des coquilles de mollusques

University of Turin

University of Burgundy-Franche-Comté

International Dual Degree *PhD* program (*Cotutelle*)

PhD program in:
Biological Sciences and Applied Biotechnologies (UNITO)
Palaeontology (UBFC)

Thesis Jury:

Jane Thomas-Oates, Professor, University of York (external reviewer)
Daniel Jackson, Professor, University of Göttingen (external reviewer)
Ursula Thun Hohenstein, Associate professor, Università degli Studi di Ferrara (examiner; president of the jury)
Philippe Ganot, Chargé de Recherche, Centre Scientifique de Monaco (examiner)
Beatrice Demarchi, Associate professor, Università degli Studi di Torino (thesis director)
Frédéric Marin, Directeur de Recherche CNRS, Université de Bourgogne-Franche-Comté (thesis director)

January 2021

Declaration

I hereby certify that the work described in this thesis is my own, except where otherwise acknowledged, and has not been submitted previously for a degree at this, or any other, university.

Acknowledgements

First of all, I would like to thank my two supervisors, Beatrice Demarchi and Frédéric Marin, with whom I share the success of this PhD project. I am grateful for all their help, support, guidance and invaluable insights that shaped my personality leading me to become the researcher that I am now. More importantly, you have also helped me to create a future vision of the person I would like to be. I am also thankful to Matthew Collins and Emmanuel Fara for valuable ideas and support throughout my PhD, as well as upcoming future research avenues.

I wish to thank all the people that helped me with my research at University of Turin: Barbara Pergolizzi and Federica Dal Bello for the help with mass spectrometry analysis, Massimo Maffei and his lab team as well as Roberto Buscaino, Nadia Barbero and Alfredo Santovito with lab instrumentation. I am thankful to colleagues Rosa Boano and Marilena Girotti for the pleasant atmosphere that has always surrounded me in the department.

I am extremely grateful for the help and support I received from the team at Biogéosciences unit and other departments of University of Burgundy: Jérôme Thomas for making beautiful pictures of my shells and archaeological samples, Rémi Laffont, Lauriane Poloni, Emilie Steimetz with CT scan analysis, Théophile Cocquerez for isotope analysis, Laurent Plasseraud for endless measurements with FTIR-ATR system. I would also like to thank Morgane Oudot, Irina Bundeleva and other Biogéosciences unit personnel for creating an always welcoming environment.

I am thankful to Meaghan Mackie at University of Copenhagen for numerous mass spectrometry analyses. I would also like to acknowledge all our team collaborators that participated in the studies: Søren H. Andersen, Paolo Biagi, André Carlo Colonese, Alberto Girod, Marion Heumüller, Hannah Koon, Giorgia Mandili, Claudio Medana, Kirsty E.H. Penkman, Helmut Schlichtherle, Sheila Taylor, Caroline Tokarski, Julie Wilson, Samantha Presslee, Cédric Broussard, Monique de Cargouët, Nicolas Potier, Solange Rigaud and Claire Manen.

During this PhD project I had a chance to be part of some of the most interesting projects which shaped my vision of research and I cannot thank the organisers enough. I am grateful to Marie Albéric, Mathieu Thoury, Jonathan Perrin and Frédéric Jamme for an exceptional experience at Soleil synchrotron. I would like to thank Davide Di Cioccio, Panagiotis Kasapidis and all of the HCMR team for making the *Spondylus* project happen, in particular to Elisavet Kaitetzidou who took care of my samples even when pandemics were taking over worldwide. I would like to acknowledge the Maristem project and their organised workshops in Tricase and Peniche where I met some of the most interesting people in the field of marine biology. I am also very grateful to Maria Borrello, Felipe Cárdenas-Arroyo and

Fabrice Monna for valuable insights and interesting discussion about archaeological projects.
Of course, I cannot be more grateful to my family which has supported me throughout all these years. I am very lucky to be surrounded by my sister and two brothers, Milda, Darius, Kajus who were always there for me, and in particular, I valued so much the insights given by my little brother who would inspire me and bring me back to what is truly important in life. I am also extremely grateful to my friends and loved ones Justina, Vaiva, Ema, Chiara and João for incredible support throughout all this time. I would also like to thank all the people I met in Turin and Dijon for helping me to explore the best of Mediterranean lifestyle that gave so much joy. I wish to thank P.B Konce and the whole Vilnius University chamber orchestra for sharing music, concerts, trips, laughs and of course long hours of rehearsals that I am pretty sure, shaped me into the way I am now.

Brief report on doctoral activities

Publications

1. **Sakalauskaite, Jorune**, Frédéric Marin, Barbara Pergolizzi, and Beatrice Demarchi. 2020. “Shell Palaeoproteomics: First Application of Peptide Mass Fingerprinting for the Rapid Identification of Mollusc Shells in Archaeology.” **Journal of Proteomics** 227 (July): 103920.
2. **Sakalauskaite, Jorune**, Laurent Plasseraud, Jérôme Thomas, Marie Albéric, Mathieu Thoury, Jonathan Perrin, Frédéric Jamme, Cédric Broussard, Beatrice Demarchi, and Frédéric Marin. 2020. “The Shell Matrix of the European Thorny Oyster, *Spondylus gaederopus*: Microstructural and Molecular Characterization.” **Journal of Structural Biology** 211 (1): 107497.
3. Evan Saitta T., Jakob Vinther, Molly K. Crisp, Geoffrey D. Abbott, Thomas G. Kaye, Michael Pittman, Ian Bull, Ian Fletcher, Xinqi Chen, Matthew Collins, **Jorune Sakalauskaite**, Meaghan Mackie, Federica Dal Bello, Marc Dickinson, Mark Stevenson A., Paul Donohoe, Philipp Heck R., Beatrice Demarchi, Kirsty Penkman EH. 2020. “Non-avian dinosaur eggshell calcite contains ancient, endogenous amino acids.” bioRxiv. https://doi.org/10.1101/2020.06.02.129999.
4. Demarchi, Beatrice, Samantha Presslee, **Jorune Sakalauskaite**, Roman Fischer, and Julia Best. 2020. “The Role of Birds at Çatalhöyük Revealed by the Analysis of Eggshell.” **Quaternary International**: The Journal of the International Union for Quaternary Research, February. https://doi.org/10.1016/j.quaint.2020.02.009.
5. **Sakalauskaite, Jorune**, Søren H. Andersen, Paolo Biagi, Maria A. Borrello, Théophile Cocquerez, André Carlo Colonese, Federica Dal Bello, et al. 2019. “‘Palaeoshellomics’ Reveals the Use of Freshwater Mother-of-Pearl in Prehistory.” **eLife** 8 (May). https://doi.org/10.7554/eLife.45644.

Attended international conferences

1. National Day of the French Society of Biology of Mineralized Tissues (SFBTM) at the **ECTS** conference, Oct 2020 (***Online presentation***; **invited speaker**).
2. 5th International Sclerochronology conference, Jun 2019, Split, Croatia (***Oral presentation***).
3. Annual meeting of French Society of Biomineralized Tissues (**JFBTM**), Jun 2019, Boulogne-sur-Mer, France (***Oral presentation + poster***): **best poster award**.
4. Environmental Archaeology (**AEA**), Nov 2018, Aarhus, Denmark, (***Oral presentation***).
5. Ancient Proteins@20 meeting (**AP@20**), Aug 2018, Copenhagen, Denmark (***Oral presentation*).**
6. 7th Postgraduate ZooArchaeology Forum (**PZAF**), Jun 2018, Palermo, Italy (***Oral presentation***).

7. 8th International Union of Prehistoric and Protohistoric Sciences World Congress (**UISPP**), Jun 2018, Paris, France (Session XIII-2) (***Oral presentation***).
8. X Congresso Nazionale di Associazione Italiana di Archeometria (**AIAr**), Feb 2018, Turin, Italy (***Oral presentation***).

Attended courses

1. English for Scientific Academic Purposes for PhD students. University of Turin. March - June 2019, (5 ECTS).
2. Workshop: Structure and functioning of mountain ecosystems. DBIOS, University of Turin, 27 June 2019, (2 ECTS).
3. Workshop: tools for research evaluation and scientific production analysis. University of Turin. 9 July, 2019, (2 ECTS).
4. Bioinformatics. Doctoral school Environnements-Santé. University of Burgundy-Franche-Comté, May 2019, (24 h).
5. Scientific Diving Course. Organized by FFESSM in the frame of UMR CNRS/UBFC 6282 activities. March-May 2019.
6. 2nd Maristem Workshop. Omic approaches to identify and characterize Marine/Aquatic Invertebrate Stem Cells. Cost Action, Peniche, Portugal. April 8-10, 2019.
7. Workshop - Entreprendre et innover. Organized by University of Burgundy-Franche-Comté (Besançon), 20 March, 2019 (16 h).
8. Synchrotron safety and DISCO beamline training. Organized by SOLEIL synchrotron, Paris, February 2019, (6 h).
9. Microorganismi al servizio della salute dell'ambiente e dell'uomo. September 2018, Orto Botanico University of Turin, (2 ESTS).
10. Training summer school in Zoology. An integrated approach to marine invertebrate biodiversity: evolutionary and functional adaptations. COST action Maristem. Avamposto Mare, Tricase, Lecce (Italy). October 2018, (42 h).
11. PhD Toolbox: basic & advanced modules of bioinformatical training, January 2018, (10 ECTS).

Research activities in external laboratories/ institutions

1. Soleil Synchrotron, DISCO beamline, Paris. Research mission as part of project: "The interplay between biochromes and biogenic calcium carbonate minerals in sea urchin spines and marine gastropod shells: implications for biomineralization and diagenesis of colored skeletal elements". Session 1: January-February 2019; session 2: December 2020.
2. HCMR-IMBBC institute in Heraklion, Greece. Research mission as part of Assemble+ project 8233 Spondylus multiOmics: bridging biomineralization and archaeology

(SpondylOmics) for the transnational access to Marine Biology station. February-March 2020.

Fellowships and Awards

1. **The Malacological Society of London yearly career research (ECR) grant** for research project *Spondylus gaederopus*: developing a new reference for amino acid racemisation dating using intra-crystalline proteins (awarded sum **£ 1500)**, 2020.
2. **Assemble+ program award for project** 8233 Spondylus multiOmics: bridging biomineralization and archaeology (SpondylOmics) for the transnational access to marine biology station HCMR-IMBBC, Crete, Greece (project awarded for the access of 30 days), 2019.
3. **Best poster award,** at JFBTM conference (French Society of Biomineralized Tissues), 2019.
4. **Eiffel excellence scholarship,** awarded from French Government to carry research work in France (awarded 12 months), 2019-2020.
5. **BGF Scholarship,** awarded from French Embassy in Italy to carry research work in France (6 months), 2017-2018.
6. **AEA student bursary** (as first qualified) for the attendance of AEA conference in Aarhus, Denmark, 2018.
7. **American Museum of Natural History Fellowship** for aProtein@20 meeting attendance, Copenhagen 2018.

Abstract

Archaeological mollusc shell artifacts and ecofacts are valuable sources to study past cultures and provide insights on how people exploited their environments. Mollusc shells were often used as raw materials to make personal ornaments and are abundantly found in archaeological sites. However, minute, heavily worked and/or fragmented shell ornaments are rarely identified at different taxonomic levels, due to limited availability of analytical approaches to determine taxon-specific (diagnostic) features. In recent years, proteomics, which exploits the high sensitivity of modern mass spectrometry techniques, has been successfully applied to the study of a variety of bioarchaeological remains, opening a new research field referred to as 'palaeoproteomics'. While modern mollusc shells represent key study models in biomineralization research to investigate the molecular mechanisms of biocalcification and mineral deposition, palaeoproteomics on archaeological shells has never been carried out before. The challenges are mainly due to the intrinsic physico-chemical features of their substrates (including low protein content), and the paucity of protein sequences for reference.

Therefore, this work presents a comprehensive proteomic-based investigation of archaeological mollusc shells, grounded in biomineralization research, and the first application of palaeoproteomics to shells. The main objective of this PhD was to investigate three key concepts: whether shell proteins carry taxonomic information, could be preserved over archaeological times scales and could be used as molecular barcodes for the taxonomic determination of ancient shell artifacts.

High performance liquid chromatography coupled to tandem mass spectrometry (HPLC-MS/MS) was used for an in-depth proteomics characterisation of a selected model, *Spondylus gaederopus*, demonstrating that its shell-associated proteins are very different from other models and likely represent lineage-specific sequences. Peptide mass fingerprinting (PMF) by MALDI-TOF mass spectrometry was employed to obtain sequence information from thirty-four different molluscan taxa. Intracrystalline shell proteins displayed clearly distinct PMFs indicating that they may encode taxonomic information and could be used as molecular barcodes to identify archaeological mollusc shells. Accelerated aging experiments were performed to mimic the diagenesis of *Spondylus* shell and the stability of intracrystalline shell proteins was investigated by a combination of immunochemistry and quantitative TMT proteomics. We concluded that *Spondylus* represents a favourable system for protein preservation, as evidenced by the thermal stability of shell peptide sequences. However, the observed degradation patterns pinpointed the complexity of the whole diagenesis process, which does not entirely follow expected trajectories from theoretical models.

Finally, a set of archaeological shell ornaments, recovered from different Mesolithic and Neolithic European prehistoric sites, were studied by proteomics coupled with structural, biomolecular and geochemical analyses. The studies showed that PMF by MALDI-TOF approach was not able to characterise these archaeological samples, which are very degraded or and/or have low protein contents, and the analysis of ancient shell proteins requires higher sensitivity, as offered by HPLC-MS/MS mass spectrometry. Indeed, palaeoproteomic analysis by HPLC-MS/MS of small “double-button” ornaments identified that they were made of freshwater mother-of-pearl shells, resolving the long-standing debate over their biological origin and provenance. The research carried out in this thesis highlights the immense potential of proteomics-based approaches to study archaeological and palaeontological shell remains.

Keywords

Mollusc shells, proteomics, biomineralization, archaeology, palaeoproteomics, shell ornaments, mass spectrometry, shell proteins, peptide mass fingerprinting

List of Abbreviations

aDNA – ancient DNA
BSMP – Basic shell matrix protein
CA – Carbonic anhydrase
CAN – Acetonitrile
CBD – Chitin-binding domain
DC – Diagenetic cleavage
EDTA – Ethylenediaminetetraacetic acid
EGF-like – Epidermal growth factor-like protein
ELISA – Enzyme-linked immunosorbent assay
ELLA – enzyme-linked lectin assay
ESI – Electrospray ionisation
EtOH – Ethanol
FASP – Filter-assisted sample preparation method
FT-MS – Fourier-transform mass spectrometry
HCCA – α-Cyano-4-hydroxycinnamic acid
HPLC – High performance liquid chromatography
HPLC-MS/MS – High-performance liquid chromatography coupled to tandem mass spectrometry
IcF – Intracrystalline
IDP – Intrinsically disordered protein
IDR – Intrinsically disordered region
IIF – (Inter+Intra)crystalline
LBK – The Linear Pottery culture (Linearbandkeramik)
LCD – Low complexity domain
MALDI-TOF-MS – Matrix assisted laser desorption ionization-time of flight mass spectrometry
NaOCl – Sodium hypochlorite
OES – Ostrich eggshell system
PCR – Polymerase chain reaction
pI – isoelectric point (the pH value at which the net charge of the protein is zero)
PTM – Post-translational modification
RLCD – Repetitive low complexity domain
SMP – shell matrix proteins
TFA – Trifluoroacetic acid
TMT – tandem mass tag
VWA – von Willebrand factor type A
ZooMS – Zooarchaeology by mass spectrometry

Table of Contents

Introduction

Premise

Mollusc shells possess exceptional mechanical properties and an intrinsic aesthetic value, which have always attracted humans. In prehistoric times, shells were one of the most prized natural resources for making personal ornaments. Hence, the study of archaeological shell ornaments provides insights into cultural traditions and symbolic practices (Vanhaeren and d'Errico, 2006; Baysal, 2019), people's mobility across time and space (Chapman and Gaydarska, 2015; Bar-Yosef Mayer, 2018), and shifts in cultural boundaries in times of lifestyle changes (Rigaud et al., 2015). However, an in-depth scientific assessment of the variety of shells that were used for ornament-making is difficult, because ornaments were heavily worked and transformed, and can be found fragmented or degraded. As a consequence, classical analytical methods (such as scanning electron microscopy) fail to identify their biological origin. This thesis proposes the use of novel biomolecular tools based on the study of ancient protein sequences – an emerging field known as "palaeoproteomics" – for the identification of archaeological shells. This approach may be an invaluable help to archaeologists, for studying mollusc shell artifacts and their remains from the past. The development of shell palaeoproteomics is based on the technological advancements in proteomic techniques and is grounded on the last decade's progress in biomineralization research. These two domains are the complementary cores of this thesis, and have been equally used in order to carry out a comprehensive investigation of archaeological shells at the molecular level. Therefore, the main focuses of the thesis are:

- In-depth biomolecular and proteomic study of a model shell system in order to determine if shell proteins hold taxonomic information.
    - The goal was to characterise the biochemistry and shell proteome of *Spondylus gaederopus* and to build a proteomics "baseline" for the identification of archaeological *Spondylus* artifacts, as explained in Chapter 1.
- Study of shell protein diagenesis patterns.
    - The goal was to conduct artificial diagenesis experiments on the same shell model system – *Spondylus gaederopus* – and to investigate the degradation patterns of its intracrystalline protein fraction, as detailed in Chapter 2.
- Development of shell palaeoproteomics.

- The first goal was to advance the methodology/analytical protocols available for the study of archaeological shells, described in Chapter 3.
- The second goal was to use shell palaeoproteomics to study prehistoric shell ornaments in order to assess protein preservation and, if possible, to obtain taxonomic identification, as shown in Chapters 4 and 5.

The first part of the manuscript includes a detailed literature review, which aims to cover the current scientific knowledge of the main themes in this field, and it is divided into four parts: the biology of mollusc shells and their interdisciplinary studies; archaeological mollusc shell ornaments; shell biomineralization; palaeoproteomics. The main thesis aims and achievements are reported in each of the thesis chapters, which consist of scientific articles (published, submitted or in preparation). Each chapter is introduced by short “digest” texts. The five chapters are followed by an integrated discussion that aims to give an overview of the main advances in shell palaeoproteomics and to provide future perspectives in the field of ancient shell proteins. Finally, the last remarks of the thesis are provided in the conclusions.

The central tenet of my work is that archaeological shell ornaments, which are heavily worked, fragmented or degraded, can be studied using palaeoproteomics, grounded on biomineralization research, with the aim of obtaining secure identification of their biological origin. Overall, this work presents an interdisciplinary investigation of mollusc shells, which brings together the fields of archaeology and biomineralization and emphasises the dynamic impact on “fossil” mollusc shell studies.

Introduction

Mollusc shells are complex composite biomaterials, which contain a small fraction of organics and are arranged into a variety of mineralised layers that exhibit different microstructures (Lowenstam and Weiner, 1989). The biomineralization of the shell, *i.e.*, mineral deposition, is regulated by a set of proteins, some of which remain occluded in the shell skeletons (Marin et al., 2016) and which can be consequently considered as molecular archives. These aspects will be presented and explained in the literature review. The analysis of shell proteins can help us to better understand the process and evolution of biomineralization and this was used to investigate my chosen model system, the bivalve shell *Spondylus gaederopus,* by using proteomic techniques and biochemical tools, as presented in Chapter 1.

The biomolecular data obtained for *Spondylus* is significant in biomineralization research because it represents one of the few studies conducted on shell model of non-commercial importance and with complex microstructure (McDougall and Degnan, 2018; Agbaje et al., 2019). The choice of *Spondylus* owes nothing to chance: in prehistoric times this shell was widely used as a raw material for making ornaments and our study lays the foundations of a large dataset usable in archaeological research (Ifantidis and Nikolaidou, 2011; Windler, 2019).

From a methodological perspective, the rise of palaeoproteomic analyses to study bioarchaeological artifacts, significantly reshaped, in recent years, the field of archaeological sciences. The analysis of ancient proteins by mass spectrometry techniques provides the ability to decode high-quality biological information from a diversity of substrates (Buckley, 2018; Welker, 2018), including bone, dental enamel, eggshell, as presented in Part 4 of the literature review. However, mollusc shells have not benefited from such advanced studies. This may be explained by the fact that molluscs represent a huge phylum (> 100 000 living species) and shell protein sequences are poorly characterised due to the lack of genetic information. In addition, mollusc shells are complex and very diversified substrates, which contain much smaller amounts of proteins compared to, for example, bones or eggshell. In this thesis, I present an in-depth investigation of methodological approaches used in palaeoproteomics in order to evaluate different sample preparation techniques, *i.e.* cleaning, extraction, purification applicable to shell substrates. These technical aspects are explained in Chapter 3. I have also used two different types of mass spectrometry approaches – peptide mass fingerprinting (PMF) by matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-TOF-MS) and high performance liquid chromatography tandem mass spectrometry (HPLC-MS/MS) to study and characterise shell proteins, which are presented in subsequent Chapters 3, 4 and 5. It is noteworthy that the two studies presented in Chapters 3 and 4 mark the first application of PMF to investigate mollusc shell sequences and expand the spectrum of mass spectrometry techniques in shell protein research.

The development of palaeoproteomic approaches for shells requires a profound understanding on how proteins degrade and “fossilise” through time – a process called diagenesis (see reviews in Demarchi et al., 2016; Demarchi, 2020). Diagenesis-induced protein breakdown, *e.g.* the hydrolysis of peptide bonds, can be studied by two approaches: i) by analysing well-dated sub-fossil/fossil shell samples or ii) by using artificial aging experiments on modern samples (with pressure and/or temperature). Diagenesis has been mainly investigated by characterising the individual amino acids and their changes in concentration and chiral properties over time (Demarchi et al., 2011; 2013; Ortiz et al., 2018), as

presented in Part 4 of the literature review. However, there is a significant lack of studies on diagenesis carried out at proteome-wide level, which would contribute to a better understanding of the pathways of protein hydrolysis and the influence of the mineral phase. Chapter 2, therefore, presents an in-depth investigation of intracrystalline *Spondylus* shell protein diagenesis carried out on artificially-aged shell samples by using biochemical and time-resolved quantitative proteomics (TMT). The study is noteworthy because it allowed to simultaneously track structural breakdown and peptide bond hydrolysis and because it is the first application of TMT quantitative proteomics to investigate protein diagenesis patterns in shells.

Finally, shell proteins can be very diverse among different species and the studies carried out in this thesis demonstrate that they can also be exploited as a source of taxonomic information, as presented in Chapters 3, 4 and 5. Most notably, the palaeoproteomic analysis of ~6000 years-old “double-button” ornaments that were found across different archaeological sites in Europe (Chapter 5), allowed us to identify that they were manufactured from the mother-of-pearl of freshwater mussels, solving a decade-long archaeological debate on their origin and provenance. Overall, the different studies presented in this thesis underline the necessity of an integrated approach that bridges different scientific disciplines, including analytical chemistry, molecular sciences, biogeochemistry as well as bioinformatics, to study proteins from archaeological shell biominerals.

Literature review

Part 1. Mollusc shells

Biology

Molluscs are remarkably diverse and, after arthropods, represent the second biggest invertebrate phylum. There are more than 160 000 species (GBIF.org, 2020) currently known, twice as that of vertebrates. In reality, the true size of the phylum is unknown, as many species are yet to be described, particularly those dwelling in remote habitats, such as the deep seabed (Bouchet et al., 2016). Molluscs also display a high variety of morphological characteristics. For example, large-size deep-ocean giant squids can be up to 12 m long, while some planktonic gastropods (pteropods) are 1 mm in length (Carter, 1990). Phylum Mollusca comprises eight extant classes, grouped into two subphyla: Conchifera, the shell-bearing molluscs and Aculifera, the shell-less molluscs (Figure 1) (Russell-Hunter, 1983). Conchifera include five classes:

- Bivalves (*e.g.* mussels, oysters, clams, cockles, scallops);
- Gastropods (*e.g.* snails, slugs, limpets, abalones, cones);
- Cephalopods (the most “evolved” molluscs, which include the coiled-shell nautilus, the cuttlefishes, the squids but also octopuses);
- Scaphopods (tusk shells);
- Monoplacophorans (*Neopilina*).

NB: note that several representatives of gastropods (slugs, sea hares) and cephalopods (octopuses) have, by evolution, secondarily lost their shells.
There are three mollusc classes in subphylum Aculifera:

- Solenogastres and caudofoveates, which are worm-like animals (they secrete calcified spicules and sclerites);
- Polyplacophorans (chiton shells), which have a series of calcified dorsal plates that are not true shells.

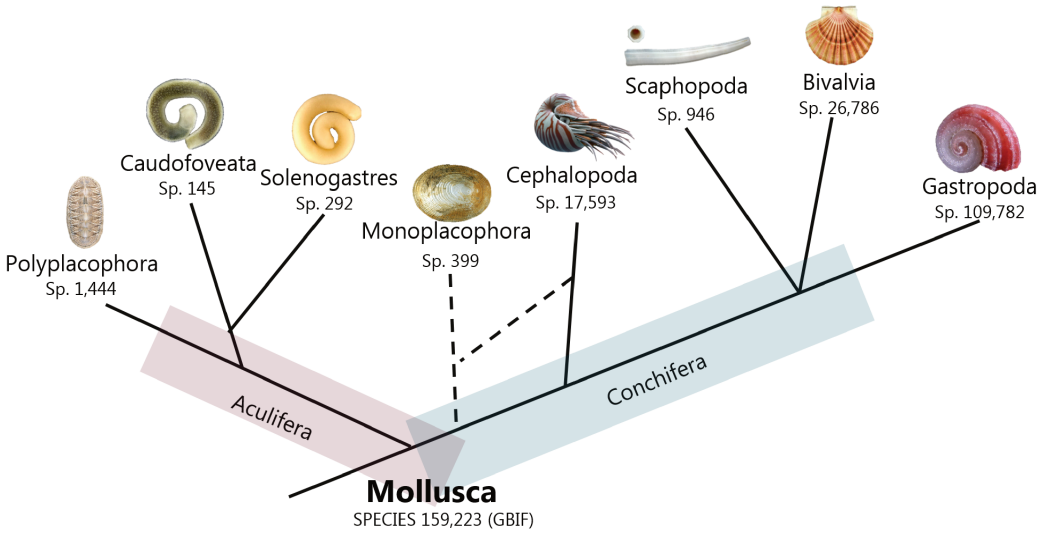


Figure 1. Current consensus of the evolutionary relationships between the major classes of the phylum Mollusca (McDougall and Degnan, 2018). The numbers in the graph also presents the current number of accepted mollusc species, based on the data published in the Global Biodiversity Information Facility (GBIF.org, May 2020). The dashed line indicates that the position for Monoplacophorans is unresolved.

Molluscs occupy a wide range of environments. In particular, gastropods, which have the highest species diversity (almost 110 000 currently known) are present in marine, freshwater and terrestrial environments, making them the most successful mollusc class (Russell-Hunter, 1983). Non marine molluscs are particularly susceptible to climatic fluctuations and their species distribution in the fossil record can therefore be used to track the changes in regional palaeoecology (Allen, 2017). Furthermore, molluscs, in their shell, record the geochemical and climatic characteristics of the environment they inhabit. Therefore, fossil mollusc shells are a valuable source to study palaeoenvironments as they can act as climate archives. More specifically, the stable oxygen and carbon isotopes of shell carbonate reflect the environmental and climatic information at the time when the animals were active and deposited the shell mineral (McConnaughey and Gillikin, 2008; Leng and Lewis, 2016). The fractionation of the oxygen isotopes ($^{18}O/^{16}O$) is governed by water temperature and salinity (Keith et al., 1964; Leng and Lewis, 2016) while carbon isotopes ($^{13}C/^{12}C$) reflect the origin of the dissolved organic content (McConnaughey and Gillikin, 2008), therefore their variations are used as proxies to investigate environments in the past (John Lowe and Walker, 2014; Allen, 2017). Marine mollusc shells are used to study seawater temperature and palaeosalinity (Maier and Titschack, 2010) and this has been especially useful with long-lived shells (*e.g.* up to several hundred years), such as *Arctica islandica* (Schöne et al.,

2005). One typical archaeological application is to study stable isotopes from archaeological shells in order to obtain information on the climatic conditions under which humans lived (Leng and Lewis, 2016). In archaeological sites with a long-term human occupation record, shells were also studied to better understand how humans have adapted in times of climate change. For example, the season of death of the mollusc can be useful to infer patterns of marine resource exploitation at various times (Colonese et al., 2011; Prendergast et al., 2016).

Molluscs are soft-bodied metazoans and their ability to biomineralize a shell skeleton, which is important for body support and protection, has provided an evolutionary advantage (McDougall and Degnan, 2018). The evolution of molluscs can be traced back to the Precambrian – Cambrian transition (Knoll, 2003; Kocot et al., 2016). More specifically, the largest diversification of shells and evolutionary wave of biomineralization took place during the Cambrian explosion, an event that occurred approximately 545-540 million years ago and is characterised by the appearance of practically all major metazoan phyla (McDougall and Degnan, 2018). Later diversification of biomineralised systems occurred during the Ordovician radiation (known as GOBE, Great Ordovician Biodiversification Event), some 40 million years after the Cambrian explosion (Knoll, 2003). The rapid diversification processes are mainly linked to the higher levels of Ca^{2+} ions (almost three-fold), increased concentration of other minerals and overall nutrient availability in seawater (Kocot et al., 2016; McDougall and Degnan, 2018). The evolution of different biomineralized structures may have also been accelerated as a defence mechanism in response to the increased pressure of existent predations (Knoll, 2003).

The calcified skeletons of mollusc shells possess exceptional material properties, which justifies the long-standing research interest in biomineralization and biomimetic sciences. For example, nacre, also known as the mother-of-pearl, alongside its eye-catching iridescence, has very interesting material properties. Nacre is extremely tough, and it is around one thousand times more resistant to fractures than the geological aragonite alone (its basic constituent form) (Addadi and Weiner, 1997; Katti et al., 2005; Gim et al., 2019). The bio-physico-chemical elements involved in shell formation are the focus of biomineralization research, which aims to understand the evolution of biomineralization and molecular features that provide some of their exceptional characteristics (Westbroek and Marin, 1998; Sarashina and Endo, 2006; Luz and Mano, 2009; Kim et al., 2016; Marin et al., 2016; Checa, 2018). However, despite a huge amount of scientific publications, the molecular mechanisms are far from understood. Thus, the possibility to replicate such exceptional biological structures under laboratory conditions, at the moment,

remains elusive. The process of shell formation and the current scientific knowledge of biomineralization will be presented in detail in further sections of this work.

Mollusc shell use

Molluscs represent an important food resource and their consumption can be traced back to early human history. In the past decades, shellfish consumption has seen a steady increase, predominantly of clams, oysters, mussels and scallops, due to rapid population growth, rising income and urbanization (Guillen et al., 2019). Molluscan aquaculture represents an active segment of the economy worldwide and is dominated by China, as well as some coastal European countries like Italy, France and Spain (*FAO Fisheries & Aquaculture*). Since ancient times, molluscs were an important and essential food resource in coastal areas, whereas in mainland sites they were often considered as more prestigious foodstuff. For example, wealthy Romans were very fond of oysters, and transported them across hundreds of kilometres, sourcing from the Atlantic coast (Andrews, 1948; Mouchi et al., 2017). Until the middle of the 17th century, oyster consumption was confined only to the wealthy. However, during the 19th century, the drop in wild oyster prices spawned the spread of oyster bars in many larger western cities, creating the so-called “oyster craze” (MacKenzie and Clyde, 1996). This oyster fashion was particularly popular in North America, however, it was not without consequences, as it resulted in an immense decline of oyster species in the coast.

In prehistoric times, mollusc played an important role in the subsistence of early humans (Stiner et al., 1999; Zilhão et al., 2020) and were an essential foodstuff for coastal populations (Colonese et al., 2011). Some scientists even suggest that molluscs, which are rich in long-chain polyunsaturated fatty acids, may have had an effect for human brain evolution (Broadhurst et al., 1998), and consequently to the dispersal of anatomically modern humans (e.g. Marean, 2014). For hunter-gatherer populations, the abundance of shellfish may have also influenced the transition to sedentary lifestyle (Colonese et al., 2011). Shell middens testify to the intensity of past mollusc exploitation and human activity on the coast (Craig and Bailey, 2007; Álvarez et al., 2011). Middens are heaps of shell debris (mostly of edible species), typically represented by oysters, clams, whelks, limpets or mussels, which have accumulated as a result of shellfish consumption and subsequent refuse disposal. Among others, some of the most famous are the Danish Late Mesolithic Ertebølle shell middens. These are ~2000 m^3 in volume and represent about 50 million oyster shells, which have accumulated over several thousands of years (Andersen, 2000; Craig and Bailey, 2007). However, it remains difficult to assess the true scale of shellfish exploitation in prehistoric Europe, and

particularly in the Mediterranean region (Colonese et al., 2011). Indeed, there are no large shell midden sites older than the mid-Holocene, *i.e.* dating from the period before the sea level reached similar levels as today's. This is because rising sea levels during the early Holocene submerged most coastal sites, which, therefore, are archaeologically "invisible" (Bailey and Flemming, 2008). The study of shell middens and other shell assemblages by stable isotope analyses enables researchers to track palaeoclimatic oscillations (Álvarez et al., 2011; Bar-Yosef Mayer et al., 2012) and provides palaeoseasonality data, which yield important information about shell gathering practices (Mannino et al., 2007). Moreover, shell assemblages are also studied for geochronological dating purposes, both by direct radiocarbon dating and by using the extent of racemisation of amino acids preserved in the shells' mineral structure (Demarchi et al., 2011; Bosch et al., 2015).

Finally, mollusc shells have excellent material and aesthetical qualities and as such have been exploited as raw materials since the Palaeolithic. The value of shells can also be attributed due to their symbolic meaning, as a link to marine environments, and their role in certain ritual practices. Large shells were used as tools or musical instruments, as for example the conch shell (Taylor et al., 1994). Cowrie shells or disc beads made of shells were frequently used as monetary objects (Trubitt, 2003). Shells were an important raw material for crafting figurines and making personal ornaments and were used to display status or for symbolic purposes (Trubitt, 2003). Shell ornaments are particularly abundant in archaeological sites and their study yields insights into cultural patterns in the past (Baysal, 2019). This will be discussed in detail in the following section of the literature review.
Pearls are probably one of the best known shell materials in modern times and are prized jewellery items. The formation of natural pearls happens when an exogenous particle (rounded or not) is entrapped in the mollusc's mantle, the shell forming tissue, and gets biomineralised as a result of a defence "encapsulation" mechanism by the pearl oyster (Landman and Mikkelsen, 2001). However, it was not until the 18-19th century that the actual mechanisms of pearl formation were known and before that, one could find pearls only just by chance. In Japan, Kokichi Mikimoto first succeeded in culturing pearls in 1893 (Nagai, 2013), and further research paved the way to the modern pearl farming industry that could satisfy the great demand of Europe's jewellery makers. Nowadays, pearls are obtained by artificial simulation, when a spherical nucleus together with a small piece of mantle tissue ("the graft") of a donor oyster are inserted into the oyster body (the gonad). The process allows a controlled and faster rate of pearl formation (approximately 18 months). The global production of pearls is worth more than 400 million US dollars and is one of the most important economic sectors in pearl farming countries, among which are Japan, French Polynesia, Australia, the Philippines and China.

However, the history of pearls and the use of mother-of-pearl shells goes back thousands of years. The oldest pearl fishing tradition is documented in the Arabian peninsula, about 7500 years ago (Charpentier et al., 2012). Pearls were highly valued by the ancient civilizations of India, China, Persia and Rome, as well as the indigenous civilizations of Pre-Columbian America (Saunders, 1999). Mother-of-pearl shell fragments were also used as dental implants by the Maya people (Bobbio, 1972), while freshwater mother-of-pearl shells, up until the 20th century in Europe and North America, were basically the sole raw material used for making buttons (Haag, 2012; Bertin, 2015).

Part 2. Archaeological shell ornaments

Shells were one of the most widely exploited raw materials for making ornaments, a craft tradition that dates back to early prehistory. Shell ornaments, an early form of jewellery, were worn on the body or embroidered on the clothes and may have been used to mark personal or group identity (Baysal, 2019). Ornaments, as the Cambridge English dictionary defines it, refers to a form of a decoration used to increase the beauty of something. However, the importance of shell ornaments in prehistory probably went far beyond the univocal view of an object with purpose of beauty, economic value or an indicator of social status (Trubitt, 2003; Baysal, 2019). Prehistoric shell ornaments embodied symbolic meanings, they were used for rituals and may have been linked to healing and protective properties (Trubitt, 2003). Therefore, the study of shell ornaments provides deeper insights into the entanglement of humans, their material cultures and the landscapes they inhabited. This section will aim to overview the trends in the use of shells for making personal ornaments, with a particular focus on their importance in Europe's prehistory.

Shell ornaments in prehistory

The practice of shell ornament-making (by archaic *Homo sapiens*) probably has its origins in Africa. The earliest examples of shell ornament use are documented in the Eastern Mediterranean region sometime around 135000-100000 years ago (Vanhaeren et al., 2006) and in North Africa around 82000 years ago (Figure 2) (Bouzouggar et al., 2007). However, Eurasian Neanderthals were equally able to exploit shells and other natural materials for symbolic purposes, as attested by a number of findings dating to roughly 125000 to 40000 years ago (Zilhão et al., 2010; Romandini et al., 2014; Hoffmann et al., 2018). Therefore, "modern behaviour" was not necessarily exclusive to *H. sapiens*. In Europe, Upper Palaeolithic (c. 50000 to 10000 years ago) *H. sapiens* hunter-gatherers frequently exploited shells as well as bone, teeth and ivory for ornament-making. Vanhaeren and d'Errico have studied almost a hundred different Upper Palaeolithic archaeological sites in Europe, spanning from the Mediterranean area to northern Europe, and found that almost half of the identified bead assemblages were made using shells (Vanhaeren and d'Errico, 2006). The use of seashells as ornaments was particularly widespread in the Mediterranean region (Peresani et al., 2019), however the knowledge on this practice is far from complete because many Palaeolithic archaeological sites were submerged due to sea level rise after the last Ice Age (Colonese et al., 2011). Shell ornaments were continuously used during the Mesolithic period, which separates the Palaeolithic and the Neolithic (Figure 2). The Mesolithic is characterised by hunter-gatherer societies with well-organised collective-hunting activities and a

slow shift towards sedentarism. During this time, there was also a gradual shift in shell diversity and bead typology used as personal ornaments (Baysal, 2019).
The most notable changes occurred during the Neolithic period, when the use of shell ornaments reached its peak (Borrello and Micheli, 2011; Chapman and Gaydarska, 2015; Rigaud et al., 2015; Windler, 2019). The Neolithic is characterised by geographically sedentary communities, which exploited domesticated plants and animals. The spread of farming, which occurred between 8800 and 5500 years ago from the Levant towards Western Europe, resulted in important technological, economical, social and cultural changes, including shifts in ornament-making practices (Rigaud et al., 2015). The archaeological record reveals that there was an increased demand for shells, especially belonging to "exotic" types (*i.e.* of non-local origin), such as *Spondylus,* and that shells were in general more frequently used as a raw material to produce highly standardised ornaments (Borrello and Micheli, 2004; Borrello, 2005; Baysal, 2019). Later on, during the Bronze Age, the use of shell ornaments gradually declined and shells were slowly replaced by other materials (Baysal, 2019).

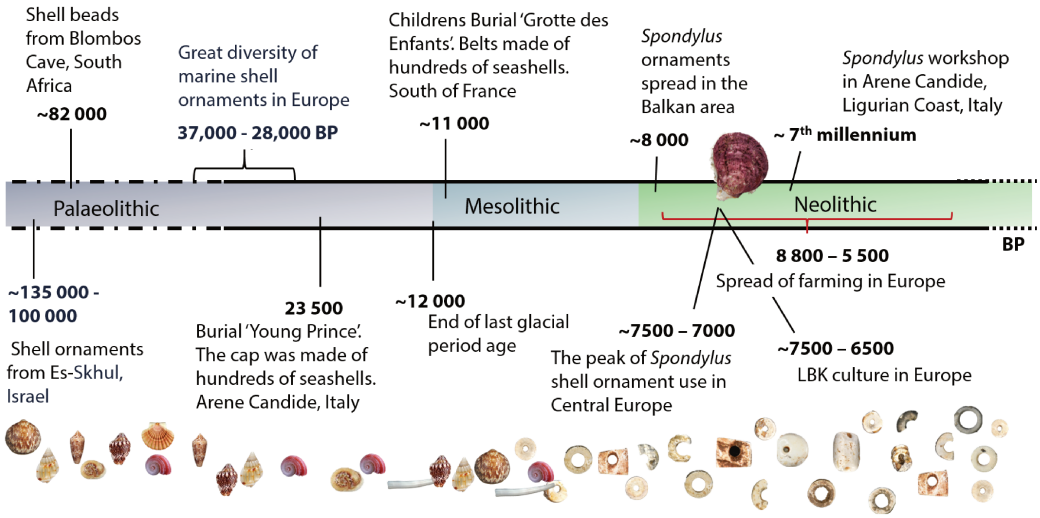


Figure 2. Shell ornaments in prehistory. A schematic timeline that displays shell use throughout the Palaeolithic, Mesolithic and Neolithic.

Archaeological shell ornaments are often associated with burial contexts. Some of the richest archaeological ornament assemblages are found as grave goods (*e.g.* Figure 3 a, b) which gives direct evidence of their symbolic meaning and provides information to study funerary practices, rituals, cultural implications and connections between humans and shells (Giacobini, 2007; Bonnardin, 2009). Ornaments recovered from household contexts are less common, but this is likely

resulting from a number of biases (Baysal, 2019). First of all, burials, especially those rich in grave goods, are more visually striking and have always attracted more attention – they were more carefully excavated, studied, investigated and published (Baysal, 2019). Such findings also enable a wider spectrum of interpretations about symbolism and meaning (Paulsen, 1974; John, 2011), therefore, studies that investigate funerary practices are more frequently found in the archaeological literature. Secondly, shell beads and their fragments are one of the smallest objects among the different archaeological findings, thus appropriate excavation methodologies (*e.g*. wet sieving, flotation techniques) are crucial for their recovery (Baysal, 2019). This means that during older archaeological excavations, which had been carried out without the use of appropriate sieving techniques (especially those in household sites), many small ornaments may have been missed, especially if they were not deposited in a specific area but scattered across the site.

However, shell ornaments in non-burial contexts provide valuable information about how they were produced, used and worn in everyday life. For example, the findings of numerous worked and semi-worked *Spondylus* shell fragments at the Arene Candide cave on the Ligurian coast (Italy) and the Dimini archaeological site next to the Aegean Sea (Greece) indicate that these sites were part of shell jewellery workshops during the Neolithic period (Borrello and Micheli, 2004; Chapman et al., 2011). Ornaments made using *Spondylus* or *Glycymeris* shells were frequently found broken, reused and recycled into other objects, which suggests that there was a continuous lifecycle of these precious artifacts (Baysal, 2019).

Small shell “double-buttons” were accidentally found during the excavations in Mesolithic shell middens in Havnø (Denmark) (Andersen, 2008) and similar types of beads were also recovered in the households of Hornstaad settlements on Lake Constance (Germany) (Heumüller, 2009). Further studies showed that the “double-buttons” were likely used to decorate clothes, when pressed into fabric and leather (Figure 3, d) (Kannegaard, 2013). Thus, research on shell ornaments from different archaeological contexts could help us to better understand the meaning that was given to shells and their lifecycle, which likely spanned several generations.

Figure 3. Examples of shell ornaments in prehistory: a) Palaeolithic burial of a young man dated to 23500 years BP discovered at Arene Candide archaeological site in Liguria, Italy. The man's head was covered with a cap, made of hundreds of perforated marine gastropod shells; b) Neolithic burial of a woman covered with hundreds of shell sequins – Corded Ware culture, Germany; c) different types of ornaments made of shells, Neolithic period; d) miniature pearl-like "double-button" shell ornaments, dated to Neolithic and Late Mesolithic (Denmark). These latter were likely pressed into leather (*e.g.* belt), as shown in the experimental archaeology-based interpretation of their use (Kannegaard, 2013).

Shell diversity

A great variety of shells were used for making ornaments, but typically, archaeological assemblages are dominated by seashells – gastropods, bivalves and scaphopods (tusk shells), with somewhat less abundant findings of freshwater

species. The following section aims to give a succinct overview of the trends in shell use and how they changed throughout the different phases of prehistory in Europe. During the Palaeolithic and Mesolithic, small marine gastropods were among the most common shell ornaments and they were mainly used in their natural form, unworked and just perforated. Over a wide geographical area, from southeast to southwest of Europe, shell beads were often made of *Columbella* sp. (most commonly by *C. rustica*), *Homalopoma sanguineu*s, *Tritia* sp. (predominantly *T. neritea*, which in older literature is also named as *Cyclope neritea*), *Conus* sp. *Nassarius* sp. and others (Borrello and Dalmeri, 2004; Giacobini, 2007; Boric and Cristiani, 2019; Laporte and Dupont, 2019; Peresani et al., 2019; Solange et al., 2019). *Dentalium* sp. (tusk) shells were also very common. Ornaments made of bivalve shells were rarer, but intact valves of *Glycymeris* and *Cardiidae* shells as well as fragments of *Pecten* have also been documented.

This set of mollusc shells represents a very small part of the species range that was actually available to coastal communities. Moreover, most of these gastropods are small, have a low nutritional value and their procurement was more difficult (Dupont, 2014). This suggests that there was a clear distinction between molluscs that were consumed as food and those that were used for ornament-making (Dupont, 2014; Perlès, 2019). Moreover, in some sites, shell selection remained constant regardless of the economic, environmental or lifestyle changes occurring through time. For example, in Franchthi Cave on the Aegean Sea in Greece, which records human occupation from the Palaeolithic to the Mesolithic, the analysis of archaeological assemblages showed that for ornaments, there was a continuous use of the same type of shells (Perlès, 2019). This indicates that the abundance, procurement or availability were not the primary reasons and that the deliberate selection of certain shell species reflected their perceived importance.

During the Neolithic there was a notable increase in the use of larger bivalve shells, in particular marine species, such as *Glycymeris* and *Spondylus,* as well as freshwater *Unio* shells (Borrello, 2005; Chapman and Gaydarska, 2015). Larger size gastropod shells such as *Charonia* sp. (Borrello, 2005) were often worked into different objects but the use of small gastropods remained constant until late Neolithic and onwards (Borrello and Dalmeri, 2004; Borrello, 2005). Yet, one of the most important shifts during the Neolithic was that shells became primarily viewed and exploited as raw material for making jewellery. Shells were worked and shaped into different objects, such as beads, bracelets, pendants (Figure 3, 5), with a final look that was completely different from their natural morphology (Borrello, 2005; Baysal, 2019; Laporte and Dupont, 2019). However, this working technique, which left almost no visible shell features, also implies that we have only a partial understanding of the diversity of shell species that were used during the Neolithic compared to that we have for the Palaeolithic.

Ornament typology

The colour of shell ornaments is an interesting feature to consider and it is likely that it had an influence over the taxa that were chosen to make ornaments/objects. Shells with vivid colours appear to dominate the archaeological ornament assemblages from Palaeolithic and Mesolithic (Peresani et al., 2019; Solange et al., 2019). The majority of the small gastropod species used as ornaments have shells with naturally bright colours, mostly yellow and red, and also display peculiar ornamentation patterns. For example, tiny Mediterranean *Homalopoma sanguineum* shells, which are just 4 mm in length, have a very intense red colour, and the brownish Mediterranean *Columbella rustica* shells have characteristic natural ornamentation patterns. Some archaeological shell ornaments, despite being thousands of years old, were recovered retaining their original colours (Solange et al., 2019), indicating that the superficial layers of the shells were not removed. Indeed, these data indicate that natural shell ornamentation was a sought-after effect and a deliberate choice.
The red colour may have had special cultural and symbolic meanings. Red *Homalopoma sanguineus* gastropod shells were recovered in many Upper Palaeolithic sites in the Italian peninsula and more generally in the Mediterranean region (Peresani et al., 2019). Moreover, shells were often coloured using ochre pigments (Baysal, 2019), giving them a deep red hue; this practice was already implemented by Neanderthals (Zilhão et al., 2010). Ochre was often found on *Glycymeris* valves (Borrello, 2005), which, however, may have also been used as pigment containers (Peresani et al., 2019). The black colour (Figure 3, c) was obtained by burning the shells, which would result in a permanent colour change (Vanhaeren and d'Errico, 2006; Boric and Cristiani, 2019; Perlès, 2019).

Ornament typologies changed dramatically during the Neolithic, when shells were exploited as raw materials, and therefore, were heavily worked and shaped into a variety of different types of beads. These were worn alone or threaded together to make larger composite jewellery items (Figure 3) (Bonnardin, 2009; Baysal, 2019). Shell ornaments became highly standardised and were produced in much larger quantities compared to earlier times. Among the most visually striking objects from the Neolithic period are the bracelets, pendants and cylindrical beads made from large *Spondylus* valves (Figure 5) (Borrello and Dalmeri, 2004; Borrello and Micheli, 2004; Borrello, 2005). Interestingly, most of the *Spondylus* bracelets had a rather small diameter – they would have been too tight to be used by adults, thus their exact position on the body, or indeed their function, remains debated (Borrello and Micheli, 2004).
The change in ornament types also changed the concept of colours, as white and light hues became preferred for shell ornaments. Naturally colourful shells, such as

Spondylus, which has a deep purple upper valve, was heavily abraded in order to obtain perfectly white finished objects (Figure 5). Freshwater-mother-of-pearl shells were worked into flat platelets (Borrello, 2005), uncovering the shiny and reflective qualities of the pearl. Moreover, marble was also used for making small beads and it may also be partially linked to the search of different types of materials characterised by a variety of white hues. The use of light reflective colours may have had a symbolic meaning to Neolithic people, and white colour might have been associated with peace, fertility, ritual purity, health and well-being (Trubitt, 2003). The findings indicate that shape and colour were an important features of shell ornaments, and were selected intentionally or achieved as a result of deliberate transformation.

Current research themes and archaeological implications

Cultural geographies

Shell ornaments can be used as a proxy to understand more about prehistoric societies and cross-cultural interactions (Baysal, 2019). Vanhaeren and d'Errico led a study of personal ornaments from different European Upper Palaeolithic sites and showed that the use of certain ornament types made of shells or other biological materials (bone, ivory, tooth), could be interpreted as the representation of biological, cultural and ethno-linguistic population groupings (Vanhaeren and d'Errico, 2006). This inference was based on the fact that availability or ease of procurement could not explain the variety of different bead types found across different European regions. The study, indeed, highlighted that these first European societies consisted of a number of different cultural entities and these may have represented distinct language families.

In another study, led by S. Rigaud, shell ornaments were employed to understand the patterns of the Neolithisation of Europe (Rigaud et al., 2015). The team used statistical analyses to classify beads from different European archaeological sites and identified an interesting pattern – Neolithic cultures in the South of Europe typically were associated to "imported" shell bead types, whereas the populations in the North of Europe were resistant to changes and continued to use their traditional (Mesolithic) ornaments. The findings exposed the complexity of two different cultural movements present during the neolithisation of Europe, which later was also confirmed by genetic studies (Hofmanová et al., 2016; Fernandes et al., 2018; Mittnik et al., 2018).

These two large studies demonstrated that the knowledge about the diversity and type of ornaments that were circulating in the past can open new perspectives in reconstructing cultural geographies, as well as tracking long distance connections.

Social prestige

Shells held an important socio-cultural value and could be considered as a proxy of prestige. The richness of burials and the abundance of shell ornaments among grave goods indicate that humans had a long lasting connection to shells. Striking ornament assemblages are documented in Palaeolithic European burials. For example, in the Grimaldi Caves (also known as Balzi Rossi), which are a group of rock shelters on the Ligurian coast near the French-Italian border, a number of graves extremely rich in shell ornaments have been uncovered (Borrello, 2005; Giacobini, 2007). In several different burials, the bodies of the deceased were decorated with shell bracelets and pendants. Some bodies were adorned with garments that were made of hundreds or thousands of seashells, including caps covering the heads and a form of a 'belt' or 'skirt' (Giacobini, 2007). These rich findings indicate social, cultural and cognitive behaviour complexity of Upper Palaeolithic European populations (Giacobini, 2007).

The archaeological findings also suggest that a great labour input was needed for collecting a large number of shells and assembling them into complex items. For example, the burial of a young male skeleton dated to around 23500 BP and found in Arene Candide archaeological site on the Ligurian coast (Italy) was decorated with thousands of perforated *Tritia* shells (Figure 3a) (Borrello, 2005; Giacobini, 2007). This young adult, indeed, is named the 'Young Prince'. Sophisticated seashell compositions were also documented in later Neolithic cultures (Bonnardin, 2009). For example, a complex garment which covered the deceased in the grave of La Balance-Ilot P in Avignon, in the South of France, was made of 158 red *Columbella rustica* shells (Zemour et al., 2017).

Worked shell ornaments had an additional value because of the craft and skills needed in order to transform the raw shell to the desired object, with different shapes and forms (Baysal, 2019). Some of the most visually striking ornaments were those made of *Spondylus* shells and similar artifacts appeared across different archaeological sites in Neolithic Europe (Borrello and Micheli, 2004; Bonnardin, 2009). Many findings have been documented in Central Europe, especially in funerary sites that are attributed to the Linear Pottery culture (LBK) (Figure 4). For example, large cylindrical beads assembled into a necklace and head decoration, as well as *Spondylus* bracelets and V-shaped pendants (Borrello and Micheli, 2004; John, 2011) were found in Aiterhofen-Ödmühle, an archaeological site in Bavaria (Germany). Large *Spondylus* cylindrical beads were also found at La Vela di Trento, one of the richest Neolithic burials in the Italian alpine region (Borrello and Dalmeri, 2004). Similar *Spondylus* ornaments are further documented in Neolithic burials close to the Paris basin, for example the *Spondylus* bracelet and disc shaped pendants that were found in Cys-la-Commune archaeological site in Aisne (France) (Borrello and Micheli, 2004; Bonnardin, 2009). Wealthy findings indicate that shell

ornaments may have been only destined to individuals with a special position in society. Also, similarities in the typology of shell jewellery may indicate exchange of cultural traditions.

Spondylus shell ornaments

Archaeological ornaments made of *Spondylus* shells represent a special case and will be discussed more in detail. *Spondylus* shells were one of the most important raw materials used for making ornaments and there are numerous *Spondylus* findings in Neolithic European archaeological sites (Moshkovitz, 1971; Chapman and Gaydarska, 2015; Windler, 2017). *Spondylus* is often considered as the "jewel of prehistory" (Ifantidis and Nikolaidou, 2011) and one of the icons of the Neolithic (Borrello and Micheli, 2004). *Spondylus gaederopus,* also known as the Mediterranean thorny oyster, is a large bivalve shell that has a bright red upper valve covered with long hollow purple spines (Figure 5b) (Galinou-Mitsoudi et al., 2019). *Spondylus* lives attached to the seabed at depths of around 5-10 m (but can be also up to 50 m) and are usually covered with the red sponge *Crambe crambe* (Figure 5a), which makes it harder to notice it underwater. The molluscs are currently mainly found on the Aegean and Adriatic coasts, and to a lesser extent, in the western part of Mediterranean Sea (Figure 4).

The use of *Spondylus* shells dates back to Palaeolithic times (Álvarez-Fernández, 2011), but a more intense exploitation was during the Neolithic (Borrello and Micheli, 2004). The first documented phase of *Spondylus* ornament use refers to the cultures in the Balkan peninsula, somewhere around 8th millennium BP (Pappa and Veropoulidou, 2011; Theodoropoulou, 2011). These cultures occupied territories that were geographically close to the Aegean Sea, where the raw material was likely sourced. Around 8000 BP *Spondylus* was already largely diffused in central Europe and especially among LBK cultures (LBK - **L**inear**b**and**k**eramik, formerly Danubian Culture, *i.e.* a group of Neolithic cultures that expanded over large areas of Europe, north and west of the Danube River, *e.g.* from Slovakia to the Netherlands). The peak was between 7500 and 7000 BP when *Spondylus* shell ornaments were widespread – all across the Danubian area, reaching the cultures in the west, as far as the Paris basin, and in the north of Europe, including present-day Poland (Chapman and Gaydarska, 2015). In Central Europe, the peak of *Spondylus* use is considered rather 'short', lasting less than a thousand of years. However, in Southeastern Europe, the Carpathian region and the Italian peninsula, *Spondylus* was continuously exploited until the later Bronze Age (Borrello and Micheli, 2004; 2011; Chapman and Gaydarska, 2015). Figure 4 shows a map of European archaeological sites where *Spondylus* artifacts were documented. It also

shows natural *Spondylus* specimen habitats, clearly highlighting shell diffusion over long distances across the European continent.

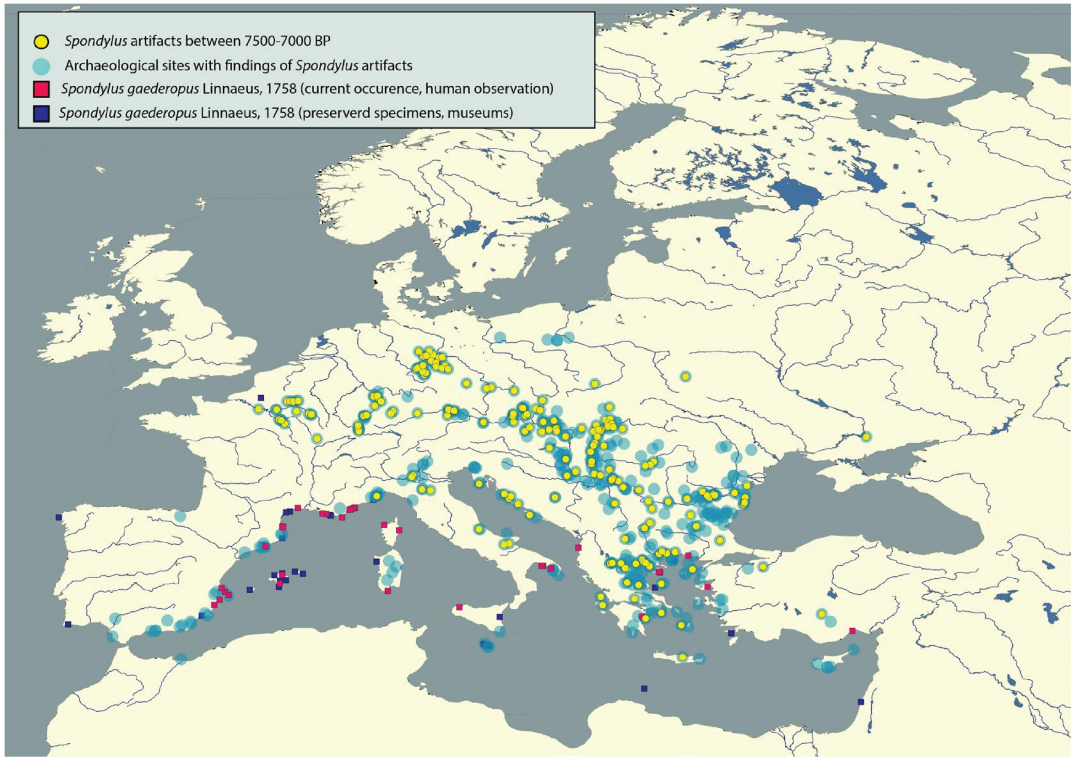


Figure 4. European map which shows Neolithic archaeological sites where *Spondylus* shell artifacts are documented (in light blue). Yellow markers highlight *Spondylus* findings that date between 7500 and 7000 BP (peak of *Spondylus* use in Europe). Markers in bright pink indicate present day occurrences of *Spondylus gaederopus* in the Mediterranean waters and in museum collections (dark blue). The distribution dataset of archaeological *Spondylus* was obtained from (Windler, 2019) and modern day occurrences were gathered from the database of GBIF.org (June, 2020).

A great variety of ornament types were made using *Spondylus* shell. The most notable forms include bracelets, pendants, bangles and annulets, as well as buttons and buckles. Some smaller and larger size beads were also assembled into necklaces or were used to embroider clothes (Figure 5) (Borrello and Micheli, 2011; Pappa and Veropoulidou, 2011; Theodoropoulou, 2011; Chapman and Gaydarska, 2015). The final objects were worked and shaped into such a way that they did not retain almost any visible features resembling that of the natural shell. Indeed, *Spondylus* ornaments which were made by fully exposing the inner white shell layers represent a striking contrast to the natural shell morphology. In addition, most of the beads (circular, tubular, disc and other shapes) were so heavily worked that

the actual determination as *Spondylus* species poses a serious challenge (Borrello and Micheli, 2004; 2011; Dimitrijević and Tripković, 2006).

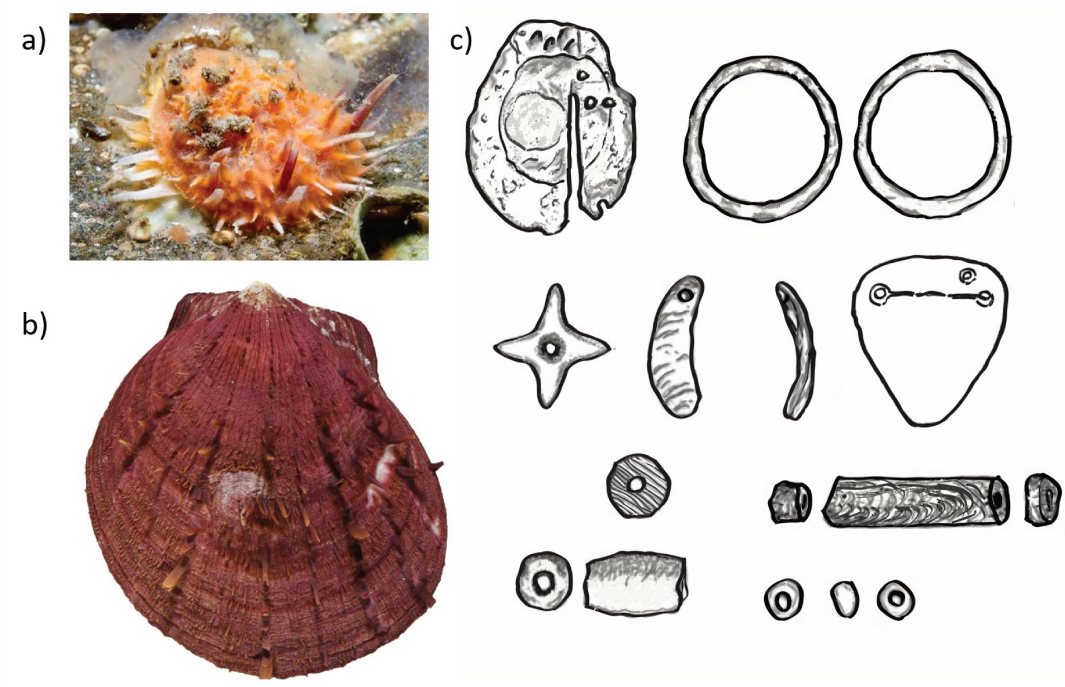


Figure 5. a) Living *Spondylus gaederopus* mollusc, covered in sponge *Crambe crambe* (*Spondylus* underwater, 2012, Sub Rimini Gian Neri, www.biologiamarina.org); b) *Spondylus* shell (photo credit J. Thomas, UMR CNRS 6282 Biogéosciences, UBFC); c) some examples of ornament types that were made of *Spondylus* shell (adapted from Borrello and Micheli, 2004; Chapman and Gaydarska, 2015).

It is interesting to consider the reasons behind material choice and the final appearance of worked ornaments. Did *Spondylus,* as species itself, held a special value, and is that why it was selected for ornament-making? Or was the choice governed by more practical reasons, such as the material and mechanical properties of the shell? First of all, the procurement of *Spondylus* was not easy. Shells found on the beach were usually eroded and thus more friable, hence their material properties were not suited for producing ornaments. To work with *Spondylus* shells, live molluscs had to be gathered, which required diving skills to depths of at least a few meters (Dimitrijević and Tripković, 2006; Chapman et al., 2008; 2011). Therefore, shell collection required time, skills and personal investment, especially in sites where production was intensive (Theodoropoulou, 2011). Secondly, shells are extremely solid materials and, compared to other substrates, *e.g.* marble, ivory or bone, they are relatively hard to work and shape

into different objects (Gurova and Bonsall, 2017; Baysal, 2019). Ornaments made of marble or other biomaterials have also been documented and found next to *Spondylus* ornaments, albeit in a much smaller scale (Dimitrijević and Tripković, 2006; Baysal, 2019). Thus, it is very likely that the use of *Spondylus* portrays reasons other than simply practical and this shell was considered as an item of prestige. The value of the material is likely due to a combination of factors, including:

i) its rarity; for example, archaeological *Spondylus* "workshops" have been documented only in a limited number of locations across the Mediterranean coast;

ii) the difficulty to procure; obtaining the shell required diving skills and knowledge where to locate it and how to tear it off its substrate;

iii) its distance from the place of origin; the "peak" of *Spondylus* shell use is documented in Central Europe and in some places it was around thousand kilometres from the nearest coast settlement on the Mediterranean Sea. Long distance routes may have added special value (and prestige) to this shell material. Interestingly, *Spondylus* sp. were also important shells in South American prehistory (Paulsen, 1974), which indicates that the prestige could also be due to its striking appearance, *i.e*. bright coloured valve with sharp and long spines.

Spondylus gaederopus represents a clear symbol of the Mediterranean Sea and could have been considered as a link to marine environments. Prehistoric societies had a long lasting relationship with marine shells, as evidenced by the continuous use of seashells since Palaeolithic times. For settled inland Neolithic populations, the use of ornaments made of *Spondylus* may have provided a personalised association with the sea (Theodoropoulou, 2011). The labour input and the skills needed in obtaining the diversity of shapes and types provided additional value to the ornaments (Baysal, 2019). Moreover, *Spondylus* shell ornaments are often considered as prestige items, because many findings are associated with rich grave goods, indicating that *Spondylus* was used for symbolic purposes or rituals. In one of the richest Copper age cemeteries in Varna (Bulgaria), *Spondylus* ornaments were found alongside a number of golden objects, suggesting that *Spondylus* was as precious as gold (Chapman et al., 2011). In addition, many broken *Spondylus* ornaments were reused or recycled into different ones, indicating a strong attachment to these "heirloom" objects and the long lifecycle of these shell artifacts (Dimitrijević and Tripković, 2006; Baysal, 2019). Considering inland archaeological sites that were further from the coast, *Spondylus* may have had an additional prestige due to its 'exotic' provenance. Indeed, the peak of *Spondylus* shell ornament use is documented in Central European sites, rather than in coastal areas with abundant material resources. However, the availability and procurement was certainly dependent on the links with the societies living on the Mediterranean coast (Borrello and Micheli, 2004; Dimitrijević and Tripković, 2006). Archaeological

research identified several *Spondylus* workshops, *e.g.* Dimini, a settlement close to the Aegean sea (Greece), a Neolithic workshop in Arene Candide on the Ligurian coast (Italy) and a production site in Orlovo, some hundred kilometres away from the Aegean Sea (Bulgaria) (Borrello and Micheli, 2004; Chapman et al., 2011; Chapman and Gaydarska, 2015). The current findings suggest that there was an active movement of *Spondylus* ornaments and there were links within different European populations.

Cultural exchanges

Shell ornaments also provide a powerful proxy to track the evolution of cultural exchange networks. *Spondylus* shell ornaments are often used as a model to study long distance networks, cross cultural interactions and even the emergence of economies in Neolithic Europe (Chapman and Gaydarska, 2015; Windler, 2017). The widespread findings of *Spondylus* shells, stretching from the Aegean Sea and as far as Northern Europe, including the Paris basin, implies that exchange routes might have stretched over 3000 km in distance (Chapman and Gaydarska, 2015). However, this theory is yet to be proven because an in-depth investigation is required in order to answer several questions (Borrello and Micheli, 2004; Dimitrijević and Tripković, 2006; Douka, 2011).

First, this impressive 3000 km *Spondylus* network, would only have existed if fresh *Spondylus* shells (*i.e.* of Neolithic age, contemporary to the makers), were sourced from the Aegean Sea, and not from elsewhere in the Mediterranean, and transported/exchanged until they reached the north of Europe. Some scholars have suggested that *Spondylus* could have also been sourced from the Black Sea, or that fossil shells were used, which were present in geological outcrops, close to Central European sites (Shackleton and Renfrew, 1970; Moshkovitz, 1971). Several studies based on geochemical isotope analyses have dismissed both of these possibilities, supporting the Mediterranean sea as the primary origin of *Spondylus* (Shackleton and Renfrew, 1970; Moshkovitz, 1971; Shackleton and Elderfield, 1990; Dimitrijević and Tripković, 2006; Bajnóczi et al., 2013). However, the distinction between the different Mediterranean locations, *i.e.* Aegean vs Adriatic Seas is not yet possible as the two marine locations do not show marked differences, as for example in isotope signatures (Bajnóczi et al., 2013). Interestingly, in archaeological research, the Ligurian Sea is rarely considered as a primary source for *Spondylus* collection, mainly because in Italian and Southern French archaeological sites, *Spondylus* findings are rather scarce. However, a Neolithic *Spondylus* workshop has been identified close to the Ligurian coast (Borrello and Micheli, 2004), suggesting that shells and/or ornaments may not have needed to travel such long distances to reach the north of Europe (Figure 6).

Moreover, there are many small-sized shell ornaments that are either *presumed* to be made of *Spondylus* or are completely unidentifiable (Dimitrijević and Tripković, 2006; Micheli and Bernardini, 2018), hence their contribution in understanding the circulation of *Spondylus* is unknown. It is highly probable that in the future, as more analyses are carried out, the long distance *Spondylus* exchange network idea may have to be revised.

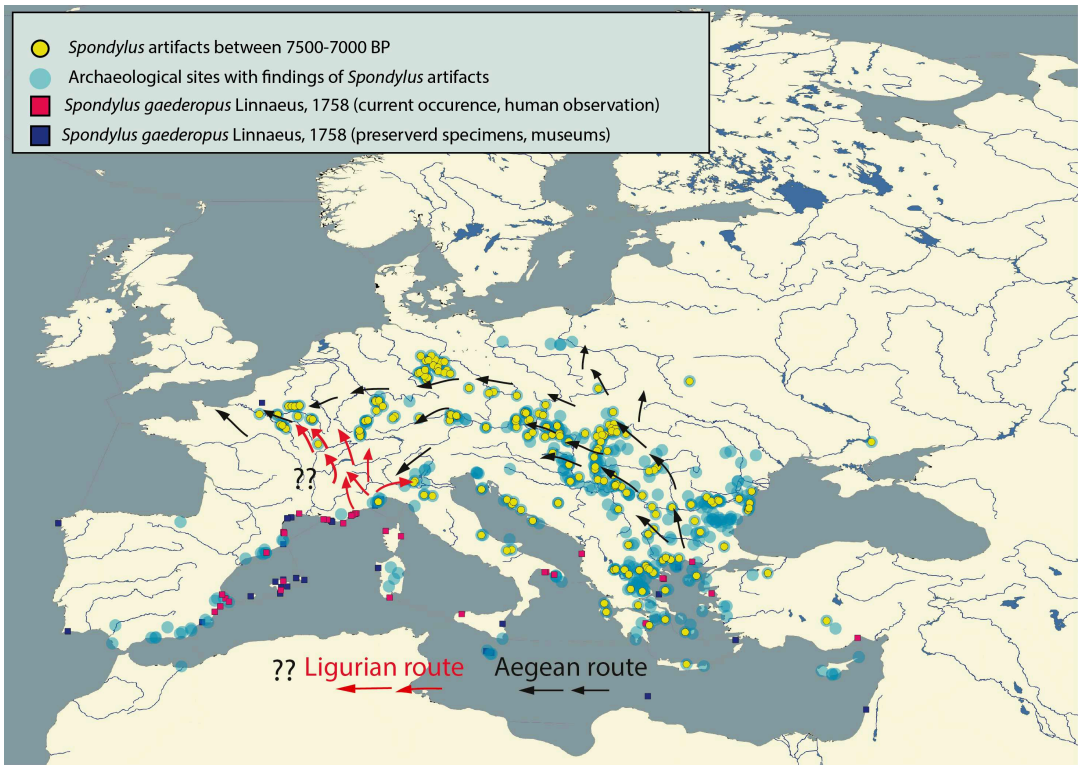


Figure 6. *Spondylus* routes in prehistoric Europe. Black arrows give a schematic representation of the 'Classical' Aegean *Spondylus* shell route directed from Mediterranean waters next to Greek coast to Central and Northern European sites. Red arrows indicate a schematic representation of a potential 'alternative' *Spondylus* route, directed from the Ligurian sea to Central and Northern European sites.

Shell ornaments. Identification?

Secure identification of archaeological worked shell items is a complex task and it is one of the greatest obstacles in archaeomalacological studies. Heavily worked, fragmented or degraded ornaments often do not retain any visible features for species discrimination (Borrello, 2005; Demarchi et al., 2014). This is particularly true for *Spondylus*, which was mostly used as a raw material. In archaeological literature, one can commonly find remarks that ornament material is undetermined,

or an assumption of shell taxon, *e.g.* "*possibly* made of *Spondylus"*. For example, sometimes ornaments are classified with just an ambiguous description, referring to a general name such as shell ornaments, cylinder, disc, irregular-type beads etc. (Nikolaidou, 2003). In archaeological literature, bracelets and pendants are noted as one of the most common types of *Spondylus* ornaments, but this may be due to a visibility bias – because of their larger size, it is easier to identify the shell. Quite often small size shell beads, which are found next to larger-size *Spondylus* ornaments, are also presumed to be made of *Spondylus*, supposing that the raw material was the same (Dimitrijević and Tripković, 2006).

Misidentifications, unsurprisingly, are quite common (Dimitrijević and Tripković, 2006; Stojanovic and Bajčev, 2016). Re-assessing and re-analysing archaeological shell ornaments from Neolithic Vinča-Belo Brdo site, the largest in the Central and Northern Balkans, Dimitrijević and Tripković showed that many bracelets previously published as *Spondylus*, were in fact made of the bittersweet clams, *Glycymeris* sp. (Dimitrijević and Tripković, 2006). Actually, *Glycymeris* bracelets were found to represent almost half of all of the studied shell bracelets. It clearly indicates that the identification of smaller artifacts, such as disc-shaped, cylindrical and oval shell beads, would be even more complex. The authors highlighted that many ornaments could be assigned as *Spondylus a priori*, without having a strong support. At the same time, the use of *Spondylus* could also be underestimated. Archaeologist David Reese notes that a small anthropomorphic figurine found in Knossos (Neolithic Crete) (Shackleton, 1968) and assigned to *Tridacna*, a giant clam which inhabits the Red Sea and thus is not native in the Mediterranean region, in fact, is probably made of *Spondylus* (personal communication, February 20, 2020). There are several other Minoan figurines among Cretan archaeological findings, which are indicated as *Tridacna* shell artifacts or, in older publications, simply described as "objects made of calcite/aragonite". These could also be made of *Spondylus* shell. The use of imported *Tridacna* shells creates a more interesting story, however it may also overshadow the persistent symbolic importance of *Spondylus* shells until protohistoric times.

The examples summarised here aimed to illustrate how species assignments are sometimes based on presumptions and thus can bias archaeological interpretations. Little is known about the full spectrum of shells used during Neolithic times. Moreover, there is a great number of small, fragmented archaeological shell artifacts that have rarely been considered for more in-depth studies. Hence, the lack of knowledge about shell diversity circulating in the past, provides only a partial picture of cultural traditions and movements in prehistoric times.

Methodological approaches for shell ornament analysis

Taxonomic shell identification of small fragmented and/or degraded archaeological shell artifacts is a complex task and remains an important issue in archaeological research. Archaeological shell ornaments are cultural heritage items and as such, require non-destructive or minimally-destructive analytical techniques. In addition, mollusc shells are complex biological substrates and have not been a subject of extensive research in archaeometry. Shells are biocomposite materials made of inorganic mineral phase (calcite and/or aragonite) and a small fraction of organics, both of which are closely associated and assembled into complex structural layers. Some of the analytical techniques which have been used to study shell materials are summarised below:

- X-ray microtomography analysis using a micro CT scan system were recently employed to study cylindrical shape ornaments, presumed to be made of *Spondylus* shell (Micheli and Bernardini, 2018). The technique is non-destructive and it works by scanning the sample with X-ray radiation. It results in a visual 3D representation of the object and its cross-section allowing to characterise the shell structure. However, the microstructural/mineralogical data has a very low taxonomic resolution because the most commonly encountered shell microstructure types, such as nacreous, prismatic, foliated and crossed-lamellar are found across many different mollusc families, comprising thousands of species (Boggild, 1930; Taylor et al., 1973; Carter, 1990).
- Isotope signatures from shell carbonate provide the information about the environment where the animals lived and formed the shell (McConnaughey and Gillikin, 2008; Leng and Lewis, 2016). However, the analysis of stable oxygen isotopes allows us to distinguish only water sources of considerably different salinity, such as freshwater vs marine (Keith et al., 1964) or Mediterranean vs Black Sea (Shackleton and Renfrew, 1970; Bajnóczi et al., 2013).
- In contrast, Demarchi et al. studied shells and archaeological ornaments using a molecular approach (Demarchi et al., 2014). The protein fraction, which is part of shell organics and is preserved over archaeological timescales, was characterised by its amino acid composition and was used to classify the different types of mollusc shells. The taxonomic resolution of the method is low (level of order), but this study showed a good future potential for molecular-based methods for shell species identification.

Palaeoproteomics, the analysis of ancient proteins preserved in sub-fossil biominerals, has been extremely successful in identifying the biological origin from fragmented or degraded archaeological and palaeontological substrates (Welker,

2018b). However, mollusc shells, and more generally all invertebrate organisms, have never been a subject of palaeoproteomic research. The development of shell palaeoproteomics may provide a new way to investigate archaeological shell ornaments and represent the central aim of this work. This interdisciplinary approach, first of all, requires in-depth understanding about the biology and biochemistry of mollusc shells, an aspect addressed in the following section.

Part 3. Shell Biomineralization

Mollusc shells and biomineralization

Mollusc shells are external calcified structures which protect and support the soft body parts of the animals. They are truly rigid substrates and are very tough biomaterials, also serving as a reserve of calcium ions, protection against predation and desiccation (Marin et al., 2012). Mollusc shells represent one of the most abundant forms of biomineralized tissues among the different metazoans, after corals (Lowenstam and Weiner, 1989; Carter, 1990). Biomineralization generally refers to a molecular and physiological process by which living systems precipitate mineral salts and produce mineralised structures (Lowenstam and Weiner, 1989). Biomineralization can be biologically induced, which means that the mineral precipitates as a result of metabolic activities and with no control on the biomineral organization (Lowenstam and Weiner, 1989). This is a typical case of some biomineralizing bacteria and algae. In contrast, the biomineralization of the mollusc shell is a biologically controlled process. The calcification occurs outside the living tissues but it is strictly specialised and genetically regulated. It is also independent from the environmental conditions and results in such a skeletal arrangement, that is completely distinct from their inorganic counterparts (McDougall and Degnan, 2018). The ability of mollusc shells to produce very hard and strongly mineralised structures have provided them an evolutionary advantage, resulting in an immensely high number of species as well as a widespread occupation of different environments (Marin et al., 2012; McDougall and Degnan, 2018). At the same time, the extremely solid structures and variety of different shell textures have won the long lasting human fascination. The formation of analogous structures *in vitro* would require elevated temperatures and specific pressure conditions, while biomineralizing organisms produce them at atmospheric settings and ambient temperature. Thus, research of shell biomineralization has a very clear aim – that is, to uncover the formula of their formation.

Shell formation

Mollusc shells are organo-mineral nanocomposite biomaterials made of calcium carbonate and a small fraction of organics (Marin and Luquet, 2004; Marin et al., 2012; 2016). Calcite and/or aragonite, the two main $CaCO_3$ polymorphs, make up the most of shell skeleton and are in close association with the organic phase (Figure 7), together assembled in a variety of different microstructures. The shell organic matrix is composed of proteins, saccharides, lipids and other metabolites (Figure 7). The biomolecules represent only a minor component of the shell (just

up to one or two percent, at best) but they are actually the main driving force of biomineralization process (Weiner and Addadi, 1997; Checa, 2018) and are the key element in maintaining the exceptional material properties. The organics have an impact on the structure hardness, provide flexibility and dissipate crack propagations, giving a superior structural stability and contributing to the overall toughness of the shell (Gilbert et al., 2005).

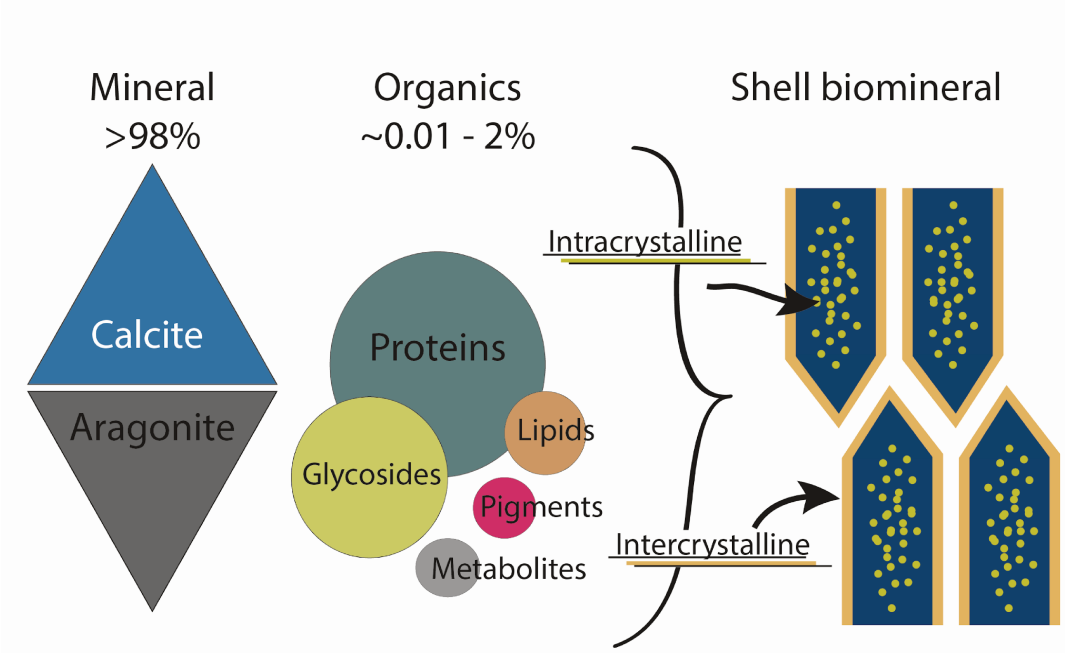


Figure 7. The graph notes the typical composition of the shell biomineral system and its mineral and organic phases. On the right hand side, the illustration displays biomineral structure and the presence of intracrystalline and intercrystalline proteins.

Shell biomineralization is regulated by the organic extracellular matrix and starts at the interface between the mantle, periostracum and the shell, *i.e.* the mineralization front (Figure 8) (Lowenstam and Weiner, 1989). The periostracum is a thin organic coating and the outermost layer of the shell. It functions by sealing the space and, during the first phases of mineral deposition, it also acts as a primary template. The mantle is a thin, ciliated and highly muscular tissue lining the inner surface of the shell and it is the main molluscan organ regulating shell formation. The mantle is composed of two epithelial cell layers, the inner and the outer. The biomineralization process is controlled by the outer epithelium and distinct zones within it are responsible for the formation of different microstructures (Marin et al., 2012). The cells of the outer epithelium secrete the macromolecular matrix and extrude the inorganic precursors of calcium carbonate (Ca^{2+} and HCO_3^-) into the space between the mineralisation front, where the precursors later assemble into

precise order. This area, called the extrapallial space, is filled with a fluid that is saturated in calcium carbonate and enriched in organic molecules (*e.g* proteins, glycoproteins). However, the full biochemical composition of this extrapallial fluid is not really understood (Marin et al., 2012).

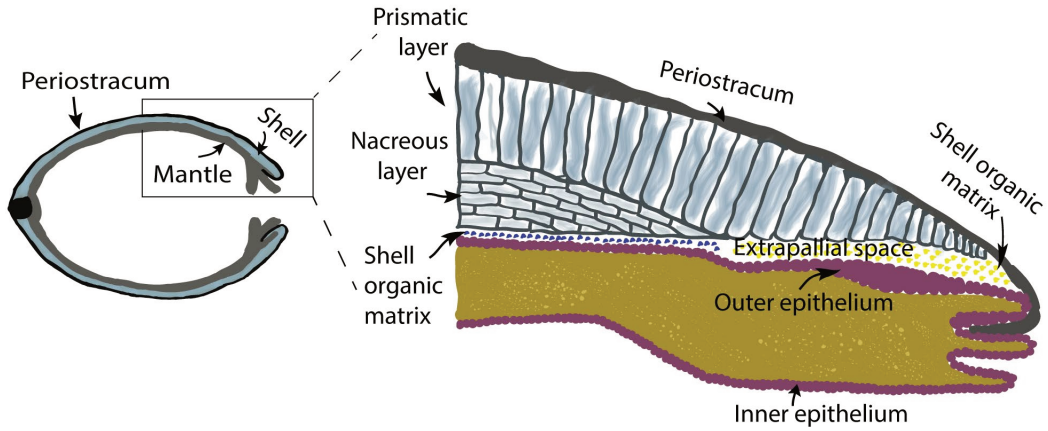


Figure 8. A representative scheme of mineralisation front where biomineralization occurs. The picture shows a schematic cross section of the shell and main molluscan parts involved in shell formation. The representation is based on nacro-prismatic shell microstructure.

The shell macromolecular matrix contains a specific set of proteins, glycoproteins, proteoglycans, polysaccharides (including chitin), lipids (Marin and Luquet, 2004) and other metabolites. This array of macromolecules interacts with the inorganic ions in a well-controlled manner and self-assemble to form the biomineralized structures (Marin et al., 2013). They are intimately associated with the mineral and have several roles in mediating crystal formation. First of all, the matrix only allows the crystallisation to occur where it is intended. The extrapallial fluid is highly saturated in inorganic precursors, hence the matrix has an inhibiting effect, aiming to prevent spontaneous crystallisation. Secondly, the organic matrix acts as the main nucleating agent, creating the microarchitectural 3D framework for crystal growth (Marin et al., 2013; Checa, 2018). It also selects the polymorph of the mineral (calcite or aragonite) and regulates the orientation of the crystal. Calcium carbonate nanoaggregates cluster and self-orientate to form bigger mineralised structures, which have multiscale structures, in the meantime entrapping some of the organic macromolecules (Weiner and Addadi, 1997; Cölfen and Antonietti, 2005; Marin et al., 2013). The biomolecules which get occluded inside the mineral skeleton are known as intracrystalline (Figure 7) (Albeck et al., 1993; 1996).

Molluscs display a high variety of biomineralized shell structures. There are more than 30 shell microstructures currently described among which the most common

are nacre (the mother-of-pearl), prismatic, crossed-lamellar (CL) and foliated (Figure 9) (Taylor et al., 1973; Carter, 1990). The microstructural patterns, in general, depend on the shell mineralogy. For example, prismatic, foliated and crossed-foliated microstructures are typically calcitic (albeit in some shells prismatic layer can also be aragonitic) while nacre and crossed-lamellar structures are exclusively aragonitic.

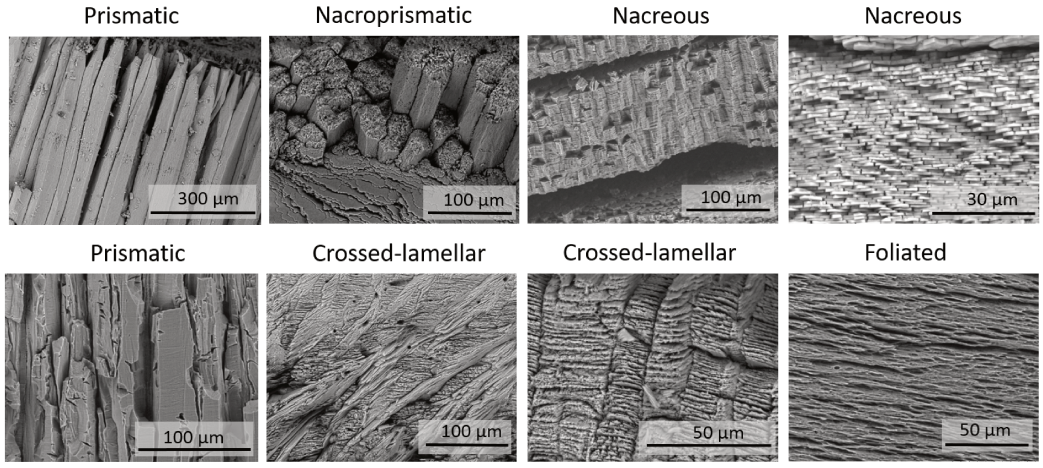


Figure 9. SEM micrographs of some of the most common types of shell microstructures: different types of prisms (*Pseudunio auricularius* and *Spondylus gaederopus*, respectively), superimposed prisms and nacre (nacro-prismatic), nacre (*Unio crassus*), crossed-lamellar and foliated (*Spondylus gaederopus*). Photographs taken by JS, at UMR, CNRS 6282 Biogéosciences, UBFC.

Nacre, also known as mother-of-pearl, is the best-known and most investigated shell microstructure from both the molecular and the structural point of view. The interest of nacre lies in its exceptional material properties (Marin et al., 2012)) and because it is intriguing from the evolutionary point of view (Jackson et al., 2010). Nacre biomineralization can be traced to the Lower Cambrian period and it is found in many different molluscan classes (Taylor et al., 1973; Carter, 1990). It is also often regarded as the ancestral shell microstructure. In simple terms nacre can be characterised by polygonal or rounded flat tablets that are arranged in well-defined, parallel sheets. It possesses high structural toughness and resistance to cracks. In fact, it is known to be at least thousand times more resistant to fractures compared to its geological mineral form, the aragonite alone (Currey, 1977; Jackson et al., 1990; Katti et al., 2005). Most of the models that describe fundamentals of shell biomineralization are based on nacre, due to its relative simplicity of its geometrical organisation ((Weiner and Traub, 1984; Lowenstam and Weiner, 1989; Addadi and

Weiner, 1997; Weiner and Addadi, 1997). However, the ‘nacre model’ fails to accurately describe the molecular features of biomineralization for other types of microstructures. Indeed, the different strategies that molluscs employ to fabricate different microstructures are mostly unknown (Checa, 2018).

Shell organic matrix

This section will provide an in-depth overview of different biomolecular components that constitute the shell organic matrix and their possible function in mineralisation. The shell organic matrix is a complex mixture of macromolecules, it makes up from 0.01 % to 4% of the total shell weight, however, in reality, for most of the shells, it rarely exceeds 1% (Marin et al., 2016). The quantity of the matrix is dependent on the type of microstructure. Nacreous and prismatic structures are known to be more rich in organics (>1 %), while foliated and crossed-lamellar shells yield very low organics content. The shell organic matrix can be present in two fractions – the intracrystalline and intercrystalline (Figure 7). As previously mentioned, during shell formation, the intracrystalline matrix gets occluded inside the shell crystallites and remains strongly bound to the mineral crystals (Albeck et al., 1993; 1996; Marin and Luquet, 2004; Marin et al., 2007b; 2014). Intercrystalline organics, on the other hand, characterise a fraction that lies between the crystals, gluing them together (Lowenstam and Weiner, 1989).

Many shell organic matrices are characterised by poor solubility in aqueous solutions. Early biochemical studies of this matrix identified the presence of acidic proteins, glycoproteins, hydrophobic silk-like proteins and chitin (Lowenstam and Weiner, 1989). Following this, a so-called sandwich model was used to describe and illustrate the first prototype of the shell matrix assembly, which was based on a nacreous shell model. The inside core was made of chitin, an insoluble saccharidic polymer made of exclusively N-acetyl-glucosamine monomers. The core was then surrounded by hydrophobic silk-like proteins, which provided flexibility to the structure. These were further coated by acidic proteins or glycoproteins (Weiner and Traub, 1984) which could directly bind the mineral via the negatively charged groups (Gilbert et al., 2005). Acidic proteins were typically considered as intracrystalline (Albeck et al., 1996). The silk-like proteins, which are rich in alanine and glycine amino acids, and are very hydrophobic, on the other hand, were mostly considered as intercrystalline. However, this simple and clear model did not fit with biochemical profiles of other shell microstructures. For example, silk-like proteins were found almost absent in crossed-lamellar shells, indicating that they are not unequivocally present in all of the shell frameworks (Lowenstam and Weiner, 1989).

The advancements in biochemistry and the development of 'omic-based approaches, in recent decades, had a significant impact on shell protein studies, providing the possibility to identify hundreds of different shell sequences (Marin et al., 2014). At the same time, it rendered the simple 'sandwich' model obsolete. Proteomic studies showed that shell matrices contain a great diversity of proteins with a variety of different functions, challenging the primary view of mineral associated sequences.

Shell proteins

Proteomics, transcriptomics and to a lesser extent genomics, are the main 'omic approaches to study proteins and genes involved in shell biomineralization and are often used together (Figure 10) (Marin et al., 2016).

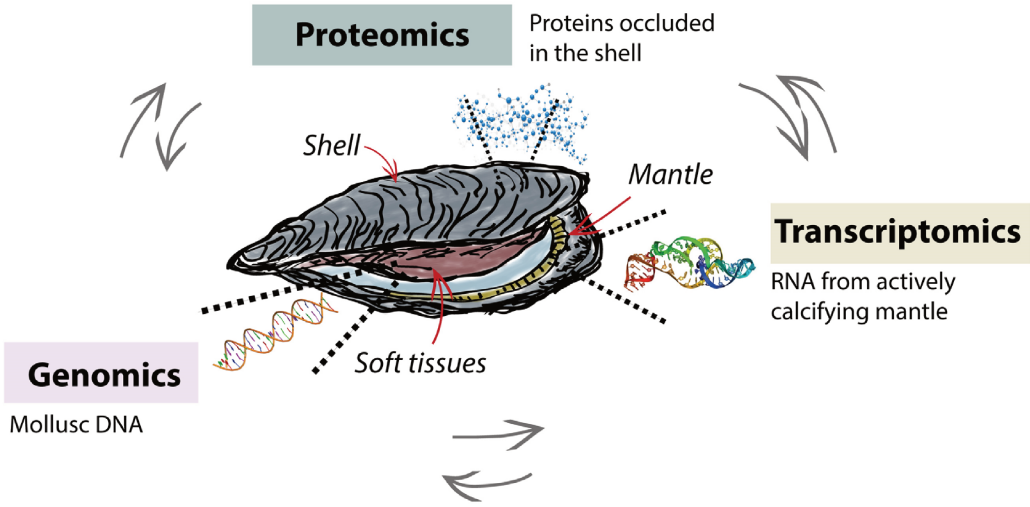


Figure 10. 'Omic' based tools used for biomolecular studies of shell biomineralization.

The analysis of the whole molluscan genome provides complete DNA sequences and allows to investigate biomineralizing genes (Powell et al., 2018). Transcriptomics, via high-throughput sequencing, is used to characterise messenger RNAs from the actively calcifying molluscan mantle, the main organ involved in shell formation. These RNAs allow the identification of transcribed genes that are directly involved in biomineralization (Jackson et al., 2010). On the other hand, proteomics is used to analyse shell matrix proteins, which are extracted from the calcified shells. When used together, the combination of these techniques enables to identify the full shell proteome, including sequences that are species specific (Marie et al., 2013b; Herlitze et al., 2018; Zhao et al., 2018). However, 'omic

techniques do not provide information about protein function and this remains a task of additional *in vitro* studies. As the *in vitro* experiments are complex and time consuming, it explains why, unfortunately, a large part of sequenced shell data, deposited in databases remain without detailed functional characterisation.

The study of shell proteins offers a close look into the molecular mechanisms of shell formation (Marin and Luquet, 2004; Sarashina and Endo, 2006; Marin et al., 2012; 2013; 2016; Kocot et al., 2016). The fundamental concept that laid the foundations of molecular biomineralization studies was that there may be a basic protein kit that molluscs use to synthesize their shell, *i.e.* known as the "molecular toolbox" of biomineralization (Marin and Luquet, 2004; Marin et al., 2012; 2013; 2016). This was the driving force of numerous 'omics-based studies targeting to characterise shell matrix proteins (SMPs) or genes, also known as the 'shellomes', and compare them between different mollusc species (Marin et al., 2013).

However, the search of the basic and universal "molecular toolbox" of biomineralization appeared to be difficult and the concept itself – perplexing. First of all, shell proteomes were found to be very distinct between different mollusc species, suggesting a rapid evolution of shell matrix proteins (Kocot et al., 2016; Marin et al., 2016). Different protein types were also identified in separate microstructural layers from the same shell (Marin and Luquet, 2005; Marie et al., 2012). For example the proteomics study of a pearl oyster shell, which has a nacro-prismatic microstructure, showed that there is a rather clear distinction between the set of proteins from nacre and the prisms (Marie et al., 2012). Less than one tenth of identified sequences were shared by the two microstructures, indicating that molluscs synthesise different protein sets to handle these organizational, functional and structural differences. However, at the same time, shell proteins did not show a simple association to the specific microstructure. For example, the study of different nacreous shells, including the pearl oyster, abalone and freshwater mother-of-pearl mussels (Marie et al., 2009a; 2009b; Jackson et al., 2010), evidenced that the protein and gene sets associated to nacre formation, were very different in all of the studied specimens. This has led to suggest that the evolution of nacre occurred independently, in different lineages and at least several times. It may also explain why many shell protein sequences are identified as species specific (Marie et al., 2013b). Thus, shell proteome studies revealed the diversity of molluscan shell sequences, which will be reviewed in detail in the following section.

Shell protein characteristics

The comparison of different shell proteomes characterised from a variety of mollusc species, brought to light the shared features, but also the differences between sequences. Typically, in one shell model around several tens and up to a hundred of shell proteins are identified, but this number can be even higher, when genomics data are available. The majority of shell proteomes currently known are characterised from nacreous shells, such as the pearl oysters (*Pinctada*) (Jackson et al., 2010; Marie et al., 2012), abalone gastropods (*Haliotis* sp.) (Jackson et al., 2010; Marie et al., 2010), marine and freshwater mussels (Mytilida, Unionoida shells) (Marie et al., 2007; Suzuki et al., 2011; Ramos-Silva et al., 2012; Gao et al., 2015). Considering other shell microstructures, proteins were identified in the giant owl limpet *Lottia gigantea* which is characterised by a very complex structural arrangement (Mann et al., 2012; Marie et al., 2013b; Mann and Edsinger, 2014). Shellomic studies were also carried out on some of the crossed-lamellar shells, such as the great pond snail (*Lymnaea stagnalis*) and Venus clam (*Venerupis philippinarum*) (Marie et al., 2011; Herlitze et al., 2018) as well as from scallop shells (*Pecten*), which have calcitic foliated structures (Sarashina and Endo, 2001; Yarra et al., 2016; Mao et al., 2018).

There are a number of conserved protein domains that are found among shellomes of different shell species and can be considered as signatures of shell matrix proteins (Figure 11). Most of the domains are associated with mineral-binding activity, interaction with carbohydrates (including chitin) as well as enzymatic activity (Marin, 2020). Von Willebrand factor-A (VWA) is a glycoprotein domain. In shell proteins it is often present together with chitin-binding domains, which are involved in the interaction with chitin, the insoluble polysaccharide polymer. The VWA likely plays an adhesion function through protein-protein interactions and it is found in many shell protein sequences, as for example in BSMP (Basic Shell Matrix Protein) and Pif proteins (Sarashina and Endo, 2006; Marin et al., 2016; Arivalagan et al., 2017). The C-type lectin-binding domains have been identified in a variety of skeletal matrix sequences, such as the nacreous protein Perlucin, and play a function in mediating glycoside binding activity in the presence of Ca^{2+} ions (Sarashina and Endo, 2006; Marin et al., 2016; Arivalagan et al., 2017). Sushi, EGF-like (epidermal growth factor) and fibronectin domains (Marin et al., 2016; Arivalagan et al., 2017) are involved in calcium binding activity and have been identified in a number of extracellular matrix-related shell proteins (Marie et al., 2013b; Arivalagan et al., 2017; Feng et al., 2017). A number of enzymes are found in shell proteomes which play an important function during the shell formation. Tyrosinase enzyme is involved in the hardening process of the exoskeleton by catalysing the protein cross-linking in order to form a stabilised structure

(Arivalagan et al., 2017). The main function of chitinase enzyme is to degrade the chitin polymer, but it is also involved in the remodelling process of the shell (Arivalagan et al., 2017). Carbonic anhydrase (typically abbreviated as CA) assists the conversion of carbon dioxide to bicarbonate, which is one of the two main precursors for shell biomineralization. Interestingly, an in-depth sequence investigation of carbonic anhydrase enzymes identified from a variety of molluscan shells suggested that CA had a complex evolution and likely underwent several independent recruitments for shell formation, in different mollusc shell lineages (Le Roy et al., 2014).

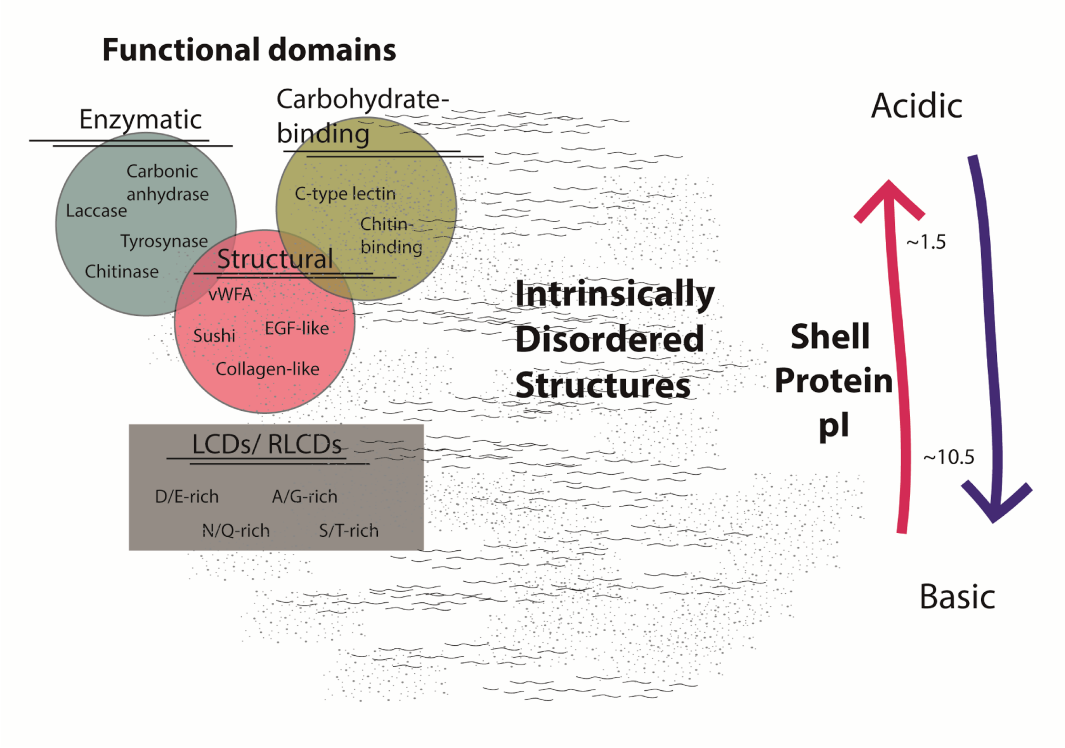


Figure 11. A schematic representation of typical shell protein features and characteristics: most common functional domains, types of low complexity domains/repetitive low complexity domains (LCDs/ RLCDs), the variation of protein isoelectric point (pI). In the middle, the scheme highlights that many shell proteins adopt intrinsically disordered structures.

Many mineral-associated shell matrix proteins share a common feature, that is their acidic nature (Albeck et al., 1996; Marin and Luquet, 2008; Samata et al., 2008; Suzuki et al., 2009). They have a low theoretical pI (the isoelectric point, the pH value at which the net charge of the protein is zero) and the sequences are characterised by a high proportion of aspartic amino acid residues. For example,

glycoprotein MSP-1, identified in the calcitic layer of *Pecten* shell, has a pI of 3.2 and Aspein, an acidic protein identified in *Pinctada* shell, has a pI value of 1.45, something unique in the living world (Sarashina and Endo, 2001; 2006). Some of these proteins can be heavily glycosylated and phosphorylated, suggesting that in their biological form, the actual pI is even lower. Acidic shell proteins and glycoproteins show a strong mineral-binding activity and may play a central role in the shell biomineralization process (Albeck et al., 1996). On the other hand, basic shell matrix proteins are also found in shells, some of which had a rather high pI value, around pI = 10 (Ramos-Silva et al., 2012; Liu et al., 2017; Jin et al., 2020). However the function of basic shell matrix proteins is poorly understood (Marin et al., 2013).

Another commonly encountered shell matrix protein feature is the preponderance of low complexity domains (*i.e.* known as LCDs) (Figure 11). LCDs are characterised by biased amino acid compositions, which are basically dominated by just one to three same amino acids (Sarashina and Endo, 2006; Kocot et al., 2016; Marin et al., 2016). Some of these LCD domains appear as repetitive motifs along the whole protein sequence, hence are noted as repetitive low complexity domains (RLCDs). Most repeats are short, with around ten amino acids per unit, although longer repeats also do occur (Kocot et al., 2016). Detailed study of mineral-associated shell proteins showed that the most common LCD domains are represented by acidic and hydrophobic amino acid residues (Marin et al., 2016). The repetitive acidic domains, usually enriched in aspartic amino acids, are probably involved in protein-mineral interaction and likely function as potential binding sites for calcium ions (Evans, 2019). The hydrophobic LCDs are characterised by long stretches of alanine and glycine amino acid residues and, in protein sequences, often appear between the acidic regions or functional domains. They likely play a structural role by creating the hydrogel-like structure which provides the flexibility for matrix assembly (Sarashina and Endo, 2006). LC/RLC domains which have glutamine, asparagine, lysine, arginine, histidine, cysteine or methionine amino acids have also been identified, however the function of these domains is not well understood (Marin et al., 2016; Marin, 2020). Overall, given the high variety and abundance of RLCD domains in shell matrix proteins, it is clear that they play an important functional role in biomineralization.

A large part of shell matrix proteins which have numerous LC/RLC domains, also adopt intrinsically disordered structures (Figure 11) (Jackson et al., 2010; Evans, 2012; Boskey and Villarreal-Ramirez, 2016). For example, a bioinformatics study of 39 different mollusc shell proteins showed that all of the sequences possessed at least several intrinsically disordered regions (IDRs) (Evans, 2012). Intrinsically disordered proteins (IDPs) are types of sequences that lack fixed or ordered three-

dimensional structure (Oldfield et al., 2019). The presence of IDRs indicate that these proteins are conformationally unstable and thus, are prone to interact with other proteins or with the mineral phase. This enables the proteins to gain some secondary structure, that is – order (Evans, 2012). The IDRs are rather flexible and their ability to adapt to multiple binding partners may explain why disordered structures are found in almost all biomineralizing systems (Evans, 2012; 2019; Boskey and Villarreal-Ramirez, 2016). Moreover, IDR regions with RLC domains often appear to be present next to conserved protein domains suggesting that shell sequences can effectively play several functions. The IDRs are known to evolve fast and this likely explains why a high number of novel shell proteins are identified as taxon specific (Kocot et al., 2016).

Shell protein sequences are also characterised by the presence of numerous post-translational modifications (PTMs), predominantly glycosylation and phosphorylation (Borbas et al., 1991; Sarashina and Endo, 2006; Samata et al., 2008). The presence of glycosylated residues acts by lowering the pI of the protein and may provide additional sites for binding the mineral or interacting with water molecules (Sarashina and Endo, 2006) (Albeck et al., 1996). Moreover, glycoproteins function in protein folding and receptor adhesion. However there is limited knowledge about the different glycoside residues and oligosaccharides present in shell matrices due to the difficulty to characterise them (Marin et al., 2016; Takeuchi et al., 2018). Phosphorylation is one of the most abundant PTM occurring in the biological world, however, the presence and extent of phosphorylation in shell proteins, as well as in skeletal proteomes of other biomineralizing invertebrates, has not been thoroughly investigated. Many shell proteins show potential sites of phosphorylation at serine amino acid residues (Sarashina and Endo, 2006; Samata et al., 2008) and phosphorus compounds have been identified in the shell matrices of some crossed lamellar-shells (Agbaje et al., 2018). The presence of phosphorylated groups suggest that the actual pI of the protein is lower than is theoretically calculated and phosphorylated proteins may also be involved in calcium-binding activity, modulating crystal growth and shape (Du et al., 2018). Bioinformatics data analysis tools were also applied to investigate the phosphorylation PTMs in shell sequences, allowing to identify additional shell proteins (Mann and Edsinger, 2014). Future application of data analysis techniques could help to systematically investigate and reassess the PTMs among shell matrix proteins.

Other biomolecular components

In contrast to proteins, the knowledge on other biomolecules that make up the shell organic matrix, including saccharides, lipids, pigments and other macromolecules, is still limited. Recent studies have shown that chitin was not a very abundant macromolecule in shells, suggesting that its role may have been overestimated (Agbaje et al., 2018). On the other hand, several shell models (Agbaje et al., 2019; Oudot et al., 2020) as well as other biomineralized invertebrate systems (Kanold et al., 2015; Takeuchi et al., 2018), showed that organic matrices contained a significant amount of saccharides. The carbohydrates could be bound to proteins and may be involved in protein-mineral interaction (Albeck et al., 1996; Marie et al., 2007) or they might be existing as free polysaccharides and involved in structuring skeletal framework.

Even less is known about the presence of lipids and pigments, mainly due to difficulties in extraction and characterization (Rousseau et al., 2006; Williams, 2017). Some of the shell pigments were found to be existing as chromoproteins, *i.e.* bound to protein sequences (Affenzeller et al., 2019; 2020). However, the precise chemical structures of these coloured compounds have not been systematically investigated and characterised and the way in which they are associated to the mineral phase, *e.g.* if they are inter- or intracrystalline, is not well understood (Cusack et al., 1992; Sun et al., 2015). Future investigation is essential to gain a holistic understanding on how shell biomolecules work together in order to create complex biomineral structures.

Analytical difficulties of shell protein characterisation

The comprehensive characterisation of shell matrix proteins is a challenging task owing to the complexity of shell matrix components and the intrinsic peculiarities of shell proteins. This section overviews the common limitations and biases that are important to consider for proteomic shell studies.

First of all, different shell species can have very diverse protein sets and thus, the full identification of shell proteome depends on how much sequence data are available in public databases. A comprehensive characterisation of shell proteomes can also be achieved by using multiple ‘omic approaches – *i.e.* when shell proteomics data are combined with genomics and/or transcriptomics. However, the genomics data may not always be available, especially for species that are rare, difficult to harvest or are even extinct. In this case, the protein identification relies on the biomolecular data of closely related species, which however, may only partially represent the true analysed proteome (Marin et al., 2013).

Secondly, the analysis of shell proteins by mass spectrometric techniques requires reviewing common analytical steps. As discussed in previous sections, shell proteins possess peculiar sequence motifs, frequently characterised by the presence of repetitive low complexity domains (RLCDs) (Marin and Luquet, 2004; Marin et al., 2013). Standard proteomic workflow procedures are poorly adapted to identify such sequences. In a typical "bottom-up" proteomics experiment (more details about the different mass spectrometry approaches are presented in Part 4) trypsin is used to digest the protein into peptides before the sample is analysed by mass spectrometry. However, trypsin cuts the sequences into peptides only after arginine and lysine, thus shell protein domains with long stretches of RLCDs that do not contain any of these two residues, cannot be accessed (Figure 12). In simple terms, this means that these types of proteins remain undetected and not identified (Marin et al., 2016). The use of different enzymes such as elastase, has shown positive results to analyse complex protein mixtures (Rietschel et al., 2009). Elastase has been applied to analyse phosphorylated proteins (Wang et al., 2008) and its use is almost a standard practise in bone proteomics (Schroeter et al., 2016). Therefore, it has also been tested for shell proteomics and will be further presented in this thesis (Chapter 5).

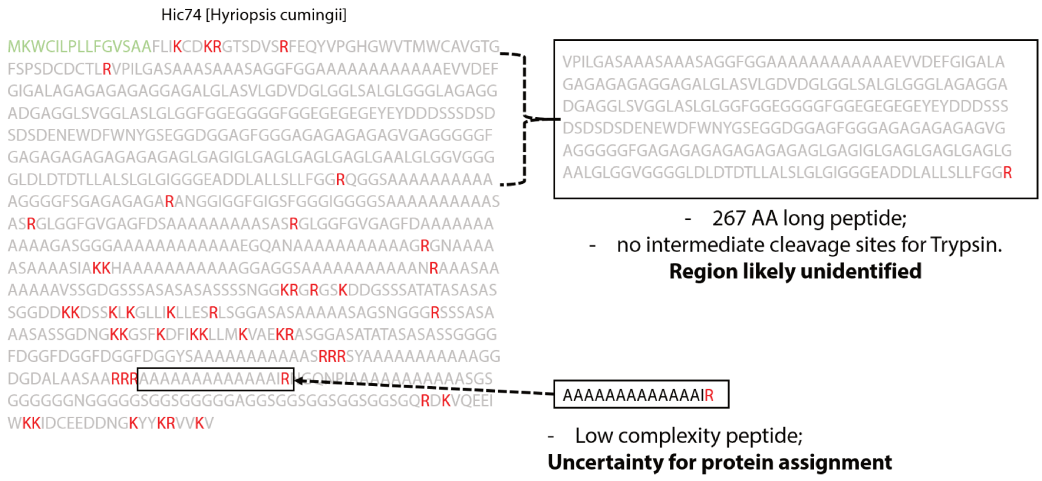


Figure 12. Shell matrix protein Hic74 from freshwater pearl mussel *Hyriopsis cumingii*. Amino acids highlighted in red mark Lys/Arg residues where protein cleavage occurs using enzyme trypsin. The figure notes biases that typically occur when analysing shell proteins: i) the lack of suitable cleavage sites using protease trypsin for "bottom-up" proteomics by LC-MS/MS. The sequence zoomed on the right shows a 247 amino acid long peptide which will likely be undetected by a typical LC-MS/MS system ii) uncertain assignment of peptides with low complexity domains. The peptide zoomed on the bottom right shows a sequence that would be theoretically identified in a proteomics experiment. The assignment of such peptide to specific protein is difficult. Green colour highlights the signal peptide.

Thirdly, there is a lack of bioinformatic data tools suitable for an in-depth characterisation and for a systematic classification of shell sequences. The identification, characterisation and homology search of shell proteins is also affected by the presence of LCD/RLCD domains. Peptides, which exclusively represent RLCD regions, due to their ubiquity, cannot be easily and securely assigned to specific protein (Figure 12) (Marin et al., 2016). For protein homology search, the bioinformatic data analysis tools, as for example BLASTp search engine, are poorly equipped in characterizing proteins with very long RLCD regions (Marie et al., 2017). This limits to gain a comprehensive and systematic classification of shell proteins and partially explains why so many shell sequences are found to be novel and species specific (Marie et al., 2011; 2013b; Marin et al., 2016; Herlitze et al., 2018; Oudot et al., 2020). New bioinformatic methods have been recently developed to investigate sequence homology of different biomineralizing shell proteins (Skeffington and Donath, 2020) and also to identify intrinsically disordered regions (Dosztányi et al., 2005; Mészáros et al., 2018). Some of these bioinformatics techniques have been investigated in this thesis and will be presented in the upcoming chapters.

Finally, there is no standardised method for shell protein extraction and there is partial ambiguity about which sequences are mineral-bound and of interest in biomineralization research. Shell biominerals are ‘open systems’ and presence of cellular and/or environmental contaminants is frequent (Marin et al., 2016). Most shellomic studies use an oxidising shell pre-treatment (by bleaching), which removes contamination and isolates a fraction that is strictly associated to calcium carbonate mineral skeleton, *i.e.* intracrystalline (Marie et al., 2013a). However the intercrystalline organics, which are thus lost, may also be of great interest in shell formation (Marin et al., 2016). Some studies have employed milder cleaning practises in order to detect extra-crystalline proteins and different PTMs, such as phosphorylation (Mann and Edsinger, 2014) or double bleaching treatments in order to compare the intracrystalline vs intercrystalline organics (Takeuchi et al., 2018; Oudot et al., 2020). However, the lack of standardisation may also hamper comparing different shell protein sets and obtaining a systematic classification of mineral-bound sequences.

Biomineralization studies of rare shell models

The majority of molecular and structural shell biomineralization studies have used nacreous shells as their model systems, which overshadows a large part of the shell species and texture diversity present in nature. The available molecular shell data are also biased towards a set of species with specific commercial importance,

for example pearl producing oysters (*Pinctada* sp.), edible molluscs such as oysters (*Crassostrea* sp., *Ostrea* sp.), mussels (*Mytilus* sp.) and scallops (*Pecten* sp.) (Samata et al., 2008; Marie et al., 2012; Gao et al., 2015; Yarra et al., 2016; Arivalagan et al., 2017; Mao et al., 2018). There are also more data for species that can be easily harvested *e.g.* pond snails, *Lymnaea* sp., land snail, *Helix aspersa* (Pavat et al., 2012; Marie et al., 2013b; Herlitze et al., 2018) compared to molluscs that live in more secluded areas. The scarcity of information is particularly evident for crossed-lamellar (CL) shells despite their interesting features. First of all, this microstructure is the most common among the Mollusca phylum. It is well-represented among clades (for example venerid bivalves) that radiated recently, in the Cenozoic times. Furthermore, crossed-lamellar microstructures exhibit a very complex arrangement (Carter, 1990; Böhm et al., 2016; Checa, 2018) and have a very high fracture toughness (Kamat et al., 2000), comparable to that of nacre.

Interestingly, the biochemistry of crossed-lamellar shells shows distinct features compared to other types of shells (Kamat et al., 2000; Mann and Jackson, 2014; Agbaje et al., 2018; 2019; Herlitze et al., 2018). The proteomic analyses of several CL shells displayed that most of the identified sequences were 'unique', *i.e.* not found in other studied shell models (Marie et al., 2011; Mann and Jackson, 2014; Herlitze et al., 2018). The organic matrices of CL shells are abundant in carbohydrates (*e.g* oligosaccharides) and are enriched in mannose, galactose, as well as galactosamine moieties (Osuna-Mascaró et al., 2014; Herlitze et al., 2018; Agbaje et al., 2019). The identification of glucosamine residues suggests the existence of chitin (Herlitze et al., 2018), however, it is likely not a rule because in other CL shells, chitin was found to be only as a minor component (Agbaje et al., 2018). In general, one of the most common features of crossed-lamellar shells is their extremely low content of organics fraction (Agbaje et al., 2018; 2019). The fraction may be just a hundredth of that found in nacreous shells. This could be due to the absence of intercrystalline silk-like fibroin proteins (Lowenstam and Weiner, 1989; Osuna-Mascaró et al., 2014; Herlitze et al., 2018), which are not identified in crossed-lamellar structure shells. Palmer has argued that in molluscs the "cost" to make the shells is mostly due to the "cost" to produce the shell organic matrix (Palmer, 1992). In the course of evolution, crossed-lamellar shells may represent such systems, evolved to assemble very tough biomineralized skeletons that do not require a very high amount of the "expensive" shell matrix, as in the case of nacre. Yet, the absence of information regarding the molecular and structural features of crossed-lamellar microstructures hinders a more complete and holistic understanding of shell biomineralization.

Shell biomineralization. Insights into multidisciplinary research.

For years, shells have been a source of inspiration. The study of shell biomineralization can provide valuable insights in different scientific fields and seed the growth of multidisciplinary research. The knowledge is now shared between different fields of biomimetic research including the search for novel bioceramics, bone regeneration and is expected to foresee even more bioinspired applications. The ability to unlock the key aspects of the biomineralization process, at least in part, one day may allow us to mimic shells and create sophisticated microarchitectures.

Shells are also a remarkable source for in-depth “palaeo” studies. The mineralised skeletons of fossil shells preserve well in palaeontological record and can provide palaeoclimatic information. Moreover, ancient shell biomolecules, entrapped in these fossil shells, may enable us to decipher the biological information. In particular, the study of ancient shell proteins could be used to investigate archaeological questions. The next chapter will overview how ancient protein studies, known as ‘palaeoproteomics’, could be applied to study archaeological and palaeontological shells and their artifacts.

Part 4. Shell Palaeoproteomics

Ancient protein studies

The study of ancient proteins has its roots in the mid-20th century, when the burgeoning field of geochemistry linked to fossil fuels exploitation paved the way to an interest into finding the original organic components (including amino acids) of fossil samples (Abelson, 1954). Mollusc shells, which are characterised by strongly mineralised structures, exhibit good protein preservation and were therefore regarded as one of the most promising systems for finding ancient biomolecules. These investigations represent the foundations of palaeobiogeochemistry. In the early phases of ancient protein research, the main techniques used to characterise biomolecules in fossil shells were based on the analysis of their bulk amino acid composition and, later, on immunochemical approaches (Figure 13). During the 1970s and 1980s, numerous studies were published, which reported the findings of million-year-old molluscan proteins and marking the beginning of a 'golden age' in fossil protein research (Jope, 1967; Matter et al., 1969; Akiyama and Wyckoff, 1970; Grégoire and Voss-Foucart, 1970; de Jong et al., 1974; Westbroek et al., 1979). For example, in 1976 Weiner and colleagues published one of the most famous works on the discovery and characterisation of 80 million-years-old proteins preserved in mollusc shells (Weiner et al., 1976). The identified "fossil" glycoproteins resembled those found in modern specimens, on the basis of their amino acid composition and biochemistry. However, at that time, no reference sequence had been published for molluscan proteins. In later studies, immunochemical methods with targeted antibodies were used to identify presence of active epitopes (antibody binding sites) in fossil samples (de Jong et al., 1974; Westbroek et al., 1979) and were also employed to detect structural fossil protein features (Collins et al., 1991).

However, from a modern day perspective, the analytical tools at that time were limited and their qualitative resolution was rather low. The fundamental issue is that there was not sufficient evidence for the authenticity of the extracted proteins. Amino acid analysis provided the composition data of the bulk protein fraction, but did not enable the secure identification of the actual fossil protein *sequences*. On the other hand, immunological analyses with antibodies, which are used to detect intact shell proteins (epitopes), can also be very sensitive to environmental or laboratory-induced contamination (Brandt et al., 2002; Buckley et al., 2017). The studies also noted that it was difficult to extract and separate original organic components from those due to environmental contamination (Westbroek et al., 1979).

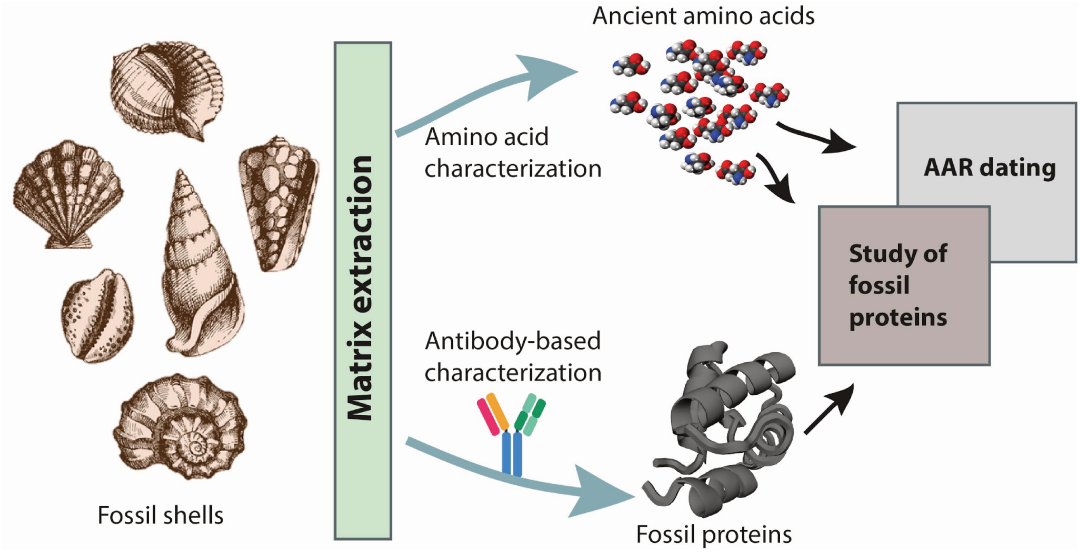


Figure 13. Early studies of fossil proteins. The figure shows the two main approaches used to identify and characterise fossil sequences in mollusc shells – characterization of fossil amino acids and immunological analyses to identify binding epitopes in fossil peptides/proteins.

The study of fossil shell proteins was also important for geochronological applications. After shell burial, the amino acids occluded in the shell undergo a slow natural degradation, *i.e.* the so-called racemisation process, which is characterised by the interconversion between the L- and D- stereoisomers. All living organisms (except some bacteria) only contain L- amino acids, while D- enantiomers are formed as a result of post mortem processes. The extent of racemisation therefore relates the diagenetic changes in amino acid stereochemistry to the time elapsed since the death of the organism (Wehmiller and Hare, 1971). Thus, the quantification of the extent of amino acid racemisation (AAR) from fossil shells can be used as a molecular clock, enabling to estimate the relative age of sub-fossil biomineralised materials and is a widely used geochronological method for archaeological and palaeontological shell substrates (Schroeder and Bada, 1976; Collins, 2000; Demarchi et al., 2011; Penkman et al., 2011).

In the last decade of the 20th century, the analysis of ancient DNA (aDNA) became the main focus of ancient biomolecule research (Pääbo, 1989). It emerged as one of the most important breakthroughs (Mulligan, 2005), a result of technological advancements in molecular biology, most importantly the development of the polymerase chain reaction (PCR) in the 1980s (Mullis, 1990). This meant that the extremely small quantities of preserved ancient DNA sequences from

archaeological and palaeontological samples, could be amplified, detected, analysed and characterised (Hofreiter et al., 2001; Cappellini et al., 2018; Lindqvist and Rajora, 2019). aDNA data were able to provide a wealth of information on ancient organisms, allowing in-depth characterisation even of extinct species. However, no such technological analogy was present for ancient proteins.

It was not until the beginning of the 21st century that modern ancient protein studies, now best-known as palaeoproteomics, took off (Collins et al., 2006; Buckley et al., 2009; Dallongeville et al., 2016). This owes to the advancements in soft-ionisation mass spectrometry techniques (Aebersold and Mann, 2003) which enabled to analyse and to detect proteins from archaeological/palaeontological samples. Palaeoproteomics is a multidisciplinary field which encompasses biomolecular sciences, geochemistry, bioinformatics, palaeontology, ecology, evolutionary biology and archaeology and aims to address questions about past societies and environments. In recent years there has been a great increase in the use of palaeoproteomics to study a range of bioarchaeological substrates and, as some scholars note, the methodology has shown excellent potential to significantly improve our understanding of the past, even for “deep time” periods (Wallace and Schiffbauer, 2016; Welker, 2018b). The next section will therefore present a detailed overview of the methodological approach and of the wide spectrum applications of palaeoproteomics to study biominerals and remains from the archaeological and palaeontological record.

Palaeoproteomics

Proteins display long-term survival potential which surpasses that of ancient DNA (aDNA). For aDNA, post-mortem irreversible degradation results in highly fragmented DNA, which is therefore irretrievable after about 100000 years, considering mild temperature conditions. In hot climates, degradation is even faster. In general, it is well accepted that the survival of DNA molecules beyond 1 million years is highly improbable (Kistler et al., 2017; Wadsworth et al., 2017; Welker, 2018b). Proteins, on the contrary, have a slower rate of fragmentation and can survive beyond this limit (Welker, 2018b). In particular, protein preservation is superior in highly mineralised substrates, *e.g.* eggshell and dental enamel, where proteins strongly interact with the mineral phase or are trapped inside the mineral crystals, therefore forming a so-called ‘closed-system’ (Demarchi et al., 2016; Dickinson et al., 2019). For example, protein sequences were retrieved from 1.77-million-years-old dental enamel of the hominin *H. erectus* (Welker et al., 2020) and an extinct rhino species (Cappellini et al., 2019), which were excavated from Dmanisi, a warm climate site in the southern Caucasus, Georgia. In another study, ancient proteins were successfully extracted, analysed and characterised from a

1.9-million-year-old dental enamel sample of the *Gigantopithecus* ape found in Chuifeng Cave in China (Welker et al., 2019). In fact, when the combined effect of temperature and time is taken into account (thermal age) and the numerical ages for these samples are normalised to a constant temperature of 10 °C, the samples date to approximately 11.8 million years ago (Myr@10 °C), *i.e*. five times older than the chronometric age. The oldest genuine protein sequence retrieved so far, comes from a 3.8 million-years-old ostrich eggshell (approximately 15 Myr@10 °C) found at the famous Laetoli site, Tanzania (Demarchi et al., 2016). These examples show that ancient protein sequences can be recovered from thousand- or million-years-old substrates and used to study different organisms, including extinct species.

Over the years, palaeoproteomics has provided the possibility to dive deep back in time and investigate some very interesting archaeological and palaeontological questions. Ancient bone is the material that has been studied the most, because of its ubiquity and good molecular-level preservation in archaeological/palaeontological sites. However, in recent years, palaeoproteomics has been rapidly expanding with an increasing number of studies focussing on different types of biological materials, such as tooth, skin, horn, eggshell as well as ancient cooking pots and tools (Richter et al., 2011; Solazzo et al., 2013; Presslee et al., 2017; Hendy et al., 2018b; 2018c; Welker, 2018b; Demarchi et al., 2020b). Ancient proteins have been used to study extinct organisms and their phylogeny, particularly in the cases where morphological features were of little help. The analysis of ancient collagen sequences obtained from bones of South American ungulates allowed to reassess their dubious evolutionary position (Welker et al., 2015). Ancient collagen sequences were also analysed in order to assess the phylogenetic relationships between different sloth species (Presslee et al., 2019). Palaeoproteomic analyses are also used for taxonomic identification of zooarchaeological remains (Buckley et al., 2009; Buckley, 2018). Large-scale studies of zooarchaeological samples from human occupation sites enable researchers to reconstruct faunal palaeoecology and to provide insights into human adaptations to climatic changes that occurred in the past (Buckley and Collins, 2011; Pothier Bouchard et al., 2020). Palaeoproteomic analyses of collagen from animal skins that had been used for making historical parchments, enabled to obtain precious information on the production processes of the first pocket Bibles (Fiddyment et al., 2015) and other parchment-based objects.

As previously mentioned, tooth enamel preserves extremely well over archaeological and palaeontological timescales and enamel proteins have been used to study extinct animals and hominids (Cappellini et al., 2019; Welker et al., 2020). Fossil dental enamel, indeed, is regarded as one of the most promising biological systems for future human evolution studies (Welker et al., 2019; 2020).

Dental calculus can also provide biological information, typically about oral health and ancient diets. For example, the study of specific protein markers obtained from the tartar of Medieval teeth allowed the researchers to investigate the diseases which affected individuals in the past (Jersie-Christensen et al., 2018). Palaeoproteomics is also used to study the diet of ancient populations (Hendy et al., 2018c). Certain food proteins can get embedded inside dental calculus and remain preserved even after the death of the person. A well-investigated example are the milk proteins which are indicators of dairy consumption (Warinner et al., 2014) and allow archaeologists to investigate dairying practices, as well as to study their diffusion during the Neolithic and Bronze Age (Jeong et al., 2018; Charlton et al., 2019). Food remains that get trapped inside the ceramic matrix of ancient cooking pots, can also sometimes be analysed by palaeoproteomics (Solazzo et al., 2008; Heaton et al., 2009). For example, the study of mineralised food proteins and lipids from Çatalhöyük, one of the largest Neolithic sites, suggested that early farmers consumed a wide variety of foods - pulses, cereals, dairy and meat (Hendy et al., 2018b).

Fragmented archaeological eggshells are remarkably abundant in archaeological sites, and can now be identified by ancient protein analyses (Presslee et al., 2017). Eggshell palaeoproteomics has been used to understand trends of increased chicken and goose egg consumption (Stewart et al., 2014; Maltby et al., 2018) and the study of eggshells from Çatalhöyük displayed the symbolical and practical importance of different wild bird taxa for this Neolithic society (Demarchi et al., 2020b). Throughout human history, it appears that eggs were consumed as food, but also used for symbolic purposes, as indicated by findings *e.g.* of chicken eggs in grave contexts (Jonuks et al., 2018), exposing the deep connection between humans and birds in the past. This overview simply exemplifies the strong impact of palaeoproteomics in modern bioarchaeology and molecular palaeontology research.

Analytical approach

The palaeoproteomics study of archaeological/palaeontological samples includes several main steps: sample preparation, *i.e*. protein extraction and purification followed by mass spectrometry and ultimately protein identification (Figure 14) (see reviews in Dallongeville et al., 2016; Cleland, 2018; Cleland and Schroeter, 2018). First of all, to prepare the samples, the proteins are extracted and this is typically done by using a mild acid solution to demineralise the substrate. Filter assisted sample preparation (FASP) (Wiśniewski et al., 2009) is the most commonly used protein purification method, although the approach may vary according to the type of sample (Schroeter et al., 2016). It is important to note that extra-clean laboratory

environments are a prerequisite for sample preparation and mass spectrometry (MS) analysis in order to prevent cross-contamination which can complicate the characterisation of endogenous sequences (Hendy et al., 2018a).

There are two principal proteomic approaches by mass spectrometry – the "bottom-up" and "top-down" (de Hoffmann and Stroobant, 2013). "Top-down" approach targets intact proteins (*i.e.* no enzymatic digestion is performed), enabling the direct detection of whole protein sequences (Catherman et al., 2014) and, potentially, their natural degradation products (*i.e.* peptides deriving from *e.g.* hydrolytic scission of the chain). While such a feature is very interesting in proteomics, this approach is much less commonly used because mass spectrometry analysis of whole proteins is less sensitive, it is technically more challenging, therefore, has been rarely employed in ancient protein studies (Dallongeville et al., 2016). In "bottom-up" proteomics the extracted proteins are degraded to their constituting peptides by using enzymes, usually trypsin, a step that produces peptides with C-terminally protonated amino acids, providing an advantage in subsequent peptide sequencing with mass spectrometry (Aebersold and Mann, 2003). It is the most commonly used proteomics approach.

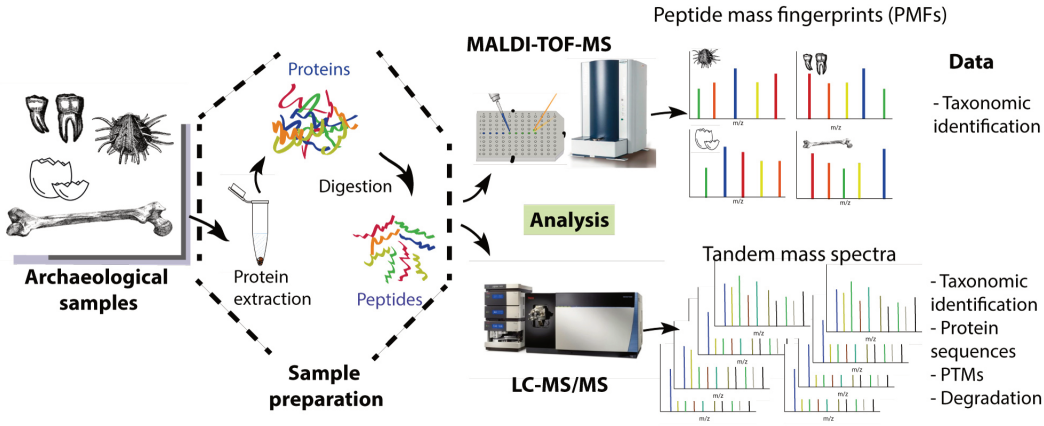


Figure 14. Scheme that depicts a typical palaeoproteomics analytical workflow. The figure displays the two main mass spectrometry techniques used in palaeoproteomics – MALDI-TOF-MS for peptide mass fingerprinting (PMF) and LC-MS/MS for proteome characterisation. On the right hand side, the figure notes the type of information obtained by each of these techniques and targeted in palaeoproteomic studies.

A mass spectrometer consists of three main elements: an ion source, a mass analyser and a detector. In the ion source the targeted molecules, *i.e.* peptides, are ionised, then introduced to a mass analyser which separates them by their mass-to-charge (*m/z*) ratios and the number of ions at each *m/z* values are registered by

the detector (de Hoffmann and Stroobant, 2013). Electrospray ionisation (ESI) and matrix-assisted laser desorption/ionisation (MALDI) are two commonly used techniques to volatilise and ionise peptides (or proteins). ESI can be directly coupled to liquid-based separation tools (for example chromatographic) and it is a preferred system for analysing complex protein samples. MALDI ionises the samples out of a dry crystalline matrix via laser pulses and it is normally used to analyse relatively simple peptide mixtures (Aebersold and Mann, 2003; El-Aneed et al., 2009). There are different types of analysers including Orbitrap, time-of-flight (TOF), ion trap, quadruple and Fourier transform ion cyclotron (FT-MS), which use different principles to separate ions, hence are used for different applications (Aebersold and Mann, 2003). The Orbitrap and TOF analysers are among the most frequently employed in current proteomic studies (Sinha and Mann, 2020).

In palaeoproteomics, the most commonly used mass spectrometry approaches are matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-TOF-MS) and “shotgun” proteomics by high performance liquid chromatography electrospray ionisation tandem mass spectrometry (LC-ESI-MS/MS) (Figure 14) (Aebersold and Mann, 2003; de Hoffmann and Stroobant, 2013; Cleland and Schroeter, 2018; Welker, 2018b). MALDI-TOF-MS is typically used to identify proteins by peptide mapping, also known as peptide mass fingerprinting (PMF), and it is widely employed in MS analysis due to its relative simplicity, good mass accuracy, high resolution and sensitivity.

Protein analysis by MALDI-TOF-MS requires several steps (de Hoffmann and Stroobant, 2013). First of all, the peptides obtained via enzymatic digestion of the proteins are dissolved in a solvent that contains small organic molecules – the MALDI matrix. It is generally a coloured compound which has a strong absorption at the laser wavelength of choice and aids both desorption and ionisation. The peptide-matrix mixture is then spotted on a special metal MALDI plate and dried, resulting in the so-called analyte-doped matrix crystal, where the peptides molecules are embedded throughout the matrix (this is the most commonly used sample preparation method although there are number of different approaches and variations (Vorm et al., 1994; O’Rourke et al., 2018)). Secondly, under vacuum conditions, the matrix-analyte spot is ablated by intense laser pulses (most commonly in UV range) of short duration, which induce heating of the crystals by the accumulation of a large amount of energy through excitation of the matrix molecules. This causes a localised sublimation of the matrix crystals carrying the intact analyte in the matrix plume. However, the exact mechanism of the MALDI process is not completely elucidated and the origin of ions produced in MALDI is still not fully understood. The selection of the matrix is a crucial factor for successful ionisation and it depends on the type of analyte. For peptide mixtures, α-Cyano-4-

hydroxycinnamic acid (CHCA) and 2,5-Dihydroxybenzoic acid (gentisic) (DHB) are the most commonly used compounds. The matrix increases the efficiency of energy transfer from the laser to the analyte and also prevents the damage from laser pulse by absorbing most of the incident energy.

MALDI is typically coupled to time-of-flight (TOF) mass analyser which separates ions according to their velocities when they drift in a free-field region, *i.e.* flight tube (de Hoffmann and Stroobant, 2013). The mass-to-charge (*m/z*) ratios are determined by measuring the time that ions take to move through this field-free region between the ion source and the detector. TOF has the ability to detect a large range of masses and it has high sensitivity. Moreover, the analysis speed of TOF is very fast. However, there are several disadvantages of the MALDI-TOF-MS approach. First of all, MALDI has a low shot-to-shot reproducibility and a strong dependence on the sample preparation method. This is because each MALDI laser shot ablates a few layers of sample-matrix film, which may cause differences in shot-by-shot spectra. In addition, sample deposition methods can lead to spectral variations. High concentrations of salts, buffers, detergents may also interfere with the desorption and ionisation process of samples.

Proteomic analysis by HPLC-ESI-MS/MS differs in the sense that the mixture of peptides is first separated using a liquid chromatography system and then the compounds are analysed by mass spectrometry (de Hoffmann and Stroobant, 2013). The peptides are separated on a chromatography column based on their chemical features, *e.g.* peptide hydrophobicity, and afterwards are ionised by an electrospray ion source (ESI). The ESI produces peptide ions at atmospheric pressure, by applying a strong electric field to a liquid passing through a capillary tube with a weak flux (de Hoffmann and Stroobant, 2013). A charge accumulates at the liquid surface located at the end of the capillary and the liquid breaks to form highly charged droplets, which are sprayed by the help of inert gas. The desorption of ions occurs with the increased electric field when the small highly charged droplets lose solvent. ESI is able to produce multiply-charged ions, which improves the sensitivity at the detector and allows the analysis of compounds of high-molecular-weight. Protonated peptides enter the mass spectrometer, where a mass spectrum is taken for each eluting peptide (MS).

ESI sources can be coupled to Orbitrap mass analyser (commonly in hybrid configurations), which is among the top high-resolution mass spectrometry analysers used in modern proteomics and is typically employed for biological applications (Zubarev and Makarov, 2013). Orbitrap belongs to the family of FT-MS mass analysers and it distinguishes ions based on their oscillation frequencies. The injected ions are trapped in the Orbitrap and then they move along the length axis

of a central metal spindle, where they can travel up to several kilometres, thus enabling very high resolutions. An 'image current' is measured, which is induced by the rapidly oscillating ions. The current is recorded in the time domain and is converted into the frequency domain using a Fourier transformation. Orbitrap analysers have very high resolutions and mass accuracies and hybrid mass spectrometry instruments with Orbitraps are often employed in tandem mass spectrometry experiments.

Tandem mass spectrometry, abbreviated MS/MS, involves at least two stages of mass analysis. The tandem mass spectrometry in space is carried out by coupling two physically distinct instruments (de Hoffmann and Stroobant, 2013). The first analyser is used to isolate a precursor ion, which then undergoes fragmentation (spontaneously or induced by some activation process), to yield product ions and neutral fragments. The second spectrometer analyses the product ions. The obtained product ion spectra can be sequenced *de novo* using mathematical algorithms (Ma et al., 2003) and the bioinformatic data analysis enables to reconstruct peptide sequences, used to identify and characterise protein sequences. It is also possible to increase the number of these steps, as for example in MS/MS/MS (or MS^3) experiments, the ions selected by the first mass analyser are fragmented, then, ions of these fragments are selected by the second mass analyser, fragmented again, and finally analysed by the third mass analyser, to obtain a product ion spectrum. The two analytical approaches differ in terms of technical set up, length and complexity of analysis, cost of operation and above all, the amount of data that can be obtained. Therefore, in palaeoproteomics, these techniques are used for different types of applications and will be discussed below.

Palaeoproteomic analysis by peptide mass fingerprinting (PMF) is predominantly used for taxonomic identification of zooarchaeological remains. The approach is commonly known as "ZooMS", that is Zooarchaeology by Mass Spectrometry (Buckley, 2018). The development of ZooMS was originally based on the analysis of type I collagen, the dominant protein in bone (Buckley et al., 2009) which, in phylogenetically distant species (for mammals, at genus level or above), shows amino acid differences in some sequence regions. The analysis of collagen by MALDI-TOF mass spectrometry therefore results in a spectrum with a unique peptide mass fingerprint (PMF). This allows researchers to identify *m/z* values from collagen that are markers of different species, and which can ultimately be used for taxonomic identification.

ZooMS is particularly useful for fragmented zooarchaeological remains, which do not retain any morphological features, allowing us to obtain taxonomic discrimination based on molecular signatures (Buckley et al., 2009; Kirby et al.,

2013). A large part of ZooMS studies are represented by analyses of bones from large mammal species and animals such as sheep, goat and cattle that have a direct link with the human environment, and can give information on *e.g.* husbandry practices (Buckley et al., 2010; Fiddyment et al., 2015; Brandt et al., 2018; Buckley, 2018). However, in recent years, more ZooMS studies are being conducted on different organisms and materials, including marine mammal bones (Buckley et al., 2014), fish bones and their scales (Richter et al., 2011; Harvey et al., 2018), bones of tortoises and bats (van der Sluis et al., 2014; Buckley and Herman, 2019). ZooMS has also been applied to keratinaceous materials (Solazzo et al., 2013) and a PMF by MALDI-TOF approach has been adapted to analyse eggshell fragments for bird species identification (Stewart et al., 2013; Presslee et al., 2017; Jonuks et al., 2018; Demarchi et al., 2020b). There are several important advantages of PMF by MALDI-TOF-MS: i) rapid analysis ii) straightforward data treatment (at least to some extent) and iii) relatively low cost. The ability to perform rapidly a very high number of analyses at the same time has been particularly useful for screening large sets of archaeological remains and to carry out palaeoecological studies (Buckley and Collins, 2011; Buckley, 2018; Buckley et al., 2018; Pothier Bouchard et al., 2020).

However, at the same time, the method is not without limits. The data obtained by MALDI-TOF-MS do not provide the information on the peptide sequences and this requires further investigation by tandem mass spectrometry (*e.g*. using a MALDI-TOF/TOF system) in order to *de novo* sequence marker peptides. However, a tandem mass spectrum obtained via MALDI-TOF/TOF can be challenging to interpret if, for example, it belongs to two or more different ion precursors or represents fragmentation patterns of salt adducts. Furthermore, some peptides may not fragment properly, which also complicates data interpretation. In MALDI-TOF-MS, peptides that are low in abundance or have a low ionisation efficiency, may not be detected, whereas the presence of post-translational modifications (PTMs) can be also difficult to interpret (Cleland and Schroeter, 2018; Welker, 2018b). In addition, another challenge from a biological point of view is that taxonomic identification cannot be achieved to species level when sequences do not possess marked differences (Cleland and Schroeter, 2018). For example, collagen peptide fingerprints do not have unique signatures for species which have had a rather short evolutionary divergence, hence they cannot be distinguished (for large mammals, at least ca. 5-million-year divergence is needed to have a taxonomic resolution at genus level). PMF analyses gets even more complex for other non-collagen based biomaterials, as in the case of eggshell. Avian eggshells have a high number of different proteins, most of which, compared to collagen, are poorly characterised (Gautron, 2019). Moreover, the phylogenetic resolution based

on eggshell protein signatures has not been fully assessed, due to lack of genomics data (Presslee et al., 2017).

Proteomics by high performance liquid chromatography coupled to tandem mass spectrometry (HPLC-MS/MS), on the other hand, gives information on the sample proteome. The latest mass spectrometry setups, *i.e.* hybrid systems with Orbitrap mass analysers (*e.g*. Q-Exactive Orbitrap mass spectrometer) provide great sensitivity and resolution and, in palaeoproteomic analysis, enables researchers to identify hundreds of protein sequences (Cleland and Schroeter, 2018) and to study the presence of different post-translational modifications (PTMs) (Cleland et al., 2015). This is important because, for example, the extent and rate of deamidation which occurs due to diagenetic alterations, can help to assess the age of the sample and authenticate protein sequences (Demarchi et al., 2016; Cleland and Schroeter, 2018; Mackie et al., 2018; Ramsøe et al., 2020). Glycosylation and phosphorylation PTMs can also provide useful information about the modifications that occur *in vivo* (Welker et al., 2019; 2020). Alongside bone, dental enamel and eggshell, tandem mass spectrometry can be used to analyse different types of biological substrates and proteins, such as plants remains (Cappellini et al., 2010), painting layers (Mackie et al., 2018), milk proteins (Tsutaya et al., 2019), mummified human skin (Demarchi et al., 2020a), bacteria from dental calculus (Fotakis et al., 2020). Compared to MALDI-TOF-MS, tandem mass spectrometry is more resource demanding. It has higher operational costs per sample (up to two orders of magnitude) and the bioinformatic data analysis requires a higher level of expertise, also being more time consuming (Cleland and Schroeter, 2018).

In general, many palaeoproteomic studies employ both mass spectrometry techniques, which allow researchers to obtain results that are complementary to each other. In a similar way, these two techniques can be used together to develop proteomic approaches to study archaeological mollusc shells and prehistoric shell artifacts, as presented in Figure 15. This will be discussed in detail in the following sections.

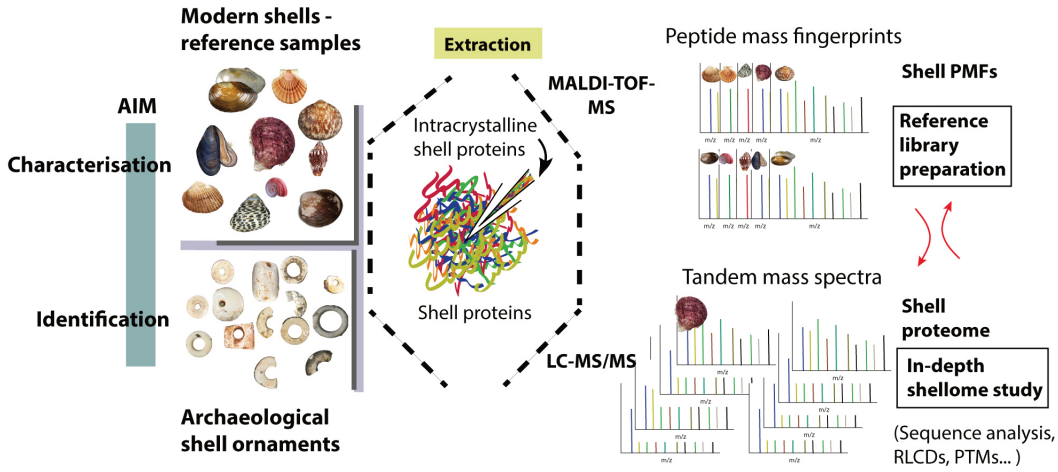


Figure 15. Schematic representation of shell palaeoproteomics for biomolecular identification of archaeological shell ornaments.

Shell palaeoproteomics – potentials and pitfalls

The key focus of palaeoproteomic research so far has mainly been the study of the past from a human perspective. This is clearly shown by the large body of published studies that investigate human evolution (Welker et al., 2016; 2019; 2020; Welker, 2018b), diets of prehistoric people (Mackie et al., 2017; Hendy et al., 2018b; 2018c; Jeong et al., 2018) and ancient diseases (Jersie-Christensen et al., 2018). Many palaeoproteomic studies also focus on animals, especially those closely related to human activities, *e.g.* domesticated species (Presslee et al., 2017; Brandt et al., 2018; Maltby et al., 2018). Furthermore, palaeoproteomics can be used to resolve long-standing questions in evolutionary biology (Cappellini et al., 2012; 2019; Welker et al., 2015; Presslee et al., 2019). Molluscs shells, and indeed most invertebrates, have not been the focus of palaeoproteomic research. This may be because, at first glance, the link between humans and mollusc shells seems more labile. From a biomolecular point of view, shell materials have also been less investigated and compared to bone or teeth, much less is known about their material, biological and molecular characteristics.

However, mollusc shells possess properties that make them a very interesting system for biomolecular research, for example the presence of intracrystalline organics. Intracrystalline proteins are characterised by a closed-system behaviour (Sykes et al., 1995) and preserve well in archaeological timescales. For example intracrystalline sequences have been retrieved from other palaeontological biominerals, most notably 3.8 million-year-old eggshell (Demarchi et al., 2016) and

1.9 million-year-old dental enamel (Welker et al., 2019). Intracrystalline mollusc shell proteins may also carry a taxonomic signal. Demarchi and colleagues have studied intracrystalline mollusc shell proteins from a variety of species by characterizing their amino acid compositions and showed that their bulk amino acid composition could provide some taxonomic discrimination (Demarchi et al., 2014). In particular, significant differences were observed between bivalves and gastropods and distinctions could be made between a subset of bivalves, such as *Arctica*, *Modiolus*, *Pecten* sp. While other shells, including *Spondylus* and *Glycymeris* were difficult to discriminate among other specimens, the study nevertheless showed that there is potential to taxonomically discriminate shells based on their intracrystalline proteins.

Ancient DNA was also successfully retrieved from different types of mollusc shells (Der Sarkissian et al., 2017; 2020). The studies showed that mollusc shells are good archives of genetic palaeoenvironmental information, including information on ancient molluscan diseases, although the possibility to use aDNA data for taxonomic identification has not been investigated. The next chapters will review current knowledge about intracrystalline protein studies from ancient shells and provide insights into the perspective of developing shell palaeoproteomics.

Intracrystalline shell proteins

Mollusc shells are nanocomposite biominerals, made of calcium carbonate and a small fraction of organics. As discussed in previous sections (Part 3), shells are characterised by the presence of intracrystalline proteins (Albeck et al., 1993; 1996). Already in early studies Towe et al. noted that the intracrystalline biomolecules preserved in calcified skeletons were likely to provide the best source of endogenous (genuine) proteins in fossil organisms (Towe, 1980; Sykes et al., 1995). Truly intracrystalline organics have been shown to approximate a so-called ‘closed-system’, which protects from contamination and rapid *in situ* degradation (Sykes et al., 1995; Penkman et al., 2008; Demarchi et al., 2016). Compared to proteins in an open system, as in the case of archaeological bone (Hendy et al., 2018a), intracrystalline proteins have the advantage of remaining stable and *“contamination-free”* over archaeological and palaeontological timescales (Demarchi et al., 2016), as long as the biominerals that contain them do not undergo dissolution or recrystallization. The operational intracrystalline shell protein fraction is defined as the one which remains unaffected when exposed to very strong oxidising agents (Sykes et al., 1995). Indeed, the standard method to isolate the intracrystalline shell proteins (and this term generally refers also to peptides, bound amino acids or other shell organics) uses a strong and prolonged bleaching treatment (Sykes et al., 1995; Penkman et al., 2008) to remove superficial

contamination and intercrystalline organics. The closed-system behaviour of intracrystalline proteins is further verified by “leaching” experiments, where bleached shell powders are heated in water in order to detect any proteins escaping from the system during accelerated diagenesis (Demarchi et al., 2013b).

In molluscs, intracrystalline shell proteins account for just 0.001-0.01% of the total shell weight. A number of mollusc shell taxa possess an intracrystalline shell protein fraction that behaves as a closed-system, including terrestrial, freshwater and marine gastropods and bivalves, and also shells with different microstructures and mineralogies (Penkman et al., 2008; 2011; Demarchi et al., 2011; 2013b; Pierini et al., 2016; Ortiz et al., 2018). Among others, the intracrystalline protein fraction has been studied from aragonitic crossed-lamellar freshwater clam *Corbicula* (Penkman et al., 2008), calcitic foliated *Pecten* (Pierini et al., 2016), nacro-prismatic freshwater mother-of-pearl mussel *Margaritifera* (Penkman et al., 2008) and nacreous marine gastropod *Phorcus* (Ortiz et al., 2018), as well as *Patella* shells, which are characterised by a very complex microstructure (Demarchi et al., 2013b). The shells of *Glycymeris*, the saltwater clams, which are common in Mediterranean archaeological and palaeontological sites, showed the presence of an intracrystalline protein fraction, however one that did not completely behave as a closed system (Demarchi et al., 2015). This may be linked to the presence of microtubules in the microstructure of *Glycymeris* (Böhm et al., 2016). Another large study isolated intracrystalline shell proteins from 29 different shell genera and characterised their amino acid composition (Demarchi et al., 2014), indicating a good potential to retrieve taxonomically-informative intracrystalline protein sequences from a variety of mollusc shells.

The make-up of intracrystalline mollusc shell proteins, however, is not well known. This is because intracrystalline shell sequences are mainly studied for geochronological dating (for a review, see Demarchi, 2020). Therefore, amino acid racemisation (AAR) analyses, which are used to quantify the extent of degradation of the intercrystalline fraction, give only information about the bulk amino acid composition. On the other hand, shell proteomic analyses in biomineralization research, obtain sequence information, but mostly considering bulk organics, *i.e.* (inter+intra)crystalline fraction. The bleaching approaches used in biomineralization studies are rather mild (Marie et al., 2013a) and are not comparable to the ones used to isolate intracrystalline (closed-system) proteins.

Intracrystalline shell proteins are often described as dominated by acidic residues (Albeck et al., 1996) and acidic domains presumably play a key role in mineral-binding (Demarchi et al., 2016). Comparing the amino acid compositions of the intracrystalline fraction and of unbleached shells, the intracrystalline sequences are

more enriched in glutamic (Glu) and aspartic (Asp) amino acid residues (Penkman et al., 2008). The analyses also show that the bleaching step significantly reduces the relative amount of hydrophobic alanine (Ala) and glycine (Gly) residues. This effect is particularly pronounced in shells with nacro-prismatic microstructures (typically organics-rich), suggesting that bleaching removes the silk-like proteins which are generally considered as intercrystalline (Lowenstam and Weiner, 1989; Penkman et al., 2008; Kocot et al., 2016). However, the relative loss of hydrophobic amino acids appears to be less evident in crossed-lamellar and foliated shells (Penkman et al., 2008; Pierini et al., 2016). For crossed-lamellar *Glycymeris,* alanine and glycine make up a considerable part of its intracrystalline protein fraction (Demarchi et al., 2015). This suggests that in crossed-lamellar and foliated shells, the intracrystalline fraction also contains hydrophobic sequences. However, the actual proteome differences between intra- and intercrystalline fractions remains to be identified. An important caveat is that occlusion of proteins inside the shell mineral skeleton may not be necessarily due to selective processes, but could result from random effects.

Shell proteins diagenesis

Understanding how shell proteins are affected by natural aging and what are the pathways of protein degradation is important in the future palaeoproteomic studies of archaeological/palaeontological shell samples. The knowledge of shell protein diagenesis can be very useful not only for dating, but also for reconstructing and authenticating ancient sequences (Brooks et al., 1990; Demarchi et al., 2016; Hendy et al., 2018a). In particular, intracrystalline shell proteins provide an excellent source to study diagenesis mechanisms, because the closed-system environment partially simplifies the system by reducing the number and type of environmental factors (Demarchi et al., 2013a).

Diagenesis is the sum of all *post mortem* processes taking place in the burial environment and is defined as the transformation of the living organism to its constituting atoms (Curry et al., 1991). In general, the diagenesis of mollusc shells is a complex process, governed by a number of different physico- chemical reactions that occur in the mineral and organic phases, interdependent to each other (Demarchi, 2020). There are number of different factors that play a role during protein diagenesis, most of which are still poorly understood (Demarchi, 2020):

- the conditions of the burial environment (*i.e.* soil pH where shells have been deposited);

- temperature and age of the sample (*i.e.* shell samples obtained from warm climate regions have been more affected by degradation and *vice versa*);
- the mineralogy and the structure of the shell (*e.g.* the calcitic polymorph is more stable than the aragonitic);
- the full composition of shell proteins occluded in the mineral skeleton (*i.e.* certain proteins or sequence domains and their association to the mineral may influence differently their rates of degradation);
- the presence and nature of other macromolecular compounds in the shell biomineral (*i.e* the interference and role of other biomolecules in shell protein diagenesis, as for example, saccharides can react with amino acids and form Maillard products (Maillard, 1912; Collins et al., 1992)).

Mineral phase transformation is one of the most pronounced effects of shell diagenesis. Aragonite is a less-stable calcium carbonate polymorph than calcite and, under natural diagenesis conditions or when subjected to heating at high temperatures (>350 °C), biogenic aragonite undergoes re-crystallisation to calcite (Yoshioka and Kitano, 1985). The process is also microstructure-dependent because studies have shown that re-crystallisation occurs at different rates for different types of aragonitic microstructures (Milano and Nehrke, 2018). Hence, the phase transformation and diagenesis is at least partially mediated by the type and characteristics of mineralised organic matrix, an effect that has also been observed in different calcium carbonate fossil samples (Marin and Gautret, 1994).

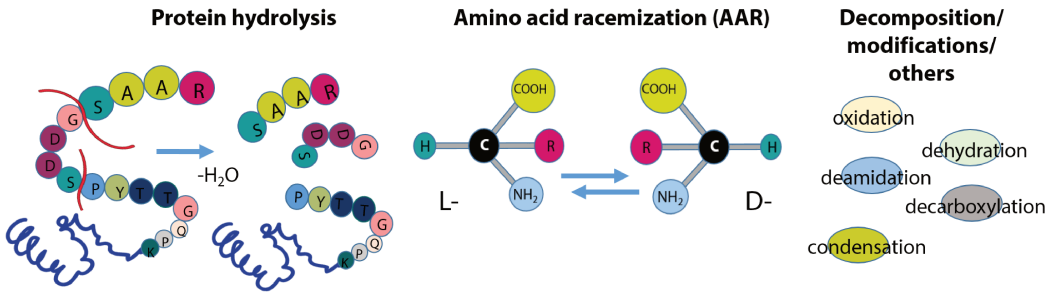


Figure 16. The main protein diagenesis reactions: hydrolysis, amino acid racemisation, decomposition and diagenesis induced modifications (figure notes some of the most common modifications).

Hydrolysis, racemisation, decomposition and diagenesis-induced modifications are the main protein diagenesis reactions (Figure 16). Decomposition refers to the decay of amino acids into their smaller molecular components (see review in

Demarchi, 2020), as for example the dehydration of Ser to Ala which is one of the better studied reactions due to their relevance for geochronological applications (Bada et al., 1978; Bada and Man, 1980; Penkman et al., 2008; Demarchi et al., 2011). Among the different diagenesis-induced modifications the most common are deamidation of asparagine and glutamine residues, as well as oxidation of methionine and tryptophan. In palaeoproteomic studies, the rates of deamidation are employed to track protein aging (Wilson et al., 2012; Buckley et al., 2018), and authenticate ancient protein sequences (Demarchi et al., 2016; Hendy et al., 2018a; Ramsøe et al., 2020). Deamidation is due to the hydrolysis of the amide side chains of Asn and Gln which are subsequently transformed into carboxylic acid groups, forming aspartic and glutamic acids (Robinson and Robinson, 2001); or it can proceed via a cyclic succinimidyl intermediate (Capasso et al., 1991). Glutamine deamidation is slower compared to asparagine and it can be used as a molecular indicator of protein preservation/degradation (Robinson and Robinson, 2001; 2004; Demarchi, 2020). However, the study of glutamine deamidation rate as a proxy of protein preservation can be complex. In the case of collagen, it has been observed that the sample preparation approach affects the outcome of analysis, as acidic extraction (which is among the most commonly employed in palaeoproteomic studies) induces partial protein degradation (Simpson et al., 2016). Moreover, deamidated peptides do not ionise well when using MALDI ion source, hence the study of peptide deamidation requires careful consideration of instrumentation type (Simpson et al., 2019).

The racemisation reaction is the post-mortem spontaneous interconversion reaction between the two chiral forms of single amino acid, the D- and L-forms (see reviews in Collins, 2000; Demarchi, 2020). L-amino acids are present in living organisms, while D-amino acids are formed by post-mortem racemisation. The ratio between the D- and L-forms in a fossil sample represents the extent of racemisation and can be used as a molecular clock for geochronological dating, as mentioned in previous sections. The analysis of amino acid racemisation (AAR) has been used to assess the state of preservation of archaeological shell ornaments (Demarchi et al., 2014) and also to directly date archaeological sites (Penkman et al., 2011; Bosch et al., 2015).

Hydrolysis is a reaction that breaks down peptide bonds in the presence of water molecules (Hill, 1965). In biomineral systems, the hydrolysis of proteins is complex, and this is reflected in the difficulty researchers consistently encounter when attempting to study and model the reaction (see reviews in Collins and Riley, 2000; Demarchi, 2020). Part of the issue is due to the almost complete lack of knowledge on the non-protein organic macromolecules present in shells and on their role in degradation. For example, proteins may react with carbohydrates present in shells,

forming condensation products (Collins et al., 1992). Recent investigations on well-preserved intracrystalline eggshell protein sequences highlighted that protein-mineral interaction may play the most important role in mediating hydrolysis (Demarchi et al., 2016). Computational modelling showed that acidic domains which bind to the mineral also result in surface stabilisation that creates an activation energy barrier for hydrolysis, “locking” the water molecules and slowing down the hydrolysis process. Future studies of shell protein diagenesis may help us to better understand the mechanisms of long term preservation and would be useful when analysing ancient sequences.

Towards the biomolecular identification of archaeological shell artifacts

The intrinsic material and biological features of mollusc shells make them excellent systems for palaeoproteomics research. First of all, the well-known presence of intracrystalline shell proteins suggests a good potential to recover sequences from the archaeological or palaeontological record. Besides, the analysis of the intracrystalline fraction selectively isolates genuine mollusc shell proteins and precludes us from analysing non-endogenous sequences (Sykes et al., 1995; Penkman et al., 2008). Biomineralization studies have also exposed the distinctiveness and diversity of mollusc shell proteins (Marin and Luquet, 2004; Marin et al., 2007a; 2016; Kocot et al., 2016), which can be exploited as biomolecular “barcodes” for shell species discrimination.

At the same time, shell palaeoproteomics research will have to face several challenges. In mollusc shells, the content of intracrystalline shell proteins is very low (as noted in the above section, can be as low as 0.001%) (Penkman et al., 2008; Pierini et al., 2016). It is considerably smaller compared to bone (Buckley and Wadsworth, 2014; Buckley, 2018) or other biomineralized closed-systems such as eggshell (Brooks et al., 1990; Crisp et al., 2013) or dental enamel (Dickinson et al., 2019). The identification and characterisation of shell proteins is also hampered by the lack of proteomics and genomics data for mollusc shells. So far, only a few tens of mollusc shell types have been thoroughly studied (Marin, 2020). Considering the size of the phylum (see Figure 1), a large part of the sequences for the moment remains completely unknown (Kocot et al., 2016; McDougall and Degnan, 2018; Evans, 2019). Finally, as discussed above, there is a lack of detailed knowledge on the precise mechanisms of degradation (Demarchi, 2020). This can hamper the identification and characterisation of ancient mollusc protein sequences, particularly from heavily degraded shell artifacts. However, fast development of new analytical approaches, advancements in mass spectrometry and bioinformatics suggest that some of the methodological issues may not take long before they are, at least in part, overcome.

Perspectives in shell palaeoproteomics

The future development of shell palaeoproteomics depends on the main research questions which we will try to address. The study of archaeological shells may help us to better understand human-environment interactions, including the value given to these materials and the subtle cultural nuances that determined the choices. The analysis of “fossil” shell proteins could also provide insights into the molecular mechanisms that help sequences to be preserved thousands or even millions-of-years (Wallace and Schiffbauer, 2016). The possibility to retrieve useful information from ancient shells depends not only on the development of novel analytical and methodological techniques, but also, and more importantly, on the human ability to interpret the data.

Shell proteomics relies on the availability of genomics/transcriptomics data and the lack of it results in a partial and incomplete picture of the mollusc shell system being studied (Marin et al., 2016). However, in the future, shell palaeoproteomics could be approached from a different perspective, using advanced bioinformatic techniques by developing database-independent tools (Welker, 2018a). For example, bioinformatic methods may be used to investigate the function of shell proteins (Skeffington and Donath, 2020) and machine learning algorithms could be employed for automated shell species discrimination based on their protein data, as it has been proposed for collagen-based samples (Gu and Buckley, 2018).

As presented in previous sections, intracrystalline shell proteins are one of the best sources for retrieving ancient phylogenetic information. The oldest mollusc shell DNA sequence was obtained from 100000 years-old bivalve shells that were retrieved from the permafrost in the very far north of Russia (Der Sarkissian et al., 2020). Considering that the DNA extraction from shells is often complicated (Der Sarkissian et al., 2017) and aDNA sequences are less stable than proteins, the findings suggest that ancient shell proteins could easily surpass the longevity limit of 100 ky. Interestingly, pigments are also highly stable biomolecules present in shells (Williams, 2017). A study demonstrated that fossil gastropod shells that were dated to the late Jurassic period (163.5-145 million years ago), under UV light, displayed original biogenic shell pigmentation patterns (Caze et al., 2015). Shell pigments may also be involved in biomineralization and can exist as chromoproteins, *i.e.* bound to shell proteins (Cusack et al., 1992; Affenzeller et al., 2020). These findings support our idea that in ancient biomolecule research, mollusc shells may represent an exceptional biological system.

References

Abelson, P.H., 1954. Amino acids in fossils. Science 119, 576–576.

Addadi, L., Weiner, S., 1997. A pavement of pearl. Nature 389, 912–913.

Aebersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. Nature 422, 198–207.

Affenzeller, S., Wolkenstein, K., Frauendorf, H., Jackson, D.J., 2019. Eumelanin and pheomelanin pigmentation in mollusc shells may be less common than expected: insights from mass spectrometry. Frontiers in zoology 16, 47.

Affenzeller, S., Wolkenstein, K., Frauendorf, H., Jackson, D.J., 2020. Challenging the concept that eumelanin is the polymorphic brown banded pigment in *Cepaea nemoralis*. Scientific reports 10, 2442.

Agbaje, O.B.A., Ben Shir, I., Zax, D.B., Schmidt, A., Jacob, D.E., 2018. Biomacromolecules within bivalve shells: Is chitin abundant? Acta biomaterialia 80, 176–187.

Agbaje, O.B.A., Thomas, D.E., Dominguez, J.G., Mclnerney, B.V., Kosnik, M.A., Jacob, D.E., 2019. Biomacromolecules in bivalve shells with crossed lamellar architecture. Journal of Materials Science 54, 4952–4969.

Akiyama, M., Wyckoff, R.W., 1970. The total amino acid content of fossil pecten shells. Proceedings of the National Academy of Sciences of the United States of America 67, 1097–1100.

Albeck, S., Aizenberg, J., Addadi, L., Weiner, S., 1993. Interactions of various skeletal intracrystalline components with calcite crystals. Journal of the American Chemical Society 115, 11691–11697.

Albeck, S., Addadi, I., Weiner, S., 1996. Regulation of calcite crystal morphology by intracrystalline acidic proteins and glycoproteins. Connective tissue research 35, 365–370.

Allen, M.J., 2017. Molluscs in Archaeology: Methods, Approaches and Applications. Oxbow Books.

Álvarez-Fernández, E., 2011. *Spondylus* shells at Prehistoric Sites in the Iberian Peninsula. In: Ifantidis, F., Nikolaidou, M. (Eds.), Spondylus in Prehistory. New Data and Approaches. Contributions to the Archaeology of Shell Technologies. Archaeopress, Oxford, pp. 13–18.

Álvarez, M., Briz Godino, I., Balbo, A., Madella, M., 2011. Shell middens as archives of past environments, human dispersal and specialized resource management. Quaternary international: the journal of the International Union for Quaternary Research 239, 1–7.

Andersen, S.H., 2000. “Køkkenmøddinger” (shell middens) in Denmark: a survey. In: Proceedings of the Prehistoric Society. Cambridge University Press, pp. 361–384.

Andersen, S.H., 2008. A report on recent excavations at the shell midden of Havnø in Denmark. Mesolithic Miscellany 19, 3–6.

Andrews, A.C., 1948. Oysters as a Food in Greece and Rome. The classical journal 43, 299–303.

Arivalagan, J., Yarra, T., Marie, B., Sleight, V.A., Duvernois-Berthet, E., Clark, M.S., Marie, A., Berland, S., 2017. Insights from the Shell Proteome: Biomineralization to Adaptation. Molecular biology and evolution 34, 66–77.

Bada, J.L., Man, E.H., 1980. Amino acid diagenesis in deep sea drilling project cores: Kinetics and mechanisms of some reactions and their applications in geochronology and in paleotemperature and heat flow determinations. Earth-Science Reviews 16, 21–55.

Bada, J.L., Shou, M.-Y., Man, E.H., Schroeder, R.A., 1978. Decomposition of hydroxy amino acids in foraminiferal tests; kinetics, mechanism and geochronological implications. Earth and planetary science letters 41, 67–76.

Bailey, G.N., Flemming, N.C., 2008. Archaeology of the continental shelf: Marine resources, submerged landscapes and underwater archaeology. Quaternary science reviews 27, 2153–2165.

Bajnóczi, B., Schöll-Barna, G., Kalicz, N., Siklósi, Z., Hourmouziadis, G.H., Ifantidis, F., Kyparissi-Apostolika, A., Pappa, M., Veropoulidou, R., Ziota, C., 2013. Tracing the source of Late Neolithic *Spondylus* shell ornaments by stable isotope geochemistry and cathodoluminescence microscopy. Journal of archaeological science 40, 874–882.

Bar-Yosef Mayer, D.E., Leng, M.J., Aldridge, D.C., Arrowsmith, C., Gümüş, B.A., Sloane, H.J., 2012. Modern and early-middle Holocene shells of the freshwater mollusc Unio, from Çatalhöyük in the Konya Basin, Turkey: preliminary palaeoclimatic implications from molluscan isotope data. Journal of archaeological science 39, 76–83.

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

Bertin, L., 2015. Exploitation of mother of pearl in the Middle Ages, Clos d'Ugnac archaeological site (Pennautier, Aude, France): Malacological study, consumption, exploitation and utilization of the nacre. Quaternary international: the journal of the International Union for Quaternary Research 375, 145–152.

Bobbio, A., 1972. The first endosseous alloplastic implant in the history of man. Bulletin of the history of dentistry 20, 1–6.

Boggild, O.B., 1930. The shell structure of the mollusks. A.F. Host, Kjobenhavn.

Böhm, C.F., Demmert, B., Harris, J., Fey, T., Marin, F., Wolf, S.E., 2016. Structural commonalities and deviations in the hierarchical organization of crossed-lamellar shells: A case study on the shell of the bivalve *Glycymeris glycymeris*. Journal of materials research 31, 536–546.

Bonnardin, S., 2009. La parure funéraire au Néolithique ancien dans les bassins

parisien et rhénan: Rubané, Hinkelstein et Villeneuve-Saint-Germain. Société préhistorique française.

Borbas, J.E., Wheeler, A.P., Sikes, C.S., 1991. Molluscan shell matrix phosphoproteins: Correlation of degree of phosphorylation to shell mineral microstructure and to in vitro regulation of mineralization. The Journal of experimental zoology 258, 1–13.

Boric, D., Cristiani, E., 2019. Special Issue: Personal Ornaments in Early Prehistory --- Taking Beads Seriously: Prehistoric Forager Ornamental Traditions in Southeastern Europe. In: Bar-Yosef Mayer, D.E., Bosch, M.D. (Eds.), PaleoAnthropology. pp. 208–239.

Borrello, M.A., 2005. Le conchiglie nella preistoria e nella protostoria. Preistoria Alpina 40, 19–42.

Borrello, M.A., Dalmeri, G., 2004. Gli ornamenti preistorici lavorati in conchiglie conservati presso il Museo Tridentino di Scienze Naturali (Trento). Preistoria Alpina.

Borrello, M.A., Micheli, R., 2004. *Spondylus gaederopus*, gioiello dell'Europa preistorica. Preistoria Alpina 40, 71–82.

Borrello, M.A., Micheli, R., 2011. *Spondylus gaederopus* in Prehistoric Italy: Jewels from Neolithic and Copper Age Sites. In: Ifantidis, F., Nikolaidou, M. (Eds.), Spondylus in Prehistory. New Data and Approaches. Contributions to the Archaeology of Shell Technologies. Archaeopress, pp. 25–37.

Bosch, M.D., Mannino, M.A., Prendergast, A.L., O'Connell, T.C., Demarchi, B., Taylor, S.M., Niven, L., Plicht, J. van der, Hublin, J.-J., 2015. New chronology for Ksâr 'Akil (Lebanon) supports Levantine route of modern human dispersal into Europe. Proceedings of the National Academy of Sciences of the United States of America 112, 7683–7688.

Boskey, A.L., Villarreal-Ramirez, E., 2016. Intrinsically disordered proteins and biomineralization. Matrix biology: journal of the International Society for Matrix Biology 52-54, 43–59.

Bouchet, P., Bary, S., Virginie, H., Marani, G., 2016. How many species of molluscs are there in the world's oceans, and who is going to describe them? In: Héros, V., Strong, E., Bouchet, P. (Eds.), Tropical Deep-Sea Benthos Volume 29. Muséum National d'Histoire Naturelle. Éditions du Muséum, Paris, pp. 9–24.

Bouzouggar, A., Barton, N., Vanhaeren, M., Errico, F. d', Collcutt, S., Higham, T., Hodge, E., Parfitt, S., Rhodes, E., Schwenninger, J.-L., Stringer, C., Turner, E., Ward, S., Moutmir, A., Stambouli, A., 2007. 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 of the United States of America 104, 9964–9969.

Brandt, E., Wiechmann, I., Grupe, G., 2002. How reliable are immunological tools for the detection of ancient proteins in fossil bones? International Journal of

Osteoarchaeology 12, 307–316.
Brandt, L.Ø., Haase, K., Collins, M.J., 2018. Species identification using ZooMS, with reference to the exploitation of animal resources in the medieval town of Odense. Danish Journal of Archaeology 7, 139–153.
Brooks, A.S., Hare, P.E., Kokis, J.E., Miller, G.H., Ernst, R.D., Wendorf, F., 1990. Dating pleistocene archeological sites by protein diagenesis in ostrich eggshell. Science 248, 60–64.
Buckley, M., 2018. Zooarchaeology by Mass Spectrometry (ZooMS) Collagen Fingerprinting for the Species Identification of Archaeological Bone Fragments. In: Giovas, C.M., LeFebvre, M.J. (Eds.), Zooarchaeology in Practice: Case Studies in Methodology and Interpretation in Archaeofaunal Analysis. Springer International Publishing, Cham, pp. 227–247.
Buckley, M., Collins, M.J., 2011. Collagen survival and its use for species identification in Holocene-lower Pleistocene bone fragments from British archaeological and paleontological sites. Antiqua 1, 1.
Buckley, M., Herman, J., 2019. Species identification of Late Pleistocene bat bones using collagen fingerprinting. International Journal of Osteoarchaeology 29, 1051–1059.
Buckley, M., Wadsworth, C., 2014. Proteome degradation in ancient bone: Diagenesis and phylogenetic potential. Palaeogeography, palaeoclimatology, palaeoecology 416, 69–79.
Buckley, M., Collins, M., Thomas-Oates, J., Wilson, J.C., 2009. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid communications in mass spectrometry: RCM 23, 3843–3854.
Buckley, M., Kansa, S.W., Howard, S., Campbell, S., 2010. Distinguishing between archaeological sheep and goat bones using a single collagen peptide. Journal of.
Buckley, M., Fraser, S., Herman, J., Melton, N.D., Mulville, J., Pálsdóttir, A.H., 2014. Species identification of archaeological marine mammals using collagen fingerprinting. Journal of archaeological science 41, 631–641.
Buckley, M., Warwood, S., Dongen, B. van, Kitchener, A.C., Manning, P.L., 2017. A fossil protein chimera; difficulties in discriminating dinosaur peptide sequences from modern cross-contamination. Proceedings. Biological sciences / The Royal Society 284. doi:10.1098/rspb.2017.0544
Buckley, M., Gu, M., Herman, J., Junno, J.-A., Denys, C., Chamberlain, A.T., 2018. Species identification of voles and lemmings from Late Pleistocene deposits in Pin Hole Cave (Creswell Crags, UK) using collagen fingerprinting. Quaternary international: the journal of the International Union for Quaternary Research 483, 83–89.
Capasso, S., Mazzarella, L., Sica, F., Zagari, A., 1991. First evidence of

spontaneous deamidation of glutamine residue via cyclic imide to α- and γ-glutamic residue under physiological conditions. Journal of the Chemical Society. Chemical communications 1667–1668.

Cappellini, E., Gilbert, M.T.P., Geuna, F., Fiorentino, G., Hall, A., Thomas-Oates, J., Ashton, P.D., Ashford, D.A., Arthur, P., Campos, P.F., Kool, J., Willerslev, E., Collins, M.J., 2010. A multidisciplinary study of archaeological grape seeds. Die Naturwissenschaften 97, 205–217.

Cappellini, E., Jensen, L.J., Szklarczyk, D., Ginolhac, A., Fonseca, R.A.R. da, Stafford, T.W., Holen, S.R., Collins, M.J., Orlando, L., Willerslev, E., Gilbert, M.T.P., Olsen, J.V., 2012. Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. Journal of proteome research 11, 917–926.

Cappellini, E., Prohaska, A., Racimo, F., Welker, F., Pedersen, M.W., Allentoft, M.E., Barros Damgaard, P. de, Gutenbrunner, P., Dunne, J., Hammann, S., Roffet-Salque, M., Ilardo, M., Moreno-Mayar, J.V., Wang, Y., Sikora, M., Vinner, L., Cox, J., Evershed, R.P., Willerslev, E., 2018. Ancient Biomolecules and Evolutionary Inference. Annual review of biochemistry 87, 1029–1060.

Cappellini, E., Welker, F., Pandolfi, L., Ramos-Madrigal, J., Samodova, D., Rüther, P.L., Fotakis, A.K., Lyon, D., Víctor Moreno-Mayar, J., Bukhsianidze, M., Jersie-Christensen, R.R., Mackie, M., Ginolhac, A., Ferring, R., Tappen, M., Palkopoulou, E., Dickinson, M.R., Stafford, T.W., Chan, Y.L., Götherström, A., Senthilvel K S, Heintzman, P.D., Kapp, J.D., Kirillova, I., Moodley, Y., Agusti, J., Kahlke, R.-D., Kiladze, G., Martínez-Navarro, B., Liu, S., Velasco, M.S., Sinding, M.-H.S., Kelstrup, C.D., Allentoft, M.E., Orlando, L., Penkman, K., Shapiro, B., Rook, L., Dalén, L., Gilbert, M.T.P., Olsen, J.V., Lordkipanidze, D., Willerslev, E., 2019. Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny. Nature 1–5.

Carter, J.G. (Ed.), 1990. Skeletal Biomineralization: Pattern, Processes, and Evolutionary Trends. Amer Geophysical Union.

Catherman, A.D., Skinner, O.S., Kelleher, N.L., 2014. Top Down proteomics: facts and perspectives. Biochemical and biophysical research communications 445, 683–693.

Caze, B., Merle, D., Schneider, S., 2015. UV Light Reveals the Diversity of Jurassic Shell Colour Patterns: Examples from the Cordebugle Lagerstätte (Calvados, France). PloS one 10, e0126745.

Chapman, J., Gaydarska, B., 2015. *Spondylus gaederopus*/*Glycymeris* Exchange Networks in the European Neolithic and Chalcolithic. In: Fowler, C., Harding, J., Hofmann, D. (Eds.), The Oxford Handbook of Neolithic Europe. Oxford University Press.

Chapman, J., Gaydarska, B., Slavchev, V., 2008. The life histories of *Spondylus* shell rings from the Varna I Eneolithic cemetery (Northeast Bulgaria):

Transformation, revelation, fragmentation and deposition. In: Acta Musei Varnaensis VI. pp. 139–162.
Chapman, J.C., Gaydarska, B.I., Skafida, E., Souvatzi, S., 2011. Personhood and the life cycle of *Spondylus* rings: an example from Late Neolithic, Greece. In: Ifantidis, F., Nikolaidou, M. (Eds.), Spondylus in Prehistory: New Data and Approaches – Contributions to the Archaeology of Shell Technologies. Archaeopress, pp. 139–160.
Charlton, S., Ramsøe, A., Collins, M., Craig, O.E., Fischer, R., Alexander, M., Speller, C.F., 2019. New insights into Neolithic milk consumption through proteomic analysis of dental calculus. Archaeological and anthropological sciences 11, 6183–6196.
Charpentier, V., Phillips, C.S., Méry, S., 2012. Pearl fishing in the ancient world: 7500 BP. Arabian Archaeology and Epigraphy 23, 1–6.
Checa, A.G., 2018. Physical and Biological Determinants of the Fabrication of Molluscan Shell Microstructures. Frontiers in Marine Science 5, 353.
Cleland, T.P., 2018. Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics. Journal of proteome research 17, 3976–3983.
Cleland, T.P., Schroeter, E.R., 2018. A Comparison of Common Mass Spectrometry Approaches for Paleoproteomics. Journal of proteome research 17, 936–945.
Cleland, T.P., Schroeter, E.R., Schweitzer, M.H., 2015. Biologically and diagenetically derived peptide modifications in moa collagens. Proceedings. Biological sciences / The Royal Society 282, 20150015.
Cölfen, H., Antonietti, M., 2005. Mesocrystals: inorganic superstructures made by highly parallel crystallization and controlled alignment. Angewandte Chemie 44, 5576–5591.
Collins, M., Cappellini, E., Buckley, M., Penkman, K., Griffin, R., Koon, H., 2006. Analytical methods to detect ancient proteins. Bio-and Material Cultures at Qumran 33.
Collins, M.J., Riley, M.S., 2000. Amino acid racemization in biominerals: the impact of protein degradation and loss. In: Goodfriend, G., Collins, M., Fogel, M., Macko, S., Wehmiller, J. (Eds.), Perspectives in Amino Acid and Protein Geochemistry. Oxford University Press, pp. 120–142.
Collins, M.J., Muyzer, G., Westbroek, P., Curry, G.B., Sandberg, P.A., Xu, S.J., Quinn, R., Mackinnon, D., 1991. Preservation of fossil biopolymeric structures: Conclusive immunological evidence. Geochimica et cosmochimica acta 55, 2253–2257.
Collins, M.J., Westbroek, P., Muyzer, G., Leeuw, J.W. de, 1992. Experimental evidence for condensation reactions between sugars and proteins in carbonate skeletons. Geochimica et cosmochimica acta 56, 1539–1544.

Colonese, A.C., Mannino, M.A., Bar-Yosef Mayer, D.E., Fa, D.A., Finlayson, J.C., Lubell, D., Stiner, M.C., 2011. Marine mollusc exploitation in Mediterranean prehistory: An overview. Quaternary international: the journal of the International Union for Quaternary Research 239, 86–103.

Craig, O.E., Bailey, G.N., 2007. Shell Middens in Atlantic Europe. Oxbow Books.

Crisp, M., Demarchi, B., Collins, M., Morgan-Williams, M., Pilgrim, E., Penkman, K., 2013. Isolation of the intra-crystalline proteins and kinetic studies in Struthio camelus (ostrich) eggshell for amino acid geochronology. Quaternary geochronology 16, 110–128.

Currey, J.D., 1977. Mechanical properties of mother of pearl in tension. Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society 196, 443–463.

Curry, G.B., Cusack, M., Walton, D., Endo, K., Clegg, H., Abbott, G., Armstrong, H., Bada, J.L., Engel, M.H., Eglinton, G., Curry, G.B., 1991. Biogeochemistry of brachiopod intracrystalline molecules. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 333, 359–366.

Cusack, M., Curry, G., Clegg, H., Abbott, G., 1992. An intracrystalline chromoprotein from red brachiopod shells: implications for the process of biomineralization. Comparative biochemistry and physiology. B, Comparative biochemistry 102, 93–95.

Dallongeville, S., Garnier, N., Rolando, C., Tokarski, C., 2016. Proteins in Art, Archaeology, and Paleontology: From Detection to Identification. Chemical reviews 116, 2–79.

Demarchi, B., 2020. Amino Acids and Proteins in Fossil Biominerals: An Introduction for Archaeologists and Palaeontologists. John Wiley & Sons.

Demarchi, B., Williams, M.G., Milner, N., Russell, N., Bailey, G., Penkman, K., 2011. Amino acid racemization dating of marine shells: A mound of possibilities. Quaternary international: the journal of the International Union for Quaternary Research 239, 114–124.

Demarchi, B., Collins, M.J., Tomiak, P.J., Davies, B.J., Penkman, K.E.H., 2013a. Intra-crystalline protein diagenesis (IcPD) in *Patella vulgata*. Part II: Breakdown and temperature sensitivity. Quaternary geochronology 16, 158–172.

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

Demarchi, B., O'Connor, S., Lima Ponzoni, A. de, Almeida Rocha Ponzoni, R. de, Sheridan, A., Penkman, K., Hancock, Y., Wilson, J., 2014. An integrated approach to the taxonomic identification of prehistoric shell ornaments. PloS one 9, e99839.

Demarchi, B., Clements, E., Coltorti, M., Locht, R. van de, Kröger, R., Penkman,

K., Rose, J., 2015. Testing the effect of bleaching on the bivalve *Glycymeris*: A case study of amino acid geochronology on key Mediterranean raised beach deposits. Quaternary geochronology 25, 49–65.

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

Demarchi, B., Boano, R., Ceron, A., Bello, F.D., Favero-Longo, S.E., Fiddyment, S., Marochetti, E.F., Mangiapane, G., Mattonai, M., Pennacini, C., Ribechini, E., Woolley, J., Zilberstein, G., Righetti, P.G., 2020a. Never boring: Non-invasive palaeoproteomics of mummified human skin. Journal of archaeological science 119, 105145.

Demarchi, B., Presslee, S., Sakalauskaite, J., Fischer, R., Best, J., 2020b. The role of birds at Çatalhöyük revealed by the analysis of eggshell. Quaternary international: the journal of the International Union for Quaternary Research. doi:10.1016/j.quaint.2020.02.009

Der Sarkissian, C., Pichereau, V., Dupont, C., Ilsøe, P.C., Perrigault, M., Butler, P., Chauvaud, L., Eiríksson, J., Scourse, J., Paillard, C., Orlando, L., 2017. Ancient DNA analysis identifies marine mollusc shells as new metagenomic archives of the past. Molecular ecology resources 17, 835–853.

Der Sarkissian, C., Möller, P., Hofman, C.A., Ilsøe, P., Rick, T.C., Schiøtte, T., Sørensen, M.V., Dalén, L., Orlando, L., 2020. Unveiling the Ecological Applications of Ancient DNA From Mollusk Shells. Frontiers in Ecology and Evolution 8, 37.

Dickinson, M.R., Lister, A.M., Penkman, K.E.H., 2019. A new method for enamel amino acid racemization dating: A closed system approach. Quaternary geochronology 50, 29–46.

Dimitrijević, V., Tripković, B., 2006. *Spondylus* and *Glycymeris* bracelets: trade reflections at Neolithic Vinča-Belo Brdo. Documenta praehistorica 33, 237–252.

Dosztányi, Z., Csizmók, V., Tompa, P., Simon, I., 2005. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. Journal of molecular biology 347, 827–839.

Douka, K., 2011. The Contribution of Archaeometry to the Study of Prehistoric Marine Shells. In: Ifantidis, F., Nikolaidou, M. (Eds.), Spondylus in Prehistory.

New Data and Approaches. Contributions to the Archaeology of Shell Technologies. Archaeopress, pp. 171–180.

Du, J., Xu, G., Liu, C., Zhang, R., 2018. The role of phosphorylation and dephosphorylation of shell matrix proteins in shell formation: an in vivo and in vitro study. CrystEngComm / RSC 20, 3905–3916.

Dupont, C., 2014. Do not mix up apples and oranges! A concept applied to shells from the Mesolithic. In: Cantillo, J.J., Bernal, D., Ramos, J. (Eds.), Moluscos Y Púrpura En Contextosarqueológicos Atlántico-Mediterráneos Nuevos Datos Y Reflexiones En Clave de Proceso Histórico. Universidad de Cadiz, pp. 41–51.

El-Aneed, A., Cohen, A., Banoub, J., 2009. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. Applied Spectroscopy Reviews 44, 210–230.

Evans, J.S., 2012. Aragonite-associated biomineralization proteins are disordered and contain interactive motifs. Bioinformatics 28, 3182–3185.

Evans, J.S., 2019. The Biomineralization Proteome: Protein Complexity for a Complex Bioceramic Assembly Process. Proteomics 19, e1900036.

FAO Fisheries & Aquaculture [WWW Document], n.d. URL http://www.fao.org/fishery/statistics/collections/en (accessed 3.31.20).

Feng, D., Li, Q., Yu, H., Kong, L., Du, S., 2017. Identification of conserved proteins from diverse shell matrix proteome in *Crassostrea gigas*: characterization of genetic bases regulating shell formation. Scientific reports 7, 45754.

Fernandes, D.M., Strapagiel, D., Borówka, P., Marciniak, B., Żądzińska, E., Sirak, K., Siska, V., Grygiel, R., Carlsson, J., Manica, A., Lorkiewicz, W., Pinhasi, R., 2018. A genomic Neolithic time transect of hunter-farmer admixture in central Poland. Scientific reports 8, 14879.

Fiddyment, S., Holsinger, B., Ruzzier, C., Devine, A., Binois, A., Albarella, U., Fischer, R., Nichols, E., Curtis, A., Cheese, E., Teasdale, M.D., Checkley-Scott, C., Milner, S.J., Rudy, K.M., Johnson, E.J., Vnouček, J., Garrison, M., McGrory, S., Bradley, D.G., Collins, M.J., 2015. Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting. Proceedings of the National Academy of Sciences of the United States of America 112, 15066–15071.

Fotakis, A.K., Denham, S.D., Mackie, M., Orbegozo, M.I., Mylopotamitaki, D., Gopalakrishnan, S., Sicheritz-Pontén, T., Olsen, J.V., Cappellini, E., Zhang, G., Christophersen, A., Gilbert, M.T.P., Vågene, Å.J., 2020. Multi-omic detection of *Mycobacterium leprae* in archaeological human dental calculus. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 375, 20190584.

Galinou-Mitsoudi, S., Imsiridou, A., Koutra, A., Samaras, D., Samara, E., 2019. Ecology, life cycle and genetic approach of *Spondylus gaederopus* Linnaeus, 1758 (bivalvia). Journal of Environmental Protection and Ecology 20, 713–722.

Gao, P., Liao, Z., Wang, X.-X., Bao, L.-F., Fan, M.-H., Li, X.-M., Wu, C.-W., Xia, S.-W., 2015. Layer-by-Layer Proteomic Analysis of *Mytilus galloprovincialis* Shell. PloS one 10, e0133913.

Gautron, J., 2019. Proteomics Analysis of Avian Eggshell Matrix Proteins: Toward New Advances on Biomineralisation. Proteomics 1900120.

GBIF.org, Global Biodiversity Information Facility. Available from: https://www.gbif.org
Spondylus GBIF Occurrence Download https://doi.org/10.15468/dl.pyss4g

Giacobini, G., 2007. Richness and Diversity of Burial Rituals in the Upper Paleolithic. Diogenes 54, 19–39.

Gilbert, P., Abrecht, M., Frazer, B.H., 2005. The Organic-Mineral Interface in Biominerals. Reviews in Mineralogy and Geochemistry 59, 157–185.

Gim, J., Schnitzer, N., Otter, L.M., Cui, Y., Motreuil, S., Marin, F., Wolf, S.E., Jacob, D.E., Misra, A., Hovden, R., 2019. Nanoscale deformation mechanics reveal resilience in nacre of *Pinna nobilis* shell. Nature communications 10, 4822.

Grégoire, C., Voss-Foucart, M.F., 1970. Proteins in shells of fossil cephalopods (nautiloids and ammonoids) and experimental simulation of their alterations. Archives internationales de physiologie et de biochimie 78, 191–203.

Guillen, J., Natale, F., Carvalho, N., Casey, J., Hofherr, J., Druon, J.-N., Fiore, G., Gibin, M., Zanzi, A., Martinsohn, J.T., 2019. Global seafood consumption footprint. Ambio 48, 111–122.

Gu, M., Buckley, M., 2018. Semi-supervised machine learning for automated species identification by collagen peptide mass fingerprinting. BMC bioinformatics 19, 241.

Gurova, M., Bonsall, C., 2017. Experimental Replication of Stone, Bone and Shell Beads from Early Neolithic Sites in Southeast Europe. In: Bar-Yosef Mayer, D., Bonsall, C., Choyke, A.M. (Eds.), Not Just for Show: The Archaeology of Beads, Beadwork and Personal Ornaments. Oxbow Books, pp. 159–168.

Haag, W.R., 2012. North American Freshwater Mussels: Natural History, Ecology, and Conservation. Cambridge University Press.

Harvey, V.L., Daugnora, L., Buckley, M., 2018. Species identification of ancient Lithuanian fish remains using collagen fingerprinting. Journal of archaeological science 98, 102–111.

Heaton, K., Solazzo, C., Collins, M.J., Thomas-Oates, J., Bergström, E.T., 2009. Towards the application of desorption electrospray ionisation mass spectrometry (DESI-MS) to the analysis of ancient proteins from artefacts. Journal of archaeological science 36, 2145–2154.

Hendy, J., Welker, F., Demarchi, B., Speller, C., Warinner, C., Collins, M.J., 2018a. A guide to ancient protein studies. Nature ecology & evolution 2, 791–799.

Hendy, J., Colonese, A.C., Franz, I., Fernandes, R., Fischer, R., Orton, D., Lucquin, A., Spindler, L., Anvari, J., Stroud, E., Biehl, P.F., Speller, C., Boivin, N.,

Mackie, M., Jersie-Christensen, R.R., Olsen, J.V., Collins, M.J., Craig, O.E., Rosenstock, E., 2018b. Ancient proteins from ceramic vessels at Çatalhöyük West reveal the hidden cuisine of early farmers. Nature communications 9, 4064.

Hendy, J., Warinner, C., Bouwman, A., Collins, M.J., Fiddyment, S., Fischer, R., Hagan, R., Hofman, C.A., Holst, M., Chaves, E., Klaus, L., Larson, G., Mackie, M., McGrath, K., Mundorff, A.Z., Radini, A., Rao, H., Trachsel, C., Velsko, I.M., Speller, C.F., 2018c. Proteomic evidence of dietary sources in ancient dental calculus. Proceedings. Biological sciences / The Royal Society 285. doi:10.1098/rspb.2018.0977

Herlitze, I., Marie, B., Marin, F., Jackson, D.J., 2018. Molecular modularity and asymmetry of the molluscan mantle revealed by a gene expression atlas. GigaScience 7. doi:10.1093/gigascience/giy056

Heumüller, M., 2009. Der Schmuck der jungneolithischen Seeufersiedlung Hornstaad-Hörnle IA im Rahmen des mitteleuropäischen Mittel- und Jungneolithikums. K. Theiss.

Hill, R.L., 1965. Hydrolysis of proteins. Advances in protein chemistry 20, 37–107.

Hoffmann, D.L., Angelucci, D.E., Villaverde, V., Zapata, J., Zilhão, J., 2018. Symbolic use of marine shells and mineral pigments by Iberian Neandertals 115,000 years ago. Science advances 4, eaar5255.

Hoffmann, E. de, Stroobant, V., 2013. Mass Spectrometry: Principles and Applications, 3rd ed. Wiley-Interscience.

Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-Del-Molino, D., Dorp, L. van, López, S., Kousathanas, A., Link, V., Kirsanow, K., Cassidy, L.M., Martiniano, R., Strobel, M., Scheu, A., Kotsakis, K., Halstead, P., Triantaphyllou, S., Kyparissi-Apostolika, N., Urem-Kotsou, D., Ziota, C., Adaktylou, F., Gopalan, S., Bobo, D.M., Winkelbach, L., Blöcher, J., Unterländer, M., Leuenberger, C., Çilingiroğlu, Ç., Horejs, B., Gerritsen, F., Shennan, S.J., Bradley, D.G., Currat, M., Veeramah, K.R., Wegmann, D., Thomas, M.G., Papageorgopoulou, C., Burger, J., 2016. Early farmers from across Europe directly descended from Neolithic Aegeans. Proceedings of the National Academy of Sciences of the United States of America 113, 6886–6891.

Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., Pääbo, S., 2001. Ancient DNA. Nature reviews. Genetics 2, 353–359.

Ifantidis, F., Nikolaidou, M. (Eds.), 2011. *Spondylus* in Prehistory: New Data and Approaches: Contributions to the Archaeology of Shell Technologies. Archaeopress, Oxford.

Jackson, A.P., Vincent, J.F.V., Turner, R.M., 1990. Comparison of nacre with other ceramic composites. Journal of Materials Science 25, 3173–3178.

Jackson, D.J., McDougall, C., Woodcroft, B., Moase, P., Rose, R.A., Kube, M.,

Reinhardt, R., Rokhsar, D.S., Montagnani, C., Joubert, C., Piquemal, D., Degnan, B.M., 2010. Parallel evolution of nacre building gene sets in molluscs. Molecular biology and evolution 27, 591–608.

Jeong, C., Wilkin, S., Amgalantugs, T., Bouwman, A.S., Taylor, W.T.T., Hagan, R.W., Bromage, S., Tsolmon, S., Trachsel, C., Grossmann, J., Littleton, J., Makarewicz, C.A., Krigbaum, J., Burri, M., Scott, A., Davaasambuu, G., Wright, J., Irmer, F., Myagmar, E., Boivin, N., Robbeets, M., Rühli, F.J., Krause, J., Frohlich, B., Hendy, J., Warinner, C., 2018. Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. Proceedings of the National Academy of Sciences of the United States of America 115, E11248–E11255.

Jersie-Christensen, R.R., Lanigan, L.T., Lyon, D., Mackie, M., Belstrøm, D., Kelstrup, C.D., Fotakis, A.K., Willerslev, E., Lynnerup, N., Jensen, L.J., Cappellini, E., Olsen, J.V., 2018. Quantitative metaproteomics of medieval dental calculus reveals individual oral health status. Nature communications 9, 4744.

Jin, C., Li, J.-L., Liu, X.-J., 2020. Teosin, a novel basic shell matrix protein from *Hyriopsis cumingii* induces calcium carbonate polycrystal formation. International journal of biological macromolecules 150, 1229–1237.

John, J., 2011. Status of Spondylus artifacts within the LBK grave goods. In: Ifantidis, F., Nikolaidou, M. (Eds.), Spondylus in Prehistory. New Data and Approaches. Contributions to the Archaeology of Shell Technologies. Archaeopress, Oxford, pp. 39–45.

John Lowe, J., Walker, M.J.C., 2014. Reconstructing Quaternary Environments, 3 edition. ed. Routledge.

Jong, E.W. de, Westbroek, P., Westbroek, J.W., Bruning, J.W., 1974. Preservation of antigenic properties of macromolecules over 70 Myr. Nature 252, 63–64.

Jonuks, T., Oras, E., Best, J., Demarchi, B., Mänd, R., Presslee, S., Vahur, S., 2018. Multi-method Analysis of Avian Eggs as Grave Goods: Revealing Symbolism in Conversion Period Burials at Kukruse, NE Estonia. Environmental Archaeology 23, 109–122.

Jope, M., 1967. The protein of brachiopod shell–II. Shell protein from fossil articulates: Amino acid composition. Comparative biochemistry and physiology 20, 601–605.

Kamat, S., Su, X., Ballarini, R., Heuer, A.H., 2000. Structural basis for the fracture toughness of the shell of the conch *Strombus gigas*. Nature 405, 1036–1040.

Kannegaard, E., 2013. Småting fra Nederst. Årbog fra museum Østjylland 207–208.

Kanold, J.M., Guichard, N., Immel, F., Plasseraud, L., Corneillat, M., Alcaraz, G., Brümmer, F., Marin, F., 2015. Spine and test skeletal matrices of the Mediterranean sea urchin *Arbacia lixula*--a comparative characterization of

their sugar signature. The FEBS journal 282, 1891–1905.
Katti, K.S., Katti, D.R., Pradhan, S.M., Bhosle, A., 2005. Platelet interlocks are the key to toughness and strength in nacre. Journal of materials research 20, 1097–1100.
Keith, M.L., Anderson, G.M., Eichler, R., 1964. Carbon and oxygen isotopic composition of mollusk shells from marine and fresh-water environments. Geochimica et cosmochimica acta 28, 1757–1786.
Kim, Y.-Y., Carloni, J.D., Demarchi, B., Sparks, D., Reid, D.G., Kunitake, M.E., Tang, C.C., Duer, M.J., Freeman, C.L., Pokroy, B., Penkman, K., Harding, J.H., Estroff, L.A., Baker, S.P., Meldrum, F.C., 2016. Tuning hardness in calcite by incorporation of amino acids. Nature materials 15, 903–910.
Kirby, D.P., Buckley, M., Promise, E., Trauger, S.A., Holdcraft, T.R., 2013. Identification of collagen-based materials in cultural heritage. The Analyst 138, 4849–4858.
Kistler, L., Ware, R., Smith, O., Collins, M., Allaby, R.G., 2017. A new model for ancient DNA decay based on paleogenomic meta-analysis. Nucleic acids research 45, 6310–6320.
Knoll, A.H., 2003. Biomineralization and Evolutionary History. Reviews in Mineralogy and Geochemistry 54, 329–356.
Kocot, K.M., Aguilera, F., McDougall, C., Jackson, D.J., Degnan, B.M., 2016. Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. Frontiers in zoology 13, 23.
Landman, N.H., Mikkelsen, P., 2001. Pearls: A Natural History, 1st ed. Harry N. Abrams.
Laporte, L., Dupont, C., 2019. Special Issue: Early Personal Ornaments --- Personal Adornments and Objects of Ornamentation: Two Case Studies From Hunter-Gatherer Burials in France (La Vergne) and Argentine (Arroyo Seco II). In: Bar-Yosef Mayer, D.E., Bosch, M.D. (Eds.), PaleoAnthropology. pp. 156–176.
Broadhurst, L.C., Cunnane, S.C., Crawford, M.A., 1998. Rift Valley lake fish and shellfish provided brain-specific nutrition for early Homo. The British journal of nutrition 79, 3–21.
Leng, M.J., Lewis, J.P., 2016. Oxygen isotopes in Molluscan shell: Applications in environmental archaeology. Environmental Archaeology 21, 295–306.
Le Roy, N., Jackson, D.J., Marie, B., Ramos-Silva, P., Marin, F., 2014. The evolution of metazoan α-carbonic anhydrases and their roles in calcium carbonate biomineralization. Frontiers in zoology 11, 75.
Lindqvist, C., Rajora, O.P. (Eds.), 2019. Paleogenomics: Genome-Scale Analysis of Ancient DNA. Springer International Publishing.
Liu, X., Pu, J., Zeng, S., Jin, C., Dong, S., Li, J., 2017. *Hyriopsis cumingii* Hic52-A novel nacreous layer matrix protein with a collagen-like structure. International

journal of biological macromolecules 102, 667–673.
Lowenstam, H.A., Weiner, S., 1989. On Biomineralization. Oxford University Press.
Luz, G.M., Mano, J.F., 2009. Biomimetic design of materials and biomaterials inspired by the structure of nacre. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences 367, 1587–1605.
Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., Lajoie, G., 2003. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid communications in mass spectrometry: RCM 17, 2337–2342.
MacKenzie, J., Clyde, L., 1996. History of Oystering in the United States and Canada, Featuring the Eight Greatest Oyster Estuaries. Marine Fisheries Review 58, 1–78.
Mackie, M., Hendy, J., Lowe, A.D., Sperduti, A., Holst, M., Collins, M.J., Speller, C.F., 2017. Preservation of the metaproteome: variability of protein preservation in ancient dental calculus. Science and technology of archaeological research 3, 74–86.
Mackie, M., Rüther, P., Samodova, D., Di Gianvincenzo, F., Granzotto, C., Lyon, D., Peggie, D.A., Howard, H., Harrison, L., Jensen, L.J., Olsen, J.V., Cappellini, E., 2018. Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. Angewandte Chemie 57, 7369–7374.
Maier, E., Titschack, J., 2010. *Spondylus gaederopus*: A new Mediterranean climate archive — Based on high-resolution oxygen and carbon isotope analyses. Palaeogeography, palaeoclimatology, palaeoecology 291, 228–238.
Maillard, L.C., 1912. Action of amino acids on sugars. Formation of melanoidins in a methodical way. Compte-Rendu de l'Academie des Sciences 154, 66–68.
Maltby, M., Allen, M., Best, J., Fothergill, B.T., Demarchi, B., 2018. Counting Roman chickens: Multidisciplinary approaches to human-chicken interactions in Roman Britain. Journal of Archaeological Science: Reports 19, 1003–1015.
Mannino, M.A., Thomas, K.D., Leng, M.J., Piperno, M., Tusa, S., Tagliacozzo, A., 2007. Marine resources in the Mesolithic and Neolithic at the Grotta dell'Uzzo (Sicily): evidence from isotope analyses of marine shells. Archaeometry 49, 117–133.
Mann, K., Edsinger, E., 2014. The *Lottia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. Proteome science 12, 28.
Mann, K., Jackson, D.J., 2014. Characterization of the pigmented shell-forming proteome of the common grove snail Cepaea nemoralis. BMC genomics 15, 249.
Mann, K., Edsinger-Gonzales, E., Mann, M., 2012. In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet Lottia gigantea. Proteome science 10, 28.

Mao, J., Zhang, W., Zhang, X., Tian, Y., Wang, X., Hao, Z., Chang, Y., 2018. Transcriptional changes in the Japanese scallop (*Mizuhopecten yessoensis*) shellinfested by Polydora provide insights into the molecular mechanism of shell formation and immunomodulation. Scientific reports 8, 17664.

Marean, C.W., 2014. The origins and significance of coastal resource use in Africa and Western Eurasia. Journal of human evolution 77, 17–40.

Marie, B., Luquet, G., Pais De Barros, J.-P., Guichard, N., Morel, S., Alcaraz, G., Bollache, L., Marin, F., 2007. The shell matrix of the freshwater mussel *Unio pictorum* (Paleoheterodonta, Unionoida). Involvement of acidic polysaccharides from glycoproteins in nacre mineralization. The FEBS journal 274, 2933–2945.

Marie, B., Marin, F., Marie, A., Bédouet, L., Dubost, L., Alcaraz, G., Milet, C., Luquet, G., 2009a. Evolution of nacre: biochemistry and proteomics of the shell organic matrix of the cephalopod *Nautilus macromphalus*. Chembiochem: a European journal of chemical biology 10, 1495–1506.

Marie, B., Le Roy, N., Marie, A., Dubost, L., Marin, F., 2009b. Nacre Evolution: A Proteomic Approach. MRS Online Proceeding Library Archive 1187. doi:10.1557/PROC-1187-KK01-03

Marie, B., Marie, A., Jackson, D.J., Dubost, L., Degnan, B.M., Milet, C., Marin, F., 2010. Proteomic analysis of the organic matrix of the abalone Haliotis asinina calcified shell. Proteome science 8, 54.

Marie, B., Trinkler, N., Zanella-Cleon, I., Guichard, N., Becchi, M., Paillard, C., Marin, F., 2011. Proteomic identification of novel proteins from the calcifying shell matrix of the Manila clam *Venerupis philippinarum*. Marine biotechnology 13, 955–962.

Marie, B., Joubert, C., Tayalé, A., Zanella-Cléon, I., Belliard, C., Piquemal, D., Cochennec-Laureau, N., Marin, F., Gueguen, Y., Montagnani, C., 2012. Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. Proceedings of the National Academy of Sciences of the United States of America 109, 20986–20991.

Marie, B., Ramos-Silva, P., Marin, F., Marie, A., 2013a. Proteomics of CaCO3 biomineral-associated proteins: how to properly address their analysis. Proteomics 13, 3109–3116.

Marie, B., Jackson, D.J., Ramos-Silva, P., Zanella-Cléon, I., Guichard, N., Marin, F., 2013b. The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties. The FEBS journal 280, 214–232.

Marie, B., Arivalagan, J., Mathéron, L., Bolbach, G., Berland, S., Marie, A., Marin, F., 2017. Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the *Unionoida Elliptio complanata* and *Villosa lienosa*. Journal of the Royal Society, Interface / the Royal Society 14. doi:10.1098/rsif.2016.0846

Marin, F., 2020. Mollusc shellomes: Past, present and future. Journal of structural biology 212, 107583.
Marin, F., Gautret, P., 1994. Les Teneurs en acides amines acides des matrices organiques solubles associees aux squelettes calcaires des demosponges et des cnidaires; une implication possible dans leur transformation diagenetique. Bulletin de la Société Géologique de France 165, 77–84.
Marin, F., Luquet, G., 2004. Molluscan shell proteins. Comptes rendus. Palevol 3, 469–492.
Marin, F., Luquet, G., 2005. Molluscan biomineralization: The proteinaceous shell constituents of *Pinna nobilis* L. Materials Science and Engineering: C 25, 105–111.
Marin, F., Luquet, G., 2008. Unusually Acidic Proteins in Biomineralization. In: Bäuerlein, E. (Ed.), Handbook of Biomineralization: Biological Aspects and Structure Formation. pp. 273–290.
Marin, F., Luquet, G., Marie, B., Medakovic, D., 2007a. Molluscan Shell Proteins: Primary Structure, Origin, and Evolution. In: Current Topics in Developmental Biology. Academic Press, pp. 209–276.
Marin, F., Pokroy, B., Luquet, G., Layrolle, P., De Groot, K., 2007b. Protein mapping of calcium carbonate biominerals by immunogold. Biomaterials 28, 2368–2377.
Marin, F., Le Roy, N., Marie, B., 2012. The formation and mineralization of mollusk shell. Frontiers in bioscience 4, 1099–1125.
Marin, F., Marie, B., Hamada, S.B., Ramos-Silva, P., Le Roy, N., Guichard, N., Wolf, S.E., Montagnani, C., Joubert, C., Piquemal, D., Saulnier, D., Gueguen, Y., 2013. “Shellome”: Proteins involved in mollusk shell biomineralization-diversity, functions. In: Watabe, S., Maeyama, K., Nagasawa, H. (Eds.), International Symposium on Pearl Research. Terrapub Tokyo, pp. 149–166.
Marin, F., Le Roy, N., Marie, B., Ramos-Silva, P., Wolf, S., Benhamada, S., Guichard, N., Immel, F., 2014. Synthesis of calcium carbonate biological materials: how many proteins are needed? Key engineering materials 614.
Marin, F., Bundeleva, I., Takeuchi, T., Immel, F., Medakovic, D., 2016. Organic matrices in metazoan calcium carbonate skeletons: Composition, functions, evolution. Journal of structural biology 196, 98–106.
Matter, P., 3rd, Davidson, F.D., Wyckoff, R.W., 1969. The composition of fossil oyster shell proteins. Proceedings of the National Academy of Sciences of the United States of America 64, 970–972.
McConnaughey, T.A., Gillikin, D.P., 2008. Carbon isotopes in mollusk shell carbonates. Geo-Marine Letters 28, 287–299.
McDougall, C., Degnan, B.M., 2018. The evolution of mollusc shells. Wiley interdisciplinary reviews. Developmental biology 7, e313.
Mészáros, B., Erdos, G., Dosztányi, Z., 2018. IUPred2A: context-dependent

prediction of protein disorder as a function of redox state and protein binding. Nucleic acids research 46, W329–W337.

Micheli, R., Bernardini, F., 2018. Ornamenti personali in conchiglia di *Spondylus*: applicazione sperimentale dell'analisi microCT per la determinazione del genere: potenzialità e limiti del metodo. In: Borgna E., C.G.P.&. C.S. (Ed.), Preistoria E Protostoria Del Caput Adriae, Studi Di Preistoria E Protostoria 5. Presented at the IIPP, pp. 265–278.

Milano, S., Nehrke, G., 2018. Microstructures in relation to temperature-induced aragonite-to-calcite transformation in the marine gastropod *Phorcus turbinatus*. PloS one 13, e0204577.

Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V., Tõrv, M., Furtwängler, A., Andrades Valtueña, A., Feldman, M., Economou, C., Oinonen, M., Vasks, A., Balanovska, E., Reich, D., Jankauskas, R., Haak, W., Schiffels, S., Krause, J., 2018. The genetic prehistory of the Baltic Sea region. Nature communications 9, 442.

Moshkovitz, S., 1971. Neolithic *Spondylus*. Nature.

Mouchi, V., Briard, J., Gaillot, S., Argant, T., Forest, V., Emmanuel, L., 2017. Reconstructing environments of collection sites from archaeological bivalve shells: Case study from oysters (Lyon, France). Journal of Archaeological Science: Reports. doi:10.1016/j.jasrep.2017.10.025

Mulligan, C.J., 2005. Isolation and Analysis of DNA from Archaeological, Clinical, and Natural History Specimens. In: Methods in Enzymology. Academic Press, pp. 87–103.

Mullis, K.B., 1990. Target amplification for DNA analysis by the polymerase chain reaction. Annales de biologie clinique 48, 579–582.

Nagai, K., 2013. A history of the cultured pearl industry. Zoological science 30, 783–793.

Nikolaidou, M., 2003. Items of Adornment. In: Elster, E.S., Renfrew, C. and (Eds.), Prehistoric Sitagroi: Excavations in Northeast Greece, 1968-1970 Volume 2: The Final Report, Monumenta Archaeologica 20. Cotsen Institute of Archaeology, Los Angeles, Los Angeles, p. 390.

Oldfield, C.J., Uversky, V.N., Dunker, A.K., Kurgan, L., 2019. Chapter 1 - Introduction to intrinsically disordered proteins and regions. In: Salvi, N. (Ed.), Intrinsically Disordered Proteins. Academic Press, pp. 1–34.

O'Rourke, M.B., Djordjevic, S.P., Padula, M.P., 2018. The quest for improved reproducibility in MALDI mass spectrometry. Mass spectrometry reviews 37, 217–228.

Ortiz, J.E., Sánchez-Palencia, Y., Gutiérrez-Zugasti, I., Torres, T., González-Morales, M., 2018. Protein diagenesis in archaeological gastropod shells and the suitability of this material for amino acid racemisation dating: *Phorcus*

lineatus (da Costa, 1778). Quaternary geochronology 46, 16–27.
Osuna-Mascaró, A., Cruz-Bustos, T., Benhamada, S., Guichard, N., Marie, B., Plasseraud, L., Corneillat, M., Alcaraz, G., Checa, A., Marin, F., 2014. The shell organic matrix of the crossed lamellar queen conch shell (*Strombus gigas*). Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology 168, 76–85.
Oudot, M., Neige, P., Shir, I.B., Schmidt, A., Strugnell, J.M., Plasseraud, L., Broussard, C., Hoffmann, R., Lukeneder, A., Marin, F., 2020. The shell matrix and microstructure of the Ram's Horn squid: Molecular and structural characterization. Journal of structural biology 107507.
Pääbo, S., 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proceedings of the National Academy of Sciences of the United States of America 86, 1939–1943.
Palmer, A.R., 1992. Calcification in marine molluscs: how costly is it? Proceedings of the National Academy of Sciences of the United States of America 89, 1379–1382.
Pappa, M., Veropoulidou, R., 2011. The Neolithic settlement at Makriyalos, Northern Greece: Evidence from the *Spondylus gaederopus* artifacts. In: Ifantidis, F., Marianna, N. (Eds.), Spondylus in Prehistory: New Data and Approaches - Contributions to the Archaeology of Shell Technologies. Archaeopress, pp. 191–208.
Paulsen, A.C., 1974. The Thorny Oyster and the Voice of God: *Spondylus* and *Strombus* in Andean Prehistory. American antiquity 39, 597–607.
Pavat, C., Zanella-Cléon, I., Becchi, M., Medakovic, D., Luquet, G., Guichard, N., Alcaraz, G., Dommergues, J.-L., Serpentini, A., Lebel, J.-M., Marin, F., 2012. The shell matrix of the pulmonate land snail *Helix aspersa maxima*. Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology 161, 303–314.
Penkman, K.E.H., Kaufman, D.S., Maddy, D., Collins, M.J., 2008. Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells. Quaternary geochronology 3, 2–25.
Penkman, K.E.H., Preece, R.C., Bridgland, D.R., Keen, D.H., Meijer, T., Parfitt, S.A., White, T.S., Collins, M.J., 2011. A chronological framework for the British Quaternary based on *Bithynia opercula*. Nature 476, 446–449.
Peresani, M., Forte, M., Romandini, M., Cilli, C., Giacobini, G., 2019. Special Issue: Early Personal Ornaments --- Marine and Freshwater Shell Exploitaiton in the Early Upper Paleolithic: Re-Examination of the Assemblages from Fumane Cave (NE Italy). In: Bar-Yosef Mayer, D.E., Bosch, M.D. (Eds.), PaleoAnthropology. pp. 61–81.
Perlès, C., 2019. Special Issue: Personal Ornaments in Early Prehistory --- Cultural Implications of Uniformity in Ornament Assemblages: Paleolithic and

Mesolithic Ornaments From Franchthi Cave, Greece. In: Bar-Yosef Mayer, D.E., Bosch, M.D. (Eds.), PaleoAnthropology. pp. 196–207.

Pierini, F., Demarchi, B., Turner, J., Penkman, K., 2016. *Pecten* as a new substrate for IcPD dating: The quaternary raised beaches in the Gulf of Corinth, Greece. Quaternary geochronology 31, 40–52.

Pothier Bouchard, G., Riel-Salvatore, J., Negrino, F., Buckley, M., 2020. Archaeozoological, taphonomic and ZooMS insights into The Protoaurignacian faunal record from Riparo Bombrini. Quaternary international: the journal of the International Union for Quaternary Research. doi:10.1016/j.quaint.2020.01.007

Powell, D., Subramanian, S., Suwansa-Ard, S., Zhao, M., O'Connor, W., Raftos, D., Elizur, A., 2018. The genome of the oyster *Saccostrea* offers insight into the environmental resilience of bivalves. DNA research: an international journal for rapid publication of reports on genes and genomes 25, 655–665.

Prendergast, A.L., Stevens, R.E., O'Connell, T.C., Hill, E.A., Hunt, C.O., Barker, G.W., 2016. A late Pleistocene refugium in Mediterranean North Africa? Palaeoenvironmental reconstruction from stable isotope analyses of land snail shells (Haua Fteah, Libya). Quaternary science reviews 139, 94–109.

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

Presslee, S., Slater, G.J., Pujos, F., Forasiepi, A.M., Fischer, R., Molloy, K., Mackie, M., Olsen, J.V., Kramarz, A., Taglioretti, M., Scaglia, F., Lezcano, M., Lanata, J.L., Southon, J., Feranec, R., Bloch, J., Hajduk, A., Martin, F.M., Salas Gismondi, R., Reguero, M., Muizon, C. de, Greenwood, A., Chait, B.T., Penkman, K., Collins, M., MacPhee, R.D.E., 2019. Palaeoproteomics resolves sloth relationships. Nature ecology & evolution 3, 1121–1130.

Ramos-Silva, P., Benhamada, S., Le Roy, N., Marie, B., Guichard, N., Zanella-Cléon, I., Plasseraud, L., Corneillat, M., Alcaraz, G., Kaandorp, J., Marin, F., 2012. Novel molluskan biomineralization proteins retrieved from proteomics: a case study with Upsalin. Chembiochem: a European journal of chemical biology 13, 1067–1078.

Ramsøe, A., Heekeren, V. van, Ponce, P., Fischer, R., Barnes, I., Speller, C., Collins, M.J., 2020. DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes. Journal of archaeological science 115, 105080.

Richter, K.K., Wilson, J., Jones, A.K.G., Buckley, M., Doorn, N. van, Collins, M.J., 2011. Fish'n chips: ZooMS peptide mass fingerprinting in a 96 well plate format to identify fish bone fragments. Journal of archaeological science 38, 1502–1510.

Rietschel, B., Arrey, T.N., Meyer, B., Bornemann, S., Schuerken, M., Karas, M., Poetsch, A., 2009. Elastase digests: new ammunition for shotgun membrane proteomics. Molecular & cellular proteomics: MCP 8, 1029–1043.

Rigaud, S., Errico, F. d', Vanhaeren, M., 2015. Ornaments reveal resistance of North European cultures to the spread of farming. PloS one 10, e0121166.

Robinson, N.E., Robinson, A., 2004. Molecular clocks: deamidation of asparaginyl and glutaminyl residues in peptides and proteins. Althouse press.

Robinson, N.E., Robinson, A.B., 2001. Prediction of protein deamidation rates from primary and three-dimensional structure. Proceedings of the National Academy of Sciences of the United States of America 98, 4367–4372.

Romandini, M., Peresani, M., Laroulandie, V., Metz, L., Pastoors, A., Vaquero, M., Slimak, L., 2014. Convergent evidence of eagle talons used by late Neanderthals in Europe: a further assessment on symbolism. PloS one 9, e101278.

Rousseau, M., Bédouet, L., Lati, E., Gasser, P., Le Ny, K., Lopez, E., 2006. Restoration of stratum corneum with nacre lipids. Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology 145, 1–9.

Russell-Hunter, W.D., 1983. 1 - Overview: Planetary Distribution of and Ecological Constraints upon the Mollusca. In: Russell-Hunter, W.D. (Ed.), Ecology. Academic Press, pp. 1–27.

Samata, T., Ikeda, D., Kajikawa, A., Sato, H., Nogawa, C., Yamada, D., Yamazaki, R., Akiyama, T., 2008. A novel phosphorylated glycoprotein in the shell matrix of the oyster *Crassostrea nippona*: A novel acidic glycoprotein from the oyster shells. The FEBS journal 275, 2977–2989.

Sarashina, I., Endo, K., 2001. The complete primary structure of molluscan shell protein 1 (MSP-1), an acidic glycoprotein in the shell matrix of the scallop *Patinopecten yessoensis*. Marine biotechnology 3, 362–369.

Sarashina, I., Endo, K., 2006. Skeletal matrix proteins of invertebrate animals: Comparative analysis of their amino acid sequences. Paleontological Research 10, 311–336.

Saunders, N.J., 1999. Biographies of Brilliance: Pearls, Transformations of Matter and being, c. AD 1492. World archaeology 31, 243–257.

Schöne, B.R., Fiebig, J., Pfeiffer, M., Gleβ, R., Hickson, J., Johnson, A.L.A., Dreyer, W., Oschmann, W., 2005. Climate records from a bivalved Methuselah (*Arctica islandica*, Mollusca; Iceland). Palaeogeography, palaeoclimatology, palaeoecology 228, 130–148.

Schroeder, R.A., Bada, J.L., 1976. A review of the geochemical applications of the amino acid racemization reaction. Earth-Science Reviews 12, 347–391.

Schroeter, E.R., DeHart, C.J., Schweitzer, M.H., Thomas, P.M., Kelleher, N.L., 2016. Bone protein "extractomics": comparing the efficiency of bone protein extractions of Gallus gallus in tandem mass spectrometry, with an eye towards

paleoproteomics. PeerJ 4, e2603.
Shackleton, J., Elderfield, H., 1990. Strontium isotope dating of the source of Neolithic European *Spondylus* shell artefacts. Antiquity 64, 312–315.
Shackleton, N., Renfrew, C., 1970. Neolithic trade routes re-aligned by oxygen isotope analyses. Nature 228, 1062–1065.
Shackleton, N.J., 1968. Knossos marine mollusca (Neolithic). Annual of the British School at Athens 63, 264–266.
Simpson, J.P., Penkman, K.E.H., Demarchi, B., Koon, H., Collins, M.J., Thomas-Oates, J., Shapiro, B., Stark, M., Wilson, J., 2016. The effects of demineralisation and sampling point variability on the measurement of glutamine deamidation in type I collagen extracted from bone. Journal of archaeological science 69, 29–38.
Simpson, J.P., Fascione, M., Bergström, E., Wilson, J., Collins, M.J., Penkman, K.E.H., Thomas-Oates, J., 2019. Ionisation bias undermines the use of matrix-assisted laser desorption/ionisation for estimating peptide deamidation: Synthetic peptide studies demonstrate electrospray ionisation gives more reliable response ratios. Rapid communications in mass spectrometry: RCM 33, 1049–1057.
Sinha, A., Mann, M., 2020. A beginner's guide to mass spectrometry–based proteomics. The biochemist 42, 64–69.
Skeffington, A.W., Donath, A., 2020. ProminTools: Shedding light on proteins of unknown function in biomineralization with user friendly tools illustrated using mollusc shell matrix protein sequences. Cold Spring Harbor Laboratory. doi:10.1101/2020.03.05.978718
Sluis, L.G. van der, Hollund, H.I., Buckley, M., De Louw, P.G.B., Rijsdijk, K.F., Kars, H., 2014. Combining histology, stable isotope analysis and ZooMS collagen fingerprinting to investigate the taphonomic history and dietary behaviour of extinct giant tortoises from the Mare aux Songes deposit on Mauritius. Palaeogeography, palaeoclimatology, palaeoecology 416, 80–91.
Solange, R., Sandrine, C., Pétillon, J.-M., Chalard, P., Laroulandie, V., Langlais, M., 2019. Special Issue: Early Personal Ornaments --- Settlement Dynamic and Beadwork: New Insights on Late Upper Paleolithic Craft Activities. PaleoAnthropology 137–155.
Solazzo, C., Fitzhugh, W.W., Rolando, C., Tokarski, C., 2008. Identification of protein remains in archaeological potsherds by proteomics. Analytical chemistry 80, 4590–4597.
Solazzo, C., Wadsley, M., Dyer, J.M., Clerens, S., Collins, M.J., Plowman, J., 2013. Characterisation of novel α-keratin peptide markers for species identification in keratinous tissues using mass spectrometry. Rapid communications in mass spectrometry: RCM 27, 2685–2698.
Stewart, J.R.M., Allen, R.B., Jones, A.K.G., Penkman, K.E.H., Collins, M.J., 2013.

ZooMS: making eggshell visible in the archaeological record. Journal of archaeological science 40, 1797–1804.

Stewart, J.R.M., Allen, R.B., Jones, A.K.G., Kendall, T., Penkman, K.E.H., Demarchi, B., O'Connor, T., Collins, M.J., 2014. Walking on Eggshells: A Study of Egg Use in Anglo-Scandinavian York Based on Eggshell Identification Using ZooMS. International Journal of Osteoarchaeology 24, 247–255.

Stiner, M.C., Munro, N.D., Surovell, T.A., Tchernov, E., Bar-Yosef, O., 1999. Paleolithic population growth pulses evidenced by small animal exploitation. Science 283, 190–194.

Stojanovic, I., Bajčev, O.S., 2016. Exotic goods in the Neolithic of the Central Balkans. *Spondylus* and other marine shell objects from Neolithic sites in the Morava Valley, Serbia. In: Slaviša, P., Korać, M., Stojanović, B. (Eds.), The Neolithic in the Middle Morava Valley: New Insights into Settlements and Economy. Institute of Archaeology, Belgrade, pp. 103–124.

Sun, X., Yang, A., Wu, B., Zhou, L., Liu, Z., 2015. Characterization of the mantle transcriptome of yesso scallop (*Patinopecten yessoensis*): identification of genes potentially involved in biomineralization and pigmentation. PloS one 10, e0122967.

Suzuki, M., Saruwatari, K., Kogure, T., Yamamoto, Y., Nishimura, T., Kato, T., Nagasawa, H., 2009. An acidic matrix protein, Pif, is a key macromolecule for nacre formation. Science 325, 1388–1390.

Suzuki, M., Iwashima, A., Tsutsui, N., Ohira, T., Kogure, T., Nagasawa, H., 2011. Identification and characterisation of a calcium carbonate-binding protein, blue mussel shell protein (BMSP), from the nacreous layer. Chembiochem: a European journal of chemical biology 12, 2478–2487.

Sykes, G.A., Collins, M.J., Walton, D.I., 1995. The significance of a geochemically isolated intracrystalline organic fraction within biominerals. Organic geochemistry 23, 1059–1065.

Takeuchi, T., Plasseraud, L., Ziegler-Devin, I., Brosse, N., Shinzato, C., Satoh, N., Marin, F., 2018. Biochemical characterization of the skeletal matrix of the massive coral, *Porites australiensis* - The saccharide moieties and their localization. Journal of structural biology 203, 219–229.

Taylor, J.D., Kennedy, W.J., Hall, A., 1973. The shell structure and mineralogy of the Bivalvia. II. *Lucinacea-Clavagellacea*, Conclusions, 9. The British Museum (Natural History), London.

Taylor, L.R., Prasad, M.G., Bhat, R.B., 1994. Acoustical characteristics of a conch shell trumpet. The Journal of the Acoustical Society of America 95, 2912–2912.

Theodoropoulou, T., 2011. *Spondylus gaederopus* in Aegean Prehistory: deciphering shapes from Northern Greece. In: Ifantidis, F., Nikolaidou, M. (Eds.), Spondylus in Prehistory: New Data & Approaches. Contributions to the Archaeology of Shell Technologies. Archaeopress, pp. 93–104.

Towe, K.M., 1980. Preserved Organic Ultrastructure: An Unreliable Indicator for Paleozoic Amino Acid Biogeochemistry. In: Hare, P.E., Hoering, T.C., King, K. (Eds.), Biogeochemistry of Amino Acids. Wiley, pp. 65–74.

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

Tsutaya, T., Mackie, M., Koenig, C., Sato, T., Weber, A.W., Kato, H., Olsen, J.V., Cappellini, E., 2019. Palaeoproteomic identification of breast milk protein residues from the archaeological skeletal remains of a neonatal dog. Scientific reports 9, 12841.

Vanhaeren, M., Errico, F. d', 2006. Aurignacian ethno-linguistic geography of Europe revealed by personal ornaments. Journal of archaeological science 33, 1105–1128.

Vanhaereny, M., Errico, F. d', Stringer, C., James, S.L., Todd, J.A., Mienis, H.K., 2006. Middle Paleolithic shell beads in Israel and Algeria. Science 312, 1785–1788.

Vorm, O., Roepstorff, P., Mann, M., 1994. Improved Resolution and Very High Sensitivity in MALDI TOF of Matrix Surfaces Made by Fast Evaporation. Analytical chemistry 66, 3281–3287.

Wadsworth, C., Procopio, N., Anderung, C., Carretero, J.-M., Iriarte, E., Valdiosera, C., Elburg, R., Penkman, K., Buckley, M., 2017. Comparing ancient DNA survival and proteome content in 69 archaeological cattle tooth and bone samples from multiple European sites. Journal of proteomics 158, 1–8.

Wallace, A.F., Schiffbauer, J.D., 2016. Proteins from the past. eLife 5. doi:10.7554/eLife.20877

Wang, B., Malik, R., Nigg, E.A., Körner, R., 2008. Evaluation of the low-specificity protease elastase for large-scale phosphoproteome analysis. Analytical chemistry 80, 9526–9533.

Warinner, C., Hendy, J., Speller, C., Cappellini, E., Fischer, R., Trachsel, C., Arneborg, J., Lynnerup, N., Craig, O.E., Swallow, D.M., Fotakis, A., Christensen, R.J., Olsen, J.V., Liebert, A., Montalva, N., Fiddyment, S., Charlton, S., Mackie, M., Canci, A., Bouwman, A., Rühli, F., Gilbert, M.T.P., Collins, M.J., 2014. Direct evidence of milk consumption from ancient human dental calculus. Scientific reports 4, 7104.

Wehmiller, J., Hare, P.E., 1971. Racemization of amino acids in marine sediments. Science 173, 907–911.

Weiner, S., Addadi, L., 1997. Design strategies in mineralized biological materials. Journal of materials chemistry 7, 689–702.

Weiner, S., Lowenstam, H.A., Hood, L., 1976. Characterization of 80-million-year-old mollusk shell proteins. Proceedings of the National Academy of Sciences of the United States of America 73, 2541–2545.

Weiner, S., Traub, W., 1984. Macromolecules in mollusc shells and their functions

in biomineralization. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 304, 425–434.
Welker, F., 2018a. Elucidation of cross-species proteomic effects in human and hominin bone proteome identification through a bioinformatics experiment. BMC evolutionary biology 18, 23.
Welker, F., 2018b. Palaeoproteomics for human evolution studies. Quaternary science reviews 190, 137–147.
Welker, F., Collins, M.J., Thomas, J.A., Wadsley, M., Brace, S., Cappellini, E., Turvey, S.T., Reguero, M., Gelfo, J.N., Kramarz, A., Burger, J., Thomas-Oates, J., Ashford, D.A., Ashton, P.D., Rowsell, K., Porter, D.M., Kessler, B., Fischer, R., Baessmann, C., Kaspar, S., Olsen, J.V., Kiley, P., Elliott, J.A., Kelstrup, C.D., Mullin, V., Hofreiter, M., Willerslev, E., Hublin, J.-J., Orlando, L., Barnes, I., MacPhee, R.D.E., 2015. Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. Nature 522, 81–84.
Welker, F., Hajdinjak, M., Talamo, S., Jaouen, K., Dannemann, M., David, F., Julien, M., Meyer, M., Kelso, J., Barnes, I., Brace, S., Kamminga, P., Fischer, R., Kessler, B.M., Stewart, J.R., Pääbo, S., Collins, M.J., Hublin, J.-J., 2016. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. Proceedings of the National Academy of Sciences of the United States of America 113, 11162–11167.
Welker, F., Ramos-Madrigal, J., Kuhlwilm, M., Liao, W., Gutenbrunner, P., Manuel, M. de, Samodova, D., Mackie, M., Allentoft, M.E., Bacon, A.-M., Collins, M.J., Cox, J., Lalueza-Fox, C., Olsen, J.V., Demeter, F., Wang, W., Marques-Bonet, T., Cappellini, E., 2019. Enamel proteome shows that *Gigantopithecus* was an early diverging pongine. Nature 576, 262–265.
Welker, F., Ramos-Madrigal, J., Gutenbrunner, P., Mackie, M., Tiwary, S., Rakownikow Jersie-Christensen, R., Chiva, C., Dickinson, M.R., Kuhlwilm, M., Manuel, M. de, Gelabert, P., Martinón-Torres, M., Margvelashvili, A., Arsuaga, J.L., Carbonell, E., Marques-Bonet, T., Penkman, K., Sabidó, E., Cox, J., Olsen, J.V., Lordkipanidze, D., Racimo, F., Lalueza-Fox, C., Bermúdez de Castro, J.M., Willerslev, E., Cappellini, E., 2020. The dental proteome of Homo antecessor. Nature 580, 235–238.
Westbroek, P., Marin, F., 1998. A marriage of bone and nacre. Nature 392, 861–862.
Westbroek, P., Meide, P.H. van der, Wey-Kloppers, J.S. van der, Sluis, R.J. van der, Leeuw, J.W. de, Jong, E.W. de, 1979. Fossil macromolecules from cephalopod shells: characterization, immunological response and diagenesis. Paleobiology 5, 151–167.
Williams, S.T., 2017. Molluscan shell colour. Biological reviews of the Cambridge Philosophical Society 92, 1039–1058.
Wilson, J., Doorn, N.L. van, Collins, M.J., 2012. Assessing the extent of bone

degradation using glutamine deamidation in collagen. Analytical chemistry 84, 9041–9048.

Windler, A., 2017. From the Aegean Sea to the Parisian Basin: *Spondylus* shell exchange in Europe during the process of Neolithisation. In: Eisenach, P., Stöllner, T., Windler, A. (Eds.), The RITaK Conferences 2013–2014. Bochum, pp. 95–110.

Windler, A., 2019. The Use of *Spondylus gaederopus* during the Neolithic of Europe. Journal of Open Archaeology Data 7, 7.

Wiśniewski, J.R., Zougman, A., Nagaraj, N., Mann, M., 2009. Universal sample preparation method for proteome analysis. Nature methods 6, 359–362.

Yarra, T., Gharbi, K., Blaxter, M., Peck, L.S., Clark, M.S., 2016. Characterization of the mantle transcriptome in bivalves: *Pecten maximus*, *Mytilus edulis* and *Crassostrea gigas*. Marine genomics 27, 9–15.

Yoshioka, S., Kitano, Y., 1985. Transformation of aragonite to calcite through heating. Geochemical journal 19, 245–249.

Zemour, A., Binder, D., Bonnardin, S., D'Ovidio, A.-M., Goude, G., Gourichon, L., Pradeau, J.-V., Sorin-Mazouni, S., Bromblet, P., Buchet, L., Cotto, K.-Y., Sénépart, I., 2017. Laboratory Excavation of a Neolithic Grave from Avignon-La Balance-Ilot P (France): Burial Practices and Garment Reconstruction. Journal of Field Archaeology 42, 54–68.

Zhao, R., Takeuchi, T., Luo, Y.-J., Ishikawa, A., Kobayashi, T., Koyanagi, R., Villar-Briones, A., Yamada, L., Sawada, H., Iwanaga, S., Nagai, K., Satoh, N., Endo, K., 2018. Dual Gene Repertoires for Larval and Adult Shells Reveal Molecules Essential for Molluscan Shell Formation. Molecular biology and evolution 35, 2751–2761.

Zilhão, J., Angelucci, D.E., Badal-García, E., Errico, F. d', Daniel, F., Dayet, L., Douka, K., Higham, T.F.G., Martínez-Sánchez, M.J., Montes-Bernárdez, R., Murcia-Mascarós, S., Pérez-Sirvent, C., Roldán-García, C., Vanhaeren, M., Villaverde, V., Wood, R., Zapata, J., 2010. Symbolic use of marine shells and mineral pigments by Iberian Neandertals. Proceedings of the National Academy of Sciences of the United States of America 107, 1023–1028.

Zilhão, J., Angelucci, D.E., Igreja, M.A., Arnold, L.J., Badal, E., Callapez, P., Cardoso, J.L., Errico, F. d', Daura, J., Demuro, M., Deschamps, M., Dupont, C., Gabriel, S., Hoffmann, D.L., Legoinha, P., Matias, H., Monge Soares, A.M., Nabais, M., Portela, P., Queffelec, A., Rodrigues, F., Souto, P., 2020. Last Interglacial Iberian Neandertals as fisher-hunter-gatherers. Science 367. doi:10.1126/science.aaz7943

Zubarev, R.A., Makarov, A., 2013. Orbitrap mass spectrometry. Analytical chemistry 85, 5288–5296.

Chapter 1

The shell matrix of the European thorny oyster, *Spondylus gaederopus*: microstructural and molecular characterisation

Digest

Mollusc shells possess exceptional material properties and are interesting systems for biomineralization research. However, the fundamental molecular mechanisms of biocalcification are, as yet, poorly understood. This question can be tackled by analysing the mineral-associated shell proteins, also known as “shellome”. The comparison of proteomics data between different mollusc species can yield a better understanding about molecular mechanisms of shell biomineralization and how these processes evolved in different molluscan lineages. However, considering the size of the Mollusca phylum and the diversity of shell microstructures, biomolecular data are still scant, and are particularly lacking for species that are not of commercial importance. The Mediterranean thorny oyster *Spondylus gaederopus* represents such a case, and was therefore chosen as a model system for an in-depth investigation. More importantly, *Spondylus* was commonly used in prehistory and, consequently, is an important shell in archaeological research.

In this study, I specifically aimed to investigate whether proteins retrieved from mollusc shells preserve taxonomic information which could be used to identify shells. *Spondylus* belongs to the Pectinida order, however the shell has a crossed-lamellar microstructure that is very different from that of other pectinoid shells which are foliated and calcitic (*e.g*. *Pecten maximus*). Moreover, very little is known about the biochemistry and proteome characteristics of crossed-lamellar microstructures.

In this work I have isolated mineral-associated biomolecules by using several-step bleaching treatments. The shell organic matrix was characterised by a set of different physical and biochemical techniques, including *in vitro* calcification assays that allowed to check whether the soluble matrix is effective or not in interacting with the precipitation of $CaCO_3$ *in vitro*. High-resolution tandem mass spectrometry was used to characterise *Spondylus* shell proteome. The study showed that *Spondylus* shell is composed of at least four different structural layers and contains very low amounts of mineral-bound organics, composed of proteins and saccharides. I have identified several mollusc shell matrix proteins which showed

features typical of biomineralized tissues and most were originally characterised in *Pecten* shells. However, protein coverages were low, suggesting that *Spondylus* shellome represents lineage-specific sequences.

This study also highlighted several limitations. First of all, a complete identification of *Spondylus* proteome is hindered by the lack of reference molecular data, such as the mantle transcriptome. This requires to have live *Spondylus* specimens which are challenging to obtain. First of all, *Spondylus* molluscs are not cultivated commercially and are not 'easy-to-harvest' animals (*e.g* the they live attached to the seabed, hence sampling tissues or transferring live animals from marine environments to laboratory aquaria needs specialised staff). *Spondylus gaederopus* is native to the Mediterranean Sea, however, in the past years, there has been a notable decline in the populations of these species (*e.g*. they are difficult to find in some of the coastal areas in the western part of the Mediterranean Sea), hence this requires assistance from well-trained marine biology specialists to locate and find species. Finally, *Spondylus* does not represent a model system in biomineralization research and has not been thoroughly studied using 'omics' approaches, therefore it requires thorough planning before conducting molecular analyses. This part of the project was ultimately supported by Assemble+ program which allowed me to gain the access to marine biology facilities at HCMR-IMBBC institute in Crete, Greece and the project it is currently ongoing. Several *Spondylus* specimens were successfully sampled in the beginning of 2020 and transcriptome sequencing is expected to be performed in the first half of 2021 at the Globe Institute of the University of Copenhagen.

Secondly, I acknowledge that the *in vitro* calcification assays which were used in this study to investigate how *Spondylus* shell matrix interacts with calcium carbonate crystals would have benefited from additional control samples. In the future, alongside with the blank controls that were used in this study, it would be useful to include a negative control, *i.e.* a non-biomineralizing protein and a positive control, *i.e.* protein or shell matrix which shows a very strong interaction with calcite crystals. A potential negative control could be commercial immunoglobulin (heavy or light chain), histones or proteinase K (pI 7.8). A good positive control could be polyaspartic acid (commercially available) or shell extracts that are known to exert a strong interaction with calcite crystals (like the soluble matrix of the calcitic prisms of *Pinna nobilis*).

Finally, I would also like to highlight that while the two bleaching treatments used to clean *Spondylus* shell and isolate mineral bound organics may have eliminated a considerable amount of shell matrix (that would also be of great interest in shell

biomineralization research), they were selected intentionally. The bleaching steps were set considering several factors:

i) first of all, we were aware of the very limited amount of genomics/transcriptomics data for *Spondylus* genus, and phylogenetically close genera. The intense bleaching step ensured that contaminant sequences were not taken into account when analysing the proteomics data of this shell;

ii) secondly, this step is particularly important when considering the study of sub-fossil shells, which was part of this PhD project. In fossil or sub-fossil materials, we look primarily at intracrystalline matrix, *i.e*., the one with the highest potential of preservation;

iii) thirdly, mineral-bound proteins have been long considered as a direct source of information when studying shell biomineralization. Therefore, bleaching allows us to "isolate" this set of core proteins that are involved in interacting with calcium ions or with calcium carbonate mineral surfaces.

In the future, as soon as *Spondylus* transcriptome is available, it, however, will be useful to re-extract *Spondylus* shell matrix without using such a harsh bleaching treatment, in order to obtain more intercrystalline sequences and have a more comprehensive characterisation of the *Spondylus* shellome.

Overall, this work is a first in-depth molecular investigation of *Spondylus* shell and the collected dataset provides additional information about the biochemistry of shells that have complex arrangement of different microstructures, dominated by layers of crossed-lamella. My study highlights the differences of shell protein contents even among species that are phylogenetically relatively close (*i.e.*, that belong to the same order or superfamily). This work is also significant in archaeological research. The distinctiveness of *Spondylus* protein sequences suggests that in the future, the molecular data can be used for palaeoproteomic analysis of ancient *Spondylus* shell artifacts. However, the development of shell palaeoproteomics, *i.e.* the use of shell proteins for species identification in archaeological samples, requires to address several methodological limitations. First of all, sample size. Analysing modern shells gives us the advantage that entire shells of any size and/or number can be devoted to extract small amounts of mineral-bound organics. The same is obviously not feasible for archaeological shell artifacts. Hence, palaeoproteomic analysis of prehistoric shells calls for an optimisation of sample preparation approaches, which will be addressed in Chapter 3. Secondly, more research is needed to understand how shell proteins age and the possibility to retrieve ancient shell sequences will be investigated in detail in Chapter 2 of this thesis.

Journal of Structural Biology 211 (2020) 107497

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

The shell matrix of the european thorny oyster, *Spondylus gaederopus*: microstructural and molecular characterization

Jorune Sakalauskaite[a,b,*], Laurent Plasseraud[c], Jérôme Thomas[b], Marie Albéric[d], Mathieu Thoury[e], Jonathan Perrin[f], Frédéric Jamme[f], Cédric Broussard[g], Beatrice Demarchi[a], Frédéric Marin[b,*]

[a] Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Turin, Italy
[b] Biogeosciences, UMR CNRS 6282, University of Burgundy-Franche-Comté (UBFC), 6 Boulevard Gabriel, 21000 Dijon, France
[c] Institute of Molecular Chemistry, ICMUB UMR CNRS 6302, University of Burgundy-Franche-Comté (UBFC), 9 Avenue Alain Savary, 21000 Dijon, France
[d] Laboratoire Chimie de la Matière Condensée de Paris, UMR, CNRS 7574, Sorbonne Université, Place Jussieu 4, 75252 Paris, France
[e] IPANEMA, CNRS, ministère de la Culture, UVSQ, USR3461, Université Paris-Saclay, F-91192 Gif-sur-Yvette, France
[f] Synchrotron SOLEIL, L'Orme des Merisiers, 91192 Gif sur Yvette Cedex, France
[g] 3P5 Proteomic Platform, University of Paris, Cochin Institute, INSERM, U1016, CNRS, UMR8104, F-75014 Paris, France

ARTICLE INFO

Keywords:
Biomineralization
Shell biochemistry
Evolution
Proteomics
Liquid chromatography-tandem mass spectrometry

ABSTRACT

Molluscs, the largest marine phylum, display extraordinary shell diversity and sophisticated biomineral architectures. However, mineral-associated biomolecules involved in biomineralization are still poorly characterised.

We report the first comprehensive structural and biomolecular study of *Spondylus gaederopus,* a pectinoid bivalve with a peculiar shell texture. Used since prehistoric times, this is the best-known shell of Europe's cultural heritage. We find that *Spondylus* microstructure is very poor in mineral-bound organics, which are mostly intercrystalline and concentrated at the interface between structural layers.

Using high-resolution liquid chromatography tandem mass spectrometry (LC-MS/MS) we characterized several shell protein fractions, isolated following different bleaching treatments. Several peptides were identified as well as six shell proteins, which display features and domains typically found in biomineralized tissues, including the prevalence of intrinsically disordered regions. It is very likely that these sequences only partially represent the full proteome of *Spondylus*, considering the lack of genomics data for this genus and the fact that most of the reconstructed peptides do not match with any known shell proteins, representing consequently lineage-specific sequences.

This work sheds light onto the shell matrix involved in the biomineralization in spondylids. Our proteomics data suggest that *Spondylus* has evolved a shell-forming toolkit, distinct from that of other better studied pectinoids – fine-tuned to produce shell structures with high mechanical properties, while limited in organic content. This study therefore represents an important milestone for future studies on biomineralized skeletons and provides the first reference dataset for forthcoming molecular studies of *Spondylus* archaeological artifacts.

1. Introduction

Biomineralization is a cellular and molecular process by which living systems precipitate mineral salts, predominantly calcium carbonate. In metazoans, this process emerged during the Precambrian-Cambrian transition in multiple lineages. Among these, molluscs represent one of the most diversified phyla, both from the morphological and ecological point of view. Their ability to biomineralize an exoskeleton, the shell, represented an innovative strategy in terms of support of soft body parts and protection against predation and desiccation in addition to constituting an abundant reserve of calcium ions (Kocot et al., 2016). This certainly explains their evolutionary success in marine, freshwater and terrestrial environments.

The biomineralization of the shell is controlled by a specialised organ, the mantle, a thin polarized epithelium that extrudes inorganic ions (mostly calcium and bicarbonate) via ion pumps and channels, and, at the same time, secretes a macromolecular matrix (Marin et al., 2012). Both inorganic and organic components assemble in well

* Corresponding authors at: Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Turin, Italy (J. Sakalauskaite) & UMR CNRS 6282 Biogéosciences, University of Burgundy-Franche-Comté, 6 Boulevard Gabriel, 21000 Dijon, France (J. Sakalauskaite, F. Marin).
E-mail addresses: jorune.sakalauskaite@unito.it (J. Sakalauskaite), frederic.marin@u-bourgogne.fr (F. Marin).

https://doi.org/10.1016/j.jsb.2020.107497
Received 27 January 2020; Received in revised form 20 March 2020; Accepted 22 March 2020
Available online 24 March 2020
1047-8477/

defined microstructures at the nano- and microscales (Marin et al., 2013) resulting in a diversity of micro-architectures – the most common being the crossed-lamellar, foliated, prismatic and nacreous (mother-of-pearl) ones (Taylor et al., 1969, 1973; Carter, 1990).

The precise molecular mechanisms controlling shell biomineralization are still poorly understood. However, extensive research on the physico-chemical properties of calcium-based biominerals (Lowenstam, 1981) and the shell biochemistry of several model organisms (Addadi and Weiner, 1997; Evans, 2008; Marin et al., 2013; Kocot et al., 2016), have shown that the secreted organic matrix, (*i.e.* a set of proteins, glycoproteins, polysaccharides and lipids), plays a crucial role in the mediation of crystal nucleation and growth (Falini et al., 1996; Addadi et al., 2006; Jackson et al., 2010; Marin et al., 2013).

Very important components are the shell matrix proteins (referred to as SMPs), which get occluded inside the shell mineral during calcification. An increasing number of SMPs have been identified in the past decade using proteomics, transcriptomics and, to a lesser extent, genomics, which allowed researchers to identify and compare the shell proteomes - abbreviated as 'shellomes' - of different groups. Although most of these studies focused on mother-of-pearl (Ramos-Silva et al., 2012; Marie et al., 2017, 2013a; Yarra et al., 2016), they shed light on peculiar characteristics of shell proteins (Sleight et al., 2016; Arivalagan et al., 2017). One interesting feature is the widespread occurrence of low complexity domains/ repetitive low complexity domains (LCDs/RLCDs) (Kocot et al., 2016; Marin et al., 2016, 2014) in nacre proteins (Jackson et al., 2010; McDougall et al., 2013; Sakalauskaite et al., 2019) and also in other microstructures (Marie et al., 2013a). This advocates for a key-role of these domains in binding mineral surfaces and in providing molecular flexibility (Sarashina and Endo, 2006). However, most importantly, 'shellomics' has also evidenced the huge diversity of the SMP repertoires between shells that have similar microstructures or species that are phylogenetically close (Marie et al., 2009a, 2009b; Jackson et al., 2010). This suggests that, beside a set of shared conserved functions in shell repertoires, the recruitment of numerous lineage-dependent molecular functions may have happened in parallel. In other words, independent evolution of SMPs may have occurred multiple times in different lineages.

Whilst there are numerous shell proteome studies of nacreous molluscs (Marin et al., 2005; Jackson et al., 2010; Marie et al., 2017, 2012, 2011, 2009a; Gao et al., 2015), only few studies have been performed on other shell microstructures, such as foliated (Yarra et al., 2016) and cross-lamellar (Osuna-Mascaró et al., 2014; Herlitze et al., 2018). This can be explained by the remarkable mechanical properties of nacre, which make it a material of choice for biomimetics (Luz and Mano, 2009; Espinosa et al., 2009; Finnemore et al., 2012), and by the commercial importance of nacreous molluscs, especially the pearl oyster and the edible mussel (Gosling, 2003; Watabe et al., 2011). In addition, nacre is considered to be an ancestral microstructural type and has been studied from an evolutionary point of view (Marie et al., 2009a; Jackson et al., 2010). On the contrary, the molecular information on the SMPs of crossed-lamellar shells is scarce (Herlitze et al., 2018; Agbaje et al., 2019), despite that it is the most common microstructure in molluscs (Carter, 1990; Wilmot et al., 1992) and possesses high fracture toughness (Kamat et al., 2000).

The thorny oyster *Spondylus gaederopus* Linnaeus, 1758 represents a typical example of this bias in "shellomic" studies. It is a member of a small family of pteriomorphid bivalves, the Spondylidae. *S. gaederopus* is endemic to the Mediterranean sea, with a fixosessil life in water depths up to 50 m. The origin and evolution of spondylids is unclear (Yonge, 1973). Based on morphology and microstructure, the Spondylidae family was thought to be derived from a Pectinid ancestor in the middle Jurassic, possibly from the genus *Spondylopecten* (Logan, 1974; Waller, 2006). This evolutionary process may have been driven by the increase of shell-crushing predators, which likely led to the development of more complex skeletal frameworks with improved density and strength (Knoll, 2003). Conversely, genetics studies suggest that Spondylidae is a sister group of the Pectinidae (Matsumoto and Hayami, 2000; Matsumoto, 2003; Barucca et al., 2004), albeit this issue has not been addressed thoroughly, since many of the molecular phylogenetic analyses use *Spondylus* as an outgroup to reconstruct the Pectinid family (Barucca et al., 2004; Puslednik and Serb, 2008). Considering that Pectinidae shells are mostly foliated calcitic, whereas *Spondylus* is mostly crossed-lamellar aragonitic, the question of evolution is puzzling and requires a careful reexamination.

Finally, *Spondylus* is of particular importance in the field of archaeology and has been regarded as an iconic shell in prehistoric times: its presence is well documented in the archaeological record and can be tracked all over Europe during the Neolithic period (Moshkovitz, 1971; Shackleton and Elderfield, 1990; Borrello and Micheli, 2011, 2004; Ifantidis and Nikolaidou, 2011; Windler, 2017), as well as in Pre-Columbian South America (Paulsen, 1974; Pillsbury, 1996). The possession of *Spondylus* was probably an indicator of wealth or high social status, and therefore reified important ritual/ symbolic messages.

In this context, our study has a dual goal:

1. characterising the shell microstructure of *Spondylus gaederopus* and the molecular signature of the associated organic matrix, including an in-depth proteomics study;
2. establishing a biomolecular dataset for the *Spondylus* shell matrix, which will constitute a reference in future studies to identify the biological origin of fragments of prehistoric shell ornaments, where morphological identification is unattainable (Demarchi et al., 2014; Sakalauskaite et al., 2019), taking into account the diagenetic processes.

2. Results

2.1. The microstructure of S. gaederopus shell

The shell of *S. gaederopus* displays a very distinctive appearance: characteristic hollow spines, a deep purple upper valve and a completely white lower valve (Fig. 1). When observed in cross-section (Fig. 2), it displays a complex combination of microstructures that includes at least 3 different layers: prismatic (P), crossed-lamellar (CL) and foliated (F). At the basis of the shell spines, the microstructure is

Fig. 1. *Spondylus gaederopus* Linnaeus, 1758 shell, analysed in this study (Greece, Saronikos. Dived at 15 m., taken live in 2010).

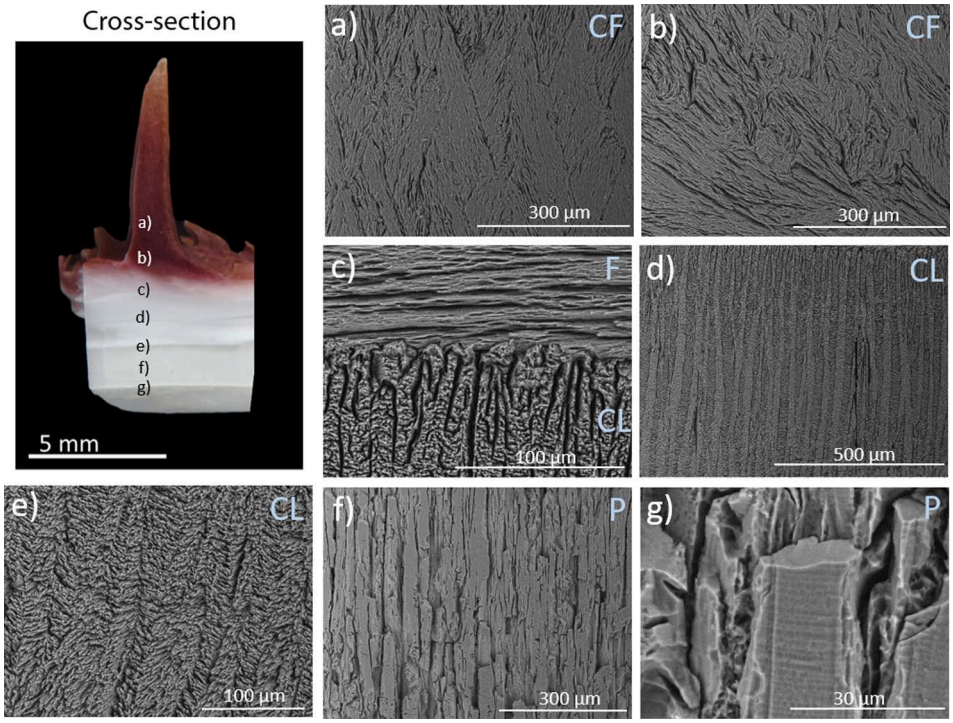


Fig. 2. SEM micrographs of the microstructures observed along a polished transversal section of *Spondylus gaederopus* shell (upper valve): a) calcitic crossed-foliated (CF) structure of the spine; b) irregular calcitic CF microstructure observed at the point of spine formation; c) transition between the calcitic foliated (F) and the aragonitic crossed-lamellar (CL) structures; d) aragonitic crossed-lamellar (CL) structure in the upper part of the shell; e) aragonitic crossed-lamellar (CL) structure in the lower part of the section; f, g) aragonitic prismatic (P) layer in the bottom part of the shell and myostracum. Microstructural abbreviations: F - calcitic foliated; P - aragonitic prismatic; CL - aragonitic crossed-lamellar, CF - calcitic crossed-foliated.

complex and irregular, but becomes crossed-foliated (calcitic CF) at later stages of spine growth (Fig. 2a, b). The outermost layer of the shell is foliated calcitic (F), similarly to what can be observed in pectinid and ostreid shells (Fig. 2c). The underlying layer is aragonitic crossed-lamellar (CL) and exhibits the typical oblique patterns that characterise this microstructure (Fig. 2d, e). This layer is traversed by a thin prismatic (aragonitic) myostracal layer. The inner and uppermost CL layers, when viewed in cross-section, have different lamellar orientations. The crossed-lamellar layer dominates the hinge area and is very thick at the central part of the shell, but at the margin, the proportion changes, and the foliated becomes the dominant microstructure. The innermost layer is composed of thin aragonitic prisms, about 15–20 µm in diameter, that exhibit a fine layering perpendicular to the elongation axis.

We note that while the changes between the different underlying aragonitic microstructures are gradual, the transition between the aragonitic crossed-lamellar and the foliated calcitic structures is abrupt and well marked (Fig. 2c). This microstructural and mineralogical transition zone was also studied by UV photoluminescence imaging. The analysis was carried out at the DISCO Beamline (SOLEIL synchrotron, Paris) using an excitation wavelength of 275 nm. The spatial distribution of shell organics was studied using a series of emission filters to detect the luminescence patterns. Fluorescence signal was detected in several different emission regions (327–353; 370–400; 420–480; 499–529 nm). It was however, the strongest in the [456–481 nm] and [535–607 nm] regions (Fig. 3). An intense luminescenting layer was observed at the interface between the aragonitic crossed-lamellar and the foliated calcitic microstructures, likely arising due to the presence of organics (however, the exact nature of fluorescent components is unknown). This indicates the existence of an intercrystalline organic layer (OL) in the transitional zone between the two layers of different polymorphs.

2.2. Extraction of the shell organic matrix

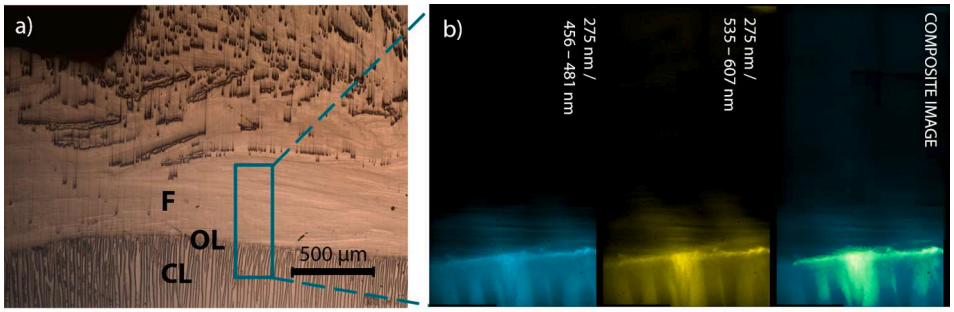
The organic matrices were extracted according to standard protocols (Marie et al., 2013b) from finely ground shell powders, which were prepared in different ways. First of all, separated valves were cleaned to remove any surface contamination (mechanically and by bleaching in NaOCl for 2 hrs, see Material and Methods for more details). Following this, we split the coarsely crushed powders into subsamples from the upper (UV) red and lower (LV) white valves. To these, two different bleaching treatments (varying in stringency) were applied to both subsamples:

a) "Mild": approach: coarsely crushed shell powders were bleached for 4 hrs (hereafter "2BL extracts"); these were further ground in fine powders (S1, S2, S3, S4).
b) "Strong" approach: obtained 2BL fine powders were bleached for extra 14 hrs (hereafter "3BL extracts") (S5, S6, S7, S8).

In total, eight shell matrices were obtained - four acid soluble (ASM) and four acid insoluble (AIM) organic fractions (Fig. 4; see also SI.1). After the 2BL treatment, the obtained organics represent ~1.5–2 wt‰ (note: per mille) of the total shell powder; more than two thirds of this matrix is the AIM. The 14-hrs additional bleaching step resulted in a severe reduction in the amount of ASM (from ~0.5 wt‰ to ~0.04 wt‰) and in an even more drastic loss of AIM (from ~1.6 wt‰ to ~0.005 wt‰). We note that for both bleaching treatments, the upper valve yielded a slightly higher amount of matrix.

2.3. FT-IR analysis of organic matrices

FTIR spectroscopy was used to characterize qualitatively the presence of functional groups in the shell organic matrix extracts. Due to low quantities of the 3BL matrices, the characterization was applied solely on the four 2BL extracts (Fig. 5). Distinctly, all samples exhibit absorption bands characteristic of proteins: the broad band between 3300 and 3260 cm^{-1} corresponding to amide A stretchings (νN–H), the amide I band (νC=O) at around 1640–1650 cm^{-1} and the amide II bands (νC–N) found between 1510 and 1529 cm^{-1}. In addition, the four spectra present also ν(C–O) absorption bands specific to carbohydrates, at around 1070 cm^{-1}, and another at around 1405–1460 cm^{-1} which can be related to carboxylate groups [$\nu_s(COO^-)$]; the latter appears clear in the two ASMs, while in the AIMs it is shifted towards higher wavenumbers and overlaps with another strong band at 1446–1460 cm^{-1}, only detected in the AIMs. A weak absorption band



F – Foliated layer (calcitic); CL – Crossed-lamellar (aragonitic); OL – Organic layer.

Fig. 3. Micro-imaging luminescence of the organo-mineral interface between the foliated calcitic and aragonitic crossed-lamellar layers in the upper valve of *Spondylus* shell, obtained using a DUV photoemission beamline of 275 nm: a) area of interest, picture taken with optical microscope; b) false color luminescence images obtained using 456–481 nm and 535–607 nm filters for the detection. Marked layers: F - foliated, OL - organic layer, CL - crossed-lamellar.

found in the four extracts at around 1227–1240 cm^{-1} can be assigned to the $\nu S=O$ vibration and indicates the presence of sulphate groups (Takeuchi et al., 2018). Moreover, we note that the two ASM spectra (S1, S3) are almost completely superimposable. They mainly differ from the two AIMs (S2, S4) by the amplitude ratios between the amide I and amide II bands. The two AIMs spectra are also very similar, in particular by exhibiting two close bands at 700–712 cm^{-1} associated with a more intense band at 854 cm^{-1}. These bands, including the intense peak at 1446–1460 cm^{-1}, slightly shifted from 1477 cm^{-1}, are characteristic of the aragonite fingerprint (internal vibration modes of CO_3^{2-}), suggesting that some mineral crystallites remain undissolved, in spite of the duration of the decalcification (overnight) and of the excess of acetic acid used for this step. In addition, the AIMs are characterized by the presence of ν(C–H) stretching vibrations at around 2920–2852 cm^{-1} whereas in ASMs this signal is weaker and shifted to 3072–2943 cm^{-1}.

2.4. SDS-PAGE

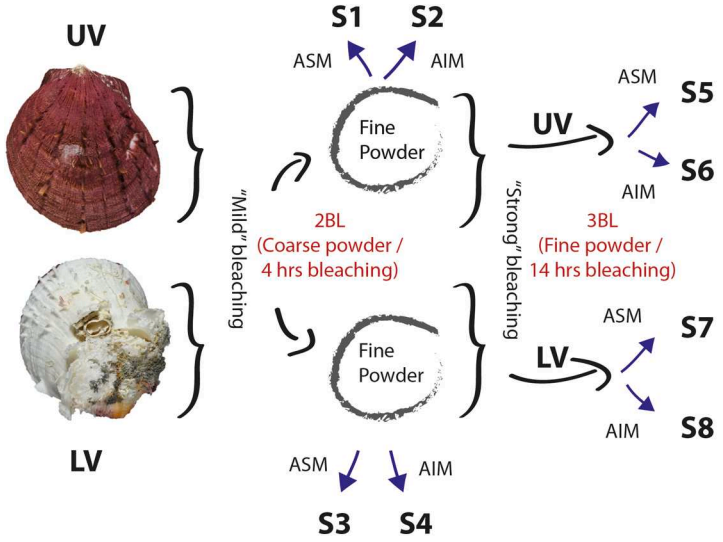
All organic shell extracts were analysed by SDS-PAGE and stained with silver nitrate and Stains-all (Fig. 6), except sample S8, which yielded a very low quantity and was entirely dedicated to proteomics. Note that the AIM extracts correspond to the Laemmli-soluble AIMs (LS-AIMs).

The silver stained gel (Fig. 6a) shows that all samples have a peculiar migration pattern, characterised by a smear from high to low molecular weights. There are no significant differences between the migration profiles from the upper (S1/2/5/6) and lower valves (S3/4/7/8); the 2BL and 3BL cleaning approaches (S1-4 and S5-7 respectively) produced similar patterns. In spite of the smear, several discrete diffuse bands can be observed in the intermediate molecular weight range: i) a set of three diffuse bands around 20, 17 and 13 kDa in all of the samples; ii) bands in the 70–30 kDa range are shared by most of the samples. In the low molecular weight range, a thick band (<11 kDa) is strongly stained brown in only half of the samples (S2/4/6/7) which migrated faster than the migration front, suggesting the occurrence of very low molecular weight compounds. At last, around 11 kDa, a relatively sharp negatively-stained band is present in all samples.

Stains-all (Fig. 6b) allows the detection of putative Ca^{2+} binding proteins (metachromatic blue), but is also efficient in staining acidic proteins; in summary it stains sialoglycoproteins in blue, Ca^{2+} binding proteins in deep blue to violet, proteins in red, and lipids in yellow-orange. All our preparations stained violet at different degrees of intensity, in contrast to the standards, which stained pink. Considering the two cleaning approaches (2BL *vs* 3BL), Stains-all shows that the signal is far less intense in the 3BL samples (S5-7). Contrary to silver staining, Stains-all does not reveal the blurred bands in the 70–11 kDa range. However, it stains intensely violet in all lanes the sharp band at 11 kDa, which is precisely the one negatively stained with silver. We also find low molecular weight components (<11 kDa stained band) in all ASMs (S1/3/5/7), while very high molecular weight (>170 kDa) components are present in the AIMs (S2/4/6).

2.5. ELLA

21 lectins were tested for their affinity to saccharidic residues present in soluble shell matrices. The analysis was performed on two ASM samples extracted after 2BL treatments from the upper (S1) and lower (S3) valves and the resulting intensity of cross-reactivity is expressed as a percentage, the most reactive lectin giving 100% (Fig. 7). Jacalin was



Organics yield (wt‰)	
S1	0.56
S2	1.58
S3	0.43
S4	1.0
S5	0.04
S6	0.2
S7	0.03
S8	0.005

Fig. 4. The scheme represents different approaches (bleaching pretreatments) by which shell organic matrix extracts were obtained (S1 - S8) and their subsequent yields. After initial bleach to remove superficial contaminants, all shell powders were bleached twice (2BL) or thrice (3BL). Bleaching procedure: 2BL - coarse powder bleached for 4 hrs; 3BL - fine sieved powder bleached for additional 14 hrs. UV - upper valve, LV - lower valve. ASM - acid soluble matrix, AIM - acid insoluble matrix.

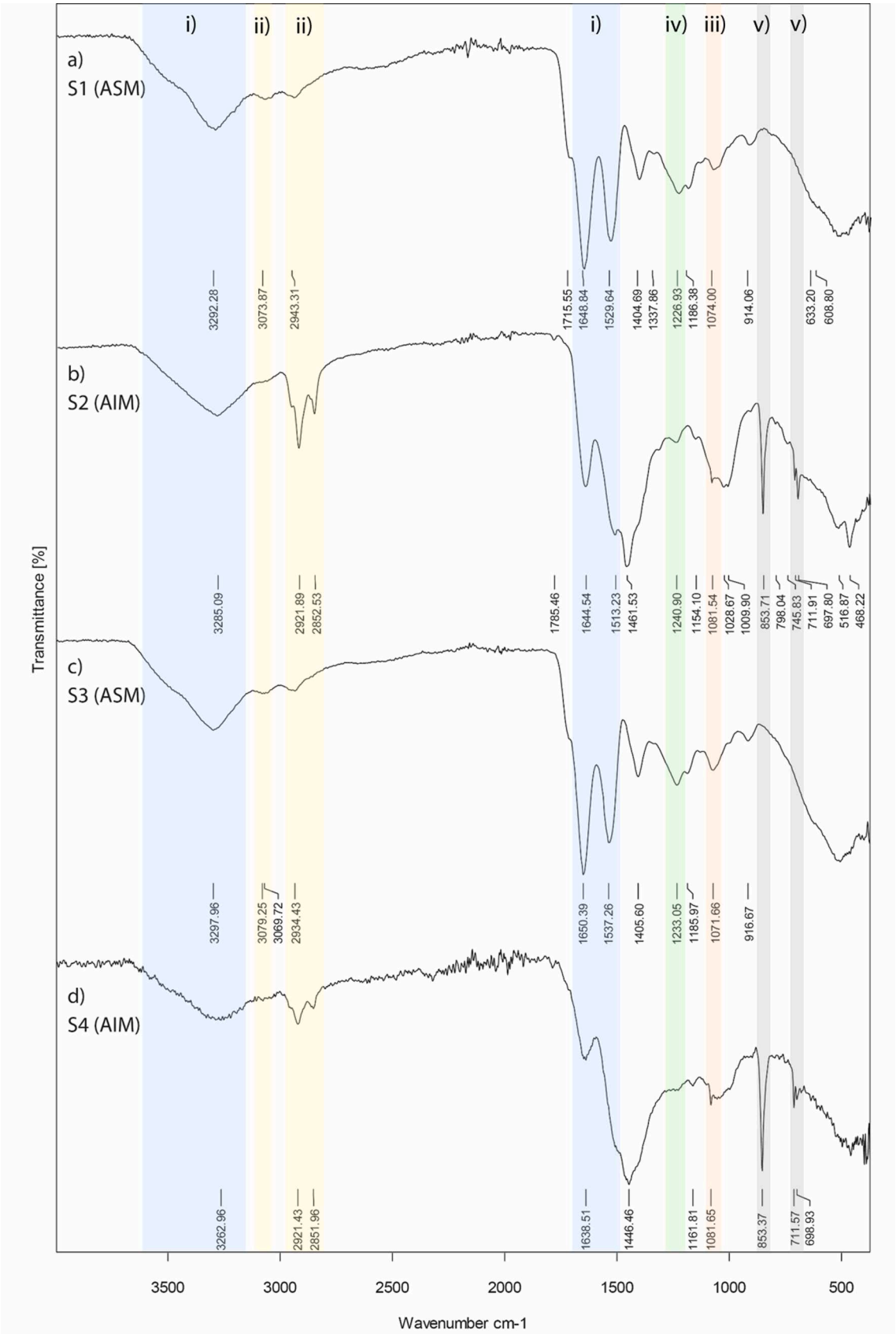


Fig. 5. FT-IR (ATR) spectra of shell organic matrices collected in the range of 4000–400 cm^{-1}: a, b) S1, S2 - ASM and AIM extracts from the shell upper valve; c, d) S3, S4 - ASM and AIM extracts from the shell lower valve. Peaks discussed in the text are marked in colour which correspond to: i) proteins; ii) organics rich in hydrocarbon groups; iii) glycosides; iv) sulphates; v) aragonite. ASM - acid soluble matrix, AIM - acid insoluble matrix.

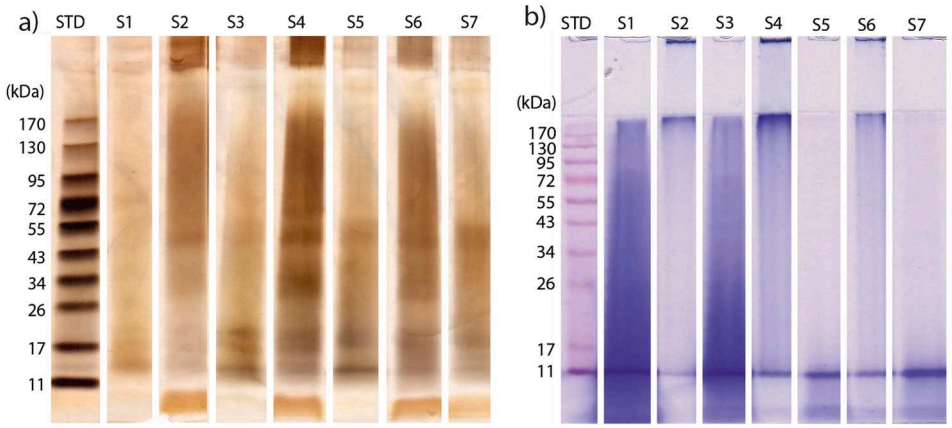


Fig. 6. SDS-PAGE gels of *Spondylus* shell organic matrices: a) stained using silver staining; b) stained using Stains-all staining. Analysed samples are as indicated in Fig. 4; STD - standard molecular weight markers.

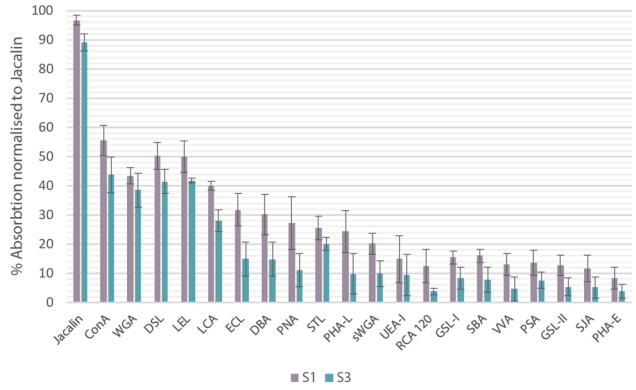


Fig. 7. Enzyme-linked lectin assay (ELLA) on two ASMs extracted from the upper (S1) and lower (S3) valves. ELLA was performed with 21 lectins, absorbance values at 405 nm were normalized to the highest value (Jacalin) corresponding to 100% reactivity (n = 4, means ± S.D).

found to be the most reactive lectin with both extracts. This lectin binds preferentially D-galactose and oligosaccharides terminating with this sugar. It also targets the galactose-β(1–3)N-acetylgalactosamine dimer, and consequently, is specific to O-linked glycoproteins. A group of four lectins including concavalin A (conA), wheat germ agglutinin (WGA), *Datura stramonium* lectin (jimsonweed, DSL) and tomato lectin (*Lycopersicon*, LEL) gave reactivities between 60 and 40%. ConA is a mannose-binding lectin of N-linked glycopeptides, while WGA, DSL and LEL bind monomers to oligomers of N-acetylglucosamine, and thus, are usually considered as chitin-binding lectins. A second group of 6 lectins, including DBA, LCA, ECL, PHA-L, PNA and STL gave moderate reactivities, between 40 and 20%. DBA binds to N-acetyl-D-galactosamine, LCA, to complex α-linked mannose-containing sequences, ECL, to N-acetyllactosamine (dimers of β-galactose and N-acetylglucosamine), PHA-L, to N-linked oligosaccharides containing galactose, mannose and N-acetyl-glucosamine, PNA, to lactose (dimer of glucose and galactose), and finally, STL, to oligomers of N-acetylglucosamine (chitin-binding). The other ten lectins were considered as unreactive with the extracts (signals below 20%).

The two soluble matrices exhibited very similar reactivity profiles, in particular with the six highest-reacting lectins and the non-reacting lectins. In the moderate reactivity range (20 to 40%), 4 lectins, namely ECL, DBA, PNA and PHA-I gave slightly different responses. In general, we noticed that in all of the cases S1 gave higher intensity than S3.

2.6. *In vitro* $CaCO_3$ *crystallisation assay*

The crystallisation assay was used to check in which manner *Spondylus* ASMs (samples S1 and S3, obtained from the upper and lower valves of the shell respectively) interact with the formation of calcite crystals grown *in vitro*. Only the ASM fractions were tested due to the insolubility of their AIM counterparts; the results were obtained at concentrations of 0.5, 2, 4, 8 and 16 $\mu g \cdot mL^{-1}$ of each extract (Fig. 8).

In blank conditions (no matrix), calcite crystals were typically rhombohedral with flat surfaces. The first signs of crystal shape modification could be observed already at low concentrations of ASMs (0.5 $\mu g \cdot mL^{-1}$), where aggregates were obtained. With both ASM extracts, the effects became more pronounced at 2 and 4 $\mu g \cdot mL^{-1}$ concentrations, at which crystals developed rounded edges. At higher concentrations (8 $\mu g \cdot mL^{-1}$ and above), crystal morphologies were completely altered. At 16 $\mu g \cdot mL^{-1}$ drop-shaped forms were produced, from the dense packing of numerous microcrystals. Overall, both ASMs

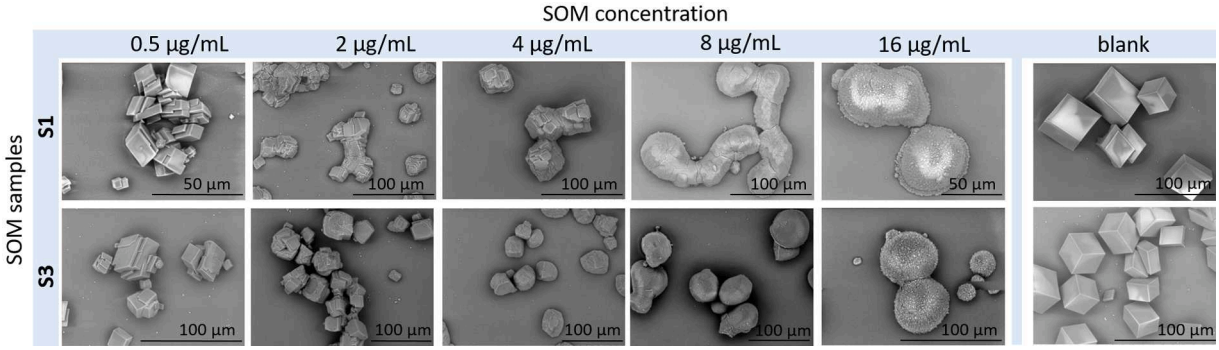


Fig. 8. SEM images of the *in vitro* $CaCO_3$ crystallisation assay in the presence of *Spondylus* ASMs (fractions S1 and S3) tested at different concentrations: the effect of ASMs was tested at increasing concentrations of 0.5, 2, 4, 8 and 16 $\mu g \cdot mL^{-1}$.

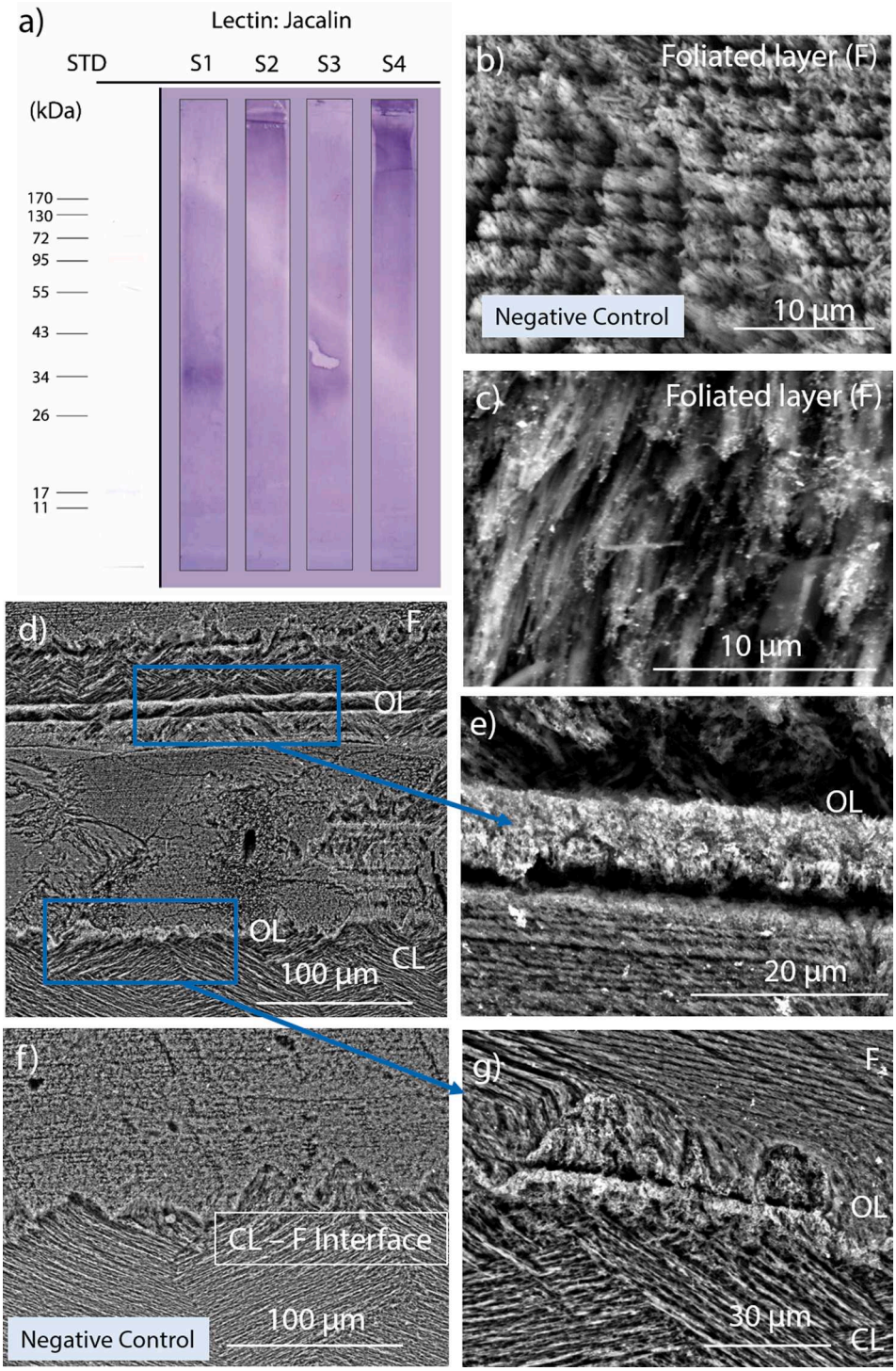


Fig. 9. Lectin-gold *in situ* localisation assay on *Spondylus gaederopus* shell sections. a) Western blot on *Spondylus* shell organic matrices (S1-4); the membrane was incubated with lectin Jacalin. b-g) *In situ* lectin-gold localization of the sugar moieties to which Jacalin binds in different zones. Images b) and f) corresponds to negative controls. F - foliated calcite layer; CL - aragonitic crossed-lamellar; OL - organic layer.

exert similar effects on the crystallisation of $CaCO_3$ and this effect is well pronounced already at low matrix concentrations.

2.7. *In situ* lectin-gold localization of sugar moieties

Lectin-gold *in situ* tests were carried out to directly localise glycans that bind Jacalin along the shell structure. This lectin previously showed the strongest signal in ELLA assay and indicated the presence of galactose/N-acetylgalactosamine residues and/ or oligosaccharides terminating with D-galactose. *In situ* localization experiments were performed on small fragments that were sampled in the region close to the margin, where foliated and crossed-lamellar layers are in contact. Jacalin was also tested on Western blots to check its binding specificity to shell matrix components (Fig. 9a). We found that it reacted mostly with a thick discrete band located around 30–34 kDa in the two tested ASMs and it gave a strong, although polydisperse, signal at very high

molecular weights in the two tested LS-AIMs.

The experiments showed that Jacalin binds to the sugars groups (possibly glycoproteins) present only in the upper, foliated calcite layer (Fig. 9c) and not in the aragonitic cross-lamellar structure. A very intense signal is observed at the interface between the crossed-lamellar (CL) and foliated (F) microstructures, (Fig. 9g) and also where irregular microstructure transits to well ordered CL layer (Fig. 9d, e). This suggests that such zones concentrate most of the glycoside-rich organic matrix (marking the OL layer). This matrix is also enriched in proteins, as the same area gave a strong signal when stained with the antibody K5090 (elicited against the soluble matrix of the nacreous layer of *Pinna nobilis*) by using immunogold technique (for further details on this see the supplementary information 1, part SI.2).

2.8. ShellOmics

We characterised the proteome of *Spondylus* shell matrix, by analysing the eight organic extracts (Tables 1,2). The tryptic digests (see Materials and Methods) were analysed by liquid chromatography coupled to a tandem mass spectrometer (LC-MS/MS). The raw mass spectrometry data were analysed using the software program PEAKS Studio 8.5 (Ma et al., 2003), searching against a protein database obtained from NCBI ("NCBI," n.d.) and restricting the taxonomy to Mollusca. We have also separately carried out data analysis using MASCOT search engine against NCBI nr and Swissprot databases of "other metazoa". However, this latter search has resulted in very few findings (only several peptides were identified), therefore, we have focused our analyses using the results obtained via PEAKS search. The number of identified proteins (with a minimum of 2 unique peptides) was higher in the AIM fractions than in the ASM (up to 21 *vs* 14 sequences respectively). This was also consistent with higher numbers of supporting peptides (up to 67 in AIM *vs* 18 in ASM) and the *de novo* tags (up to ~9100 *vs* 5600) (Table 1).

Considering sequences identified with at least one unique peptide, a total of 46 proteins were found (SI data 2). Many of these, while poorly covered, were detected in several of the eight extracts and some of them may potentially be relevant for calcification, *e.g.*, alkaline phosphatase, mucin-5AC-like, insoluble matrix protein, extensin-like, ADAM family mig-17, fibronectin type III domain-containing protein 2-like, glycine-rich cell wall structural protein-like, testis-expressed protein 9-like, 15 kDa selenoprotein-like, brevican core protein, and biotin-protein ligase-like.

However, we discuss in detail only the identifications supported by a minimum of two unique peptides, in at least one of the matrices, hence restricting the pool to 6 proteins (Table 2). No significant differences were observed between the lower and the upper valves of the shell in terms of proteome composition. However, the duration of the bleaching treatment appears to have an effect on the quality of the identification: counterintuitively, we obtained more supporting peptides and higher coverages from the 3BL extracts than from the 2BL.

2.9. Protein identifications

Five out of six shell proteins identified in *Spondylus* were originally identified from the japanese scallop *Mizuhopecten yessoensis* (family Pectinidae), for which the full genome has been published in 2017 (Wang et al., 2017).

- Uncharacterized protein LOC110461617 (gi|1207960950|ref|XP_021370834.1) was found in all samples except S1, and with 9% coverage in sample S8 (30 supporting peptides and 234 identified *de novo* tags sharing at least 6 AA residues). It is a high molecular weight (289.152 kDa) and basic (theoretical pI 8.25) shell protein, enriched in glycine residues (20.1%), with a notable presence of serine (7.5%), glutamine (7.1%), threonine (6.2%), alanine (6%) and asparagine (6%). The protein exhibits four von Willebrand factor type A domains (VWA) and three chitin-binding domains. In sample S8, these domains are partially covered (Fig. 10). The protein shows the potential occurrence of glycosylation, phosphorylation and N-myristoylation sites (however only a few of the first two modifications have been actually identified in the peptides by the PEAKS software, using the Spider search); there are also predicted NHL repeats (commonly found in a variety of enzymes). Intrinsically disordered regions (IDRs) dominate the C-terminal half of the protein sequence (predicted by ANCHOR2 and IUPred2), but short disordered motifs are also located between the protein family (PFAM) domains (Fig. 10). Overall, an NCBI BLASTp search revealed that this sequence is similar to that of PIF protein (with a query coverage of 68%) and collagen alpha-6(VI) chain (both from *Mizuhopecten yessoensis*), collagen alpha-4(VI) chain (*Crassostrea gigas*) and BMSP (from *Pinctada fucata* and *Mytilus galloprovincialis*), all showing E-values of 0. The homologous BMSP protein from *Mytilus galloprovincialis* (gi|347800228|dbj|BAK86420.1|), a basic calcium-binding protein occurring in nacreous shells (Suzuki et al., 2011), which similarly has four VWA and two chitin-binding domains interspersed by IDRs (SI data 1, Fig. SI.4), was also identified in three of our *Spondylus* AIM extracts (samples S2/6/8).
- Uncharacterized protein LOC110452388 (gi|1207934318|ref|XP_021356561.1|; gi|1207934321|ref|XP_021356563.1|) is a small (280 AA, 30.58 kDa), very basic protein (theoretical pI 11.33) rich in proline (11.4%), arginine (9.6%), glycine (9.3%), serine (8.2%), threonine (8.2%) and asparagine (7.9%). This protein is predicted to have several glycosylation, phosphorylation and N-myristoylation sites (predicted with Interpro (Mitchell et al., 2019)), but none of these modifications have been identified in our sequences by the PEAKS software. Most of the protein is predicted to be intrinsically disordered, with an ordered structure limited solely to the C-terminus (supplementary information 1 Fig. SI.4). A BLASTp search indicates that it is homologous to other shell proteins such as ubiquilin-like isoform X3 and heterogeneous nuclear ribonucleoprotein 87F-like isoform X2 proteins (both from *Crassostrea virginica*, E-values 2e-05 and 3e-05 respectively), although the maximum query coverage is only around 30%.

Table 1
Summary of proteomics results: number of identified proteins, total number of supporting peptides, total number of de novo only tags present in each sample. The thresholds for peptide and protein identification were set as follows: protein false discovery rate (FDR) = 0.5%, protein score 10lgP > 20, unique peptides 2, de novo sequences scores (ALC%) > 50.

Data	2BL				3BL			
	Upper valve		Lower valve		Upper valve		Lower valve	
	ASM S1	AIM S2	ASM S3	AIM S4	ASM S5	AIM S6	ASM S7	AIM S8
Identified proteins / protein families:	2	18	5	6	4	12	14	20
Supporting peptides	2	41	7	17	4	25	15	54
De novo only tags	5864	9186	6651	8189	5610	8260	5340	8046

Table 2
Skeletal matrix proteins identified in *Spondylus* shell. The presented sequences are identified with a minimum of 2 unique peptides (numbers in bold) in at least one of the eight matrices (S1-S8). Cell colour corresponds to the sequence coverage in the sample. The numbers indicate the total number of supporting peptides (Pep), unique peptides (Un) and the identified de novo tags (Dn.t) which share at least 6 amino acid residues (see legend for details).

Samples				Protein Hits	Name	Uncharacterized protein LOC110461617 [Mizuhopecten yessoensis]	Uncharacterized protein LOC110452388 [Mizuhopecten yessoensis]	Laccase [Mizuhopecten yessoensis] & Laccase isoforms	G-protein coupled receptor 112 / Uncharacterized protein LOC110443881 [Mizuhopecten yessoensis]	BMSP [Mytilus galloprovincialis]	Acidic mammalian chitinase, chitinase-3-like protein & isoforms [Mizuhopecten yessoensis]
					Protein Profiles	BMSP-like protein; pI=8.25; Domains: 4 von Willebrand A, 3 CB, RLCD	pI=11.33	Enzyme L-ascorbate oxidase superfamily; pI=6.5; Domains: CuRO_1-3	pI=4.36; Domains: 7tmB2_Adhesion, GPS, MSCRAMM_SdrD/C, CLECT.	pI=8.92; Domains: 4 von Willebrand A, 1 CB, RLCD	Enzyme chitinase; pI=9.03; Domains: GH18_chitolectin, CBM_14, ChtBD2.
2BL	Upper valve	ASM	S1	Pep. (Un)			1 (1)	1 (1)			
				Dn.t			10	2			
		AIM	S2	Pep. (Un)		**22 (22)**	1 (1)	1 (1)	1 (1)	**2 (2)**	**2 (2)**
				Dn.t		140	8	4	22	36	4
	Lower valve	ASM	S3	Pep. (Un)		1 (1)	**3 (3)**				
				Dn.t		31	5				
		AIM	S4	Pep. (Un)		**12 (12)**			1 (1)		
				Dn.t		204			12		
3BL	Upper valve	ASM	S5	Pep. (Un)		1 (1)	1 (1)				
				Dn.t		35	19				
		AIM	S6	Pep. (Un)		**13 (13)**		1 (1)	1 (1)	**2 (2)**	
				Dn.t		180		7	13	41	
	Lower valve	ASM	S7	Pep. (Un)		**3 (3)**	**2 (2)**	1 (1)			
				Dn.t		35	6	4			
		AIM	S8	Pep. (Un)		**30 (29)**	1 (1)	**3 (3)**	**3 (3)**	**3 (2)**	
				Dn.t		234	7	9	13	53	
					NCBI accession number*	XP_021370834.1	XP_021356561.1; XP_021356563.1	OWF48064.1; XP_021358299.1 & of other isoforms	OWF34702.1; XP_021343995.1	BAK86420.1	OWF36273.1; XP_021340846.1

COVERAGE: 0 - 1 % | 2 - 4 % | 5 - 8 % | ≥9 %

LEGEND
• **In bold** - proteins that were identified with more than 2 peptides • Pep. (Un) - All supporting peptides (Un - unique peptides)
• *Dn.t* - de novo tags sharing 6 amino acid residues • *Some of the proteins have different accession numbers in the NCBI database

- Laccase (gi|1205898913|gb|OWF48064.1|) and laccase-7-like isoforms X1, X2 (all from *Mizuhopecten yessoensis*) were identified in five of the samples (S1/2/6/7/8), albeit with a very weak coverage (1–3%). It is a slightly acidic protein (theoretical pI 6.5) with a molecular weight of 104.16 kDa and is enriched in serine (9.3%), glycine (8.1%), proline (7.4%) and histidine (7.3%). This enzyme belongs to L-ascorbate oxidase superfamily (E-value 7.44e-51) and exhibits three cupredoxin domains (Cu-oxidase_3, Cu-oxidase and Cu-oxidase_2). This protein possesses intrinsically disordered regions, located between the first two cupredoxin domains and at the C-terminal part of the sequence (supplementary information 1 Fig. SI.4). It is homologous to other shell proteins including L-ascorbate oxidase [*Crassostrea gigas*], oxidoreductase OpS5-like [*Crassostrea virginica*] and other shell laccase-like proteins such as laccase-2-like [*Crassostrea virginica*], laccase-1-like [*Pomacea canaliculata*] or laccase-4-like from the brachiopod *Lingula anatina*.
- G-protein coupled receptor 112 from *Mizuhopecten yessoensis* (gi|1205812537|gb|OWF34702.1|) was identified only in the AIM matrices (samples S2/4/6/8). It is a very acidic (theoretical pI 4.36) high molecular weight (302.80 kDa) protein, with a composition dominated by serine (11.5%), glutamic (10.9%) and aspartic (8.4%) acids. Threonine (7.5%), leucine (7%), asparagine (6.7%) and isoleucine (6%) are also abundant. It belongs to the GPCR, family 2, secretin-like protein family, which is homologous to the C-type lectin-like superfamily. It possesses immunoglobulin-like domains, GPS motifs and GPCR domains. Most of its sequence is intrinsically disordered, except at the C-terminus (supplementary information 1 Fig. SI.4). The sequence shows similarities to several other uncharacterized shell proteins and adhesion G-protein coupled receptor G proteins from shells, such as *Crassostrea virginica*, *Pomacea canaliculata*, or the brachiopod *Lingula anatina*.
- Acidic mammalian chitinase (gi|1205814209|gb|OWF36273.1|) and its isoforms from *Mizuhopecten yessoensis* were identified in sample S2 only. This is a basic protein (pI 9.03) with a molecular weight of 78.79 kDa. It has chitinase II and chitin-binding domains as well as predicted intrinsically disordered region in the last third part of the sequence (supplementary information 1 Fig. SI.4) and shows similarity to the glycoside hydrolase superfamily proteins. The protein is homologous to many other molluscan chitinases, among them those from *Hyriopsis cumingii* (the triangle shell mussel), *Crassostrea gigas* (the Pacific edible oyster) and *Octopus vulgaris* (the common octopus). Chitinases are enzymes that cleave glycosidic bonds of chitin, and are consequently involved in the remodelling of chitinous scaffolds.

2.10. Peptidome analysis

Given the low number of proteins identified in the *Spondylus* shell organic matrix using standard database searches, we implemented a different characterisation strategy, focusing on the analysis of the several thousands peptide sequences which can be reconstructed by assisted *de novo* algorithms (using PEAKS Studio 8.5). First, all of the raw MS files were searched against a database of known common laboratory contaminants (cRAP). In all of the samples, between 51 and 226 peptide sequences were identified belonging to contamination (corresponding to human keratin, trypsin, wool keratin, beta-casein) and such contaminant peptides were discarded. We then produced a list of genuine *de novo* peptides (ALC > 80%) (SI data 3) and searched these

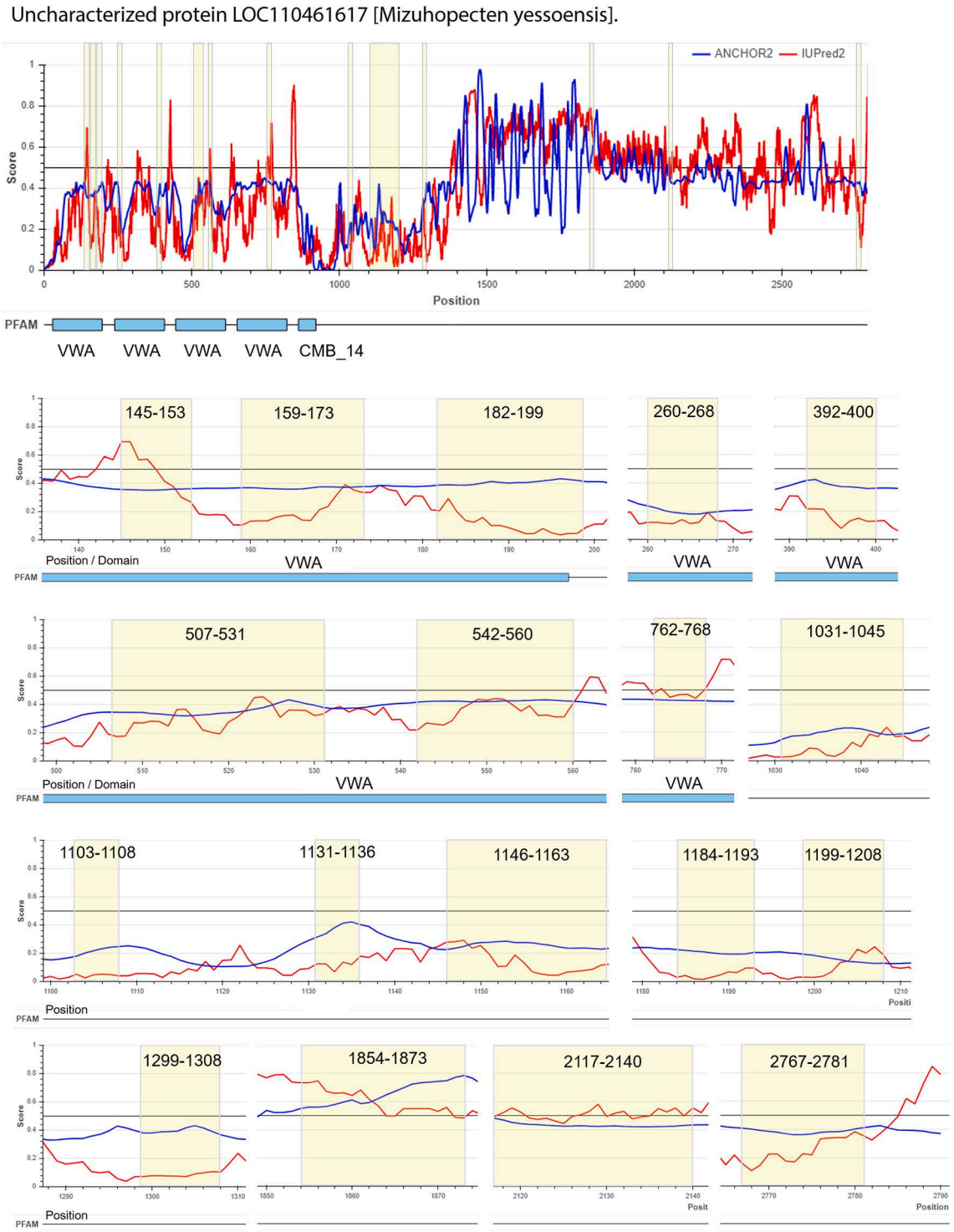


Fig. 10. Prediction of intrinsically disordered regions (IDRs) for Uncharacterized protein LOC110461617 [*Mizuhopecten yessoensis*] and the coverage of this protein in sample S8. Structure prediction was obtained using the IUPred2A software (IUPred2 and ANCHOR2 tools). The graph shows the disorder tendency of the residues, where a higher score (y axis) corresponds to a higher probability of disorder (values from 0 to 1). The top graph shows this tendency throughout the full sequence; the regions where covering peptides have been identified (in sample S8) are marked in yellow and detailed in the figures below. Where present, protein domain families (PFAM) are indicated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

"peptidomes" for any pattern emerging in the different extracts (Fig. 11). In agreement with previous findings we observed that the AIMs contain more peptides than the ASMs (~1930 *vs* ~1150). A number of these sequences display amino acid patterns that are typically observed in other shell proteins, such as poly-Gly (predominantly in AIMs), poly-Ser, poly-Gln (predominantly in ASMs) and acidic motifs (EE, DD), but also unusual motifs, like poly-Leu and poly-Pro.

3. Discussion

This study presents a comprehensive biomolecular and microstructural investigation of the Mediterranean thorny oyster *Spondylus gaederopus*. The bivalve, which underwent a relatively recent evolution, shows divergent molecular features compared to other well studied species. Our discussion examines different aspects, including

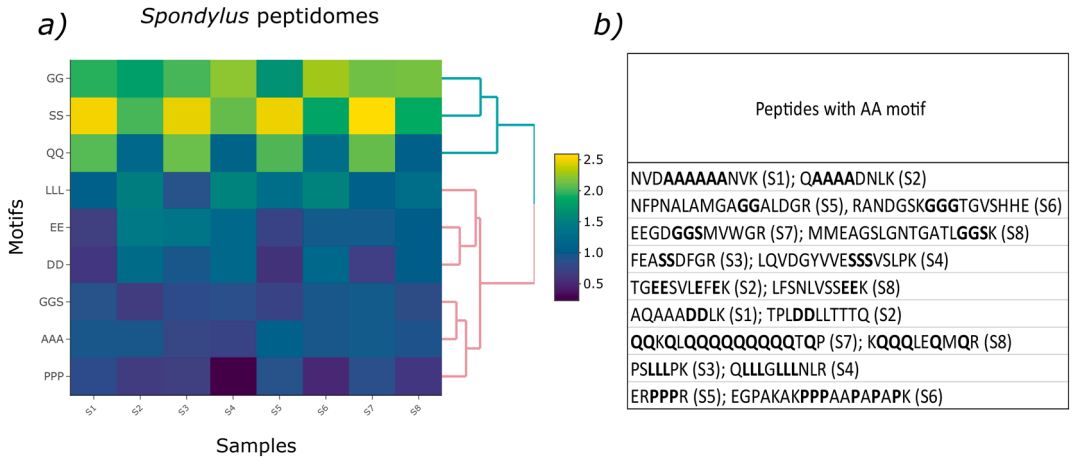


Fig. 11. *Spondylus* shell peptidomes (samples S1-S8) characterized by occuring motifs: a) heatmap that represents the occurrence of amino acid motifs (vertical axis) found in peptides of samples S1-S8 (horizontal axes); b) example peptides with these motifs. The full table with peptidome sets can be found in SI Data 3.

microstructure, matrix biochemistry and proteomics.

3.1. Microstructure

Spondylus and the typical representatives of the Pectinidae family, *e.g. Pecten,* belong to the same superfamily (Pectinoidea), however their microstructures are different; the only shared characteristic being their upper foliated calcitic layers. Compared to other pectinids (Carter, 1990) *S. gaederopus* shells are certainly more complex in their texture combination, especially with regard to the middle and inner layers, where at least three different microstructures are observed (Logan, 1974): aragonitic prismatic, crossed-lamellar (with also different lamellae orientations) and calcitic foliated. Furthermore, the thickness ratio of these three microstructures changes along the growth line, an effect that has been discovered to be temperature-dependent rather than taxon-specific (Nishida et al., 2012; Nishida et al., 2015). We also observed that the change between different aragonitic microstructures is very gradual, while the switch between the layers of two different polymorphs (cross-lamellar aragonite and foliated calcite) is abrupt and well marked, confirming earlier descriptions (Wilmot et al., 1992). The long hollow spines, a morphological shell character of spondylids and which function both as an inducement for the settlement of epibionts on the shell and as antipredator-defense (Logan, 1974; Feifarek, 1987), also display a complex microstructure, which is rather disordered at their basis, morphing into crossed-foliated. This microstructural feature has been evidenced in gastropod shells (Taylor and Reid, 1990; Fuchigami and Sasaki, 2005), but, interestingly, has not been reported in bivalves so far.

3.2. Biochemistry

The amount of extracted organic fraction in *S. gaederopus* shell is very low. It decreases even further with progressive bleaching times (in 3BL samples, only 0.04–0.25 wt‰ of organics, noted as "intracrystalline", is obtained).

One may argue that the shell cleaning with NaOCl is the first responsible for a low amount of organic matrix. This is true after the strong 3BL step, but not after 2BL, which remains mild: a similar treatment performed a few years ago on nacre coarse powder clearly showed that the extracted matrix was still abundant, even after an overnight incubation in NaOCl (Parker et al. 2015).

We also note that the isolation of an "operational" intracrystalline fraction of organics is usually carried out using concentrated NaOCl (12% w/v) and exposure times of 48 hrs or more (Penkman et al., 2008). However, when considering the 2BL extracts, the reactivity of Jacalin (the most reactive lectin with the sugar moieties present in the shell organics) is comparable to that observed on 48 hrs bleached extracts (see supplementary data 1, section SI.5); this indicates that the 3BL step isolates a fraction that is in all likelihood "intracrystalline".

The finding that *Spondylus* matrices are rather organic-poor is consistent with the data obtained on other cross-lamellar (Marie et al., 2013a; Osuna-Mascaró et al., 2014; Demarchi et al., 2015; Agbaje et al., 2019) and foliated (Pierini et al., 2016) mollusc shells; this contrasts to what is observed in organic-rich nacreous and prismatic microstructures (Marie et al., 2009b; Marin et al., 2013; Sakalauskaite et al., 2019). As spondylids represent a bivalve family that emerged rather recently, in the Middle Jurassic (Waller, 2006), our observations corroborate the pioneering idea of Palmer that, in molluscs, the "cost" for making a shell is mostly driven by the "cost" to synthesize the organic matrix (Palmer, 1992); we can argue that evolution was directed to strike a balance between the mechanical properties of a biomineral and the actual "cost" required to produce it. For example, nacre is very "expensive", while crossed-lamellar microstructures, albeit exhibiting only slightly inferior mechanical properties (Kamat et al., 2000), are "cheaper" to produce. In the course of the Phanerozoic, mollusc evolution has favoured the latter solution; the Spondylidae may illustrate the emergence of very tough microstructures that do not require a lot of mineral-bound shell matrix for their synthesis.

The biochemical composition of *Spondylus* shell matrix is typical of mollusc shells, *i.e.* a mixture of proteins and saccharide moieties that display different solubilities in acidic conditions (hence, the distinction between AIM and ASM). In our study, no compelling differences were observed between the shell matrices of the upper red and lower white valves.

On gels, the ASM set of macromolecules, with the exception of few, show mostly non-discrete band profiles - an observation that is similar to most, if not all, shell matrices (Marin et al., 2007). While typical calcium-binding proteins (like calmodulin, which exhibits the EF-hand motif) stain blue with Stains-all, *Spondylus* ASMs stain purple. This feature had already been observed for other acidic shell matrices, such as that extracted from the calcitic prisms of the fan mussel *Pinna nobilis* (Marin et al., 2005). We have also observed that *Spondylus* ASMs (which represent the sum of inter- and intracrystalline matrices) induced strong modifications of the crystal formation during the *in vitro* $CaCO_3$ crystallization; the effect was much more prominent compared to ASMs extracted from other aragonite shells, *e.g.* the land snail *Helix aspersa maxima* (Pavat et al., 2012). This result is interesting because *in vitro* assays typically show that acidic macromolecules (proteins and glycoproteins) induce the strongest effect and are usually considered to be intracrystalline (Albeck et al., 1993; Albeck et al., 1996). Indeed, glycoproteins are identified in our *Spondylus* organic matrices, yet, more interestingly the proteomics data reveals that basic proteins are also

abundant in the intracrystalline fraction, a finding that expands the knowledge on mineral-bound intracrystalline organics (Demarchi et al., 2016).

Skeletal matrices have their own saccharide signature (Kanold et al., 2015) and the matrix carbohydrates are suspected to perform several functions related to biomineralization (Albeck et al., 1996; Sarashina and Endo, 2001; Marie et al., 2007; Samata et al., 2008). However, they are largely ignored in biomineralization studies, whether they are present as free polysaccharides or glycoproteins. In our case, the lectin assay of the ASMs of *S. gaederopus* underlines the diversity of the sugar residues and of the glycosylation pattern:

1) The preponderance of O-linked glycoproteins and/ or oligosaccharides with Gal, Gal-GalNAc dimers (as observed by the strong affinity to Jacalin);
2) The presence of N-linked mannose containing oligopeptides;
3) The presence of N-acetylglucosamine (or oligomers of this sugar residue), suggesting that chitin is most likely part of the shell organic matrix. As chitin is usually insoluble, it may be either that the glycans in the ASMs are constituted of oligomeric forms, or that the ASM contains a soluble form of chitin, partially cleaved by the bleaching treatment prior to the extraction.

It is interesting to note that Jacalin on Western blots shows different binding profiles in the ASM and LS-AIM matrices; this implies that either the glycan targets are different in their monosaccharide sequences, or that they are identical but differently crossed-linked, *i.e.* the soluble monomeric form being present in the ASM while the crossed-linked one in the LS-AIM. Finally, *in situ* gold-mapping shows that the glycans targeted with Jacalin are not evenly distributed throughout the shell cross-section, but are specific to the upper foliated calcitic layer and are particularly concentrated at the junction with the underlying aragonitic crossed-lamellar layer (Fig. 9), as seen also by luminescence imaging. The data suggest that this organic layer plays an important role in this structural and mineralogical transition - switching from foliated calcite to crossed-lamellar aragonite, while maintaining these two microstructures glued together (Wilmot et al., 1992).

3.3. Proteomics

Shotgun proteomics was employed in order to characterise the shell proteome composition. It is however important to stress that the results obtained will be forcibly biased, due to several limitations, including the intrinsic peculiarities of shell matrix proteins and the paucity of -omics reference data for shell matrices, particularly evident in poorly studied organisms such as spondylids.

1) Shell matrix proteins are frequently characterized by the presence of repeated low complexity domains (LCDs or RLCDs, frequently found as long poly-Ala/Asp/Ser/Gly blocks (Kocot et al., 2016; Marie et al., 2017; Sakalauskaite et al., 2019), and post translational modifications such as phosphorylation (Du et al., 2018) and glycosylation (Mann and Edsinger, 2014; Boskey and Villarreal-Ramirez, 2016), making the protein sequences difficult to cleave using standard proteases and thus reducing the chance of identification.
2) The identification is database-dependent. Mollusc shell proteomes are likely to be very different across taxa and microstructures (Jackson et al., 2010; Marie et al., 2013a) but only a small part of these "shellomes" (mainly those of commercially-relevant species) have been studied so far in a phylum that comprises between 85,000 and 100,000 shell-forming species (Marie et al., 2012, 2011; Yarra et al., 2016; Feng et al., 2017). Crossed-lamellar microstructures are particularly common among bivalves, gastropods and scaphopods and yet, they are the most understudied.

In our work - the first on spondylids - we had to rely on the omics data available for closely related species (in particular pectinoid shells). However, it is likely that spondylids "shellomes" are much more diversified than presumed, given the complex evolution of Pectinoidea (Smedley et al., 2019).

3.4. Comparison between shell extracts

Despite the low organics yield obtained from the bleached shell powders, high-resolution tandem mass spectrometry was successful in obtaining several thousands of product ion spectra for each sample; their sequences were reconstructed using the software PEAKS Studio 8.5.

Spondylus ASMs and AIMs exhibit similar sets of proteins, but more sequences were identified in the AIMs. Interestingly, this is particularly evident for the intracrystalline AIM matrix (the 3BL sample) indicating that by exposing shell to longer bleaching times and removing a large part of the intercrystalline organics, we can facilitate their detection.

The use of PEAKS software resulted in the identification of between 5000 and 9000 *de novo* sequences for each sample (Table 1). Several of these reconstructed peptide sequences show interesting features that are commonly observed in other shell proteomes (Fig. 11). These include glycine and serine repeats as well as acidic blocks (such as EE, DD) in the AIM extracts and polyQs in ASMs. Another motif, commonly encountered in nacreous shell proteins, poly-Ala blocks, was rare in *Spondylus* proteins.

While, at first sight, *Spondylus* peptidomes resemble "typical" shell protein sequences, the majority of *de novo* sequences could not be assigned to any known molluscan proteins. This clearly indicates that the six identified proteins represent only the partial picture of the full *Spondylus* shell associated proteome and the true number of proteins is likely much higher. Five out of the six identified sequences belong to the pectinoid *Mizuhopecten yessoensis*, the japanese scallop, which is the phylogenetically closest relative with a full genome analysed (Wang et al., 2017). The coverages were low for all of the identified sequences (1–9%), but most of these proteins possess features typical of shell biomineralization, such as WVA and chitin-binding domains present in Uncharacterized protein LOC110461617 [*Mizuhopecten yessoensis*] and BSMP. Both of these proteins are homologues to PIF, suggesting that these sequences found in *Spondylus* are important in calcium binding activity and possibly play a role in the biomineralization of the aragonitic structures (Suzuki et al., 2011; Arivalagan et al., 2017; Feng et al., 2017).

The two identified enzymes have functions related to shell formation: i) chitinase, a chitin binding protein that catalyses the cleavage of chitin, could be involved in the (re)modeling of the chitinous scaffold (Yonezawa et al., 2016); ii) laccase [*Mizuhopecten yessoensis*], a copper oxidase enzyme, could be involved in cross-linking of the matrix during biomineralization. In the prism matrix of the pearl oyster, an enzyme with similar function was identified (Marie et al., 2012), while laccase-like genes have been shown to be involved in shell immunity, biomineralization and pigmentation (Mao et al., 2018; Yue et al., 2019).

We note that all six identified shell matrix proteins exhibit intrinsically disordered regions, IDRs (Fig. 10, Fig. SI.4 (Supplementary data 1)). An important characteristic of IDRs lies in their flexibility, which enables matrix assembly and protein-mineral interaction (Kalmar et al., 2012; Boskey and Villarreal-Ramirez, 2016). This view is in agreement with the preponderance of these regions in many skeletal matrix proteins (Evans, 2012). The uncharacterized protein LOC110461617 identified in *Spondylus* matrices shows the presence of the IDRs located in between the functional domains (the WVAs in the N-terminal part of the sequence) and dominate almost half of the C-terminal part of the sequence (Fig. 10). Interestingly, the highest coverage of this protein in our samples coincides with the conserved regions, while IDRs have very low and sporadic coverage; since IDRs are believed to evolve fast (Kocot et al., 2016) and to be lineage-specific, this may explain why these domains are poorly covered, when using the

sequence data from other genera as reference. These sequence gaps will be certainly filled by the acquisition of a transcriptome of actively calcifying *S. gaederopus* specimens.

4. Conclusion

Our paper describes the first molecular characterisation of the shell of *S. gaederopus*, a representative of spondylid family, with a particularly complex microstructure. We found that *Spondylus* possesses a remarkably small amount of mineral-bound skeletal matrix. This latter is rich in glycosidic moieties and is mainly intercrystalline, in particular concentrated at the transition between different microstructural layers. We carried out proteomic analysis, which enabled to partly characterize the shell protein repertoire of *Spondylus*. Even though almost all of the identified sequences were most closely matched to pectinoid shell proteins (belonging to *Mizuhopecten yessoensis*), the low numbers of identified proteins and their scant coverages suggest that the *S. gaederopus* shellome is quite different from that of *Pecten*. The lack of genomics data for spondylids implies that most of the *Spondylus* shell proteins remain largely unknown. Our key finding therefore is that *S. gaederopus* (and this might be extended to the whole spondylid family) is likely to have adopted distinctive molecular strategies to biomineralize its shell. In the future, this may result in a reassessment of the phylogenetic relationships of Spondylidae within the pteriomorphid bivalves clade.

This unique molecular fingerprint, albeit puzzling from an evolutionary point of view, is rich in perspectives for "palaeoshellomics", *i.e.* the study of ancient shell proteins (Sakalauskaite et al., 2019). This is particularly interesting for biomolecular archaeologists aiming to obtain taxonomic identification of fragmented archaeological shell objects, where biological origin cannot be assigned via classical means (microstructural observations). Our findings suggest that the molecular identification of precious *Spondylus* shell fragments among many other shells in the palaeontological/ archaeological records, is now made possible.

5. Materials and methods

5.1. Sampling and preparation

Spondylus gaederopus Linnaeus, 1758 (Mollusca; Bivalvia; Pteriomorphia; Pectinida; Pectinoidea; Spondylidae; *Spondylus*) is endemic to the Mediterranean sea and is not listed in the International Union for Conservation of Nature (IUCN's) list of threatened species. Adult shells (around 100 mm size) were purchased from Conchology, Inc ("Conchology, Inc," n.d.). The specimens were collected live at 15 m depth (by diving, as listed by the vendors), in Saronikos, Greece, in 2010. In the laboratory, the shell was cleaned using abrasive tools and immersed in diluted NaOCl (~1–1.5% active chlorine) for 2 hrs (we call this step as 1BL − 1st bleach treatment). The upper and lower valves were rinsed with water, air-dried and crushed into ~2 mm fragments with a Jaw-crusher (Retsch BB200) separately, obtaining two sample subsets of the lower valve (LV) and the upper valve (UV). The obtained fragments were bleached in ~1–1.5% NaOCl for 4 hrs (2BL), rinsed with water (5x), ethanol (1x) and air-dried. Coarsely crushed samples were powdered with a mortar grinder (Frisch Pulverisette 2) to particle size < 200 μm (sieving). Half of both subsamples (UV and LV) were set aside (2BL samples) and the second batch was bleached for the third time in ~1–1.5% NaOCl for 14 hrs (3BL), thoroughly rinsed with water (5x) and air-dried. In total, 4 powdered samples were obtained: 2 batches of different cleaning approach (2Bl and 3BL) each with 2 subsamples - upper and lower valves (UV and LV).

5.2. Extraction of shell organic matrix (SOM)

The extraction was carried out using a well-established protocol developed in our lab (Ramos-Silva et al., 2012; Takeuchi et al., 2018). In short, powdered samples were decalcified using cold diluted acetic acid (10% v/v, 100 μL every 5 s) overnight and then centrifuged (3893 G, 30 min.) to separate the acid soluble (ASM) and acid insoluble matrices (AIM). The AIM matrix was rinsed (5x) with ultra pure water and freeze-dried. The ASM was ultrafiltered using 10 kDa cut-off membrane (Millipore, Ref. PLGC07610), dialysed against 1L of MiliQ water, with 6 changes over 2 days, and freeze-dried.

5.3. Biochemical characterization

5.3.1. FT-IR spectroscopy

FT-IR analysis in Attenuated Total Reflectance (ATR) mode was used to check the overall composition of ASM and AIM matrices of the 2BL samples. The analysis was performed with a Bruker Vector 22 FT-IR spectrometer (BrukerOptics Sarl, France, Marne la Vallée) fitted with a GoldenGate ATR device (SpecacLtd, Orpington, UK) in the 4000–400 cm^{-1} range (twelve scans with a spectral resolution of 4 cm^{-1}). Spectra analysis and assignment of absorption bands were performed by comparison with previously published data (Kanold et al., 2015; Takeuchi et al., 2018).

5.3.2. SDS-PAGE

ASMs were fully dissolved in ultrapure water to a final concentration of 4 μg/mL and 4x Laemmli buffer was added (ratio 3:1). AIMs were resuspended and treated similarly. Denaturation was carried out at 99 °C for 5 min, then samples were cooled on ice. Because of limited AIM solubility, only the Laemmli-soluble supernatants, referred to as LS-AIM, were analysed by SDS-PAGE. Proteins were run on precast (4–15%) gradient (Bio-Rad) and 12% gels, then stained with silver nitrate and Stains-all (Campbell et al., 1983; Takeuchi et al., 2018) respectively.

5.3.3. Western-blot

Samples S1 and S2 were run on a 12% SDS-PAGE gel as reported above, then electro-transferred on a PVDF Immobilon P membrane (Millipore) (Matsudaira, 1987). To detect the glycosylated complex, the membrane was incubated in Carbo-Free blocking solution (Vector Laboratories, Peterborough, UK, ref. SP-5040) then in a solution that contained the biotinylated lectin Jacalin (Vector Labs. Ref. B-1155, diluted in TBS/Tween20, ratio 1:100) and finally, after rinsing, in a solution containing avidin-AP conjugate (Sigma A7294, dilution 1/70000). Thorough rinsing steps with TBS/Tween20 were performed. The membrane was stained with NBT/BCIP (Sigma B5655).

5.3.4. Enzyme-Linked lectin assay (ELLA)

The enzyme-linked immuno assay was carried out on two ASM fractions following previously described procedures (Kanold et al., 2015; Takeuchi et al., 2018). In short, ASM aqueous solutions (100 μL per well, approx. ~200 ng/ well) were deposited in 96-well microplates (MaxiSorp, Nunc/Thermo Scientific, Nunc A/S, Roskilde, Denmark), incubated for 90 min at 37 °C, washed with TBS/Tween20 and blocked with Carbon-Free solution. Three sets of 7 biotinylated lectins were tested (Vector Laboratories, Peterborough, UK, Ref. BK-1000, BK-2000, BK-3000; diluted in TBS/Tween-20 with ratio 1:200 for kits 1,2 and 1:100 for kit 3); lectins were incubated for 90 min at 37 °C, washed, and a solution of alkaline phosphatase-conjugated avidin (A7294 Sigma, 1:70000) was added for another 90 min at 37 °C. Microplates were washed and incubated with the ELISA (aqueous diethanolamine solution, pH 9.8) substrate solution, containing phosphatase substrate (p-nitrophenylphosphate) at 37 °C. Optical density was read using a BioRad Model 680 microplate reader at 405 nm; background values (blank tests) were subtracted and the values were converted to percentage of reactivity, with the highest OD values at 100%. A total of 4 replicates were performed.

5.3.5. *In vitro $CaCO_3$ crystallisation*

The capacity of ASM to interact *in vitro* with $CaCO_3$ was tested in the diffusion assay, as described previously (Kanold et al., 2015; Takeuchi et al., 2018). 200 µL of 10 mM $CaCl_2$ solutions containing ASMs at increasing concentrations (0.25 – 16 µg/ml) were placed in 16-well culture slides (Lab-Tek, Nunc/ Thermo Scientific, Rochester, NY, USA). The plastic cover was pierced beforehand to allow the reaction with ammonium bicarbonate vapours, placed on top of the slide and sealed with parafilm. It was placed in a desiccator under vacuum which contained ammonium bicarbonate crystals and incubated at 4 °C for 72 hrs. Afterwards, the solution was carefully removed with a blunt-end needle and the slide was air-dried. The 16-well culture slide was disassembled and $CaCO_3$ crystals were observed directly (without any carbon sputtering) using a tabletop scanning electron microscope (TM 1000, Hitachi).

5.4. *In situ localization*

5.4.1. *Lectin-gold assay*

Using a diamond saw, a transverse section, perpendicular to the growth line, was cut from the upper valve of the shell (area closer to the margin). The section was polished using a fine-grain sandpaper, followed by an alumina suspension (0.05 µm) and cut into small square fragments. They were cleaned using sonication, then bleached for 10 min with NaOCl (~1–1.5% active chlorine) and rinsed with ultrapure water. Fragments were etched with 1% EDTA (v/v) in the ultrasound bath for 5 min and rinsed again with ultrapure water. The rest of the procedure follows that reported in (Takeuchi et al., 2018): incubation in Carbo-Free blocking solution (1 h); overnight incubation with Jacalin solution (prepared in TBS/Tween-20, diluted 1:10; plus NaN_3, 1/10000 to prevent bacterial growth); rinsing step with TBS/Tween-20 (6x times, 10 min each); incubation with goat anti-biotin antibody conjugated to ultra-small gold particles (0.8 nm) (GABio, Ultra Small, Ref. 800.088; Aurion, Wareningen, The Netherlands) for 2 hrs; after rinsing step and gentle drying, incubation in the silver enhancement solution (British Biocell International, Ref. SEKL15) for about 10–15 min; rinsing in milli-Q water and drying at 37 °C. All incubations were performed at room temperature with gentle shaking. Samples were observed with SEM (Hitachi, TM1000) under back-scattered electron mode

To preclude false positive signals, 3 negative controls were performed in parallel using the same conditions but omitting one or two of the following steps: 1) incubation with Jacalin (negative control 1); 2) incubation with gold-coupled anti-biotin (negative control 2); 3) both incubations with Jacalin and the gold-coupled anti-biotin (negative control 3). All other steps were done similarly. Total of 4 replicas of the experiments were carried out to validate the observations.

5.4.2. *Luminescence imaging by DUV synchrotron photoemission beamline*

The Deep-Ultraviolet (DUV–visible) photoluminescence imaging was applied to trace the distribution of organics in the biomineral of a selected area. Fluorescence maps were recorded on the DISCO beamline (Jamme et al., 2013) using the setup TELEMOS build around an Olympus IX71 inverted microscope stand and a Peltier iDus CCD detector with 26x26 µm pixel size. A small fragment was cut from the upper valve of the shell, that also included a formed spine, and was mirror-polished. The area was scanned using a x40 objective and the luminescence maps were reconstructed and corrected using an in-house script in Matlab, developed at DISCO beamline by one of the authors (F.J.).

5.5. *Proteomic analysis*

For proteomic analyses, digestion with trypsin of each of the eight shell organic extracts was carried out in-gel, after a short migration, around one centimeter in the separating 8% acrylamide gel, stained with Coomassie blue. After the gel runs, samples were destained twice with a mixture of 100 mM ammonium bicarbonate (ABC) and 50% (v/v) acetonitrile (ACN) for 30 min at 22 °C and then dehydrated using 100% ACN for 15 min, before being reduced with 25 mM ABC containing 10 mM DTT for 1 h at 56 °C and alkylated with 55 mM iodoacetamide in 25 mM ABC for 30 min in the dark at 22 °C. Gel pieces were washed twice with 25 mM ABC and dehydrated (twice, 20 min) with 100% ACN. Gel cubes were incubated with sequencing grade modified trypsin (Promega, USA; 12.5 ng/µl in 40 mM ABC with 10% ACN, pH 8.0) overnight at 37 °C. After digestion, peptides were extracted twice with a mixture of 50% ACN – 5% formic acid (FA) and then with 100% ACN. Extracts were dried using a vacuum centrifuge Concentrator plus. Resulting peptides were resuspended in 10 µL of 10% ACN − 0.1% trifluoroacetic acid (TFA) and subjected to analysis by LC-MS/MS.

Mass spectrometry (MS) was performed using an Ultimate 3000 Rapid Separation Liquid Chromatographic (RSLC) system (Thermo Fisher Scientific) online with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific). 1 µL of peptides was loaded, concentrated and washed on a C_{18} reverse phase pre-column (3 µm particle size, 100 Å pore size, 75 µm i.d., 2 cm length, Thermo Fisher Scientific). The loading buffer contained 98% H_2O, 2% ACN and 0.1% TFA. Peptides were then separated on a C_{18} reverse phase resin (2 µm particle size, 100 Å pore size, 75 µm i.d., 25 cm length, Thermo Fisher Scientific) with a 1 h gradient from 99% solvent A (0.1% FA and 100% H_2O) to 90% solvent B (80% ACN, 0.085% FA and 20% H_2O).

The mass spectrometer acquired data throughout the elution process and operated in a data dependent scheme with full MS scans acquired with the Orbitrap, followed by MS/MS HCD fragmentations acquired with the Ion Trap on the most abundant ions detected. Mass spectrometer settings were: full MS (AGC: 2x10e5, resolution: 6x10e4, *m/z* range 350–1500, maximum ion injection time: 60 ms) and MS/MS (HCD collision energy: 30%, AGC: 2x10e4, resolution:3x10e4, maximum injection time: 100 ms, isolation windows: 1.6 *m/z* Da, dynamic exclusion time setting: 30 s). The fragmentation was permitted for precursors with a charge state of 2, 3, 4 and up.

5.6. *Bioinformatic search*

Product ion spectra were analysed using PEAKS Studio (v. 8.5, Bioinformatics Solutions Inc. (BSI) (Ma et al., 2003)) and searched against a "Mollusca protein" database, downloaded from NCBI ("NCBI," n.d.) on 19/12/2018, restricting the taxonomy to Mollusca, including the search of common laboratory contaminants (cRAP; common Repository of Adventitious Proteins: http://www.thegpm.org/crap/). Search parameters were defined assuming no enzyme digestion (in order to also detect naturally cleaved peptides, if any), fragment ion mass tolerance of 0.05 Da and a parent ion tolerance of 10 ppm. Results obtained by SPIDER searches (*i.e.* including all possible modifications) were used for peptide identification and protein characterization, choosing the following threshold values for acceptance of high-quality peptides: false discovery rate (FDR) threshold 0.5%, protein scores − 10lgP ≥ 20, unique peptides ≥ 2, *de novo* sequences scores (ALC %) ≥ 50. An additional full list of protein sequences identified with just one unique peptide were also exported. The peptide sequences identified in shell proteins were individually checked using the BLASTp tool (https://blast.ncbi.nlm.nih.gov/), and any sequences that were homologous to common laboratory contaminants were excluded from any further analysis.

5.7. *Protein profiles*

Protein physical and chemical parameters (MW, theoretical PI, amino acid composition) were assessed using ProtParam tool on the EXPASY server. Protein classification, prediction of domains and motifs was carried using Interpro database (Mitchell et al., 2019) and Motif

scan tool on the EXPASY server. Intrinsically disordered regions in proteins were predicted using IUPred2A (Mészáros et al., 2018) and the BLASTp tool was used to identify homologous proteins.

5.8. Peptidome analysis

For the shell peptidome analysis, the reconstructed *de novo* tags in PEAKS were searched against a database of common laboratory contaminants (cRAP, search parameters defined above). The results obtained by SPIDER search were used by selecting these threshold values: false discovery rate (FDR) threshold 0.5%, protein scores − 10lgP ≥ 20, unique peptides ≥ 0. Only the very high quality *de novo* sequences were taken into account, by selecting the local confidence score ALC % ≥ 80; exported sequences were grouped in Supplementary Data 3.

6. Data deposition

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al., 2013) partner repository with the dataset identifier PXD016760.

Credit authorship contribution statement

Jorune Sakalauskaite: Conceptualization, Investigation, Methodology, Data curation, Writing - original draft. **Laurent Plasseraud:** Investigation, Methodology, Data curation, Writing - review & editing. **Jérôme Thomas:** Investigation, Methodology. **Marie Albéric:** Investigation, Methodology, Resources, Writing - review & editing. **Mathieu Thoury:** Resources. **Jonathan Perrin:** Resources, Writing - review & editing. **Frédéric Jamme:** Investigation, Methodology. **Cédric Broussard:** Investigation, Methodology. **Beatrice Demarchi:** Conceptualization, Methodology, Data curation, Writing - original draft, Project administration, Funding acquisition. **Frédéric Marin:** Conceptualization, Investigation, Methodology, Data curation, Writing - original draft, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

JS, FM and BD would like to thank Emmanuel Fara (University of Burgundy-Franche-Comté) and Matthew Collins (Universities of Copenhagen and Cambridge) for valuable insights and discussions.

JS, FM and BD are supported by the PHC Galilée programme (France, Italy), organised by Italo-French University (UIF/UFI) (project G18-464 / 39612SB) and Campus France; JS acknowledges the support of the Campus France fund obtained from French embassy in Italy (BGF) and program Eiffel (France).

BD is funded by the "Giovani Ricercatori - Rita Levi Montalcini" Programme (MIUR; Ministero dell'Istruzione dell'Università e della Ricerca, Italy).

The work performed at UMR CNRS 6282 Biogéosciences (JS, FM), Dijon was financed via the annual recurrent budget of this research unit allocated to JS and FM, and via extra CNRS funding (INTERRVIE).

The work performed at SOLEIL synchrotron (JS, FM, MA, MT, JP, FJ) was financed via proposal no. 20180213 at DISCO IMAGING beamline.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsb.2020.107497.

References

Addadi, L., Joester, D., Nudelman, F., Weiner, S., 2006. Mollusk shell formation: a source of new concepts for understanding biomineralization processes. Chemistry 12, 980–987.

Addadi, L., Weiner, S., 1997. A pavement of pearl. Nature 389, 912–913.

Agbaje, O.B.A., Thomas, D.E., Dominguez, J.G., Mclnerney, B.V., Kosnik, M.A., Jacob, D.E., 2019. Biomacromolecules in bivalve shells with crossed lamellar architecture. J. Mater. Sci. 54, 4952–4969.

Albeck, S., Addadi, I., Weiner, S., 1996. Regulation of calcite crystal morphology by intracrystalline acidic proteins and glycoproteins. Connect. Tissue Res. 35, 365–370.

Albeck, S., Aizenberg, J., Addadi, L., Weiner, S., 1993. Interactions of various skeletal intracrystalline components with calcite crystals. J. Am. Chem. Soc. 115, 11691–11697.

Arivalagan, J., Yarra, T., Marie, B., Sleight, V.A., Duvernois-Berthet, E., Clark, M.S., Marie, A., Berland, S., 2017. Insights from the Shell Proteome: Biomineralization to Adaptation. Mol. Biol. Evol. 34, 66–77.

Barucca, M., Olmo, E., Schiaparelli, S., Canapa, A., 2004. Molecular phylogeny of the family Pectinidae (Mollusca: Bivalvia) based on mitochondrial 16S and 12S rRNA genes. Mol. Phylogenet. Evol. 31, 89–95.

Borrello, M.A., Micheli, R., 2011. *Spondylus gaederopus* in Prehistoric Italy: Jewels from Neolithic and Copper Age Sites. In: Ifantidis, F., Nikolaïdou, M. (Eds.), *Spondylus* in Prehistory: New Data and Approaches. Archaeopress, Contributions to the Archaeology of Shell Technologies, pp. 25–37.

Borrello, M.A., Micheli, R., 2004. *Spondylus gaederopus*, gioiello dell'Europa preistorica. Preistoria Alpina 40, 71–82.

Boskey, A.L., Villarreal-Ramirez, E., 2016. Intrinsically disordered proteins and biomineralization. Matrix Biol. 52–54, 43–59.

Campbell, K.P., MacLennan, D.H., Jorgensen, A.O., 1983. Staining of the Ca2+-binding proteins, calsequestrin, calmodulin, troponin C, and S-100, with the cationic carbocyanine dye "Stains-all" J. Biol. Chem. 258, 11267–11273.

Carter, J.G. (Ed.), 1990. Skeletal Biomineralization: Pattern, Processes, and Evolutionary Trends. Amer Geophysical Union.

Conchology, Inc [WWW Document], n.d. URL https://www.conchology.be/?t=1 (accessed 11.22.19).

Demarchi, B., Clements, E., Coltorti, M., van de Locht, R., Kröger, R., Penkman, K., Rose, J., 2015. Testing the effect of bleaching on the bivalve *Glycymeris*: A case study of amino acid geochronology on key Mediterranean raised beach deposits. Quat. Geochronol. 25, 49–65.

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

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

Du, J., Xu, G., Liu, C., Zhang, R., 2018. The role of phosphorylation and dephosphorylation of shell matrix proteins in shell formation: an *in vivo* and *in vitro* study. CrystEngComm 20, 3905–3916.

Espinosa, H.D., Rim, J.E., Barthelat, F., Buehler, M.J., 2009. Merger of structure and material in nacre and bone – Perspectives on *de novo* biomimetic materials. Prog. Mater Sci. 54, 1059–1100.

Evans, J.S., 2012. Aragonite-associated biomineralization proteins are disordered and contain interactive motifs. Bioinformatics 28, 3182–3185.

Evans, J.S., 2008. "Tuning in" to mollusk shell nacre- and prismatic-associated protein terminal sequences. Implications for biomineralization and the construction of high performance inorganic-organic composites. Chem. Rev. 108, 4455–4462.

Falini, G., Albeck, S., Weiner, S., Addadi, L., Others, 1996. Control of aragonite or calcite polymorphism by mollusk shell macromolecules. Science-AAAS-Weekly Paper Edition 271, 67–69.

Feifarek, B.P., 1987. Spines and epibionts as antipredator defenses in the thorny oyster *Spondylus americanus* Hermann. J. Exp. Mar. Bio. Ecol. 105, 39–56.

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

Finnemore, A., Cunha, P., Shean, T., Vignolini, S., Guldin, S., Oyen, M., Steiner, U., 2012. Biomimetic layer-by-layer assembly of artificial nacre. Nat. Commun. 3, 966.

Fuchigami, T., Sasaki, T., 2005. The shell structure of the Recent Patellogastropoda (Mollusca: Gastropoda). Paleontol. Res. 9, 143–168.

Gao, P., Liao, Z., Wang, X.-X., Bao, L.-F., Fan, M.-H., Li, X.-M., Wu, C.-W., Xia, S.-W., 2015. Layer-by-Layer Proteomic Analysis of *Mytilus galloprovincialis* Shell. PLoS One 10, e0133913.

Gosling, E., 2003. Bivalve Molluscs - Biology, Ecology and Culture. Fishing New Books, Blackwell Science, Oxford, UK.

Herlitze, I., Marie, B., Marin, F., Jackson, D.J., 2018. Molecular modularity and

asymmetry of the molluscan mantle revealed by a gene expression atlas. Gigascience 7. https://doi.org/10.1093/gigascience/giy056.
Ifantidis, F., Nikolaidou, M., 2011. Spondylus in Prehistory: New Data and Approaches: Contributions to the Archaeology of. Shell Technologies.
Jackson, D.J., McDougall, C., Woodcroft, B., Moase, P., Rose, R.A., Kube, M., Reinhardt, R., Rokhsar, D.S., Montagnani, C., Joubert, C., Piquemal, D., Degnan, B.M., 2010. Parallel evolution of nacre building gene sets in molluscs. Mol. Biol. Evol. 27, 591–608.
Jamme, F., Kascakova, S., Villette, S., Allouche, F., Pallu, S., Rouam, V., Réfrégiers, M., 2013. Deep UV autofluorescence microscopy for cell biology and tissue histology. Biol. Cell 105, 277–288.
Kalmar, L., Homola, D., Varga, G., Tompa, P., 2012. Structural disorder in proteins brings order to crystal growth in biomineralization. Bone 51, 528–534.
Kamat, S., Su, X., Ballarini, R., Heuer, A.H., 2000. Structural basis for the fracture toughness of the shell of the conch *Strombus gigas*. Nature 405, 1036–1040.
Kanold, J.M., Guichard, N., Immel, F., Plasseraud, L., Corneillat, M., Alcaraz, G., Brümmer, F., Marin, F., 2015. Spine and test skeletal matrices of the Mediterranean sea urchin Arbacia lixula–a comparative characterization of their sugar signature. FEBS J. 282, 1891–1905.
Knoll, A.H., 2003. Biomineralization and Evolutionary History. Reviews in Mineral. Geochem. 54, 329–356.
Kocot, K.M., Aguilera, F., McDougall, C., Jackson, D.J., Degnan, B.M., 2016. Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. Front. Zool. 13, 23.
Logan, A., 1974. Morphology and Life Habits of the Recent Cementing Bivalve *Spondylus Americanus* Hermann from the Bermuda Platform. Bull. Mar. Sci. 24, 568–594.
Lowenstam, H.A., 1981. Minerals formed by organisms. Science 211, 1126–1131.
Luz, G.M., Mano, J.F., 2009. Biomimetic design of materials and biomaterials inspired by the structure of nacre. Philos. Trans. A Math. Phys. Eng. Sci. 367, 1587–1605.
Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., Lajoie, G., 2003. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun. Mass Spectrom. 17, 2337–2342.
Mann, K., Edsinger, E., 2014. The *Lottia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. Proteome Sci. 12, 28.
Mao, J., Zhang, W., Zhang, X., Tian, Y., Wang, X., Hao, Z., Chang, Y., 2018. Transcriptional changes in the Japanese scallop (*Mizuhopecten yessoensis*) shellinfested by Polydora provide insights into the molecular mechanism of shell formation and immunomodulation. Sci. Rep. 8, 17664.
Marie, B., Arivalagan, J., Mathéron, L., Bolbach, G., Berland, S., Marie, A., Marin, F., 2017. Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the *Unionoida Elliptio complanata* and *Villosa lienosa*. J. R. Soc. Interface 14. https://doi.org/10.1098/rsif.2016.0846.
Marie, B., Jackson, D.J., Ramos-Silva, P., Zanella-Cléon, I., Guichard, N., Marin, F., 2013a. The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties. FEBS J. 280, 214–232.
Marie, B., Joubert, C., Tayalé, A., Zanella-Cléon, I., Belliard, C., Piquemal, D., Cochennec-Laureau, N., Marin, F., Gueguen, Y., Montagnani, C., 2012. Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. Proc. Natl. Acad. Sci. U. S. A. 109, 20986–20991.
Marie, B., Le Roy, N., Marie, A., Dubost, L., Marin, F., 2009a. Nacre Evolution: A Proteomic Approach. MRS Online Proceeding Library Archive 1187. https://doi.org/10.1557/PROC-1187-KK01-03.
Marie, B., Le Roy, N., Zanella-Cléon, I., Becchi, M., Marin, F., 2011. Molecular evolution of mollusc shell proteins: insights from proteomic analysis of the edible mussel *Mytilus*. J. Mol. Evol. 72, 531–546.
Marie, B., Luquet, G., Pais De Barros, J.-P., Guichard, N., Morel, S., Alcaraz, G., Bollache, L., Marin, F., 2007. The shell matrix of the freshwater mussel *Unio pictorum* (Paleoheterodonta, Unionoida). Involvement of acidic polysaccharides from glycoproteins in nacre mineralization. FEBS J. 274, 2933–2945.
Marie, B., Marin, F., Marie, A., Bédouet, L., Dubost, L., Alcaraz, G., Milet, C., Luquet, G., 2009b. Evolution of nacre: biochemistry and proteomics of the shell organic matrix of the cephalopod *Nautilus macromphalus*. Chembiochem 10, 1495–1506.
Marie, B., Ramos-Silva, P., Marin, F., Marie, A., 2013b. Proteomics of CaCO3 biomineral-associated proteins: how to properly address their analysis. Proteomics 13, 3109–3116.
Marin, F., Amons, R., Guichard, N., Stigter, M., Hecker, A., Luquet, G., Layrolle, P., Alcaraz, G., Riondet, C., Westbroek, P., 2005. Caspartin and calprismin, two proteins of the shell calcitic prisms of the Mediterranean fan mussel *Pinna nobilis*. J. Biol. Chem. 280, 33895–33908.
Marin, F., Bundeleva, I., Takeuchi, T., Immel, F., Medakovic, D., 2016. Organic matrices in metazoan calcium carbonate skeletons: Composition, functions, evolution. J. Struct. Biol. 196, 98–106.
Marin, F., Le Roy, N., Marie, B., 2012. The formation and mineralization of mollusk shell. Front. Biosci. 4, 1099–1125.
Marin, F., Le Roy, N., Marie, B., Ramos-Silva, P., Wolf, S., Benhamada, S., Guichard, N., Immel, F., 2014. Synthesis of calcium carbonate biological materials: how many proteins are needed? Key Eng, Mater, pp. 614.
Marin, F., Luquet, G., Marie, B., Medakovic, D., 2007. Molluscan Shell Proteins: Primary Structure, Origin, and Evolution, in: Current Topics in Developmental Biology. Academic Press, pp. 209–276.
Marin, F., Marie, B., Hamada, S.B., Ramos-Silva, P., Le Roy, N., Guichard, N., Wolf, S.E., Montagnani, C., Joubert, C., Piquemal, D., Saulnier, D., Gueguen, Y., 2013. "Shellome": Proteins involved in mollusk shell biomineralization-diversity, functions, in: Watabe, S., Maeyama, K., Nagasawa, H. (Eds.), . Terrapub Tokyo, p. 149:166.
Matsudaira, P., 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. 262, 10035–10038.
Matsumoto, M., 2003. Phylogenetic analysis of the subclass Pteriomorphia (Bivalvia) from mtDNA COI sequences. Mol. Phylogenet. Evol. 27, 429–440.
Matsumoto, M., Hayami, I., 2000. Phylogenetic analysis of the family Pectinidae (Bivalvia) based on mitochondrial cytochrome c oxidase subunit I. J. Molluscan Stud. 66, 477–488.
McDougall, C., Aguilera, F., Degnan, B.M., 2013. Rapid evolution of pearl oyster shell matrix proteins with repetitive, low-complexity domains. J. R. Soc. Interface 10, 20130041.
Mészáros, B., Erdos, G., Dosztányi, Z., 2018. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 46, W329–W337.
Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.-Y., El-Gebali, S., Fraser, M.I., Gough, J., Haft, D.R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A.P., Paysan-Lafosse, T., Pesseat, S., Potter, S.C., Qureshi, M.A., Rawlings, N.D., Redaschi, N., Richardson, L.J., Rivoire, C., Salazar, G.A., Sangrador-Vegas, A., Sigrist, C.J.A., Sillitoe, I., Sutton, G.G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Yong, S.-Y., Finn, R.D., 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 47, D351–D360.
Moshkovitz, S., 1971. Neolithic *Spondylus*. Nature.
NCBI [WWW Document], n.d. URL https://www.ncbi.nlm.nih.gov/protein/?term=%22molluscs%22%5Borgn%3A_txid6447%5D (accessed 12.19.18).
Nishida, K., Ishimura, T., Suzuki, A., Sasaki, T., 2012. Seasonal changes in the shell microstructure of the bloody clam, *Scapharca broughtonii* (Mollusca: Bivalvia: Arcidae). Palaeogeogr. Palaeoclimatol. Palaeoecol. 363–364, 99–108.
Nishida, K., Suzuki, A., Isono, R., Hayashi, M., Watanabe, Y., Yamamoto, Y., Irie, T., Nojiri, Y., Mori, C., Sato, M., Sato, K., Sasaki, T., 2015. Thermal dependency of shell growth, microstructure, and stable isotopes in laboratory-reared *Scapharca broughtonii* (Mollusca: Bivalvia): Thermal control on shell microstructure. Geochem. Geophys. Geosyst. 16, 2395–2408.
Osuna-Mascaró, A., Cruz-Bustos, T., Benhamada, S., Guichard, N., Marie, B., Plasseraud, L., Corneillat, M., Alcaraz, G., Checa, A., Marin, F., 2014. The shell organic matrix of the crossed lamellar queen conch shell (*Strombus gigas*). Comp. Biochem. Physiol. B Biochem. Mol. Biol. 168, 76–85.
Palmer, A.R., 1992. Calcification in marine molluscs: how costly is it? Proc. Natl. Acad. Sci. U. S. A. 89, 1379–1382.
Parker, A., Immel, F., Guichard, N., Broussard, C., Marin, F., 2015. Thermal stability of nacre proteins of the Polynesian oyster: a proteomic study. Key Engineer. Mater. 672, 222–231.
Paulsen, A.C., 1974. The Thorny Oyster and the Voice of God: *Spondylus* and *Strombus* in Andean Prehistory. Am. Antiq. 39, 597–607.
Pavat, C., Zanella-Cléon, I., Becchi, M., Medakovic, D., Luquet, G., Guichard, N., Alcaraz, G., Dommergues, J.-L., Serpentini, A., Lebel, J.-M., Marin, F., 2012. The shell matrix of the pulmonate land snail *Helix aspersa maxima*. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 161, 303–314.
Penkman, K.E.H., Kaufman, D.S., Maddy, D., Collins, M.J., 2008. Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells. Quat. Geochronol. 3, 2–25.
Pierini, F., Demarchi, B., Turner, J., Penkman, K., 2016. Pecten as a new substrate for IcPD dating: the Quaternary raised beaches in the Gulf of Corinth. Greece. Quat. Geochronol. 31, 40–52.
Pillsbury, J., 1996. The Thorny Oyster and the Origins of Empire: Implications of Recently Uncovered *Spondylus* Imagery from Chan Chan, Peru. Latin American Antiquity 7, 313–340.
Puslednik, L., Serb, J.M., 2008. Molecular phylogenetics of the Pectinidae (Mollusca: Bivalvia) and effect of increased taxon sampling and outgroup selection on tree topology. Mol. Phylogenet. Evol. 48, 1178–1188.
Ramos-Silva, P., Benhamada, S., Le Roy, N., Marie, B., Guichard, N., Zanella-Cléon, I., Plasseraud, L., Corneillat, M., Alcaraz, G., Kaandorp, J., Marin, F., 2012. Novel molluskan biomineralization proteins retrieved from proteomics: a case study with Upsalin. Chembiochem 13, 1067–1078.
Sakalauskaite, J., Andersen, S.H., Biagi, P., Borrello, M.A., Cocquerez, T., Colonese, A.C., Dal Bello, F., Girod, A., Heumüller, M., Koon, H., Mandili, G., Medana, C., Penkman, K.E., Plasseraud, L., Schlichtherle, H., Taylor, S., Tokarski, C., Thomas, J., Wilson, J., Marin, F., Demarchi, B., 2019. "Palaeoshellomics" reveals the use of freshwater mother-of-pearl in prehistory. Elife 8. https://doi.org/10.7554/eLife.45644.
Samata, T., Ikeda, D., Kajikawa, A., Sato, H., Nogawa, C., Yamada, D., Yamazaki, R., Akiyama, T., 2008. A novel phosphorylated glycoprotein in the shell matrix of the oyster *Crassostrea nippona*: A novel acidic glycoprotein from the oyster shells. FEBS J. 275, 2977–2989.
Sarashina, I., Endo, K., 2006. Skeletal matrix proteins of invertebrate animals: Comparative analysis of their amino acid sequences. Paleontol. Res. 10, 311–336.
Sarashina, I., Endo, K., 2001. The complete primary structure of molluscan shell protein 1 (MSP-1), an acidic glycoprotein in the shell matrix of the scallop *Patinopecten yessoensis*. Mar. Biotechnol. 3, 362–369.
Shackleton, J., Elderfield, H., 1990. Strontium isotope dating of the source of Neolithic European *Spondylus* shell artefacts. Antiquity 64, 312–315.
Sleight, V.A., Marie, B., Jackson, D.J., Dyrynda, E.A., Marie, A., Clark, M.S., 2016. An Antarctic molluscan biomineralisation tool-kit. Sci. Rep. 6, 36978.
Smedley, G.D., Audino, J.A., Grula, C., Porath-Krause, A., Pairett, A.N., Alejandrino, A., Lacey, L., Masters, F., Duncan, P.F., Strong, E.E., Serb, J.M., 2019. Molecular phylogeny of the Pectinoidea (Bivalvia) indicates Propeamussiidae to be a non-monophyletic family with one clade sister to the scallops (Pectinidae). Mol. Phylogenet. Evol. 137, 293–299.

Suzuki, M., Iwashima, A., Tsutsui, N., Ohira, T., Kogure, T., Nagasawa, H., 2011. Identification and characterisation of a calcium carbonate-binding protein, blue mussel shell protein (BMSP), from the nacreous layer. Chembiochem 12, 2478–2487.
Takeuchi, T., Plasseraud, L., Ziegler-Devin, I., Brosse, N., Shinzato, C., Satoh, N., Marin, F., 2018. Biochemical characterization of the skeletal matrix of the massive coral, *Porites australiensis* - The saccharide moieties and their localization. J. Struct. Biol. 203, 219–229.
Taylor, J.D., Kennedy, W.J., Hall, A., 1973. The shell structure and mineralogy of the Bivalvia. II. Lucinacea-Clavagellacea, Conclusions, 9. The British Museum (Natural History), London.
Taylor, J.D., Kennedy, W.J., Hall, A., 1969. The shell structure and mineralogy of Bivalvia: introduction, *Nuculacea-Trigonacea*, Bulletin of the British Museum (Natural History), Zoology. Supplement, The British Museum (Natural History), London.
Taylor, J.D., Reid, D.G., 1990. Shell microstructure and mineralogy of the Littorinidae: ecological and evolutionary significance. Hydrobiologia 193, 199–215.
Vizcaíno, J.A., Côté, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., Ríos, D., Wang, R., Hermjakob, H., 2013. The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 41, D1063–D1069.
Waller, T.R., 2006. Phylogeny of families in the Pectinoidea (Mollusca: Bivalvia): importance of the fossil record. Zool. J. Linn. Soc. 148, 313–342.
Wang, S., Zhang, J., Jiao, W., Li, J., Xun, X., Sun, Y., Guo, X., Huan, P., Dong, B., Zhang, L., Hu, X., Sun, X., Wang, J., Zhao, C., Wang, Y., Wang, D., Huang, X., Wang, R., Lv, J., Li, Y., Zhang, Z., Liu, B., Lu, W., Hui, Y., Liang, J., Zhou, Z., Hou, R., Li, X., Liu, Y., Li, H., Ning, X., Lin, Y., Zhao, L., Xing, Q., Dou, J., Li, Y., Mao, J., Guo, H., Dou, H., Li, T., Mu, C., Jiang, W., Fu, Q., Fu, X., Miao, Y., Liu, J., Yu, Q., Li, R., Liao, H., Li, X., Kong, Y., Jiang, Z., Chourrout, D., Li, R., Bao, Z., 2017. Scallop genome provides insights into evolution of bilaterian karyotype and development. Nat Ecol Evol 1, 120.
Watabe, S., Maeyama, K., Nagasawa, H. (Eds.), 2011. Recent Advances in Pearl Research—Proceedings of the International Symposium on Pearl Research 2011. Terrapub, Tokyo, Japan.
Wilmot, N.V., Barber, D.J., Taylor, J.D., Graham, A.L., 1992. Electron microscopy of molluscan crossed-lamellar microstructure. Philos. Trans. R. Soc. Lond. B Biol. Sci. 337, 21–35.
Windler, A., 2017. From the Aegean Sea to the Parisian Basin: Spondylus shell exchange in Europe during the process of Neolithisation, in: Eisenach, P., Stöllner, T., Windler, A. (Eds.), The RITaK Conferences 2013–2014. Bochum, pp. 95–110.
Yarra, T., Gharbi, K., Blaxter, M., Peck, L.S., Clark, M.S., 2016. Characterization of the mantle transcriptome in bivalves: *Pecten maximus*, *Mytilus edulis* and *Crassostrea gigas*. Mar. Genomics 27, 9–15.
Yonezawa, M., Sakuda, S., Yoshimura, E., Suzuki, M., 2016. Molecular cloning and functional analysis of chitinases in the fresh water snail, *Lymnaea stagnalis*. J. Struct. Biol. 196, 107–118.
Yonge, M., 1973. Functional morphology with particular reference to hinge and ligament in *Spondylus* and *Plicatula* and a discussion on relations within the superfamily Pertinacea (Mollusca: Bivalvia). Philos. Trans. R. Soc. Lond. B Biol. Sci. 267, 173–208.
Yue, X., Zhang, S., Yu, J., Liu, B., 2019. Identification of a laccase gene involved in shell periostracal tanning of the clam *Meretrix petechialis*. Aquat. Biol. 28, 55–65.

Chapter 2

The diagenesis of intracrystalline *Spondylus* shell proteins: a proteomics study

Digest

Mollusc shells possess good protein preservation potential in the sub-fossil record and constitute fascinating biological systems to study how proteins degrade in a biomineral – a process called shell protein diagenesis. Understanding shell proteins ageing processes and the mechanisms of their degradation may help us to interpret fossil sequences, authenticate ancient proteins from prehistoric samples (*e.g.* archaeological *Spondylus* shell ornaments) and provide a "calibrated" molecular clock for protein dating. The main protein degradation reactions are amino acid racemisation, decomposition, modification (*e.g.* deamidation) and peptide bond hydrolysis. Generally speaking, the diagenesis reactions are known, but the order and pathways according to which they occur are poorly understood. Information is particularly limited on the patterns of shell protein hydrolysis, which is often the key reaction initiating diagenesis.

In this work, I have used immunochemistry and quantitative proteomics to study how intracrystalline *Spondylus* shell proteins degrade. I have used accelerated aging experiments to simulate the diagenesis by heating shell powders at four different temperatures (80, 95, 110, 140 °C) and for different durations. The samples were characterised by immunochemistry and tandem mass tag (TMT) labelled quantitative proteomics, which allowed to track both matrix structure loss (based on reactive carbohydrate and protein groups) and peptide bond hydrolysis.

The study showed that structural degradation occurs together with hydrolysis, although the former reaction appeared to be slightly faster. Immunochemistry and proteomics revealed that mild heating does not induce instant diagenesis, but actually contributes to the uncoiling (=denaturing) of the matrix structure and helps to better detect carbohydrate groups and peptides. I have also identified a number of peptide bonds that were prone to hydrolysis. In the literature, some of these bonds are however reported as stable, indicating that there is not a simple correlation between a theoretical protein stability in solution and in a biomineralised system. This work, therefore, highlights the complexity of the intracrystalline *Spondylus* shell matrix proteins and suggests that the overall microchemical environment plays an important role in mediating protein degradation.

It is important to note that the diagenesis patterns obtained in this study provide only a partial picture of all the degradation processes, as a consequence of limited molecular data for *Spondylus* sp. and phylogenetically close molluscs. As *Spondylus* transcriptome becomes available (which is an ongoing line of research, as detailed in the digest section of the previous chapter), it will become possible to identify additional hydrolysis patterns. I would like to highlight, that we do not consider that the data are lost without the transcriptome – for the moment, it remains "undiscovered" and as soon as we have more proteomics data for this species, the data will be re-examined using bioinformatic analysis, simply *in silico*.

In spite of some limitations, this work is among the first attempts to study shell protein diagenesis by simultaneously tracking matrix structure loss and peptide bond hydrolysis using TMT labelled quantitative proteomics. The results will be of great interest in palaeoproteomics research, *e.g.* the study of shell proteins from sub-fossil and fossil samples; it may also have an impact in biomineralization and biomimetic research. Finally, this work also indicates that *Spondylus* shell is a robust system for preserving intracrystalline proteins and therefore archaeological *Spondylus* artifacts can be analysed and interpreted based on their molecular signatures.

Provisional list of authors:
Jorune Sakalauskaite[1,2], Meaghan Mackie[3,4], Matthew Collins[3,5], Frédéric Marin[2], Beatrice Demarchi[1]

Affiliations
1. Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Turin, Italy;
2. Biogéosciences, UMR CNRS 6282, University of Burgundy-Franche-Comté, 6 Boulevard Gabriel, 21000 Dijon, France.
3. Section for Evolutionary Genomics, GLOBE Institute, Faculty of Health and Medical Science, University of Copenhagen, Øster Farimagsgade 5, 1353 Copenhagen, Denmark.
4. Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Blegdamsvej 3b, 2200 Copenhagen N, Denmark.
5. McDonald Institute for Archaeological Research, University of Cambridge, West Tower, Downing St, CB2 3ER, Cambridge, UK.

To be submitted to: Organic Geochemistry, Geochimica Cosmochimica Acta

Abstract

The *post-mortem* breakdown of intracrystalline mollusc shell proteins has been extensively investigated, particularly with regard to its use as a “molecular clock” for dating fossil substrates, since this fraction of proteins represents an excellent system to retrieve genuine biomolecules from archaeological/palaeontological samples. Despite numerous ancient protein studies, the fundamental aspects on how protein structures and sequences are affected by diagenesis are far from being understood. In this study we investigate the degradation of intracrystalline *Spondylus* shell proteins by performing artificial aging experiments and by using immunochemistry and quantitative proteomic techniques to determine patterns of structural loss and peptide bond hydrolysis. *Spondylus* was selected due to its importance in archaeological research. Powdered and bleached shell samples were subjected to accelerated aging by heating in water at four different temperatures – 80, 95, 110, 140 °C for diverse time durations. The structural loss of carbohydrate and protein groups was tracked by enzyme-linked lectin assay (ELLA) and enzyme-linked immunosorbent assay (ELISA). Tandem mass tag (TMT) quantitative proteomics was performed to track changes in peptide/protein

relative abundances over time and the data were used to investigate peptide bond hydrolysis patterns.

The results obtained in this study confirm the overall complexity of intracrystalline shell matrix degradation and add new knowledge on the subtle interplay between diagenetic processes affecting macromolecular structures and those which affect sequences. More specifically, we find that heating does not induce instant organic matrix decay, but firstly aids in uncoiling crossed-linked structures and enhances matrix detection. Furthermore, the calculated apparent activation energies of the loss of the immunochemical signal, that is E_a (carbohydrate groups) = 112.36 kJ/mol, E_a (protein groups) = 110.62 kJ/mol, suggest that secondary matrix structure degradation may go hand-in-hand with protein hydrolysis. We also find that proteins do not undergo complete hydrolysis after prolonged heating (10 days at 110 °C), whereas the structural signal is not further observed. Eight peptide bonds that were prone to hydrolysis were identified in the sequence of the protein displaying the best identification scores (uncharacterised protein LOC117318053 from *Pecten maximus*). These breakages involved unexpected pairs of amino acids, challenging the existence of a simple correlation to theoretical bond stabilities proposed in the literature. Our work presents the first approach to simultaneously track the structural and peptide bond diagenesis in shells, including the novel use of TMT quantitative proteomics. The data obtained in this study therefore represent a major stepping stone towards future palaeoproteomic studies on archaeological shells.

Introduction

The interest in studying ancient mollusc shell proteins dates back to the middle of the 20th century, when it became possible to analyse free amino acids following an acid hydrolysis preparation step (Abelson, 1954). There are numerous published works that sought to characterise million-year-old sequences (de Jong et al., 1974; Weiner et al., 1976; Westbroek et al., 1979) or to use amino acids for geochronological applications (Hare, 1968; Schroeder and Bada, 1976; Endo et al., 1995). Mollusc shell substrates display interesting features because they are compact, highly mineralised and possess a small fraction of intracrystalline organics that get occluded inside the mineral skeleton during shell formation (Lowenstam and Weiner, 1989; Albeck et al., 1993). These organics are often considered to behave like in a closed-system (Sykes et al., 1995), *i.e.* are protected from environmental interferences and can be isolated by extensive bleaching (Sykes et al., 1995; Penkman et al., 2008). Intracrystalline proteins provide a source of genuine biological information; for example, the analysis of their amino acid composition and extent of racemisation have been extensively studied as a reliable molecular clock, in determining the relative age of fossils (Collins and Riley, 2000; Penkman et al., 2008; Demarchi et al., 2013b; Pierini et al., 2016).

At proteomics level, a secure and confident identification of ancient sequences (Collins et al., 2006; and see a recent review in Welker, 2018) became possible only in the recent decade, with the advent of soft ionisation mass spectrometry techniques, most notably, the use of liquid chromatography coupled to electrospray ionisation (ESI) and tandem mass spectrometry (LC-ESI-MS/MS) (de Hoffmann and Stroobant, 2013). Palaeoproteomics, the analysis of ancient proteins by mass spectrometry, has enabled the identification of 3.8-million-year-old sequences from ostrich eggshell (OES) (Demarchi et al., 2016), as well as the characterisation of 1.77- and 1.9-million-years-old dental enamel proteomes (Welker et al., 2019; 2020). Recently, we have also been able to retrieve phylogenetic information from archaeological mollusc shell artifacts (Sakalauskaite et al., 2019). However, despite the growing number of ancient protein studies, the mechanisms of protein preservation and degradation in biominerals, including mollusc shells, are still poorly understood (Marin et al., 2007; Demarchi et al., 2013c; 2016; Buckley and Wadsworth, 2014; Parker et al., 2015).

Diagenesis is the sum of biological, chemical and physical processes that occur after the death of the organism. In shells, and in general in all biominerals, diagenesis is complex, governed by a number of different physico-chemical reactions that occur in the mineral and organic phases, interdependently (Hare and

Hoering, 1977; Demarchi, 2020). Key factors include: pH and temperature conditions of the burial environment; age of the sample; mineralogy and structure (calcite vs aragonite arranged into nacreous, prismatic, foliated, crossed-lamellar or other microstructures) (Milano and Nehrke, 2018); the composition of the shell proteome and the presence of other macromolecular compounds, such as saccharides, lipids (Collins et al., 1992), as well as the presence of water (linked or not to the carbonate phase). There are two main approaches to study molecular diagenesis in shells – it can be done by using independently dated shells (*e.g.* by radiocarbon) or by artificial diagenesis experiments, *i.e.* heating modern shells at high temperatures (Collins, 2000; Penkman et al., 2008; Crisp et al., 2013; Demarchi et al., 2013a; 2016). While degradation induced by laboratory conditions does not reflect the slow aging process occurring naturally (Tomiak et al., 2013), such experiments provide however a good base for understanding some of the diagenetic pathways, and have been exploited by a number of different kinetic studies (*e.g.* Penkman et al., 2008; Crisp et al., 2013; Demarchi et al., 2013a; 2016).

The main protein diagenesis reactions are hydrolysis, amino acid racemisation and decomposition as well as diagenesis-induced modifications. Decomposition refers generally to the decay of individual amino acids. Some of the most well investigated diagenesis-induced modifications are the deamidation of asparagine (Asn) or glutamine (Gln) amide side chains (Robinson and Robinson, 2004) as well as oxidation of tryptophan (Trp) and methionine (Met) (for a recent review see Demarchi, 2020). The rate of Asn/Gln deamidation can be used as a molecular clock (Robinson and Robinson, 2001; 2004), and a proxy to date and authenticate ancient sequences, often used for collagen-based samples (Wilson et al., 2012; Demarchi et al., 2016; Schroeter and Cleland, 2016; Procopio and Buckley, 2017; Ramsøe et al., 2020). The racemisation of amino acids describes the post-mortem spontaneous interconversion between the D- and L-chiral forms of amino acids (Hare, 1968; Schroeder and Bada, 1976). The extent of amino acid racemisation (AAR) rate determined on “fossil” intracrystalline mollusc shell proteins (Demarchi et al., 2011; 2013a) can be used as a relative geochronological tool (Penkman et al., 2011; Bosch et al., 2015; Demarchi et al., 2015; Pierini et al., 2016). Protein hydrolysis is the breakdown of peptide bonds into constituting smaller fragments in the presence of water molecules (Hill, 1965; Qian et al., 1993; Demarchi, 2020). This process is particularly interesting for palaeoproteomic research but, at the same time, is perplexingly difficult to study (Hare, 1971; Mitterer, 1993).

Modelling protein hydrolysis in laboratory conditions, particularly in biological systems, is difficult because the degradation of the same protein sequence differs significantly when compared in water solution and inside the biomineral (Demarchi, 2020). The degradation studies of pure synthetic peptides in aqueous solutions

showed the advantage of having a simplified and well controlled system, enabling an in-depth characterisation of peptide fragmentation patterns (Demarchi et al., 2013c). However, these studies fail to reflect protein behaviour in a biomineralized system due to the presence of structural factors, such as protein-mineral interactions (Demarchi et al., 2016). For example, a synthetic struthiocalcin (the main protein of ostrich eggshell – OES) peptide, when artificially aged in water solution, showed that the single Ser residues investigated displayed racemisation rates 6 times faster than those calculated for bulk Ser within the OES biomineral (Demarchi et al., 2013c). Similarly, a struthiocalcin peptide which was found to be preserved in several 3.8 million-year-old OES fragments, when heated in water solution, showed fast degradation rates and was not found to be particularly stable (Demarchi et al., 2016). There is also a limited knowledge on the presence and role of other (non proteinaceous) organic components, *e.g.* glycosides. These may also play an active role in diagenesis by causing matrix cross-linking with sugar components, a process known as the Maillard reaction (Maillard, 1912; Collins et al., 1992). Proteomics studies targeting biomineralizing shell proteins could help us to gain a better understand of hydrolysis processes (Parker et al., 2015) and provide an extremely useful molecular tool for the direct dating of ancient biological materials in the future.

In this work we aim to investigate the diagenesis process of intracrystalline shell matrix proteins at structural and proteomic level by employing immunochemistry and quantitative proteomics. We have chosen *Spondylus gaederopus* as a model system due to its importance and prevalence in archaeological research as underlined in part two of the literature review (Borrello and Micheli, 2011; Ifantidis and Nikolaidou, 2011; Chapman and Gaydarska, 2015; Windler, 2017). We specifically targeted to investigate intracrystalline shell proteins, that are obtained via extensive bleaching step. First of all, intracrystalline proteins provide a ‘simplified’ system to study patterns of protein degradation and what is the effect of the mineral binding activity. Secondly, this treatment is especially useful for the study of degraded or sub-fossil samples. Indeed, in many shell models, intracrystalline proteins show a “closed-system” behaviour, as demonstrated in the studies on gastropods and bivalves, *e.g.* by Penkman et al., 2008, Demarchi et al., 2013; Pierini et al., 2016. In particular, it has been shown that “truly” intracrystalline proteins do not ‘leach’ from the system in ‘leaching experiments’ carried out at high temperatures in aqueous environments and consequently these proteins exhibit the highest potential for being retained (intact or degraded) in the system. This enables the effective study of protein diagenesis in biomineral systems. On the contrary, intercrystalline sequences are affected by the action of percolating water, which accelerates protein hydrolysis, and bacteria or other microbial agents, which rapidly and unpredictably degrade organic matter in this “open-system” environment

(Demarchi, 2020). Considering future proteomic analysis of sub-fossil samples, it is important to note that ancient shells are particularly susceptible to environmental effects (*e.g.* bacterial proteins migrating from the soil or due to natural contaminants of living molluscs), hence the bleaching step ensures that only endogenous mollusc shell proteins are analysed.

Therefore, in this study, we performed accelerated diagenesis experiments by subjecting powdered and bleached *Spondylus* samples to heating at four different temperatures – 80, 95, 110, 140 °C and for various time durations, ranging from several hours and even up to two hundred days (see Figure 1 for details on the experimental design). The heated shell matrices were extracted and characterised by immunochemistry and one heated set was analysed by quantitative proteomics using isobaric tag labelling (TMT – tandem mass tag). The TMT quantitative proteomics approach is based on labelling the samples heated at multiple time-points with different tags. The samples are then mixed together and measured in a single run using high performance LC-MS/MS (Erdjument-Bromage et al., 2018). This provides temporally resolved measurements of peptide (protein) abundance in time and has also been used to study protein degradation kinetics in other biological tissues (Welle et al., 2016).

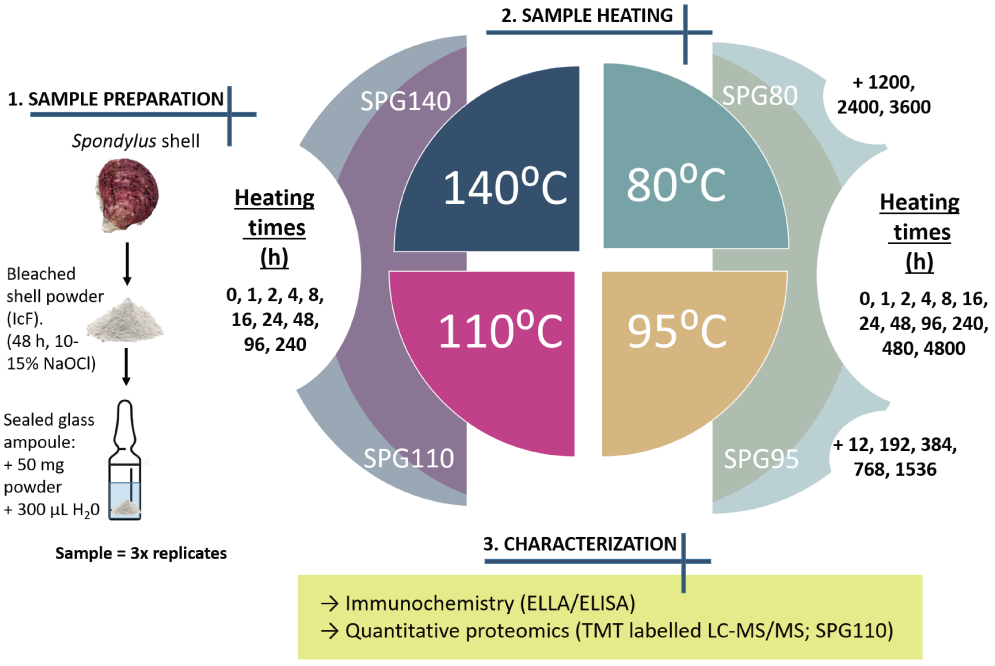


Figure 1. Experimental design of the methodology used in this study. The scheme displays the main steps of sample preparation, heating and matrix characterisation. Powdered shell

samples were bleached with concentrated NaOCl (10-15%) for 48h to isolate the intracrystalline fraction of organics (IcF). The powdered samples (50 mg) were added to the glass vials, covered with 300 µL of water, sealed and heated at different temperatures and time lengths – sets SPG80, SPG95, SPG110, SPG140. Each sample was prepared in triplicate. The samples were characterised by the immunochemical techniques (ELLA/ELISA) and the set SPG110 was analysed by quantitative proteomics.

Materials and Methods

Set up of the heating experiments and sample preparation

The modern *Spondylus gaederopus* shell used in this study was available as a finely ground powder (particle size < 200 µm). It was previously prepared by cleaning the crushed shell with diluted bleach (1.0-1.5% NaOCl, 4 h of soaking) (Sakalauskaite et al., 2020b). The intracrystalline fraction (IcF) was isolated by further bleaching the shell powder for 48 h with concentrated NaOCl (10-15%; 1:50 w/v of powder/bleach ratio). The bleach solution was centrifuged, the powder was cleaned with water (HPLC-grade, performing 5x washes) and rinsed with ethanol (HPLC-grade, 1x wash) before being air dried. The glass ampules (Wheaton, 1 mL, clear) were also cleaned by bleaching in concentrated NaOCl (10-15%) for 48 h, rinsed with H_2O (HPLC-grade) and dried in the oven. 50 mg of *Spondylus* powder was weighed and placed into glass ampoules, filled with 300 µL of H_2O (HPLC-grade) and flame-sealed. For blank samples, only water was added. The samples were placed in the oven and heated at four different temperatures: 80, 95, 110, 140 °C and for different times (Figure 1 and SI table 1 in Appendix 3). Three replicates were prepared for all the time points. After heating, the samples were cooled down at room temperature and stored at -20 °C. The samples were defrosted at room temperature to separate the supernatant solutions. The powders were air-dried and weighed into separate eppendorfs: 5 mg aliquots were set aside for immunochemical analysis (two replicates were prepared for each time point and temperature) and 35 mg of powders were used for proteomics. Full list of samples and the analyses that were carried out are presented in the Appendix 3, SI table 1.

Immunochemistry

The enzyme-linked lectin (ELLA) and enzyme-linked immunosorbent (ELISA) assays were carried out to characterise the EDTA-soluble extracts of the shell samples heated at different temperatures and times. We also analysed the shell samples that correspond to (inter+intra) crystalline fraction (IIF, obtained by extracting the organics from the shell powder that was not subjected to 48-h bleaching treatment) in order to compare the IIF and the IcF matrices. Lectin jacalin

and shell antibody K5090 were used for ELLA and ELISA respectively, as they were previously shown to have a strong cross reactivity with *Spondylus* shell organic matrix (Sakalauskaite et al., 2020b). Lectin jacalin is a carbohydrate-binding protein that recognises galactose/N-acetylgalactosamine residues and/or oligosaccharides terminating with D-galactose. Shell antibody K5090 was elicited against the soluble matrix of the nacreous layer of *Pinna nobilis* (Marin et al., 1994), however, the exact binding epitopes of this antibody are undetermined. 5 mg of shell powders were demineralised with 200 µL 20% EDTA solution (pH ~8.2) at room temperature, under constant agitation (1300 rpm). Upon demineralisation, 200 µL of TBS were added to reach a total volume of 400 µL. 100 µL of the soluble extracts (2x of the aliquots were used for ELLA and 2x for ELISA plates) were incubated in a 96-well microplates (MaxiSorp, Nunc/Thermo Scientific, Nunc A/S, Roskilde, Denmark) for 90 min at 37 °C. The plates were washed thrice with TBS/Tween-20 and blocked for 30 min at 37 °C with Carbo-Free blocking solution (ref. SP 5040, Vector Laboratories) for the ELLA and 1% gelatin solution in TBS for the ELISA assay. Following this, either 100 µL solutions of lectin jacalin (Vector Laboratories, Peterborough, UK, Ref. BK-3000; diluted in TBS/Tween-20 with ratio of 1:100) or antibody K5090 (diluted in 0.2% gelatin TBS/Tween-20 with ratio of 1:50) were added. The plates were incubated for 90 min at 37 °C and afterwards washed with TBS/Tween-20. Either 100 µL of alkaline phosphatase-conjugated avidin solution (A7294 Sigma, 1:70000, ELLA) or 100 µL of GAR/AP solution (A3687 Sigma, 1:30000, ELISA) were added and left for another 90 min at 37 °C. Microplates were thoroughly washed with TBS/Tween-20 and incubated with aqueous diethanolamine solution (pH 9.8), containing phosphatase substrate (p-nitrophenylphosphate) at 37 °C. Optical density (OD) was read using a BioRad Model 680 microplate reader at 405 nm. Blank values were obtained by analysing samples with no added shell matrix. For calculations, background and blank absorptions were subtracted and for the heated shell samples (*i.e.* different SPG sets) the obtained intensity values were converted to the percentage of reactivity with the non-heated sample (time point 0 h) representing 100% reactivity. For the characterisation of IcF matrix, the percentage of reactivity was calculated in respect to IIF (OD values of IIF given 100%). In total, 104 heated shell matrix samples were analysed by immunochemistry: 30 samples from set SPG80, 34 samples from set SPG95, 20 samples from set SPG110, 20 samples from set SPG140 (see Appendix 3 SI table 1 for full list of samples).

Proteomics

Quantitative proteomics with tandem mass tag (TMT) labelling was used to analyse sample set SPG110, which represents one of the four temperature experiments carried out in this study (Figure 1). Eleven SPG110 samples were selected for TMT

labelling: nine samples that were heated at 110 °C for different times (1, 2, 4, 8, 16, 24, 48, 96, 240 h), one unheated sample (0 h) and one internal blank. The full sample list is as follows: spg01 (unheated sample), spg04, spg10, spg14, spg17, spg20, spg24, spg24, spg27, spg28 (internal blank), spg31, spg36; for more details, see also SI table 1).

Extraction

First of all, the proteins were extracted from all of the samples using the FASP sample preparation method (Sakalauskaite et al., 2020a). In short, shell powders were demineralised using cold acetic acid (10% v/v) which was added in 200 µL aliquots to a total volume of 800 µL. Demineralisation was carried out in cold conditions and with frequent agitation (~100 rpm). Demineralised extracts were centrifuged at 13.4k rpm for 10 min to separate acid soluble (ASM) and acid insoluble (AIM) matrices. The ASM extracts were loaded to PALL Nanosep centrifugal devices (3kDa, 0.5 mL), concentrated and desalted with water (HPLC-grade, 0.5 mL aliquots, 5x washes), and finally exchanged to buffer Ambic (50 mM ammonium bicarbonate, pH 7.5-8). The AIM extracts were rinsed with water (HPLC grade, 1.5 mL aliquots, 5x washes) and mixed with the ASM extracts. The protein 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, 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.

TMT Labelling

Dried peptides were sent to The Globe Institute, University of Copenhagen, Denmark for TMT labelling and tandem mass spectrometry analysis. The samples were resuspended in 50 µL 50% acetonitrile (ACN). Due to the error caused by low amounts of protein detectable by spectrophotometer analysis (at 280 nm and 205 nm), the samples were normalised by the starting weight of the shell used. Therefore, a portion of the sample (34-40 µl) was taken forward for TMT labelling. HEPES buffer and ACN were then added for a total concentration of 50% ACN and 30 mM HEPES, with the pH checked to be around 8. Thermo-Scientific TMT labels (11-plex) were prepared by resuspending in anhydrous ACN and then 0.02 mg of label was added to each sample, which was then vortexed and incubated at room temperature (covered) for 1 hour. The reaction was quenched by adding 1% hydroxylamine and incubated at room temperature for a further 15 min. The samples were pooled and then cleaned using an in-house made C18 StageTip (Rappsilber et al., 2007). Label details are as follows: 126 – spg01, 127N – spg04,

127C – spg10, 128N – spg14, 128C – spg17, 129N – spg20, 129C – spg24, 130N – spg27, 130C – spg28, 131 – spg31, 131C – spg36.

Mass Spectrometry (MS)

The StageTip was eluted using 20 µl each 40% then 60% ACN and vacuum centrifuged at 40 °C until approximately 3 µl remained. It was then resuspended with 10 µL 0.1% trifluoroacetic acid (TFA) and 5% ACN solution. 5 µL was then analysed by an EASY-nLC 1200 (Thermo Fisher Scientific, Bremen, Germany) coupled to a Q-Exactive HF-X orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) on a 77 min gradient. Chromatographic and MS parameters were then performed based on previously published methods for ancient and degraded samples (Mackie et al., 2018), adjusted for TMT analysis. Therefore, the isolation window was narrowed to 0.8 *m/z* and the normalised collision energy raised to 33. Other MS parameters were set as follows: MS1- 120,000 resolution at *m/z* 200 over the *m/z* range 350–1400, target of 3e6, maximum injection time (IT) of 25 ms; MS2- top 10 mode, 60,000 resolution, target of 2e5, maximum IT of 118 ms, and dynamic exclusion of 20 s.

Bioinformatics

Bioinformatic analyses were performed using PEAKS Studio X software (Bioinformatic Solutions Inc (Ma et al., 2003)). A “Mollusca protein” database was created on 10/06/2020 by downloading all the sequences from the NCBI protein repository with taxonomy restricted to phylum Mollusca. It was used to search the product ion spectra obtained by the *de novo* sequencing of the PeaksX algorithm. Search parameters were selected as follows: fragment ion mass tolerance of 0.05 Da and a parent ion tolerance of 10 ppm and no enzyme digestion was selected to detect diagenetically cleaved peptides. Results obtained by SPIDER search (*i.e.* including all possible modifications) were used for peptide identification and protein characterisation. The thresholds for peptide and protein identification were set as follows: false discovery rate (protein FDR) = 0.5%, protein score −10lgP ≥ 30, unique peptides ≥ 1, *de novo* sequences scores (ALC%) ≥ 50. The search also included a database of common laboratory contaminants (cRAP; common Repository of Adventitious Proteins: http://www.thegpm.org/crap/), which were excluded from further data interpretation. The peptide sequences identified in shell proteins were also individually checked using BLASTp tool (https://blast.ncbi.nlm.nih.gov/) to prevent any misidentifications with exogenous sequences and peptides that were homologous to common contaminants or bacterial proteins were excluded from further analyses. TMT quantification is performed by measuring the intensities of fragment ion reporter ions released from the different labels in the tandem MS mode during peptide fragmentation.

Quantitation was performed by PeaksX using the selected parameters: quantification mass tolerance: 0.2 Da, FDR Threshold (%): 1.0, spectrum filter ≥ 31.9, quality ≥ 0, reporter ion intensity ≥ 0E0, Protein significance ≥ 0, significance method – PEAKSQ, Unique peptides ≥ 1. Relative quantification of peptides/proteins in each of the samples was obtained by calculating the relative reporter ion intensity in respect to values found in blank samples (spg28, tagged with TMT11-130C label). The data were used to create relative abundance profiles of intracrystalline shell proteins and peptides, represented by bar plots and line graphs that show abundance in real time scale (presented in Table 3, 4 and Figure 7).

Results

Intracrystalline shell matrix. Biochemistry and proteome characterization.

In this study we have characterized the biochemical and proteome composition of *Spondylus gaederopus* intracrystalline shell matrix. The IcF *Spondylus* shell proteins were analysed by LC-MS/MS (Table 1). For biochemical characterisation, ELLA and ELISA immunochemistry assays were employed to assess the cross-reactivity of the intracrystalline shell matrix fraction (IcF) with lectin jacalin and with shell antibody K5090 (Figure 2). In *Spondylus*, jacalin was previously found to bind a 30-34 kDa protein as well as insoluble compounds of high molecular weight (Sakalauskaite et al., 2020b). Shell antibody K5090 cross-reacts with proteinaceous components of *Spondylus* organic matrix and in western blots it was shown to bind ~50 kDa molecular weight proteins as well as some weakly soluble and high molecular weight components (for more details see Appendix 3, SI material 1).

With regard to the previously published biomolecular data on *Spondylus* (Sakalauskaite et al., 2020b) we present two main updates: i) first of all, in this work we specifically sought to characterise the mineral-bound organics, *i.e.* the operational intracrystalline fraction (IcF), whereas in the previous shellomics study, milder bleaching treatments were used (48 h bleaching step with ~10-15% NaOCl vs 4-14 h with ~1.0-1.5% NaOCl); ii) secondly, the proteomics identification was done using an updated protein database in which more than 200000 molluscan protein sequences were included (database as of June, 2020 vs that of December, 2018).

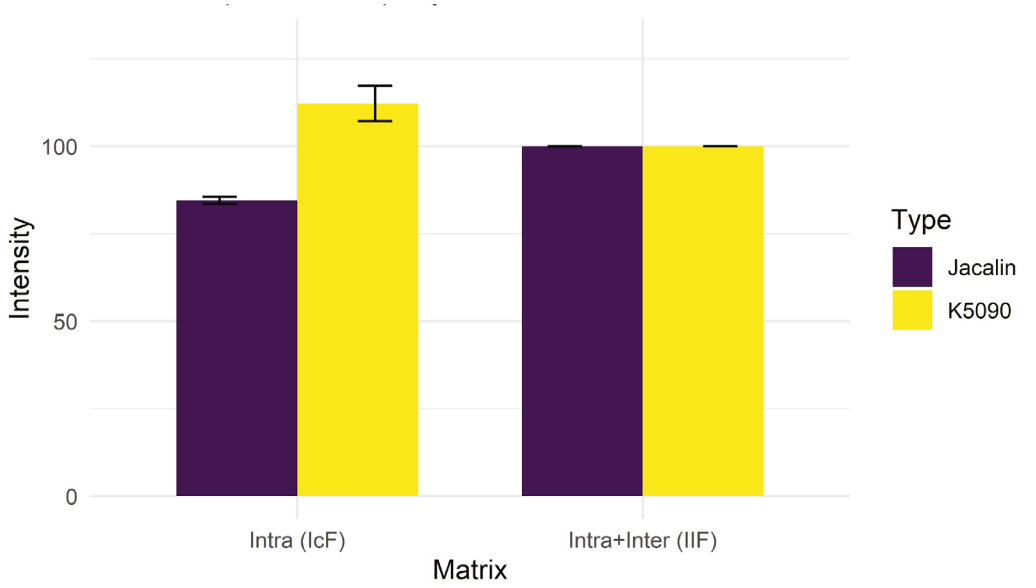


Figure 2. ELLA and ELISA assays, which compare the signal intensity between the two *Spondylus* shell matrices – intracrystalline (IcF) and (inter+intra)crystalline (IIF). Lectin jacalin was used for ELLA and antibody K5090 for ELISA. The IcF matrix is obtained via 48-h bleaching of fine shell powder using concentrated NaOCl (~10-15%). The values on the y axis correspond to absorbance at 405 nm and the relative intensity is calculated considering the reactivity of IIF matrix to be 100% (n = 4 (extractions), means ± S.D).

Figure 2 shows that the cross-reactivity of IcF matrix obtained after 48 h of bleaching does not display major changes with both lectin jacalin and antibody K5090. The IcF affinity to jacalin drops by ~15±1%, whereas with K5090, the signal increases by ~12±5%.

The proteomics analysis resulted in the identification of five protein sequences: uncharacterized protein LOC117318053, laccase-2-like, carbonic anhydrase 2-like, mucin-2-like isoform X1, aquaporin-10-like isoform X1. Three of the identified proteins are supported by more than two unique peptides (Table 1) and all sequences belong to the great scallop shell *Pecten maximus* (Bivalvia, Pectinida, Pectinidae). We find that the protein sequence coverage ranges from 1% up to maximum of 10% (as in uncharacterized protein LOC117318053 [*Pecten maximus*]) and the number of supporting peptides varies from 1 to 42. All of the proteins show a relatively high number of *de novo* tags, *e.g.* up to 155 are found in uncharacterized protein LOC117318053.

Table 1. List of the intracrystalline shell proteins identified in *Spondylus gaederopus*. Threshold values for peptide and protein identification: false discovery rate (protein FDR) = 0.5%, protein score −10lgP ≥ 30, unique peptides ≥ 1, *de novo* sequences scores (ALC%) ≥ 50. The proteins were filtered by verifying that each supporting peptide does not match any exogenous sequences (contamination, bacterial proteins etc.) using BLASTp tool.

Protein	**Accession number**	**Coverage (%)**	**Supporting peptides**	**Unique peptides**	***De novo* tags**	**Sequence similarity: (BLASTp analysis)**
Uncharacterized protein LOC117318053 [*Pecten maximus*]	>gi|1835552843|ref|XP_033728946.1|	10	42	41	155	Uncharacterized protein LOC110461617, protein PIF, collagen alpha-6(VI) chain, matrilin-2 [*Mizuhopecten yessoensis*], BMSP [*Pinctada fucata, Mytilus galloprovincialis*], COL6A [*Mytilus coruscus*], uncharacterized protein LOC105319624 isoform X3 [*Crassostrea gigas*].
Laccase-2-like [*Pecten maximus*]	>gi|1835484459|ref|XP_033755522.1	3	2	2	37	Laccase-7-like isoform X2 [*Mizuhopecten yessoensis*], uncharacterized protein LOC111134409 isoform X1 [*Crassostrea virginica*], laccase-1-like [*Crassostrea gigas*], laccase abr2 [*Drosophila albomicans*].

Carbonic anhydrase 2-like [*Pecten maximus*]	gi|1835482085|ref|XP_033754304.1|	3	2	2	19	Carbonic anhydrase 12 [*Mizuhopecten yessoensis*], carbonic anhydrase 3-like isoform X2 [*Crassostrea virginica*], carbonic anhydrase-related protein [*Crassostrea gigas*].
Mucin-2-like isoform X1 [*Pecten maximus*]	gi|1835474913|ref|XP_033725463.1|	1	1	1	78	Mucin-17-like isoform X2 [*Pecten maximus*], mucin-5AC-like [*Mizuhopecten yessoensis*].
Aquaporin-10-like isoform X1 [*Pecten maximus*]	gi|1835493482|ref|XP_033750578.1|	4	1	1	24	Aquaporin-9-like isoform X2 [*Mizuhopecten yessoensis*], aquaporin-10 isoform X1 [*Crassostrea gigas*], aquaporin-7 [*Lingula anatina*], aquaporin-7-like [*Pomacea canaliculata*], aquaporin-3 [*Strongylocentrotus purpuratus*], aquaporin-7-like isoform X1 [*Acropora millepora*].

Table 2. Sequence characteristics of the intracrystalline (IcF) *Spondylus* shell matrix proteins.

Protein	**pI**	**Acidic AAs (Asp/ Glu)**	**Basic AAs (Arg/ Lys)**	**Instability index (II)**	**Aliphatic index**	**GRAVY**
Uncharacterized protein LOC117318053 [*Pecten maximus*]	8.8	8%	8.9%	stable	65.84	-0.402
Laccase-2-like [*Pecten maximus*]	6.4	9.9%	7.8%	unstable	65.55	-0.543
Carbonic anhydrase 2-like [*Pecten maximus*]	6.6	14.5%	13.9%	stable	60.71	-0.852
Mucin-2-like isoform X1 [*Pecten maximus*]	6.2	7.4%	6.9%	unstable	56.58	-0.52
Aquaporin-10-like isoform X1 [*Pecten maximus*]	6.3	7.8%	6.9%	stable	93.95	0.259

Considering the characteristics of the protein sequence (Table 2) we find that most of the identified intracrystalline *Spondylus* shell proteins have a slightly acidic pI, between ~6.2 and ~6.6, except for uncharacterized protein LOC117318053 which is basic (pI 8.8). In all of the proteins, except for the carbonic anhydrase 2-like, the percentage of acidic/basic amino acids accounts for less than 10% (for carbonic anhydrase 2-like the percentage is closer to 15%). The majority of the sequences are classified as hydrophilic (based on a negative grand average of hydropathicity index – GRAVY), except for the aquaporin-10-like, which is hydrophobic. Finally, according to the Guruprasad protein instability index (II) (Guruprasad et al., 1990), three proteins (uncharacterized protein LOC117318053, carbonic anhydrase 2-like,

aquaporin-10-like isoform X1) are classified as stable and two (laccase-2-like, mucin-2-like isoform X1) as unstable.

The uncharacterized protein LOC117318053 from *Pecten maximus* was the best covered protein (10%) identified in the *Spondylus* intracrystalline shell matrix. The protein has high sequence similarity with uncharacterized protein LOC110461617 from the Yesso scallop [*Mizuhopecten yessoensis*], which was also found in our previous proteomics study of *Spondylus* (Sakalauskaite et al., 2020b). These two proteins are characterized by the presence of four von Willebrand factor A domains (vWA), a chitin binding domain and intrinsically disordered regions (IDRs) displaying putative vWF-collagen binding, glycoprotein binding and metal ion-dependant adhesion sites. They are also homologous to other mollusc shell proteins such as BMSP proteins from *Pinctada fucata* and *Mytilus galloprovincialis*, protein PIF, collagen alpha-6(VI) chain and matrilin-2, all from *Mizuhopecten yessoensis* and COL6A from *Mytilus coruscus*. Two enzymes were identified: laccase-2-like and carbonic anhydrase (CA) 2-like. The laccase-2-like protein has three copper oxidase domains and belongs to cupredoxin superfamily proteins. It displays copper ion binding and oxidoreductase activity and may be involved in oxidation-reduction processes, but its role in biomineralization is fairly unknown. On the other hand, the carbonic anhydrase enzymes which belong to carbonic anhydrase alpha-class family proteins are ubiquitous in many shell matrices of other molluscan species (Le Roy et al., 2014). CAs are involved in catalysing the conversion of carbon dioxide and water into carbonic acid, protons and bicarbonate ions. Both of the identified enzyme sequences are accordingly homologous with similar laccase-like and CA proteins of other mollusc shell species, including that from *Mizuhopecten yessoensis*, *Mytilus coruscus* and *Crassostrea* sp. In the intracrystalline *Spondylus* protein fraction, we also identify mucin-2-like and aquaporin-10-like proteins, both of which were not detected in the first proteomics study of *Spondylus* (Sakalauskaite et al., 2020b). Mucin-2-like isoform X1 from *Pecten maximus* is particularly rich in threonine (17.6%) and its sequence is homologous to the other mucin-like proteins, such as mucin-17-like and mucin-5AC-like. Mucin-like sequences are characterized as high molecular weight, heavily glycosylated (up to 70–80% their structure), gel-forming proteins with tandem repeat units (McGuckin et al., 2015). They have been identified in other matrices from molluscan shells (Marin et al., 2000; Sleight et al., 2016). In contrast, aquaporin-10-like isoform X1 has not been well documented and investigated in the shell matrices of molluscs. Aquaporin-10-like belongs to the aquaporin-like superfamily, whose members are involved in transmembrane transport processes facilitating water diffusion. The identified sequence is homologous to aquaporin-like proteins of other invertebrates, including molluscs (*Crassostrea gigas*, *Pomacea*

canaliculata), brachiopods (*Lingula anatina*) and echinoderms (*Strongylocentrotus purpuratus*, *Asterias rubens*).

Diagenesis of the intracrystalline shell matrix

In this part of the study we investigate how artificial aging affects the intracrystalline organic matrix of *Spondylus gaederopus* shell as well as the patterns of matrix degradation. The artificially-aged shell samples, *i.e.,* heated at different high temperatures and durations, were subjected to matrix extraction and characterised by immunochemistry and proteomics.

Immunochemistry

Immunochemistry assays were used to assess the loss of reactive groups and epitope activity of the intracrystalline organics during shell diagenesis in real time. Lectin jacalin and shell antibody K5090, which previously showed a strong cross-reactivity with the intracrystalline shell matrix of *Spondylus*, were used to track structural degradation processes after exposure to heating (for more details see material and methods and Figure 1). Figure 3 shows four line charts that depict the evolution of cross reactivity: plots a), c) display the results obtained using lectin jacalin, and plots b), d) display the results obtained with shell antibody K5090. The a) and b) graphs show the diagenesis profiles for the full length of the experiment (time plotted in logarithmic scale), while the c) and d) graphs show the diagenesis profiles during the first 48 hours of heating. The different line colours/types correspond to different temperature experiments: 80 °C – purple; 95 °C – blue; 110 °C – green; 140 °C – yellow (see also legend on the right).

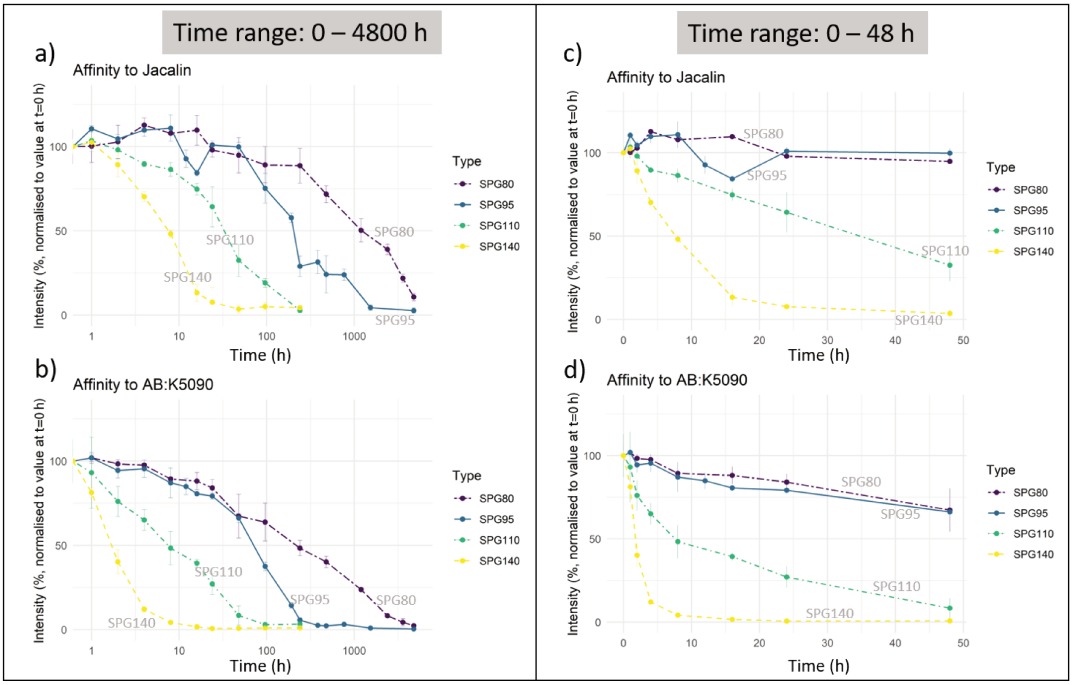


Figure 3. Line plots showing the diagenesis of the intracrystalline (IcF) *Spondylus* shell matrix tracked by ELLA and ELISA using lectin jacalin and shell antibody K5090 respectively. The plots show the cross-reactivity changes of the IcF shell matrix when the shell samples are subjected to heating at four different temperatures – 80, 95, 110, 140 °C and to different time intervals. The x axis corresponds to the heating time (hours) and the y axis corresponds to the intensity of the assay signal in percentage (% normalised to an unheated sample, *i.e.*, sample at time point 0 h). Graphs a) and b) show diagenesis patterns over full time range (time plotted in logarithmic scale); graphs c), d) present results in real time zoomed at a time range of 1-48 hours. More details on heating times are indicated in Figure 1. (n = 4, means ± S.D)

The graphs show that the structural degradation of the intracrystalline shell matrix is temperature dependent. Prolonged heating times decrease the shell matrix affinity to both lectin jacalin and antibody K5090, but at different rates. Figure 3 shows that at the highest temperature, 140 °C (set SPG140, yellow dashed line), the binding affinity of heated samples with lectin jacalin and antibody K5090 exponentially decreases with time. The reactivity of samples heated at 110 °C (set SPG110, in green dash dot line) also decreases exponentially. However, during the first 48 hours (early diagenesis) the matrix cross-reactivity with lectin jacalin shows almost a linear decrease and with K5090 – semi-linear (Figures 3c and 3d). At lower temperatures, 80 and 95 °C (sets SPG80 and SPG95 in purple long-dash dot and blue lines respectively) during the first 48 hours, only a gradual linear decline of reactivity is observed for both jacalin and K5090. Interestingly, the matrix cross-

reactivity with K5090 in the first 48 hours is almost superimposable for samples heated at the two temperature experiments under 100 °C, and a notable drop is only seen afterwards for samples heated at 95 °C. We also note that at 95 °C the pattern of matrix cross-reactivity with jacalin displays higher variability. Comparing the different sets, it is clear that the loss of signal with jacalin is much slower than with K5090:

- At 140 °C jacalin signal disappears after 24 hours of heating. This is almost 3 times slower than what is observed for K5090.
- At 80 °C K5090 signal starts to decrease after a few hours and it is not detected after 3600 h (150 days) of heating. In contrast, jacalin signal starts to decrease after 24 hours and it remains present even after 4800 h (200 days).

We also observe that in the first hours of heating, the cross-reactivity with jacalin shows an increased intensity (Figure 3 a, c). After one hour of heating at 110 °C and 140 °C the signal is higher compared to non-heated samples, whereas at 80 and 95 °C, the signal remains more than 100% for 24 hours. However, the same effect is not observed with antibody K5090.

The datasets were used to derive apparent kinetic parameters of the structural loss. As a first approximation, we presume that the rate of degradation is a first order, non-reversible reaction. The natural logarithm of the cross-reactivity intensity with jacalin and K5090 was plotted versus time (Figure 4). We note that not all of the obtained data were used for kinetic calculations and several time points were excluded to obtain the best linear fitting. The observed reaction rate constants (k_{obs}) were calculated assuming first order irreversible kinetics, for the four different temperatures and for both of the cross-reactive substrates separately. These values were subsequently used to derive activation energies (E_a) of structural degradation reactions (Figure 5). The apparent activation energy of reactive carbohydrate group loss (assessed with jacalin) is 112.36 kJ/mol; the degradation of epitope activity with K5090 is calculated to be E_a = 110.62 kJ/mol.

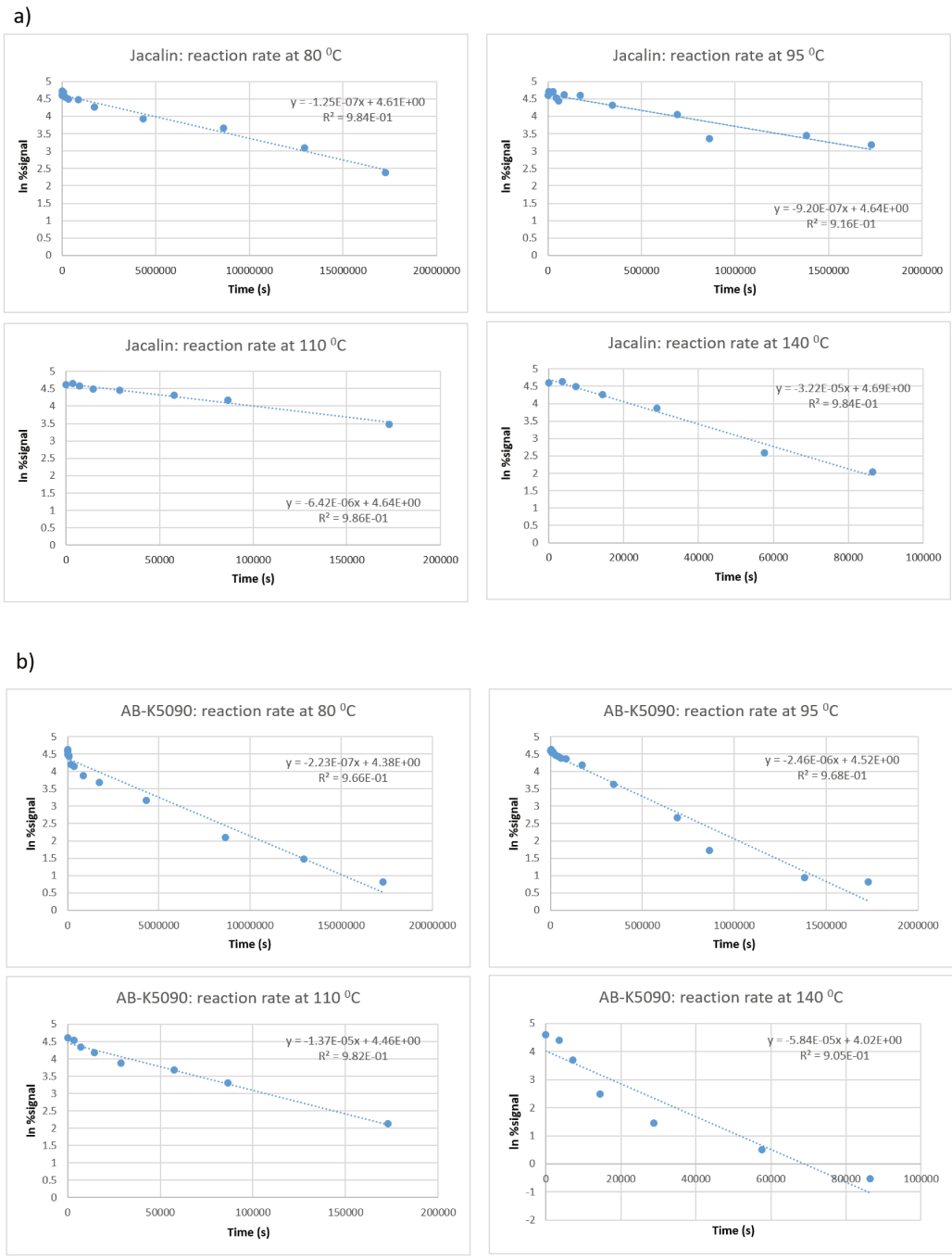
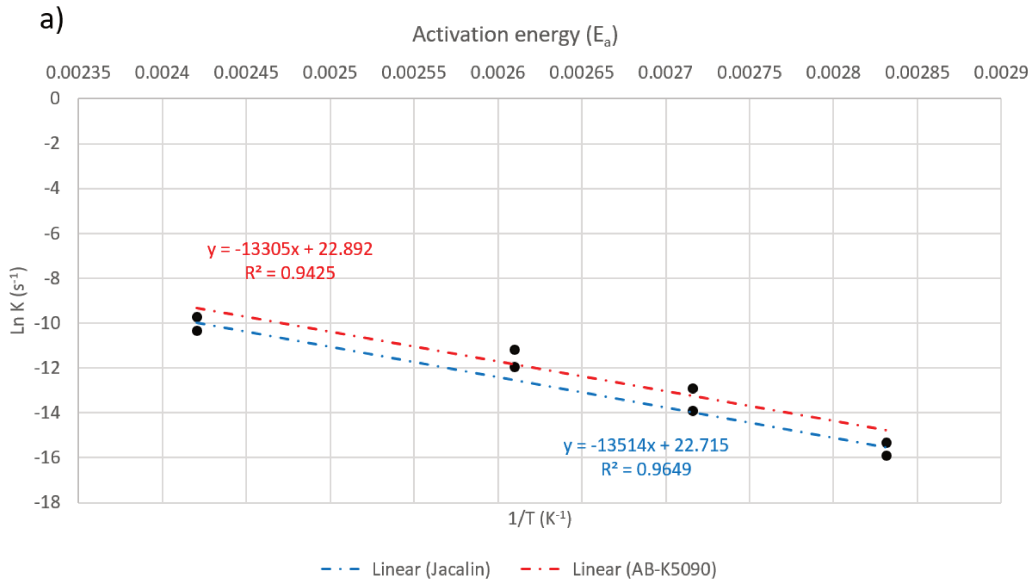


Figure 4. First order, non-reversible kinetics fitted to structural degradation reaction of intracrystalline shell matrix determined by ELLA/ELISA assays. The fittings were obtained for all four different temperature experiments: a) plots correspond to the loss of reactive carbohydrate groups that bind to lectin jacalin (ELLA); b) plots correspond to the loss of reactive proteinaceous epitopes that cross-react with shell antibody K5090 (ELISA). The

obtained kinetic constants were used to calculate activation energies of structural degradation presented in Figure 5.

a)



b)

Temperature (°C)	Lectin jacalin k_{obs} (s^{-1})	AB: K5090 k_{obs} (s^{-1})
80	1.25E-07	2.23E-07
95	9.20E-07	2.46E-06
110	6.42E-06	1.37E-05
140	3.22E-05	5.84E-05
E_a (kJ/mol)	112.36	110.62

Figure 5. a) Arrhenius plot which corresponds to the loss of the intracrystalline matrix structure. b) Calculated apparent activation energies (E_a) of structural degradation reactions. The kinetic parameters were obtained from ELLA and ELISA assays which tracked the loss of epitope and/or reactive carbohydrate group activity in the heated shell samples. The affinity to the shell matrix was separately studied with lectin jacalin and shell antibody K5090 which target the degradation of carbohydrate groups and proteinaceous epitopes respectively.

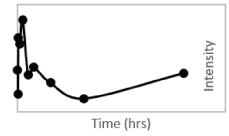
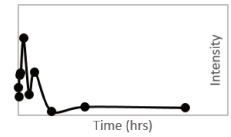
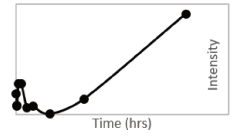
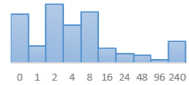
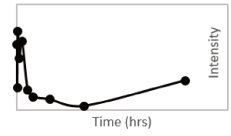
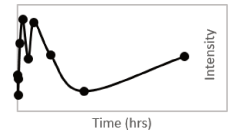
Quantitative proteomics

Intracrystalline *Spondylus* shell matrix proteins

Proteomic analysis with TMT labelling was used to obtain qualitative and quantitative information on shell samples that were heated at 110 °C. It was used to track the changes in protein/peptide abundance in shell samples heated for nine

different time intervals – 1, 2, 4, 8, 16, 24, 48, 96, 240 h and one unheated sample (the set also included one blank sample for a total of 11 samples). In short, the intracrystalline shell proteins were extracted from the selected samples, labelled with 11 different isobaric TMT tags, pooled together and analysed at the same time by LC-MS/MS (one injection). For all of the heated/unheated samples, the abundance of different peptides and proteins was calculated relative to the internal blank sample (*i.e.,* no shell proteins). The obtained values were plotted against heating time (h). Tables 3 and 4 display the relative abundance profiles represented in bar plots and line graphs, the latter displaying changes in real time. We note that the use of line plots is due to simplicity in visualization but they may not represent real trend lines and are interpreted cautiously. The tables also list all identified and quantified proteins, number of sequences that were used to create abundance profiles and the supporting peptides.

Table 3. Qualitative and quantitative proteomics data from the shell samples heated at 110 °C (set SPG110) obtained by TMT proteomics. The set includes samples heated for 9 different time points – 1, 2, 4, 8, 16, 24, 48, 96, 240 h; and one unheated sample (marked as 0 h). The 2nd and 3rd columns list the identified sequences used to create quantification graphs. The bar plots show relative protein abundance after different heating time intervals, line graphs display these changes in real time; the calculated intensity is relative to the internal blank sample. x-axis corresponds to heating time; the y-axis shows relative protein intensity at the specific time point.

Protein	Quantified sequenced	Peptides	Relative abundance profiles	
Uncharacterized protein LOC117318053 [Pecten maximus]	33 seq.: 25/25 peptides/unique	Listed in Table 4		
Carbonic anhydrase 2-like [Pecten maximus]	2 seq.: 1/1 peptides/unique	FGN(+.98)NRPIQR FGNNRPIQR		
Mucin-2-like isoform X1 [Pecten maximus]	1 seq.: 1/1 peptides/unique	GSVNVALLNILPELR		
Aquaporin-10-like isoform X1 [Pecten maximus]	1 seq.: 1/1 peptides/unique	SFVASILVFLVYF		
Laccase-2-like [Pecten maximus]	1 seq.: 1/1 peptides/unique	ADGLFGALVIR		

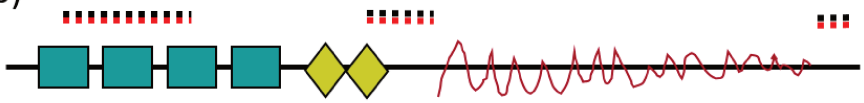
The quantification of uncharacterized protein LOC117318053 from *Pecten maximus*, which had the highest protein coverage (Figure 6), was based on 25 unique peptide sequences. The abundance profiles of the remaining four shell sequences were reconstructed based on only one (unique) peptide, which implies that the data represent only a single peptide, but not the whole protein.

a) Uncharacterized protein LOC117318053 [*Pecten maximus*]

```
   1  MKLPISLVLL ALMVHNAYQQ VLNCYVPADI MILIDGSDSI QDQNWREIKN FVSQLVANFD IGRDAIHVGF
  71  VVYSSDVGDH IGLQPYKPKN VLRTLSGILR QPKASTNTAK GIEYARNEFK TRGRPGVPKI LIVITDGSSD
                                                                1.
 141  NPRETRTQAN IAKIEGTRVI AVGIGQTFRD ELRQIASRPE KVYTAASFAT LQTLVFEIQR MVCQVITTTT
 211  TTTTAIPVPT KPPVVIPVPT DKICDVPGDI VFVMDGSDSI DDADFVRQKL FVANLIDNFE ISTEAIHVGL
 281  VVYSTIIGDT VGLQPPKNKE LLKILARNLR HPKVGTNTAL GIERARNMIR KEGRAMAPKL IVVITDGRSS
 351  SPKLTVAQAN MAKVEGITMV AVGVGTQIFS DELSQIASSS RKVFEVSDFR SLELIITSMR NLLCQAITTT
 421  TSTTTTTQKP TYIPPPDDLF CKVPADIGIL LDGSDSIADA DWVKQKHFVA SLINNLDVGR ETIHVGVVVF
 491  STIIGETVGL TPFKPKELLM ILSNNLKQPK VGTNTALGIQ RMRQMLATQG RANAPKVMII ITDGKSSSPK
 561  KTISQAGLAK REGITVIAVG VGSQLFREEL SQIATNDRKL FTVSDFQGLQ QIIVTLRNLI CQVITTTSTT
 631  QLVTTPQPPP VTIPPYNGCE VPAEIVFMIH GSDDIKEDNW GLNKKFVSGL VRNLYVTRGA HHIGVVVYSK
 701  GVGDHMSLQP YKEVFDVQNH INSFTHPRGG ANTAAALAKL RESFNTHARS GAPKIGVLIT DQTSTTPADT
 771  KLQAQLAKKE GVDFLTLGIG SNVNVGELAN IATDSSKLFR AMSFNRLDSM MPRIRDMICK IVTTPKTTPP
 841  PPPVSTPAPD FSKLCAGCLM NSNSGFNPYP GDCTKYVQCW RDGGNIRGVI KDCPYGQFWD EDAITCRPSM
 911  NVMCPQDPCI NAPDGFTYGM PDHGCRSHWV CVKGYSVASC CPEGTYYVEG MGCMRGMPCK TPCPPDGGLI
 981  ISPNCEKEVH WDGRFYMENV PGKGKMVRRC APGTIFDKGY CTCVIASSNY VPVTPNDKCM ESVHLNFDVN
1051  MKDKSGKRTQ INNVGVRKTK YGTAHFNGAS YFNIWNLGSI DYSDKFALKF KFKFDRWENR FDMGGYPVDA
                                  2.                                        3.
1121  GWKGDFLWKW DSQTGWNRNY TVGGSWNVVT LMNMFKRMFA ATSENEMETI MVEIAESPDF QILLYHLGFM
1191  PQSTNAIQIL RQLMSGNGAV FIRQILRQIG NMKGGMEIRR LLLQLLMTRE VNDWVRDHSG THVDITMKIW
1261  RKFLMSLNLT DSSWDIGGGW NNGMDDRTFD IIADLWYQFV RDQGMTGPEW EGKLKWALGN DRQKGSLDKR
1331  WIFDILFATK NNTIKNIWLD YLNKRNYSID WISTEWQIDD GKWWNWWWNR RERWNGSESM ESLAQLADDH
1401  LWTRNGNRTG RPRVEQWHGW DYGMTGPEWQ GKLKWALSNN RRKGSLDKRW ILDIMFAMRN TTIKDVWLDY
1471  ITKRNFQINW IGTGWQIGVG MGGTGGGIGG GGGMAQSQWN AWLNWLMTTY GGQGGSGMGT GQGGQGWGQG
1541  WGQGWGQGGG QGGQGSFTFD NFFHWLLGSG SFGNDFSFSG GSGTNTGGQG QGNGQDQGWF WGGQSNGQGQ
1611  GWFWGGQGNG GSGGNDDWML RINGSSQTGG SGGNFGTNDY WNSGTGKDAW FFGGSAGGNG IGGSQGGTNG
1681  GIWGGAGSSA GSWGGNVGQD GSIFGGQGHG QFGSQGGFNG ESNWRAGGVF GGAAGHGGFG EQWQDTSQTG
1751  YDVWGNQQGQ GGFGGTHGNG NGGTDAGGQG QVSVGTGGQW GILSGGQGGM GIDGHGGFQG GTGGVANGGG
1821  ELGTSGIYGG AGSKGQTKGQ WDMNSVSNGQ WGTLGGEVEG GGFWSNSAGR NGNGGTGGMA KGSAWPNGFE
1891  QGGSGSLVGG AGAQSGFGQG MDGGLAVGAA WQNGFGQGMA GGVAGQDDFG KGGSGGLARQ AGAQNGFGQS
1961  MTGGAAWQKG FGQDLVGGLA GAAGARNSFG QGVDGGLAGG TAWQHGVGQG MTGGASLQKG FGQDLVGGLA
2031  GGAGAQNSFG QGVDGGLAGG NAWQQGFGQG MSGGVATQNN FGQGMDGGLA GGTGAKNSFG QVIAGGAATQ
2101  NGFGQDDARG LTFGEGAGQI KYGQSGFGGI DARFGGQMGH VVNDMGGSGK TWRLKLDANQ VNRDDSWSVQ
2171  GNTVKGINGK QGTNSMSASS ANSDQGSALS SSGRGGSMSV AFGVNGDNGK SGSVVSGSIG GYGGSGNGTD
2241  LYGLIHGGKK QGNQNAGRGG SGMSSGTGSN DDGNMNGRVP TLYDSEGGSV GGTIDIGQRM NGQSTTRRVG
2311  QNVHFNVNEQ LVNSENSIRM NGQRGHKING ASIDWQSKIL ANDDIGRQGV GWSKSHMSGQ WNTRANGGDI
2381  NGQTVRIVQG GINSIGANRQ SVGSDLGQGQ VGSSSQGQGS IRSQEPGDVG AFLNVNGMSN TDISALYDQY
2451  KKTFGGTGIV QNKVVPRLQD LPRGETLGQF LKQTKLGSGV IGRKGLWGFD QGGTGFGGGA SGGSFGFGGS
2521  GGNSGEVDVG DINTILNNLG GAVGSQSGSS SSSSSQESGE MQIGGGGSSM SGGSSSTGSA STANKQSSSQ
2591  SSSQEGMRTG GNWQYGQNRL HMPFEFDGLD FPDLFRRKRA APSGDIFQYQ TLLSNCQSSG QGASVKIEAN
2661  ERSIKLGMLT EGKQKMKVMK ENIVPGWNEV TMTYDGKKLK SKIQNWRGAQ ERSVPLKGNI QKRQGFKIGG
2731  ACGASANFKG QMDDILLYNC LPSKPQQGPR GTGNNK
                4.
```

d Deamidation (NQ) (+0.98)

b)

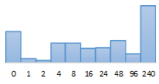
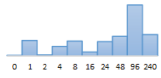
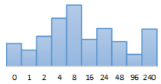
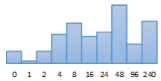
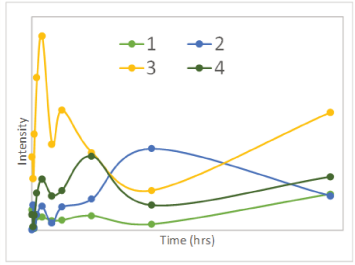
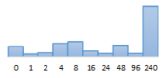
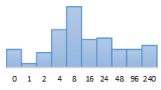
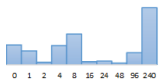
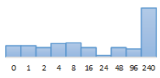
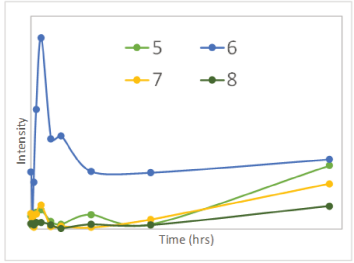
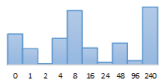
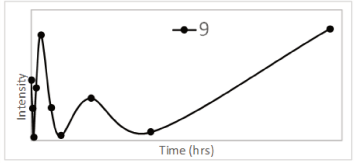
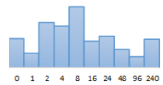
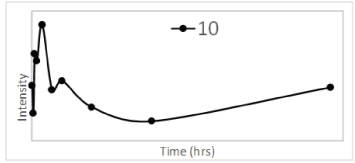
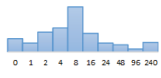
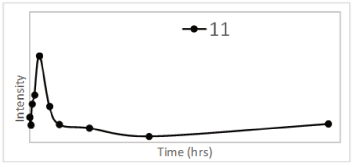


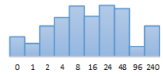
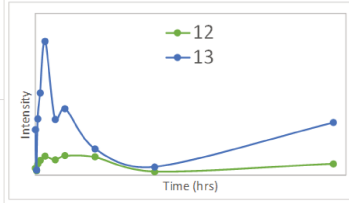
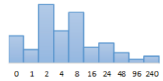
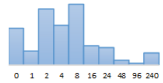
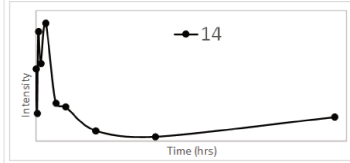
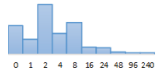
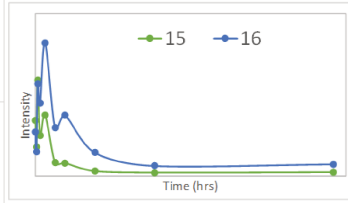
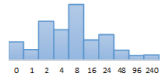
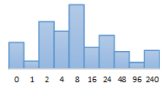
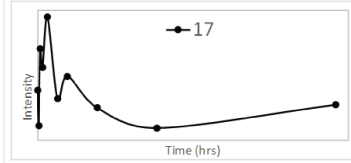
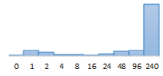
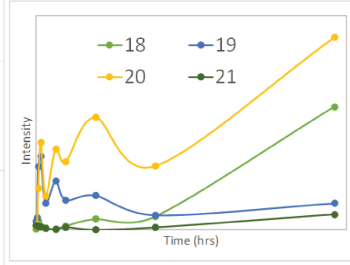
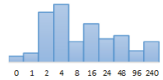
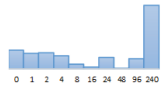
Protein coverage

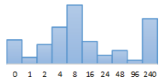
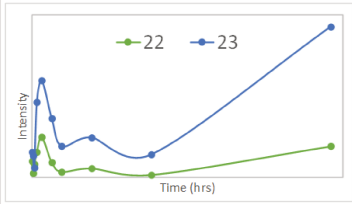
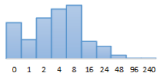
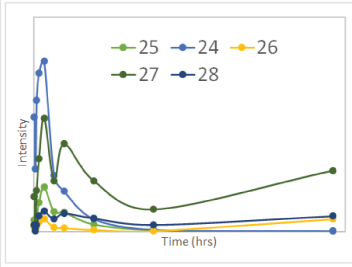
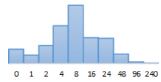
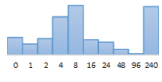
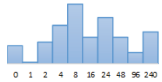
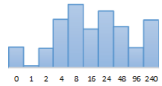
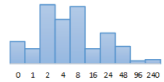
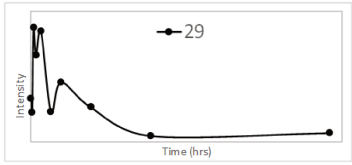
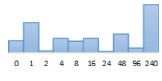
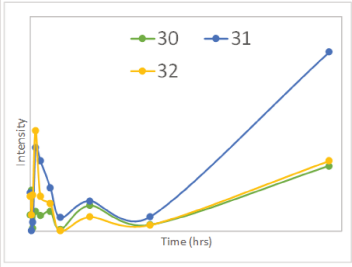
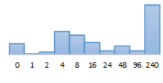
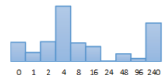
Figure 6. Sequence of uncharacterized protein LOC117318053 [*Pecten maximus*]. This protein was found to have the highest coverage and number of supporting peptides in the *Spondylus* shell IcF matrix; subsequently it was used to obtain an in-depth investigation of peptide degradation patterns. a) Full protein sequence. Grey colour marks 14 protein regions that were covered by identified peptides and for which, quantification was obtained. Sequence fragments that are marked by the red dashed lines (numbered from 1 to 4) indicate regions where multiple different peptide fragments were identified (analysed in detail and presented in Figure 5). b) Schematic representation of protein domains and the obtained sequence coverage: VWA – von Willebrand factorA; CBD – chitin binding domain; IDR – intrinsically disordered region.

The data show that during the first hours of heating the relative abundance of uncharacterized protein LOC117318053 actually increases and only after several hours the protein starts to degrade. For example, at 110 °C the protein is most abundant after 8 hours of heating and its intensity starts to gradually decline after 16 hours. However, after 240 h the protein does not completely disappear. Similar trends are also observed for peptide sequences of carbonic anhydrase 2-like, laccase 2-like, aquaporin 10-like and mucin-2-like proteins. However, CA peptides appear to be fully lost after 48 h of heating. In contrast, the peptide that belongs to the mucin-2-like sequence shows a notable increase in abundance after the longest heating interval (240 h). At the same time point, other sequences, except that of CA, also show a gradual rise in intensity. Hence, their relative intensities increased twice – first after several hours of heating (mainly after ~8 h) and then after around 10 days of heating (which is the longest heating time used in this experiment and for this temperature). However, it remains difficult to confirm this 'secondary' effect and more heating data points are needed to fully understand what happens just before and after this secondary increase.

Table 4. Quantified peptides that belong to uncharacterized protein LOC117318053 [*Pecten maximus*], obtained by analysing heated shell samples (at 110 °C). The table displays the identified peptide sequences and their relative abundances in samples heated for different time durations – 0, 1, 2, 4, 8, 16, 24, 48, 96, 240 h. The table also notes the position of these peptides in the protein sequence (see also Figure 6) and their type: tryptic sequences (T), peptides that have an deamidation modification (PTM.D; deamidation is marked with the mass shift +0.98 Da) and diagenetic cleavages (DC 1-4). Detailed analysis of DC peptides is presented in Figure 7. The bar plots show relative protein abundance after different heating time intervals, line graphs display these changes in real time. Note that the line graphs are used to have a better visualisation, but may not accurately represent the actual diagenesis trends. The intensity is calculated relative to the internal blank sample. x-axis corresponds to the heating time; the y-axis shows relative protein intensity at the specific time point.

Position	No.	Type	Peptide	Quantification profiles → heating time →
159 - 173	1	T, PTM. D	VIAVGIGQTFRDELR	
	2		VIAVGIGQ(+.98)TFRDELR	
	3		VIAVGIGQTFR	
	4		VIAVGIGQ(+.98)TFR	 
182 - 199	5	DC 1	VYTAASFATLQTLVFEIR	
	6		ATLQTLVFEIR	
	7		VYTAASF	
	8		ATLQTLVF	 
282 - 297	9	T	STIIGDTVGLQPPK	 
340 - 348	10	T	LIVVITDGR	 
364 - 384	11	T	VEGITMVAVGVGTQIFSDELR	 

392 - 400	12	T	KVFEVSDFR		
	13	T	VFEVSDFR		
481 - 504	14	T	ETVHVGVVVFSTIIGETLGLTPFK		
521 - 531	15	T	VGTNTALGIQR		
	16	T, PTM. D	VGTN(+.98)TALGIQR		
676 - 682	17	T	FVSGLVR		
1147 - 1156	18	DC 2	N(+.98)VVTLMN(+.98)MFK		
	19		VVTLMNMFK		
	20		VVTLMN(+.98)MFK		
	21		VVTLMNMF		

1182 - 1201	22	DC 3a	ILLYHLGFMPQ(+.98)STNAIQILR		
	23		HLGFMPQSTNAIQILR		
	24	DC 3	HLGFQPQ(+.98)STNAIQILR		
	25		HLGFQPQ(+.98)STN(+.98)AIQILR		
	26		STNAIQILR		
	27		STN(+.98)AIQILR		
	28		STN(+.98)AIQ(+.98)ILR		
1288 - 1301	29	T	TFDIIADLWYQFLR		
2740- 2754	30	DC 4	GQMDDILLYNCLPEK		
	31		GQMDDILLYN		
	32		GQMDDILLY		

The diagenesis of uncharacterized protein LOC117318053

A detailed investigation of protein degradation and peptide bonds hydrolysis patterns was carried out for the uncharacterized protein LOC117318053 [*Pecten maximus*]. The protein is mostly covered in the N-terminus part of the sequence, which coincides with the two conserved protein domains, *i.e.* vWA and CBD (Figure

6b). Several peptides are also identified in the C-terminus portion, while the uncovered part of the protein coincides with intrinsically disordered regions (IDRs). 32 supporting peptides were labelled and quantified (Table 4). The blue bar plots represent the relative abundance in samples heated for different time durations and line plots display the abundance changes in real time (used for visualisation purposes and, considering that they may not represent accurate diagenesis trends, should be interpreted cautiously). Here we highlight some of the most interesting observations and features:

a) A great number of peptides (at least 25) show an increase in intensity in the first hours of heating compared to the unheated samples. For the majority of these sequences, the relative abundance reaches maximum values after 8 hours of heating before starting to decrease.

b) Around fourteen peptides show a significant increase in relative abundance after 240 h hours of sample heating (the longest time point for this 110 °C set).

c) It was possible to directly investigate the deamidation reaction of N/Q amino acid residues. This is because some peptides were quantified in both the unmodified and the N/Q deamidated forms (for example, peptides 24, 25 – HLGFQPQ(+.98)STNAIQILR and HLGFQPQ(+.98)STN(+.98)AIQILR respectively, presented in Table 4).

d) It was also possible to investigate diagenesis-induced peptide bonds hydrolysis events. We have determined four sequence segments which were identified by multiple peptides of different lengths and were not the result of the enzymatic digestion (peptides DC1-4, in Figure 7 and also in Table 4). The relative abundance data of heated samples were used to investigate peptide bond fragmentation (Figure 7) discussed in detail below.

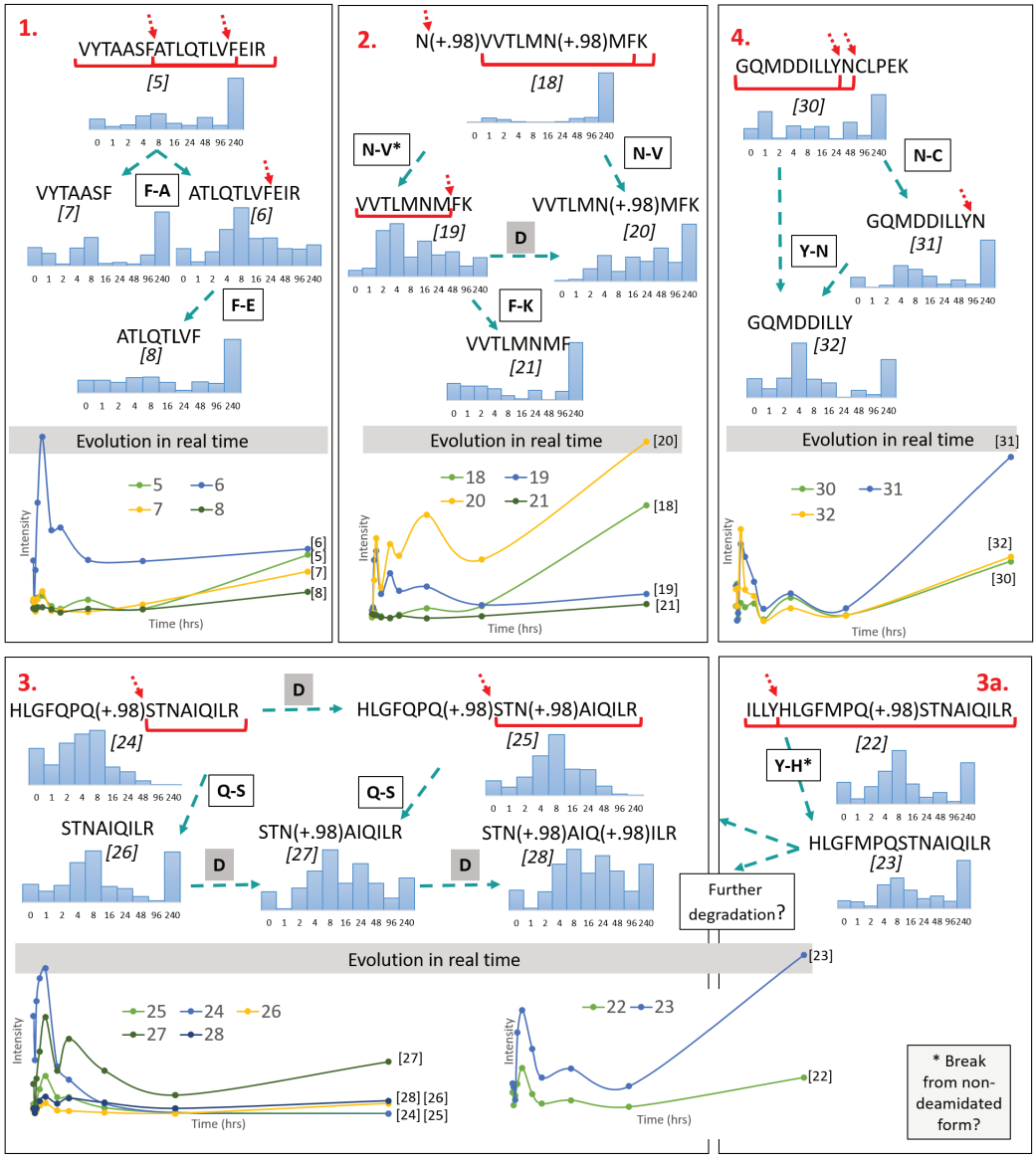


Figure 7. The scheme displays possible pathways of peptide bond fragmentation during shell protein diagenesis based on quantification data of multiple diagenetically fragmented sequences from uncharacterized protein LOC117318053 (sequences DC 1-4, Table 4; peptides are numbered underneath the sequence). The quantification graphs display peptide abundance and are represented as bar and line plots. x-axis corresponds to the sample heating time – 0, 1, 2, 4, 8, 16, 24, 48, 96, 240 h; y-axis — relative abundance, normalised against the internal blank sample.

The hydrolysis of the VYTAASFATLQTLVFEIR parent peptide (peptide no. 5, sequence segment 1) results in two smaller fragments, VYTAASF and ATLQTLVFEIR (no. 6 & 7). The peptide bond breakage occurs between F-A residues. Accordingly, the daughter fragment ATLQTLVFEIR shows a notable increase in abundance after the first hours of heating. This is followed by further hydrolytic damage, causing a F-E breakage and resulting in a daughter peptide with sequence ATLQTLVF (no. 8). The relative abundance of ATLQTLVF is highest in the sample heated for the longest time, 240 h.

The hydrolysis of peptide N(+.98)VVTLMN(+.98)MFK (no. 18, sequence segment 2), which has two deamidated Asn residues, occurs via N-V breakage resulting in shorter daughter peptides VVTLMNMFK and VVTLMN(+.98)MFK (no. 19 & 20). The appearance of an undeamidated peptide and the fact that its relative abundance increases during the first few hours of heating may indicate that N-V hydrolysis also occurs from an undeamidated 'parent' peptide form (not observed in our samples). This could imply that the N-V bond hydrolysis occurs faster than the deamidation of all of the Asn residues. However, this seems unlikely given that Asn deamidation is a very facile reaction compared to the peptide bond hydrolysis. The undeamidated fragment VVTLMNMFK (19) further loses C-terminal Lys (K), resulting from peptide bond breakage (F-K) and ending to the shorter fragment VVTLMNMF (no. 21). The smallest peptide fragment VVTLMNMF, the mono-deamidated fragment VVTLMN(+.98)MFK and the double-deamidated 'parent' sequence N(+.98)VVTLMN(+.98)MFK reach all their maximum relative abundances after 240 h of heating.

Peptide HLGFQPQ(+.98)STNAIQILR (no. 24, sequence segment 3) has one deamidated glutamine residue and HLGFQPQ(+.98)STN(+.98)AIQILR (no. 25) is the same peptide but with one extra deamidated asparagine residue. The relative abundance profiles suggest that in this sequence, the second Gln residue is more easily deamidated than the other Gln and Asn residues. This may be aided by the presence of a hydrophilic Ser residue downstream of the second Gln. Ser is known to undergo in-chain racemisation probably via a two-water-mediated cyclisation and bond reorganisation (Demarchi et al., 2013c) and it may contribute to the overall instability of the surrounding residues. The three daughter peptides STNAIQILR, STN(+.98)AIQILR and STN(+.98)AIQ(+.98)ILR (no. 26, 27, 28), which differ in the number of deamidated sites (undeamidated, N-deamidated and N,Q-deamidated respectively), are the result of Q-S peptide bond hydrolysis. After 240 h of heating, the N-deamidated peptide sequence STN(+.98)AIQILR displays the highest relative abundance.

We have also identified peptide ILLYHLGF**M**PQ(+.98)STNAIQILR (no.22, sequence segment 3a), which is almost identical to peptide sequence no. 25, except for the Gln → Met substitution. However, due to the lack of protein sequence data for *Spondylus*, we cannot exclude the possibility that these peptides cover two

different protein regions that share similar sequences. The Y-H peptide bond hydrolysis can be detected, but as the daughter peptide HLGFMPQSTNAIQILR is undeamidated, we hypothesise that the parent fragment was also undeamidated. The highest relative abundance of these peptides are observed after 8 and 24 h of heating.

The hydrolysis of GQMDDILLYNCLPEK (no. 30, sequence segment 4) occurs at the N-C bond resulting in a daughter fragment GQMDDILLYN (no. 31). This fragment breaks further (Y-N bond), by losing Asn at the C terminus, which results in fragment GQMDDILLY (no. 32). The highest relative abundance of these peptides are observed after first hours of heating (around 4 h) and also after 240 h.

By analysing these four different sequence fragments we identified eight peptide bonds that are unstable and especially prone to hydrolysis. These are:

1) Phe-Ala;
2) Phe-Glu;
3) Asn(+.98?)-Val;
4) Phe-Lys;
5) Gln(+.98)-Ser;
6) Tyr-His;
7) Asn-Cys;
8) Tyr-Asn.

The relative abundance values of the studied peptide sequences were also used to derive apparent reaction rates. We assumed that peptide bond hydrolysis and deamidation reactions follow first order kinetics (equation 1):

Equation 1: $\ln[A]=-kt+\ln[A]_o$

where [A] represents the relative abundance at time point t; $[A]_o$ – relative abundance at time point 0; k – reaction rate, t – time. We plotted the natural logarithm of different peptide relative abundance values versus time. The obtained graphs are presented in Appendix 3, SI material 1. In general, we observed poor fitting of the data to a linear regression line. Table 5 summarises only the regressions which yielded $R^2 > 0.9$. These include:

1) Hydrolysis reaction between the N(+.98)-V peptide bond (from sequence **N(+.98)V**VTLMN(+.98)MFK). The observed rate constant for the reaction was calculated as follows: k_{obs} = 9.30E-05 (s^{-1}).
2) The decrease in abundance of peptide HLGFQP**Q(+.98)S**T**N**AIQILR represents the deamidation of Asn but it could also be due to hydrolysis between Q(+.98)-S. The observed rate constant of this reaction was calculated as follows: k_{obs} = 8.41E-06 (s^{-1}).
3) Asn deamidation in peptide ST**N**AIQILR. The observed rate constant of this reaction was calculated as follows: k_{obs} = 1.04E-05 (s^{-1}).

Table 5. Kinetics data obtained for some of the hydrolysis and deamidation reactions which were found to yield an acceptable fit to first order kinetics ($R^2 > 0.9$). The table reports the reaction for which the kinetics were modelled, the observed kinetic rate constant (k_{obs}) and additional details regarding the linear regression fitting.

Peptide → reaction	**Reaction type**	**k_{obs} (s^{-1})**	**Details**
N(+.98)VVTLMN(+.98)MFK → VVTLMN(+.98)MFK	Hydrolysis between N(+.98)-V	9.30E-05	Peptide no.: 18 → 20 R^2 = 0.97 Used data points (h): 1, 2, 4, 8, 16.
HLGFQPQ(+.98)STNAIQILR ?→ HLGFQPQ(+.98)STN(+.98)AIQILR ?→ STN(+.98)AIQILR	Deamidation of Asn or hydrolysis Q(+.98)-S?	8.41E-06	Peptide no.: 24 → 25/ 27 R^2 = 0.93 Used data points (h): 8, 16, 24, 48, 96, 240 **Other hydrolysis reactions could occur in parallel or afterwards.*
STNAIQILR → STN(+.98)AIQILR	Deamidation of Asn	1.04E-05	Peptide no.: 26 → 27 R^2 = 0.94 Used data points (h): 8, 16, 24, 48, 96.

Discussion

Structural degradation

In this study we have used immunochemistry to characterise the structural degradation of the intracrystalline *Spondylus* shell matrix by tracking the activity loss of carbohydrate groups and protein epitopes. First of all, ELLA and ELISA assays with lectin jacalin and shell antibody K5090 showed that the intracrystalline and the (inter+intra)crystalline *Spondylus* matrices displayed similar cross-reactivity (Sakalauskaite et al., 2020b). This suggests that certain carbohydrate groups are actually incorporated inside the lattices of mineral crystals (Lang et al., 2020). We also observed that K5090, which recognises protein epitopes, in the intracrystalline *Spondylus* shell matrix showed slightly higher cross-reactivity (+ ~12%). This suggest that intense bleaching acts by breaking the insoluble organic matrix components (*e.g.* chitin or other crossed-linked structures) and helps to expose the reactive matrix epitopes, as has been also observed on other proteomics studies of *Spondylus* shell (Sakalauskaite et al., 2020a; 2020b).

Artificial diagenesis experiments show that the degradation of matrix structure is temperature dependent. The degradation at 140 °C is very fast, and both carbohydrate and protein structures are lost after around 8-16 hours, whereas at 80 °C , the epitopes remain active for up to 150 days of heating, similar to what has been observed in other shell models (Marin et al., 2007; Parker et al., 2015). Interestingly, glycoside structures targeted in this study appear more stable, as the lectin signal is still observed 200 days after heating. We note an interesting feature, *i.e.* that at the lower temperatures (below 100 °C) and in the first hours of heating an increased crossed-reactivity with lectin jacalin was observed. This suggests that heating first helps to uncoil and denature complex and cross-linked matrix structures allowing a better exposure of target sugar motifs to the lectin, before the actual degradation processes occur. The degradation pattern obtained by ELLA and ELISA analyses showed a good fit with first order irreversible reaction kinetics, which are also usually considered for peptide hydrolysis (Collins and Riley, 2000). The datasets were used to calculate the observed reaction rate constants at different temperatures and further derive apparent activation energies for structure degradation: E_a (carbohydrate groups) = 112.34 kJ/mol and E_a (protein structures) = 110.62 kJ/mol. Both structures show similar activation energies, albeit the loss of carbohydrate groups has slightly higher E_a values, in accordance with the degradation pattern observed for different heating temperatures. While we were not able to obtain activation energies for peptide bond hydrolysis in *Spondylus* shells (as only one temperature was investigated by proteomics), these values are found

comparable with the E_a of intracrystalline amino acid hydrolysis in other biomineral systems, *e.g*. limpet shell (*Patella*) and ostrich eggshell (OES) systems, for which first order hydrolysis reaction kinetics have been studied (Crisp et al., 2013; Demarchi et al., 2013a). For example, the E_a of Ala and Ser amino acid hydrolysis is around ~100 kJ/mol in *Patella* and 107-119 kJ/mol in OES (ostrich eggshell system); for Asx (aspartate/asparagine) it was found to be around 108 kJ/mol in *Patella* shell. In the same *Patella* shell*,* the racemization activation energies were higher, ranging from ~130 to ~150 kJ/mol for Ala, Asx, Ile, Glx (glutamine/glutamic acid), Val, Leu. The observed E_a decomposition reactions of Ser and Asx amino acid were found to be 131, 146 kJ/mol. Hence, assuming that intracrystalline shell protein diagenesis follows similar models in different biomineral types, the data indicate that structural loss of proteins/glycoproteins/carbohydrates may happen synchronously to peptide bond hydrolysis.

We are aware that the information obtained by ELLA and ELISA analysis represent only a partial picture and the results may be biased to some extent. First of all, this is because we lack the knowledge on the full biomolecular composition of the intracrystalline *Spondylus* shell matrix. In this study we have investigated the cross-reactivity only with one lectin jacalin. However, *Spondylus* matrix shows a moderate cross-reactivity with other ten lectins (as seen in previous study (Sakalauskaite et al., 2020b) and presented in Chapter 1), which also indicates the presence of other types of carbohydrate structures, *e.g.* N-linked mannose or N-acetylglucosamine. Secondly, we do not precisely know the target epitopes of antibody K5090 and we do not know which proteins (or glycoproteins) it binds to. And thirdly, in general, ELLA and ELISA assays can only be performed on soluble matrices, hence we have no access to the data on the insoluble fraction, which typically represents a notable part of shell organics (Sakalauskaite et al., 2020b) and may show a different diagenesis pattern (Parker et al., 2015). Nevertheless, the use of the relative rates of cross-reactivity changes provides a useful proxy to monitor structural degradation.

Protein degradation

Quantitative proteomics was used to study the degradation processes of intracrystalline shell proteins and it is the first application of TMT labelled proteomics to track diagenesis in a biomineral system. The advantage of this approach is that the qualitative information on protein/peptide fragmentation pathways (Demarchi et al., 2013c; 2016) is combined with the quantitative data in “real time”, thus yielding novel information on some pathways of peptide bond hydrolysis and deamidation. We note that the quantitative proteomics experiment did not include a replicate and this may be reconsidered in future analyses.

However, the relative abundance of identified peptides showed consistent results when compared to the data obtained by immunochemistry, suggesting that variation of data is minimal.

In all samples, we identified five pectinoid (*Pecten maximus*) mollusc shell proteins: an uncharacterized protein LOC117318053; two enzymes – carbonic anhydrase and laccase-like; aquaporin-like, a water channel protein and mucin-like, a gel-forming protein. However, only uncharacterized protein LOC117318053 had significant coverage and was supported by 25 unique peptides. It is important to highlight that the protein identification, characterization and the subsequent analysis of degradation patterns is limited due to the lack of 'omics' data (genomics, transcriptomics), particularly because *Spondylus* shell appears to have a very distinct proteome (Sakalauskaite et al., 2020b). This implies that the diagenesis patterns investigated in this study represent only a partial picture. However, uncharacterized protein LOC11731805, which had a good sequence coverage, was identified by multiple peptides of different sequence lengths, allowing an in-depth investigation of peptide bond hydrolysis. In the future, as more molecular data are collected for different molluscan species, including spondylid bivalves, the same proteomic dataset will provide additional information simply by re-analysing the data *in silico*.

The quantitative proteomic data showed complex shell protein and peptide degradation pathways. First of all, the majority of identified peptides had higher relative abundance in heated samples and were most abundant after 8 hours of heating at 110 °C. An analogous effect was observed in the structural degradation dataset obtained by immunochemistry. This suggests that heating contributes to denature/uncoil shell matrix by partially degrading cross-linked protein-carbohydrate structures and likely results in better matrix solubilisation as well as higher efficiency of digestive enzymes used in proteomics (*i.e*. trypsin) (Marin et al., 2016; Agbaje et al., 2019). For example, in our study we find that a mucin-2-like peptide was most abundant in the sample heated for 240 hours. Mucin proteins are typically heavily glycosylated (can be up to 80% of their sequence) and are characterized by insoluble structures when polymerized (Marin et al., 2000; 2013; McGuckin et al., 2015). A similar effect has been also observed in other subfossil shells, as, for example, the protein coverage or the concentration of analysed amino acids was found to be higher to that found in modern mollusc shell samples (Demarchi et al., 2013a; Sakalauskaite et al., 2019). Interestingly, some peptides showed a pattern of diagenesis that included a 'secondary' increase of their relative intensity – first after several hours of heating and then after around 10 days of heating. We interpret the secondary increase as a result of uncoiling, denaturation

and solubilisation of highly crossed-linked (and initially insoluble) shell matrix components that require more energy (*i.e.* prolonged heating) to be degraded. Finally, we find that after 240 hours of heating, many peptides remained high in abundance, indicating that the proteins did not undergo complete hydrolysis. However, at the same time point, the immunochemistry data show that both the reactive carbohydrate groups and the protein epitopes were already lost and inactive. In this case, it implies that peptide bond hydrolysis reaction is slower than the structural matrix loss. Referring to the discussion above, the data clearly show that the activation energy of hydrolysis reaction should be considered for each protein and hydrolytic site separately.

The study of multiple peptide sequence segments and peptide bond fragmentation of uncharacterized protein LOC117318053 [*Pecten maximus*] revealed a complex and slightly unexpected patterns of hydrolysis. First of all, we find that most abrupt changes in peptide relative abundance (*e.g.* decrease for larger peptides and increase for their smaller fragments) occurred after 8-16 h of heating. During the same time interval, immunochemistry data show that the targeted carbohydrate group activity decreases to around 80-75% of its initial intensity, whereas the proteinaceous signal falls to 50-40%. In the future, as other temperature sets are analysed by proteomics, it may be possible to obtain the activation energies (E_a) of protein/peptide hydrolysis and compare the kinetic parameters with structure degradation as well as with other shell systems (Demarchi et al., 2013a).

We have identified eight different peptide bond sites that were prone to hydrolysis. These were next to the amino acids that were hydrophobic (Ala, Val, Phe), polar with uncharged (Gln, Asn, Ser, Tyr) and charged (Glu, Lys, His) side chains. Typically, in aqueous solution, hydrophilic amino acids with polar (Ser, Tyr, Asn, Gln) or electrically charged side chains (Arg, His, Lys, Asp, Glu) are hotspots for hydrolysis. On the other hand, aliphatic (Ala, Val, Ile, Leu) and 'bulky' amino acids (Phe, Tyr) are typically harder to break, hence are considered more stable (Hill, 1965; Demarchi, 2020). Therefore, the hydrolysis of intracrystalline proteins in *Spondylus* shell system does not show a simple correlation to the theoretical stability of peptide bonds in solution, which is reported in the literature (Demarchi et al., 2013c). Similar findings have also been observed for other shell proteins when studying their thermal stability (Marin et al., 2007). We find that three out of eight identified peptide bond breakages in our data actually occur between dipeptides with phenylalanine – the bulky hydrophobic residue. Moreover, in peptide VYTAAS**FA**TLQTLVFEIR the Phe-Ala bond (in bold), which involves two aliphatic amino acids, is the first to break followed by another Phe-Glu cleavage from fragment ATLQTLVFEIR. In another study, Gurusprasad et al. used

computational methods and proposed that protein stability/instability *in vivo* was correlated with the presence of certain dipeptides (Guruprasad et al., 1990) and for example pairs with Trp (*e.g.* W-T; W-G; W-A), Asn (N-W, N-F, N-Q, N-T, N-G), and Met (M-Q, M-R) were classified as unstable. Our dataset suggests that the opposite is true here – seven out of eight peptide bond breakage sites involve dipeptides that, according to his method, would be classified as stable.

Structural features, *i.e.* matrix cross-linkage, ligand binding and protein-mineral interaction are known to play an important role in mediating protein diagenesis in biomineral systems (Demarchi, 2020). The effect of surface stabilization was found as the main factor enabling the preservation of a 3.8-million-year-old protein sequence in the calcite biomineral system of ostrich eggshell (Demarchi et al., 2016). Computational modelling suggested that protein-mineral binding involves acidic amino acids (Asp and Glu) that stabilize the adjacent water molecules and results in a higher energy barrier for the hydrolysis compared to the energy that would be required in solution (Demarchi, 2020). Yet, in our *Spondylus* shell samples, among the different identified and quantified peptides, acidic domains are actually rare (*e.g.* poly-E and poly-D that were only observed in peptide sequence GQMDDILLY). Indeed, we find that the intracrystalline *Spondylus* peptide sequences have a mix of basic and acidic pIs, suggesting that peptide-mineral interaction may occur in multiple ways. For example, *in vitro* studies with synthetic peptides showed that sequences with uncharged or with basic side chains, as well as those rich in Ser/Thr, had a strong affinity to calcite/aragonite mineral and affected calcite nucleation (Li et al., 2002; Gebauer et al., 2009). Phosphorylated Ser/Thr/Tyr residues can create additional coordination sites with calcium carbonate and phosphorylated proteins have been shown to be involved in the process of shell mineralisation (Borbas et al., 1991; Samata et al., 2008; Du et al., 2018). In addition, phosphorylated peptides at Ser residues have been identified in 1.77-million-year-old enamel proteomes indicating that phosphoproteins are stable in time (Welker et al., 2020). We suggest that in *Spondylus* shell matrix, it is the overall microchemical and structural environment that mediates protein diagenesis.

The peptide quantification data also show a consistently complex pattern of deamidation which increases with the heating time. The deamidation of Asn residues is known to be more than 60 times faster than that of Gln (Robinson and Robinson, 2004), but we also observe the opposite in *Spondylus*. For example, in peptide sequence HLGFQPQ(+.98) STNAIQILR, Asn undergoes deamidation after Gln. This either suggests that the 3D structure of the protein slows down Asn deamidation (Robinson and Robinson, 2001), or that the Gln deamidation is accelerated by the adjacent Ser which can undergo in-chain racemization

(Demarchi et al., 2013c). It may also be the combination of both of these effects. We also observe that the deamidation co-occurs with hydrolysis. For example, peptide GQMDDILLYNCLPEK, which has two potential sites for deamidation, Asn and Gln, first breaks into smaller peptides GQMDDILLYN and GQMDDILLY. Both of these do not exhibit any modifications, showing that in that case, direct peptide bond hydrolysis is faster than deamidation. We were able to calculate the reaction rate constants (k, s^{-1}) of one of the Asn deamidation reaction (ST**N(+.98)**AIQILR; k = 9.30E-05 s^{-1}) and one hydrolysis (**N(+.98)V**VTLMN(+.98)MFK → VVTLMN(+.98)MFK; k = 1.04E-05 s^{-1}). The obtained values show that at 110 °C the reaction rates of Asn deamidation and N(+.98?)-V peptide bond hydrolysis amino acids are very similar. When compared to the immunochemistry data, the hydrolysis/deamidation rate constants also fall within a similar range as the loss of protein structure (k = 1.37E-05 s^{-1}) and the disappearance of carbohydrate groups (k = 6.42E-06 s^{-1}). Finally, in our dataset we observe that peptides with multiple deamidated sites appear to be more stable in time. For example, after 240 h of heating, the double-deamidated peptide N(+.98)VVTLMN(+.98)MFK has a higher relative abundance compared to its subsequent smaller fragment VVTLMNMFK (a result of the N(+.98)-V or N-V bond breakage). Asn residues are primarily deamidated in the flexible regions (Kossiakoff, 1988) and the loss of amide side chain may contribute to make the structure more rigid, whereas the change in net charge may also impact structure stability (Silva et al., 2005; Zapadka et al., 2017). Overall, the data indicate that protein degradation processes in *Spondylus* shells - and, by extension, in all mollusc shells - are complex and pose challenges to interpretation when using currently available theoretical models. However, our work represents the first attempt to investigate diagenesis experimentally at proteomics level and, as more data will be collected, shows a future potential for uncovering the mechanisms that underlie the stability (or lack thereof) of biomineralized proteomes.

Conclusions

This study presents the first diagenesis investigation of intracrystalline shell proteins by simultaneously tracking aging-induced changes affecting both the structural properties of proteins and their sequences. Overall, we find that the intracrystalline *Spondylus* shell matrix, which consists of proteins and carbohydrate structures, is considerably stable over the prolonged heating. TMT proteomics shows that intracrystalline proteins do not undergo complete hydrolysis after 10 days of heating at 110 °C, while at the same time immunochemical proxies suggest that the matrix structure is largely lost. We also observe that heating does not induce instant matrix degradation, but, as we hypothesise, mild or short

temperature exposures firstly allow uncoiling (denaturing) cross-linked structures and facilitate the detection of proteins and glycosidic groups. A more in-depth sequence investigation was carried out for uncharacterized protein LOC117318053 [*Pecten maximus*] which enabled us to identify eight peptide bonds that were especially prone to hydrolysis. We observed patterns of peptide bond hydrolysis that did not fit with the theoretical stability of individual amino acids. Our data support the idea that structural organization between proteins, carbohydrates and the mineral surfaces as well as the overall microchemical environment, all play key roles in stabilizing the protein sequences and regulating degradation pathways. Our work further exposes the intrinsic peculiarities of protein aging in biomineralised tissues, but also suggests that *Spondylus* may provide a good system for protein preservation in fossil/sub-fossil records. Finally, this work also presents an analytical advancement, as it is the first application of TMT labelled proteomics for protein diagenesis studies in biominerals, enabling to track changes and pathways of degradation of individual shell proteins, peptides and their fragments in real time.

References

Abelson, P.H., 1954. Amino acids in fossils. Science 119, 576–576.

Agbaje, O.B.A., Thomas, D.E., Dominguez, J.G., Mclnerney, B.V., Kosnik, M.A., Jacob, D.E., 2019. Biomacromolecules in bivalve shells with crossed lamellar architecture. Journal of Materials Science 54, 4952–4969.

Albeck, S., Aizenberg, J., Addadi, L., Weiner, S., 1993. Interactions of various skeletal intracrystalline components with calcite crystals. Journal of the American Chemical Society 115, 11691–11697.

Borbas, J.E., Wheeler, A.P., Sikes, C.S., 1991. Molluscan shell matrix phosphoproteins: Correlation of degree of phosphorylation to shell mineral microstructure and to in vitro regulation of mineralization. The Journal of experimental zoology 258, 1–13.

Borrello, M.A., Micheli, R., 2011. *Spondylus gaederopus* in Prehistoric Italy: Jewels from Neolithic and Copper Age Sites. In: Ifantidis, F., Nikolaidou, M. (Eds.), Spondylus in Prehistory. New Data and Approaches. Contributions to the Archaeology of Shell Technologies. Archaeopress, pp. 25–37.

Bosch, M.D., Mannino, M.A., Prendergast, A.L., O'Connell, T.C., Demarchi, B., Taylor, S.M., Niven, L., Plicht, J. van der, Hublin, J.-J., 2015. New chronology for Ksâr 'Akil (Lebanon) supports Levantine route of modern human dispersal into Europe. Proceedings of the National Academy of Sciences of the United States of America 112, 7683–7688.

Buckley, M., Wadsworth, C., 2014. Proteome degradation in ancient bone: Diagenesis and phylogenetic potential. Palaeogeography, palaeoclimatology, palaeoecology 416, 69–79.

Chapman, J., Gaydarska, B., 2015. *Spondylus gaederopus*/*Glycymeris* Exchange Networks in the European Neolithic and Chalcolithic. In: Fowler, C., Harding, J., Hofmann, D. (Eds.), The Oxford Handbook of Neolithic Europe. Oxford University Press.

Collins, M., Cappellini, E., Buckley, M., Penkman, K., Griffin, R., Koon, H., 2006. Analytical methods to detect ancient proteins. Bio-and Material Cultures at Qumran 33.

Collins, M.J., Riley, M.S., 2000. Amino acid racemization in biominerals: the impact of protein degradation and loss. In: Goodfriend, G., Collins, M., Fogel, M., Macko, S., Wehmiller, J. (Eds.), Perspectives in Amino Acid and Protein Geochemistry. Oxford University Press, pp. 120–142.

Collins, M.J., Westbroek, P., Muyzer, G., Leeuw, J.W. de, 1992. Experimental evidence for condensation reactions between sugars and proteins in carbonate skeletons. Geochimica et cosmochimica acta 56, 1539–1544.

Crisp, M., Demarchi, B., Collins, M., Morgan-Williams, M., Pilgrim, E., Penkman,

K., 2013. Isolation of the intra-crystalline proteins and kinetic studies in Struthio camelus (ostrich) eggshell for amino acid geochronology. Quaternary geochronology 16, 110–128.

Demarchi, B., 2020. Amino Acids and Proteins in Fossil Biominerals: An Introduction for Archaeologists and Palaeontologists. John Wiley & Sons.

Demarchi, B., Williams, M.G., Milner, N., Russell, N., Bailey, G., Penkman, K., 2011. Amino acid racemization dating of marine shells: A mound of possibilities. Quaternary international: the journal of the International Union for Quaternary Research 239, 114–124.

Demarchi, B., Collins, M.J., Tomiak, P.J., Davies, B.J., Penkman, K.E.H., 2013a. Intra-crystalline protein diagenesis (IcPD) in *Patella vulgata*. Part II: Breakdown and temperature sensitivity. Quaternary geochronology 16, 158–172.

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

Demarchi, B., Collins, M., Bergström, E., Dowle, A., Penkman, K., Thomas-Oates, J., Wilson, J., 2013c. New experimental evidence for in-chain amino acid racemization of serine in a model peptide. Analytical chemistry 85, 5835–5842.

Demarchi, B., Clements, E., Coltorti, M., Locht, R. van de, Kröger, R., Penkman, K., Rose, J., 2015. Testing the effect of bleaching on the bivalve *Glycymeris*: A case study of amino acid geochronology on key Mediterranean raised beach deposits. Quaternary geochronology 25, 49–65.

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

Du, J., Xu, G., Liu, C., Zhang, R., 2018. The role of phosphorylation and dephosphorylation of shell matrix proteins in shell formation: an in vivo and in vitro study. CrystEngComm / RSC 20, 3905–3916.

Endo, K., Walton, D., Reyment, R.A., Curry, G.B., 1995. Fossil intra-crystalline biomolecules of brachiopod shells: diagenesis and preserved geo-biological information. Organic geochemistry 23, 661–673.

Erdjument-Bromage, H., Huang, F.-K., Neubert, T.A., 2018. Sample Preparation for Relative Quantitation of Proteins Using Tandem Mass Tags (TMT) and

Mass Spectrometry (MS). Methods in molecular biology 1741, 135–149.
Gebauer, D., Verch, A., Börner, H.G., Cölfen, H., 2009. Influence of Selected Artificial Peptides on Calcium Carbonate Precipitation - A Quantitative Study. Crystal growth & design 9, 2398–2403.
Guruprasad, K., Reddy, B.V., Pandit, M.W., 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein engineering 4, 155–161.
Hare, E.P., 1968. Racemization of amino acids in fossil shells. Carnegie Inst. Wash. Yearb. 66, 526–528.
Hare, P.E., 1971. Effect of hydrolysis on the racemization rate of amino acids. Carnegie Institute of Washington Yearbook 70, 256–258.
Hare, P.E., Hoering, T.C., 1977. The organic constituents of fossil mollusc shells. Year book - Carnegie Institution of Washington 78, 625–631.
Hill, R.L., 1965. Hydrolysis of Proteins. In: Anfinsen, C.B., Anson, M.L., Edsall, J.T., Richards, F.M. (Eds.), Advances in Protein Chemistry. Academic Press, pp. 37–107.
Ifantidis, F., Nikolaidou, M. (Eds.), 2011. *Spondylus* in Prehistory: New Data and Approaches: Contributions to the Archaeology of Shell Technologies. Archaeopress, Oxford.
Jong, E.W. de, Westbroek, P., Westbroek, J.W., Bruning, J.W., 1974. Preservation of antigenic properties of macromolecules over 70 Myr. Nature 252, 63–64.
Kossiakoff, A.A., 1988. Tertiary structure is a principal determinant to protein deamidation. Science 240, 191–194.
Lang, A., Mijowska, S., Polishchuk, I., Fermani, S., Falini, G., Katsman, A., Marin, F., Pokroy, B., 2020. Acidic monosaccharides become incorporated into calcite single crystals. Chemistry A European Journal 2020.08.03.234310.
Le Roy, N., Jackson, D.J., Marie, B., Ramos-Silva, P., Marin, F., 2014. The evolution of metazoan α-carbonic anhydrases and their roles in calcium carbonate biomineralization. Frontiers in zoology 11, 75.
Li, C., Botsaris, G.D., Kaplan, D.L., 2002. Selective in Vitro Effect of Peptides on Calcium Carbonate Crystallization. Crystal growth & design 2, 387–393.
Lowenstam, H.A., Weiner, S., 1989. On Biomineralization. Oxford University Press.
Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., Lajoie, G., 2003. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid communications in mass spectrometry: RCM 17, 2337–2342.
Mackie, M., Rüther, P., Samodova, D., Di Gianvincenzo, F., Granzotto, C., Lyon, D., Peggie, D.A., Howard, H., Harrison, L., Jensen, L.J., Olsen, J.V., Cappellini, E., 2018. Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. Angewandte Chemie 57, 7369–7374.

Maillard, L.C., 1912. Action of amino acids on sugars. Formation of melanoidins in a methodical way. Compte-Rendu de l'Academie des Sciences 154, 66–68.

Marin, F., Muyzer, G., Dauphin, Y., 1994. Caractérisations électrophorétique et immunologique des matrices organiques solubles des tests de deux Bivalves Ptériomorphes actuels, *Pinna nobilis* L. et *Pinctada margaritifera* (L.). Comptes rendus de l'Academie des Sciences. Serie 2. Fascicule a. Sciences de la terre et des planetes/Earth and planetary sciences. Montrouge 318, 1653–1659.

Marin, F., Corstjens, P., Gaulejac, B. de, Vrind-De Jong, E. de, Westbroek, P., 2000. Mucins and molluscan calcification. Molecular characterization of mucoperlin, a novel mucin-like protein from the nacreous shell layer of the fan mussel *Pinna nobilis* (Bivalvia, pteriomorphia). The Journal of biological chemistry 275, 20667–20675.

Marin, F., Morin, V., Knap, F., Guichard, N., 2007. Caspartin: Thermal stability and occurrence in mollusk calcified tissues. In: Arias, J.L.A., Fernández, M.S. (Eds.), Biomineralization: From Paleontology to Materials Science. Presented at the 9th International Symposium on Biomineralization, researchgate.net, pp. 281–288.

Marin, F., Marie, B., Hamada, S.B., Ramos-Silva, P., Le Roy, N., Guichard, N., Wolf, S.E., Montagnani, C., Joubert, C., Piquemal, D., Saulnier, D., Gueguen, Y., 2013. "Shellome": Proteins involved in mollusk shell biomineralization-diversity, functions. In: Watabe, S., Maeyama, K., Nagasawa, H. (Eds.), International Symposium on Pearl Research. Terrapub Tokyo, pp. 149–166.

Marin, F., Bundeleva, I., Takeuchi, T., Immel, F., Medakovic, D., 2016. Organic matrices in metazoan calcium carbonate skeletons: Composition, functions, evolution. Journal of structural biology 196, 98–106.

McGuckin, M.A., Thornton, D.J., Whitsett, J.A., 2015. Chapter 14 - Mucins and Mucus. In: Mestecky, J., Strober, W., Russell, M.W., Kelsall, B.L., Cheroutre, H., Lambrecht, B.N. (Eds.), Mucosal Immunology (Fourth Edition). Academic Press, Boston, pp. 231–250.

Milano, S., Nehrke, G., 2018. Microstructures in relation to temperature-induced aragonite-to-calcite transformation in the marine gastropod *Phorcus turbinatus*. PloS one 13, e0204577.

Mitterer, R.M., 1993. The Diagenesis of Proteins and Amino Acids in Fossil Shells. In: Engel, M.H., Macko, S.A. (Eds.), Organic Geochemistry: Principles and Applications. Springer US, Boston, MA, pp. 739–753.

Parker, A., Immel, F., Guichard, N., Broussard, C., Marin, F., 2015. Thermal stability of nacre proteins of the Polynesian pearl oyster: a proteomic study. In: Marin, F., Brümmer, F., Checa, A., Furtos, G., Lesci, I.G., Šiller, L. (Eds.), Key Engineering Materials. Trans Tech Publ, pp. 222–231.

Penkman, K.E.H., Kaufman, D.S., Maddy, D., Collins, M.J., 2008. Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells.

Quaternary geochronology 3, 2–25.
Penkman, K.E.H., Preece, R.C., Bridgland, D.R., Keen, D.H., Meijer, T., Parfitt, S.A., White, T.S., Collins, M.J., 2011. A chronological framework for the British Quaternary based on *Bithynia opercula*. Nature 476, 446–449.
Pierini, F., Demarchi, B., Turner, J., Penkman, K., 2016. *Pecten* as a new substrate for IcPD dating: The quaternary raised beaches in the Gulf of Corinth, Greece. Quaternary geochronology 31, 40–52.
Procopio, N., Buckley, M., 2017. Minimizing Laboratory-Induced Decay in Bone Proteomics. Journal of proteome research 16, 447–458.
Qian, Y., Engel, M.H., Macko, S.A., Carpenter, S., Deming, J.W., 1993. Kinetics of peptide hydrolysis and amino acid decomposition at high temperature. Geochimica et cosmochimica acta 57, 3281–3293.
Ramsøe, A., Heekeren, V. van, Ponce, P., Fischer, R., Barnes, I., Speller, C., Collins, M.J., 2020. DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes. Journal of archaeological science 115, 105080.
Rappsilber, J., Mann, M., Ishihama, Y., 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nature protocols 2, 1896–1906.
Robinson, N.E., Robinson, A., 2004. Molecular clocks: deamidation of asparaginyl and glutaminyl residues in peptides and proteins. Althouse press.
Robinson, N.E., Robinson, A.B., 2001. Prediction of protein deamidation rates from primary and three-dimensional structure. Proceedings of the National Academy of Sciences of the United States of America 98, 4367–4372.
Sakalauskaite, J., Andersen, S.H., Biagi, P., Borrello, M.A., Cocquerez, T., Colonese, A.C., Dal Bello, F., Girod, A., Heumüller, M., Koon, H., Mandili, G., Medana, C., Penkman, K.E., Plasseraud, L., Schlichtherle, H., Taylor, S., Tokarski, C., Thomas, J., Wilson, J., Marin, F., Demarchi, B., 2019. “Palaeoshellomics” reveals the use of freshwater mother-of-pearl in prehistory. eLife 8. doi:10.7554/eLife.45644
Sakalauskaite, J., Marin, F., Pergolizzi, B., Demarchi, B., 2020a. Shell palaeoproteomics: First application of peptide mass fingerprinting for the rapid identification of mollusc shells in archaeology. Journal of proteomics 227, 103920.
Sakalauskaite, J., Plasseraud, L., Thomas, J., Albéric, M., Thoury, M., Perrin, J., Jamme, F., Broussard, C., Demarchi, B., Marin, F., 2020b. The shell matrix of the european thorny oyster, *Spondylus gaederopus*: microstructural and molecular characterization. Journal of structural biology 211, 107497.
Samata, T., Ikeda, D., Kajikawa, A., Sato, H., Nogawa, C., Yamada, D., Yamazaki, R., Akiyama, T., 2008. A novel phosphorylated glycoprotein in the shell matrix of the oyster *Crassostrea nippona*: A novel acidic glycoprotein from the oyster

shells. The FEBS journal 275, 2977–2989.
Schroeder, R.A., Bada, J.L., 1976. A review of the geochemical applications of the amino acid racemization reaction. Earth-Science Reviews 12, 347–391.
Schroeter, E.R., Cleland, T.P., 2016. Glutamine deamidation: an indicator of antiquity, or preservational quality? Rapid communications in mass spectrometry: RCM 30, 251–255.
Silva, T., Kirkpatrick, A., Brodsky, B., Ramshaw, J.A.M., 2005. Effect of deamidation on stability for the collagen to gelatin transition. Journal of agricultural and food chemistry 53, 7802–7806.
Sleight, V.A., Marie, B., Jackson, D.J., Dyrynda, E.A., Marie, A., Clark, M.S., 2016. An Antarctic molluscan biomineralisation tool-kit. Scientific reports 6, 36978.
Sykes, G.A., Collins, M.J., Walton, D.I., 1995. The significance of a geochemically isolated intracrystalline organic fraction within biominerals. Organic geochemistry 23, 1059–1065.
Tomiak, P.J., Penkman, K.E.H., Hendy, E.J., Demarchi, B., Murrells, S., Davis, S.A., McCullagh, P., Collins, M.J., 2013. Testing the limitations of artificial protein degradation kinetics using known-age massive Porites coral skeletons. Quaternary geochronology 16, 87–109.
Weiner, S., Lowenstam, H.A., Hood, L., 1976. Characterization of 80-million-year-old mollusk shell proteins. Proceedings of the National Academy of Sciences of the United States of America 73, 2541–2545.
Welker, F., 2018. Palaeoproteomics for human evolution studies. Quaternary science reviews 190, 137–147.
Welker, F., Ramos-Madrigal, J., Kuhlwilm, M., Liao, W., Gutenbrunner, P., Manuel, M. de, Samodova, D., Mackie, M., Allentoft, M.E., Bacon, A.-M., Collins, M.J., Cox, J., Lalueza-Fox, C., Olsen, J.V., Demeter, F., Wang, W., Marques-Bonet, T., Cappellini, E., 2019. Enamel proteome shows that *Gigantopithecus* was an early diverging pongine. Nature 576, 262–265.
Welker, F., Ramos-Madrigal, J., Gutenbrunner, P., Mackie, M., Tiwary, S., Rakownikow Jersie-Christensen, R., Chiva, C., Dickinson, M.R., Kuhlwilm, M., Manuel, M. de, Gelabert, P., Martinón-Torres, M., Margvelashvili, A., Arsuaga, J.L., Carbonell, E., Marques-Bonet, T., Penkman, K., Sabidó, E., Cox, J., Olsen, J.V., Lordkipanidze, D., Racimo, F., Lalueza-Fox, C., Bermúdez de Castro, J.M., Willerslev, E., Cappellini, E., 2020. The dental proteome of Homo antecessor. Nature 580, 235–238.
Welle, K.A., Zhang, T., Hryhorenko, J.R., Shen, S., Qu, J., Ghaemmaghami, S., 2016. Time-resolved Analysis of Proteome Dynamics by Tandem Mass Tags and Stable Isotope Labeling in Cell Culture (TMT-SILAC) Hyperplexing. Molecular & cellular proteomics: MCP 15, 3551–3563.
Westbroek, P., Meide, P.H. van der, Wey-Kloppers, J.S. van der, Sluis, R.J. van der, Leeuw, J.W. de, Jong, E.W. de, 1979. Fossil macromolecules from

cephalopod shells: characterization, immunological response and diagenesis. Paleobiology 5, 151–167.

Wilson, J., Doorn, N.L. van, Collins, M.J., 2012. Assessing the extent of bone degradation using glutamine deamidation in collagen. Analytical chemistry 84, 9041–9048.

Windler, A., 2017. From the Aegean Sea to the Parisian Basin: *Spondylus* shell exchange in Europe during the process of Neolithisation. In: Eisenach, P., Stöllner, T., Windler, A. (Eds.), The RITaK Conferences 2013–2014. Bochum, pp. 95–110.

Zapadka, K.L., Becher, F.J., Gomes Dos Santos, A.L., Jackson, S.E., 2017. Factors affecting the physical stability (aggregation) of peptide therapeutics. Interface focus 7, 20170030.

Chapter 3

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

Digest

Mollusc shells are an important natural resource and this has been the case throughout human history. Indeed, shells represent one of the best known raw materials that were used to make personal ornaments. In archaeological research the study of shell ornaments can tell us more about human-environment interaction, peoples' mobility across time and space, exchanges of materials and expertise, and shifting cultural boundaries. However, current knowledge about mollusc shells that were used in prehistory is far from complete, because there is a lack of suitable analytical approaches to study heavily worked or fragmented shell remains. Palaeoproteomics, the analysis of ancient proteins by mass spectrometry, nowadays is one of the most widely used approaches to obtain biological information from a diversity of archaeological samples. In particular, peptide mass fingerprinting (PMF) by MALDI-TOF mass spectrometry has the advantage of being relatively cheap and simple, which allows researchers to rapidly conduct large-scale screening analyses. Palaeoproteomics by PMF has been applied to archaeological bones, as well as other biominerals such as ivory, horn and eggshell, but not to mollusc shells.

In the two following studies (Chapter 3 and Chaper 4), I aimed to develop a biomolecular approach to characterise mollusc shell proteins by peptide mass fingerprinting (PMF) and find out if the obtained shell PMFs could be used as molecular barcodes for shell identification in archaeological samples. The main objective of the first study (Chapter 3) was to optimise sample preparation methodology and to verify that intracrystalline sequences are distinct for molecular barcoding of shells. In the second study (Chapter 4) I used the optimised method to build a reference library of shell PMFs and tested the approach on a set of archaeological shell (presumably) ornaments from different Neolithic sites in France.

Chapter 3 presents a detailed investigation of different sample preparation techniques used in palaeoproteomics to find an optimal approach for shell protein

analysis from small-size mollusc shell samples. I demonstrated that intracrystalline proteins extracted from less than 20 mg of heavily bleached shell powders can be characterised by MALDI-TOF mass spectrometry, resulting in high-quality PMF spectra. I found that the single-pot, solid-phase-enhanced sample preparation method (SP3), which is not yet fully established in palaeoproteomics, significantly increased the recovery of intracrystalline shell proteins and was well suited for mollusc shell samples. In addition, the study demonstrated that different shells showed distinct PMFs, including taxa that are phylogenetically close and possess the same microstructures.

Chapter 4 presents the second study in which I successfully used the optimized method to extract intracrystalline proteins from other 27 different mollusc shells in order to build an in-house reference library of shell PMFs. The analysed shell taxa belonged to the four major shell-bearing mollusc classes – bivalves, gastropods, cephalopods and scaphopods. They also represented the main microstructures found in mollusc shells, *i.e.* nacreous, prismatic, foliated, crossed-lamellar and homogeneous. The obtained results highlighted the uniqueness of shell PMFs and it was also possible to investigate shared PMF features between phylogenetically close specimens. The PMF analysis of archaeological ornaments resulted in the identification of several individual peaks, but many of these could not be easily assigned to reference markers, collected in the study. This may be due to low protein content or may indicate that archaeological shell samples were too degraded.

These two studies mark a significant advancement in archaeological sciences, palaeoproteomics and the field of biomineralization. First of all, the data showed the diversity and distinctiveness of intracrystalline mollusc shell proteins which encode biological information, and are therefore useful for taxonomic identification. In the future, as more shells are analysed by proteomics, it will become possible to expand the knowledge on shell protein variability. Furthermore, this study provides the first attempt to carry out a molecular investigation of archaeological shell ornaments using PMF by MALDI-TOF mass spectrometry. The studies show a good potential for future studies involving the rapid survey of modern shell fragments but also indicate that the analysis of archaeological samples may require further optimisation. In the future, the samples will be reanalysed by high performance liquid chromatography coupled to tandem mass spectrometry (HPLC-MS/MS). The potential of this technique in shell palaeoproteomics will be presented and discussed in the fifth and last chapter of this thesis.

Journal of Proteomics 227 (2020) 103920

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

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

Jorune Sakalauskaite[a,b,*,1], Frédéric Marin[b], Barbara Pergolizzi[c], Beatrice Demarchi[a,**]

[a] Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Turin, Italy
[b] Biogéosciences, UMR CNRS 6282, University of Burgundy-Franche-Comté, 6 Boulevard Gabriel, 21000 Dijon, France
[c] Department of Clinical and Biological Sciences, University of Turin, AOU S. Luigi, 10043 Orbassano, TO, Italy

ARTICLE INFO

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.

Significance: We characterise for the first time the peptide mass fingerprints of the intracrystalline shell protein fraction isolated from different molluscan taxa. We demonstrate that these proteins yield distinctive PMFs, even for shells that are phylogenetically related and/or that display similar microstructures.

Furthermore, we extend the range of sample preparation approaches for "shellomics" by testing the SP3 method, which proved to be well-suited to shell protein extraction from small-size and protein-poor samples.

This work thus lays the foundations for future large-scale applications for the identification of mollusc shells and other invertebrate remains from the archaeological and palaeontological records.

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].

* Correspondence to: J. Sakalauskaite, Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Turin, Italy.
** Correspondence to: B. Demarchi, Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Turin, Italy.
E-mail addresses: jorune.sakalauskaite@unito.it (J. Sakalauskaite), beatrice.demarchi@unito.it (B. Demarchi).
[1] Second address: Biogéosciences, UMR CNRS 6282, University of Burgundy-Franche-Comté, 6 Boulevard Gabriel, 21000 Dijon, France.

https://doi.org/10.1016/j.jprot.2020.103920
Received 2 April 2020; Received in revised form 12 July 2020; Accepted 20 July 2020
Available online 24 July 2020

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 within the mineral crystals. These are known as "intracrystalline proteins" and may represent a so-called 'closed system', remaining inaccessible to environmental contamination and protected from rapid degradation processes (e.g., microbial) and thus persisting 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* 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 amino acids. 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. Materials 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* shells also retain 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, 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 (Fig. 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 h; 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 h; 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 h - 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-h-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.

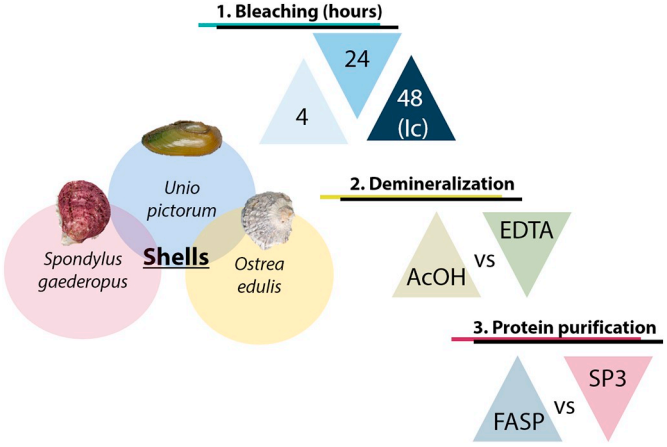


Fig. 1. Scheme showing the different approaches tested for shell protein extraction. "Mild" and "Intermediate" bleaching steps (4 and 24 h) were performed using diluted ~1.0–1.5% NaOCl; "strong" 48-h bleaching was performed using concentrated ~10–15% NaOCl, which isolates the "operational" intracrystalline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

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 h) were carried out using diluted NaOCl (1.0–1.5%), while the "strong" bleaching step (for a duration of 48 h) was carried out to isolate intracrystalline fraction (Ic) using concentrated NaOCl (10–15%).

No.	Method annotation	Bleaching (h)	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)		

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, ~0.5 M) by adding 500 μL to each of the powdered samples and thoroughly mixing with a vortex for ~4 h.

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).

2.2.3.1. FASP extraction. The extracts were concentrated using PALL Nanosep centrifugal devices (3 kDa, 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 1 M DL-dithiothreitol (Sigma, Canada) for 1 h at 65 °C, alkylated with 0.5 M 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.

2.2.3.2. 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 (3×), exchanged to buffer (50 mM ammonium bicarbonate, pH 7.5–8) and the mixture sonicated for 30 s. 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).

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 Supplementary material 1). All the spectra are reported in Supplementary material 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 material 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 h 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 Sections 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 (Fig. 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

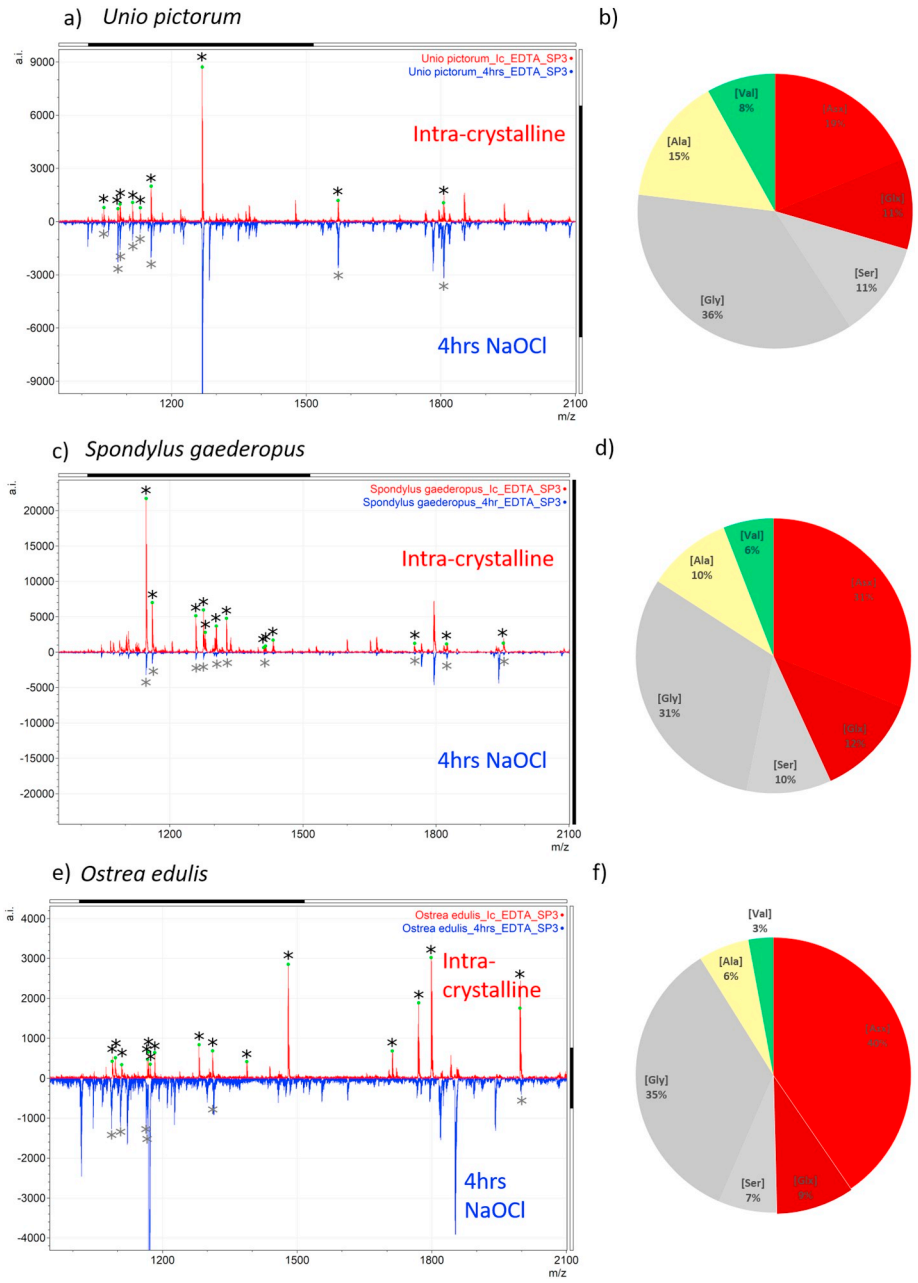


Fig. 2. *Unio pictorum, Spondylus gaederopus* and *Ostrea edulis* (inter + intra)crystalline vs intracrystalline (Ic) peptide mass fingerprints (PMFs) (a, c, e) and bulk amino acid (AA) compositions corresponding to the Ic fraction (b, d, f) obtained from previously published work [14]. Intracrystalline 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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-h bleaching step using concentrated NaOCl (~12%). Previous studies have shown that this treatment to be effective for all shell taxa tested thus far, regardless of their age (modern vs fossil) or provenance [14,36–40,45,57].

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* (Fig. 2). Comparing the PMFs of the intracrystalline shell protein fraction (Fig. 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* (Fig. 2a), while for *Spondylus* and *Ostrea* (Fig. 2c, e), the spectra of the fraction obtained after "mild" bleaching were of lower quality. This was particularly evident for *Spondylus* (Fig. 2c), for which the PMF of the Ic fraction was significantly better than that obtained from both the 4-h (Fig. 2c, in blue) and the 24-h bleached samples (Supplementary material 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 (Fig. 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 (Fig. 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

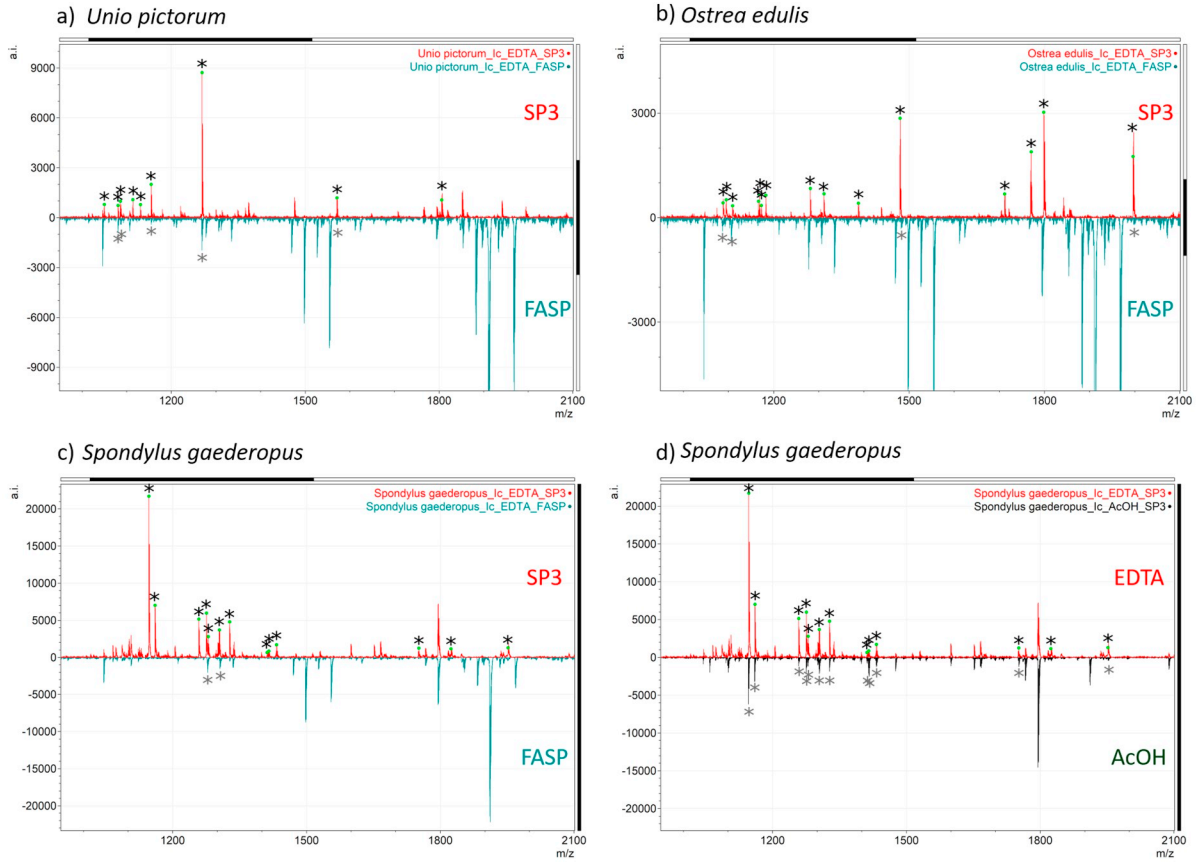


Fig. 3. Peptide mass fingerprints (PMFs) of the intracrystalline shell protein fraction extracted from *Spondylus gaederopus, Unio pictorum* and *Ostrea edulis.* Spectra a-c show PMFs of a) *Unio*, b) *Spondylus* and c) *Ostrea,* obtained by single-pot, solid-phase sample preparation (SP3, in red) or 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 h 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.

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* (Fig. 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 (Fig. 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 the spectra of samples demineralised with EDTA or acetic acid (Fig. 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

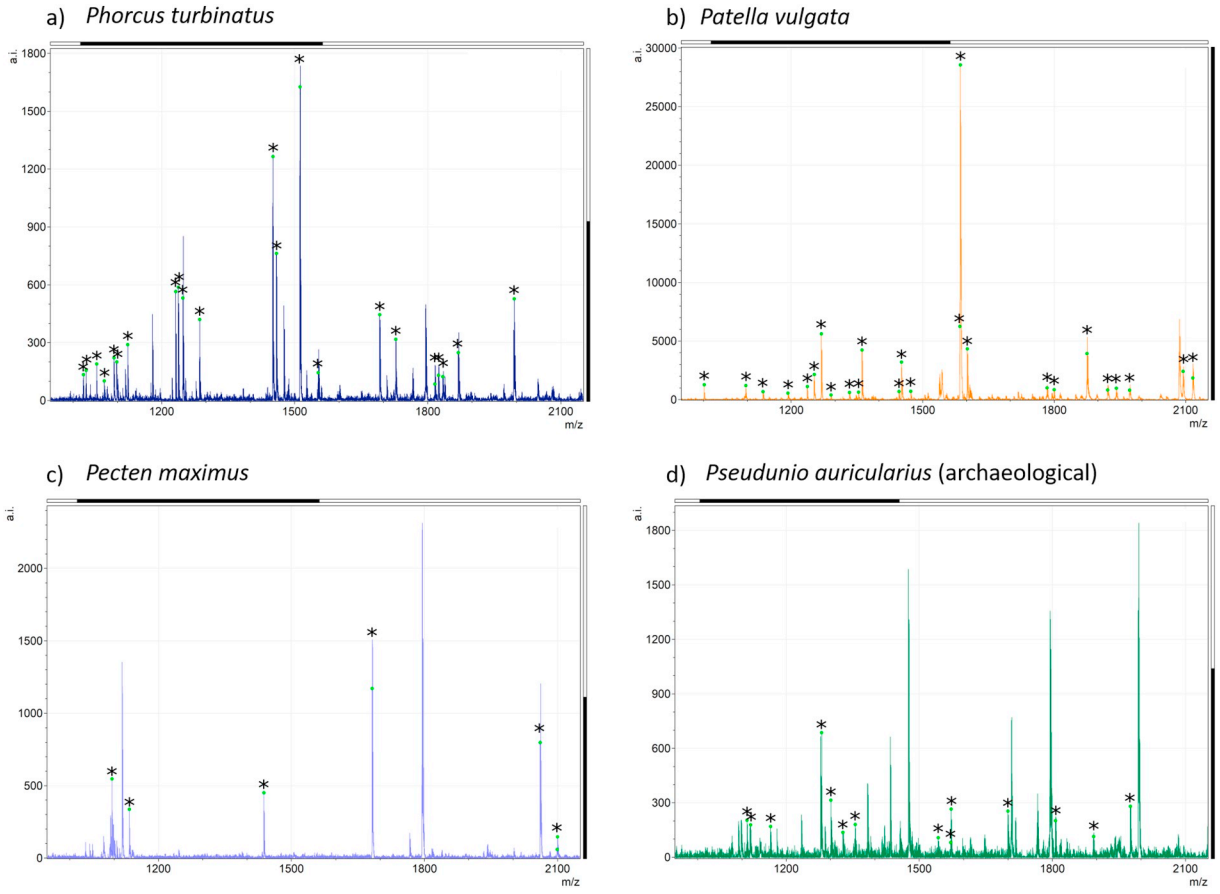


Fig. 4. Intracrystalline PMFs of the four different shell species that were used to validate the protein extraction method (Ic_EDTA_SP3): a) *Phorcus turbinatus* (modern), b) *Patella vulgata* (modern), c) *Pecten maximus* (modern), *Pseudunio auricularius* (Neolithic, 5226–5023 cal BCE). Asterisks indicate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

fractions - the acid soluble (ASM) and acid insoluble (AIM) matrices.

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 h 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*.Fig. 4 shows the PMFs obtained; as noted for *Spondylus, Ostrea* and *Unio* (Fig. 2), most of the marker peptides were observed in the 1000–2000 *m/z* range (Fig. 4a-d, markers represented by asterisks).

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

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			

thirteen) peptide markers after 96 h continuous heating at 80 °C and of three after 4800 h (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 [64], 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 (Fig. 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*).

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 h), 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).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2020.103920.

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, Proc. Natl. Acad. Sci. 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, J. Archaeol. Res. 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] E. Borgna, P.C. Guida, S. Corazza, Spondylus gaederopus: applicazione sperimentale dell'analisi microCT per la determinazione del genere, 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, 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 (1) (2011).
[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, Sci. Technol. Archaeol. Res. 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, K.S. Senthilvel, 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, 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, Intracrystalline 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, Biomacromolecules in bivalve shells with crossed lamellar architecture, J. Mater. Sci. 54 (2019) 4952–4969.
[44] F. Pierini, B. Demarchi, J. Turner, K. Penkman, *Pecten* as a new substrate for IcPD dating: the quaternary raised beaches in the Gulf of Corinth, Greece, Quat. Geochronol. 31 (2016) 40–52.
[45] J. Sakalauskaite, L. Plasseraud, J. Thomas, M. Albéric, M. Thoury, J. Perrin, F. Jamme, C. Broussard, B. Demarchi, F. Marin, The shell matrix of the European thorny oyster, Spondylus gaederopus: microstructural and molecular characterization, J. Struct. Biol. 211 (1) (2020) 107497.
[46] T. Jonuks, E. Oras, J. Best, B. Demarchi, R. Mänd, S. Presslee, S. Vahur, Multimethod analysis of avian eggs as grave goods: revealing symbolism in conversion period burials at Kukruse, NE Estonia, Environ. Archaeol. 23 (2018) 109–122.
[47] B.M. Sichert, P. Rentzel, B. Demarchi, J. Best, A. Negri, S. Deschler-Erb, Incubated eggs in a Roman burial? A preliminary investigation on how to distinguish between the effects of incubation and taphonomy on avian eggshell from archaeological sites, J. Archaeol. Sci. Rep. 26 (2019) 101845.
[48] F. Marin, B. Marie, S.B. Hamada, P. Ramos-Silva, N. Le Roy, N. Guichard, S.E. Wolf, C. Montagnani, C. Joubert, D. Piquemal, D. Saulnier, Y. Gueguen, "Shellome": Proteins involved in mollusk shell biomineralization-diversity, functions, in: S. Watabe, K. Maeyama, H. Nagasawa (Eds.), Terrapub Tokyo, vol. 149, 2013, p. 166.
[49] B. Marie, D.J. Jackson, P. Ramos-Silva, I. Zanella-Cléon, N. Guichard, F. Marin, The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties, FEBS J. 280 (2013) 214–232.
[50] F. Marin, I. Bundeleva, T. Takeuchi, F. Immel, D. Medakovic, Organic matrices in metazoan calcium carbonate skeletons: composition, functions, evolution, J. Struct. Biol. 196 (2016) 98–106.
[51] K. Mann, E. Edsinger, The Lottia gigantea shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis, Proteome Sci. 12 (2014) 28.
[52] A. Girod, Lavorazioni su conchiglia di *Pseudunio auricularius* (Mollusca, Bivalvia) nel Neolitico antico a Isorella (BS), Italia settentrionale, Atti Della Società Italiana Di Scienze Naturali E Del Museo Civico Di Storia Naturale Di Milano, vol. 151, (2010), pp. 89–98.
[53] E. Starnini, P. Biagi, N. Mazzucco, The beginning of the Neolithic in the Po plain (northern Italy): problems and perspectives, Quat. Int. 470 (2018) 301–317.
[54] M.D. Bosch, M.A. Mannino, A.L. Prendergast, T.C. O'Connell, B. Demarchi, S.M. Taylor, L. Niven, J. van der Plicht, J.-J. Hublin, New chronology for Ksâr 'Akil (Lebanon) supports Levantine route of modern human dispersal into Europe, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 7683–7688.
[55] Conchology, Inc, https://www.conchology.be/?t=1, (2020) (accessed November 22, 2019).
[56] B. Marie, P. Ramos-Silva, F. Marin, A. Marie, Proteomics of CaCO3 biomineral-associated proteins: how to properly address their analysis, Proteomics. 13 (2013) 3109–3116.
[57] B. Demarchi, Amino Acids and Proteins in Fossil Biominerals: An Introduction for Archaeologists and Palaeontologists, John Wiley & Sons, 2020.
[58] C.S. Hughes, S. Moggridge, T. Müller, P.H. Sorensen, G.B. Morin, J. Krijgsveld, Single-pot, solid-phase-enhanced sample preparation for proteomics experiments, Nat. Protoc. 14 (2019) 68–85.
[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*, Aquac. Fish. 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, Contrib. Archaeol. Shell Technol. (2011) 25–37.
[66] B.P. Carter, P. Benjamin, *Spondylus* in South American prehistory, 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, J. Open Archaeol. Data 7 (2019) 19.

Chapter 4

Shell palaeoproteomics II: peptide mass fingerprinting of mollusc shells in archaeology

Provisional list of authors:

Jorune Sakalauskaite[1,2], Barbara Pergolizzi[3], Jérôme Thomas[2], Rémi Laffont[2], Emilie Steimetz[2], Laurent Plasseraud[4], Solange Rigaud[5], Claire Manen[6], Monique de Cargouët[7], Nicolas Potier[7], Frédéric Marin[2], Beatrice Demarchi[1]

Affiliations

1. Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Turin, Italy;
2. Biogéosciences, 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.
4. Institute of Molecular Chemistry, ICMUB UMR CNRS 6302, University of Burgundy-Franche-Comté (UBFC), 9 Avenue Alain Savary, 21000 Dijon, France
5. UMR 5199 – PACEA, CNRS, Université de Bordeaux, Bâtiment, B8, Allée Geoffroy Saint Hilaire, 33615, Pessac, France.
6. UMR TRACES, CNRS, Université Toulouse Jean Jaurès, Toulouse, France.
7. Cerep-Musées, Société archéologique de Sens et Musées de Sens, 5, rue Rigault, 89100 Sens, France.

To be submitted to: Journal of Archaeological Science

Abstract

Mollusc shells are ubiquitous findings in archaeological sites, but a large part of shell remains, including artifacts, cannot be identified due to limited analytical tools applicable to worked and fragmented samples. We have recently published a simple and cost-effective approach that uses shell peptide mass fingerprints (PMFs), obtained by MALDI-TOF-MS, as molecular barcodes for taxonomic shell determination. The present work, which represents the second part of the shell PMF study, aims to advance the technique by analysing a larger diversity of mollusc shells and to test the method on unidentified archaeological ornaments presumably

made of shells. Intracrystalline proteins were characterized from 27 different mollusc shell taxa in order to create an in-house shell PMF library. The studied shells belong to four main molluscan classes (bivalves, gastropods, cephalopods and scaphopods), and represent the most commonly encountered shell microstructures (nacre, prismatic, foliated, homogeneous, crossed-lamellar). Above all, the studied shells are commonly found in archaeological sites. In this study, individual good quality PMFs were obtained for the majority of shells analysed. Some shell markers were found to be shared by several taxa and this allowed us to study relatedness of shell PMFs between phylogenetically related species. The shell PMF method was further applied to study sixteen archaeological ornament samples that came from different Neolithic sites in the South and North of France. Whilst at least seven archaeological samples showed mollusc shell structures by mineralogical and microstructural analyses (by FTIR-ATR and SEM) and one was made of a bone-like material, most of the identified peaks could not be assigned to known reference markers collected in this study. The PMF spectra also suggest that some of the archaeological samples could be poorly preserved. Overall, this work represents the first application of PMF to characterise a wide variety of modern and archaeological mollusc shells. It highlights the complexity of shell systems but also suggests that, in the future, with additional optimization, the approach could be a useful tool in molecular shell barcoding.

Introduction

Since prehistoric times, mollusc shells have been an important raw material used for making ornaments, and shell artifacts are remarkably abundant in archaeological sites. In archaeological research, as highlighted in the literature review (Part 2), shell ornaments provide a valuable proxy to study cultural patterns, people's mobility and cross-cultural interactions (Vanhaeren and d'Errico, 2006; Chapman and Gaydarska, 2015; Rigaud et al., 2015; Baysal, 2019). However, the information about the deliberate choice of specific shells that were used to make prehistoric ornaments, and therefore the knowledge on the shell diversity circulating in the past, remains limited, mainly because there is a lack of suitable analytical tools for taxonomic shell identification when archaeological specimens are found to be small, heavily worked, fragmented or degraded.

Palaeoproteomics, the analysis of ancient proteins by mass spectrometry, has emerged as one of the most promising molecular approaches to study bioarchaeological remains (see Part 4 of the literature review for more details and Welker, 2018 for a recent review). The two main mass spectrometric techniques used in palaeoproteomics are tandem mass spectrometry coupled to liquid chromatography (HPLC-MS/MS) and peptide mass fingerprinting (PMF) by matrix-

assisted laser desorption/ionization time-of-flight mass-spectrometry (MALDI-TOF-MS) (and for recent review, see Cleland and Schroeter, 2018 and Welker, 2018). Both of these approaches provide specific advantages and so are effectively applied to different types of studies (also presented in detail in Part 4 of the literature review). In a brief, tandem mass spectrometry by HPLC-MS/MS (when used for "shotgun" approaches) provides higher resolution data and information on the whole proteome (Aebersold and Mann, 2003; de Hoffmann and Stroobant, 2013). It is typically used for studying complex biological samples and to carry out detailed investigation of tens or hundreds of protein sequences, including in heavily degraded sub-fossil samples (Demarchi et al., 2016; Welker et al., 2020)
In contrast, the result of a MALDI-TOF-MS analysis is a peptide mass fingerprint spectrum, which should be characteristic of specific biological samples (*i.e*. proteins/mix of proteins). The resolution is lower, but the technique has two main advantages: the cost of the analysis is significantly lower (up to one hundred times per sample) and, from a bioinformatics point of view, the data are much less complex to treat (Aebersold and Mann, 2003; de Hoffmann and Stroobant, 2013). Therefore, it has been particularly useful for large scale screening of archaeological/palaeontological samples (Welker, 2018). In archaeological research, palaeoproteomic analysis by PMF has been applied to study species palaeodiversity, human diet and to gain insights into past cultures (Buckley and Collins, 2011; Stewart et al., 2014; Jonuks et al., 2018; Demarchi et al., 2020; Pothier Bouchard et al., 2020). The PMF technique is well established for taxonomic classification of bone (also known as ZooMS – Zooarchaeology by Mass Spectrometry (Buckley et al., 2009), for an in-depth review see also Buckley, 2018) and has also been successfully used to discriminate different types of archaeological materials, such as eggshell (Stewart et al., 2014; Presslee et al., 2017; Demarchi et al., 2020). However, peptide mass fingerprinting studies on archaeological mollusc shells have not elicited significant scientific attention and are in their infancy (Sakalauskaite et al., 2020a).

The previous chapter presented our recently developed methodology of PMF for rapid identification of mollusc shells in archaeology (Sakalauskaite et al., 2020a). The study showed that intracrystalline mollusc shell proteins, which represent an extremely small fraction of the biomineral (Penkman et al., 2008), can be extracted and characterised by MALDI-TOF mass spectrometry from miniature shell samples (an important constraint for archaeological applications). The method was successfully applied to seven different shell specimens that possessed distinct microstructures (including one sub-fossil shell) and demonstrated that intracrystalline shell proteins have unique peptide mass fingerprints (PMFs) useful for taxonomic shell barcoding. However, the application of shell palaeoproteomics by PMF to archaeological projects requires to study a wider variety of mollusc shells

and to test method applicability to archaeological samples. These two objectives are addressed in this work which therefore represents the second part of shell palaeoproteomics study by PMF.

The first aim of this study was to create an in-house molecular “library” of intracrystalline shell protein PMFs and to test the viability of the method for molecular shell barcoding. For this purpose, 27 shell taxa were analysed, which belong to the four main shell bearing molluscan classes:

1) Bivalves: bittersweet clam *Glycymeris glycymeris* (Arcida); pearl oyster *Pinctada margaritifera* (Ostreida); four freshwater mother-of-pearl mussels *Potamida litoralis, Anodonta cygnea, Unio crassus, Margaritifera margaritifera* (Unionida); four marine ‘mussels’ *Mytilus galloprovincialis, Mytilus californianus, Lithophaga lithophaga* (Mytiliida) and the fan mussel *Pinna nobilis* (Ostreida); edible oyster *Crassostrea gigas* (Ostreida), three venus clams *Ruditapes philippinarum*, *Mercenaria mercenaria*, *Venus verrucosa* (Venerida) and saltwater clams *Cerastoderma edule*, *Tridacna* sp. (Cardiida); (*the taxonomic order of the bivalves is indicated in the brackets and it is according to the World Register of Marine Species (WoRMS, 2020);

2) Gastropods: *Aliger gigas*; *Cornu aspersum*; *Littorina littorea*; three abalones *Haliotis asinina, Haliotis tuberculata, Haliotis discus*; the limpet shell *Lottia gigantea*;

3) Scaphopodes: *Dentalium* and *Antalis* sp.;

4) Cephalopod *Nautilus macromphalus*.

This selection of taxa encompasses the main calcitic and aragonitic microstructures found in mollusc shells – nacre, prismatic, foliated, homogeneous and crossed-lamellar (Boggild, 1930; Taylor et al., 1969; Carter, 1990). Among the different species analysed, the majority are marine molluscs, several are freshwater bivalves (Unionida) and one is the terrestrial garden snail *Cornu aspersum*. Marine molluscs were selected in order to include different geographic regions: European waters, including both the Atlantic Ocean and the Mediterranean Sea (*e.g. Glycymeris glycymeris, Venus verrucosa, Mytilus galloprovincialis, Cerastoderma edule*, *Pinna nobilis*), as well as subtropical regions in the Pacific Ocean (*e.g. Ruditapes philippinarum, Pinctada margaritifera, Crassostrea gigas, Haliotis* sp.) and tropical waters in Southern and Central America (*e.g. Aliger gigas*, *Mercenaria mercenaria*) (data from GBIF.org, 2020 and WoRMS, 2020). A large proportion of the taxa analysed in this work are edible species and the majority of those represent molluscs that were an important food resource in prehistory (Colonese et al., 2011). Among these are *G. glycymeris*, commonly known as bittersweet clam, Venerida, the saltwater clams, mussels in the family Mytilidae (e.g *Mytilus* sp.), cardiida clams, also known as the cockle shells (*C. edule*), the common periwinkle *Littorina littorea* and abalones *Haliotis.* sp. Some of the species are widely consumed today,

as for example the oyster *C. gigas* and the manilla clam *R. philippinarum* are of high commercial importance in the aquaculture economy. High economic value is also attributed to *P. margaritifera*, also known as the black-lip pearl oyster, for its ability to produce pearls. Interestingly, pearl oysters were also exploited in prehistory and in Arabia peninsula, the pearl fishing tradition dates back to approximately 7500 BP (Charpentier et al., 2012).

The shell set was also selected in order to include species which were valued since prehistoric times and are frequently documented among archaeological findings. Among others, the shells of *Glycymeris*, Venerid clams, cockles (*Cerastoderma edule*) as well as *Dentalium* and *Antalis* sp. were used intact or as a raw material for making ornaments (Bar-Yosef, 2005; Bonnardin, 2009; Bar-Yosef Mayer, 2018; Baysal, 2019). Thick and almost circular *Glycymeris* valves were also employed as pigment containers (Zilhão et al., 2010). Freshwater mother-of-pearl Unionida shells were an important raw material in crafting shell artifacts (Cartwright, 2003; Mayer, 2005; Bertin, 2015) and “button-like” ornaments made of Unionoida mother-of-pearl are also investigated in detail in this thesis work (Chapter 5). The use of noble pen shell *Pinna nobilis*, the large bivalve with nacreous and prismatic microstructures, which at the moment is classified as a critically endangered species (IUCN 2020), is also well documented in archaeological sites in the Mediterranean region. For example, fragments of worked *Pinna* shells were found in Bronze Age sites in the south of Italy (Minniti, 2005) and mother-of-pearl figurines were also documented in Early Bronze Age sites in Cyprus (Ridout-Sharpe, 2017). Besides, the byssus of *Pinna* shells, a silk-like thread that the animals use to attach themselves to sandy substrates, is commonly known as “sea silk” because it was a very valuable item and was used to make precious textiles (Kirby, 2018). *Tridacna* shells, which are found in shallow waters of coral reefs and are characterized by large and thick valves, were used to make engraved shell discs and also figurines (Brandl, 1984). *Aliger gigas* gastropod shell (also known as *Strombus gigas* or the queen conch) was not only used as raw material for making shell beads but also as a musical instrument, the trumpet, and it is a symbolically important shell in South American prehistory (Paulsen, 1974). As already mentioned in the literature review (Part 2), in archaeological research, misidentification of shell species used to make ornaments do occur. For example, bracelets made of *Glycymeris* shells can be often mistaken with those made of *Spondylus* shells (Dimitrijević and Tripković, 2006) and *Tridacna* figurines found in Mediterranean archaeological sites, could actually be made of *Spondylus* shells as well. The Mediterranean *Spondylus gaederopus* shell was widely used in prehistory and it was already investigated in the first study of shell palaeoproteomics by PMF technique. It showed that *Spondylus* provides a distinct PMF (Sakalauskaite et al., 2020a). Overall, the development of this shell PMF library was directed to have a substantial

representation of taxonomic, structural and typological diversity including the shell species that are commonly studied in archaeology. By all means, the set of mollusc species investigated in this study represents only a fraction of shell diversity used in prehistoric times. In the future, the library can be updated by more shell types that are of interest in archaeomalacological studies.

The second aim of this study was to carry out an integrated structural and proteomics study of archaeological ornament samples. The goal was to test protein preservation and investigate the possibility of obtaining their biological identification. The archaeological ornaments considered here were presumed to be made of molluscan shell and come from different Neolithic sites in France (Figure 1). The first set of ornaments was provided by collaborators Solange Rigaud and Claire Manen from PACEA institute ("from Prehistory to Present Time: Culture, Environment, Anthropology" UMR 5199 CNRS – University of Bordeaux, France). The set includes three small rounded beads, 5-6 mm in diameter, which come from the Le Taï cave archaeological site in the Occitanie region, South of France (Figure 1a, c). The site is situated in the small valley of Sartanette, near the Remoulins village, and was occupied at various times during the Neolithic era. The archaeological ornaments date back to the Early Neolithic period (~7000 BP), which is characterized by the first farmer communities appearing in the South of France (Caro et al., 2014). The site is located less than 100 km away from the Mediterranean Sea, suggesting that any ornaments could be made of marine shells. The second set of ornaments was provided by collaborators Monique de Cargouët and Nicolas Potier from collections of CEREP and archaeological museum of Sens (Société archéologique de Sens et Musées de Sens, France). It includes a variety of worked beads of different shapes and sizes obtained from several different Neolithic archaeological sites in the Yonne valley (Passy, Villeneuve-la-Guyard, Villeneuve-sur-Yonne, Vinneuf), which are near the Paris basin, in Northern Burgundy, North of France (Figure 1b, c). In these sites, a wide variety of ornament types are documented. Among these, several are described to be made of shell, dentine and stone (Prestreau, 1992; Bonnardin, 2009). For example, in the necropolis of Passy, around one hundred different shell objects were found, presumably made of marine gastropod shells *Nucella lapillus*, *Euspiria catena*, *Littorina littorea*, but also intact valves of *Pecten maximus*, as well as freshwater mussels *Pseudunio auricularius* and *Unio* sp. *(Bonnardin, 2014)*. In the Neolithic village of Villeneuve-la-Guyard, an impressive adornment made of large cylindrical shell beads was found and it is presumed to be made of *Spondylus* shell (Prestreau, 1992). The Yonne valley is located approximately 300 km away from the Manche sea, 400 km from the Atlantic Ocean, and more than 500 km in direct line from the Mediterranean Sea. The presence of marine shells is therefore

especially intriguing and hints at the possibility that the studied ornaments could also be made of marine shells.

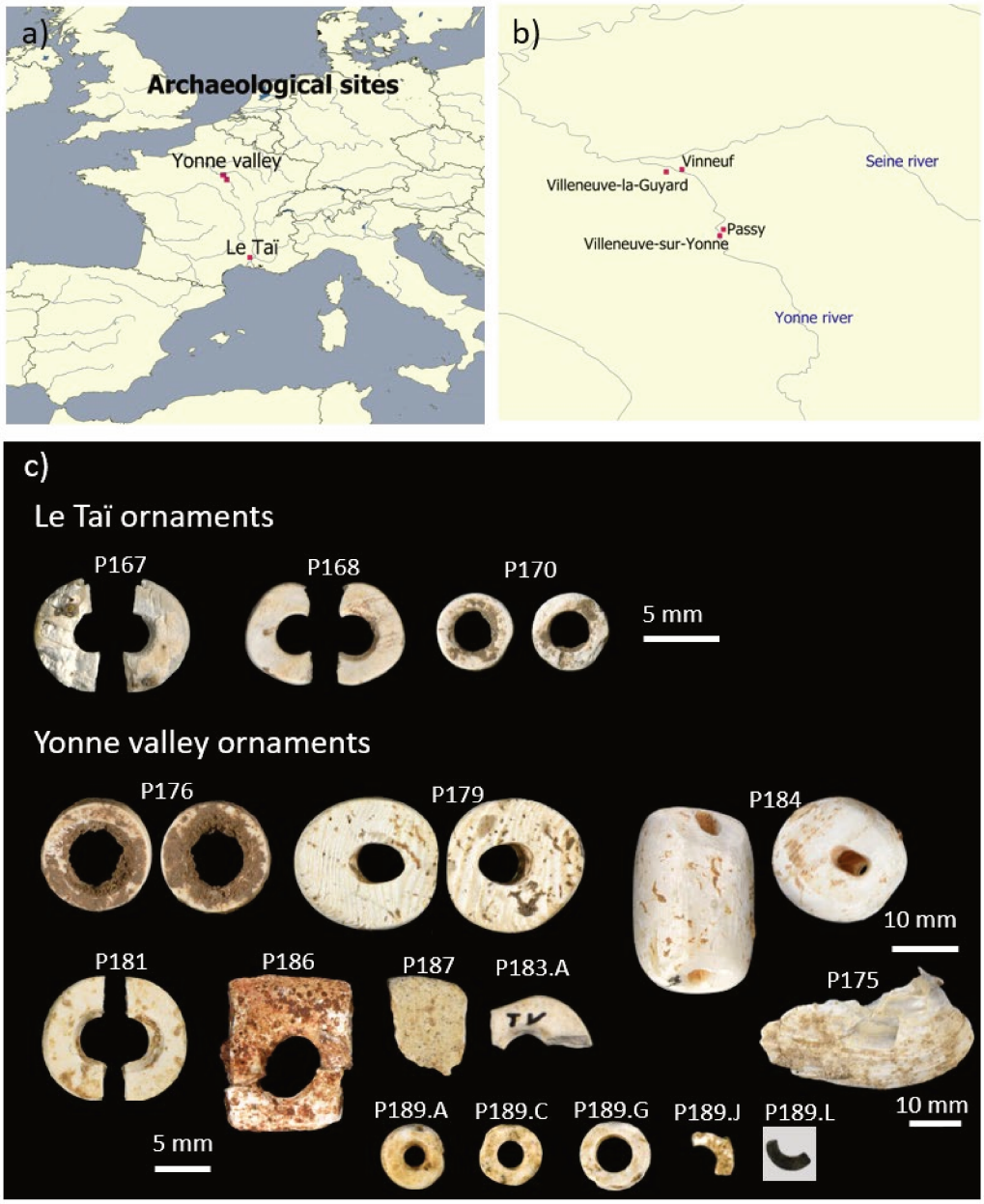


Figure 1. a) The map displays the location of Le Taï and Yonne valley archaeological sites where the ornament samples have been sourced; b) detailed map of archaeological sites located in the Yonne valley, Burgundy region, North of France; c) archaeological ornaments analysed in this study. Ornaments from Le Taï archaeological site (Occitanie, South of France) and ornaments from archaeological sites in the Yonne valley, Burgundy, North of France (photos, J. Thomas, Biogéosciences Dijon).

The two sets of ornament samples provide a good example to test the viability of shell PMF approach and they also present an interesting case study to determine the shell type/origin.

Materials and methods

Samples: modern and archaeological

Modern shells that were used to create shell PMF library were obtained from personal collection of Frédéric Marin (Biogéosciences, UMR CNRS 6282, University of Burgundy-Franche-Comté, France) and Beatrice Demarchi (Department of Life Sciences and Systems Biology, University of Turin, Italy). The shells were available as finely grained powders (particle size: 200-500 µm). A detailed list of all of the shells is presented in Table 2 (shell PMF library). Archaeological samples P167, P168, 170 from Le Taï archaeological site (Occitanie, South of France) were obtained from collaborators in PACEA institute, UMR 5199 CNRS, University of Bordeaux, France. Archaeological samples P175, P176, P179, P181, P183.A, P184, P186, P187, P189.A, P189.C, P189.G, P189.J, P189.L from archaeological sites in Yonne valley were obtained from collections present at CEREP and archaeological museum of Sens (Société archéologique de Sens et Musées de Sens), France. All the details of archaeological samples are presented in Table 1.

Table 1. Archaeological samples analysed in this study. Abbreviations: S - sector, Sq - square (carré), L - layer.

Sample	Type	Site	Site excavation details	Mass (g)
P167	Circular bead (fragmented)	Le Taï, (Remoulins, France)	S: PRL132 Sq: M9d L: d23	0.56
P168	Circular bead (fragmented)		S: T04 Sq: M11 L: d37	0.57
P170	Circular bead		S: PRL138 Sq: M10c L: dd25	0.24
P175	Shell valve (fragmented)	Passy (Yonne, France)	Sablonnière "Grossgatal" sép.4.1	1.96

P176	Circular bead	Villeneuve-la-Guyard (Yonne)	St.90 03.08.88	0.73
P179	Tubular triangular bead	Villeneuve-la-Guyard (Yonne)	Pass249, Carré2.	1.41
P181	Circular bead (fragmented)	Villeneuve-sur-Yonne (Yonne)	F9, Carré SB, Object 151	0.30
P183.A	Circular bead (?) (fragmented)	Villeneuve-sur-Yonne (Yonne)	F9 TV	0.19
P184	Tubular bead	Villeneuve-la-Guyard (Yonne)	Sépulture 251	14.40
P186	Trapezoid bead	Vinneuf (Yonne)	Vinneuf la sablière de Misy	0.78
P187	Flat fragment	Passy (Yonne, France)	Sépulture 3, "Césarine"	0.98
P189.A	Circular bead	Passy (Yonne, France)	Passy Moise sep 14.3	0.24
P189.C	Circular bead			0.09
P189.G	Circular bead			0.18
P189.J	Circular bead (fragmented)			0.02
P189.L	Circular bead (fragmented)			0.02

Physical and structural characterization of archaeological samples

The archaeological ornaments were analysed by X-ray computed tomography (CT-scan), scanning electron microscopy (SEM) and Fourier transform infrared spectroscopy in attenuated total reflectance mode (FTIR-ATR) in order to characterise their morphology, microstructure and mineralogy, respectively. They were first scanned with a Micro CT scanning system (SkyScan 1174, Bruker) which reconstituted 3D representations in a non-destructive way. Scans were obtained for all samples except for P175 (the intact shell valve) and heavily fragmented samples P187, P189.J, P189.L. The microstructure was then characterized using an environmental scanning electron microscope (SEM, Hitachi TM1000 Tabletop Microscope). In order to keep the integrity of the samples, no mechanical or chemical cleaning was used and no carbon coating was applied prior to the analysis. The mineralogy of the archaeological samples was characterized using

infrared spectroscopy in attenuated total reflectance mode (FTIR-ATR). Small grains of the material were carefully sampled with a scalpel and the spectra were acquired with a Bruker Vector 22 instrument (BrukerOptics Sarl, France, Marne la Vallée) fitted with a GoldenGate attenuated total reflectance (ATR) device (SpecacLtd, Orpington, UK) in the 4000–500 cm^{-1} range (twelve scans at a spectral resolution of 4 cm^{-1}). Spectral analyses were performed with the OPUS software provided by the instrument manufacturer (BrukerOptics Sarl). The assignment of the different absorption bands was obtained by comparison with previous spectra descriptions available in the bibliography (Loftus et al., 2015; Henry et al., 2017; Marques et al., 2018).

Shell Palaeoproteomics

Sample extraction and preparation

All modern shell and archaeological samples were extracted for proteomics and analysed by MALDI-TOF mass spectrometry. For modern samples, 20 mg of finely ground shell powder was used for extracting the intracrystalline shell proteins and protein extraction was carried out in duplicate, *i.e.* two powder aliquots were used separately. For the archaeological samples, around 20 mg of powder was carefully grained from the surface (or interior of the ornament) with a scalpel. Sample preparation and protein extraction was carried out following previously developed protocol (Sakalauskaite et al., 2020a). In short, all powdered samples from modern shells and archaeological samples were bleached using concentrated NaOCl (10-15%) for 48 h to isolate the operational intracrystalline protein fraction. Samples were thoroughly rinsed with ultrapure water (5x rinses). Bleached powders were demineralised with a 0.5 M EDTA solution (Sigma-Aldrich, pH=8) in a thermo shaker overnight (1500 rpm, at RT). The solubilized shell extracts were reduced using 1M DL-Dithiothreitol (Sigma, Canada) for 1 h at 65 °C and alkylated with 0.5 M iodoacetamide (Sigma, USA) for 45 min at room temperature in the dark. Shell proteins were extracted using single-pot, solid-phase sample preparation method (SP3). For this, 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 (final EtOH/sample solution ratio 1:1 v/v) and incubated at 24 °C for 5 min at ~1000 rpm. The eppendorfs were placed on a magnetic rack for separation, the supernatants were removed and discarded. The proteins bound to beads were cleaned with 80% EtOH (3x rinses), exchanged to buffer (50 mM ammonium bicarbonate, pH 7.5-8) and the mixture sonicated for 30 secs to homogenise. The protein digestion was carried using trypsin (0.5 µg, Promega, proteomics grade) at 37 °C, overnight and with light shaking (~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.

Sample P189.L (bone-like substrate) was prepared using filter-aided sample preparation method (FASP). Around 24 mg of the sample was demineralised with cold 0.6 M hydrochloric acid overnight, at 4 °C. The sample was centrifuged (13 000 rpm for 10 min) and the solubilised fraction was separated to a new eppendorf. The insoluble fraction was treated with Guanidine·HCl 6 M solution at 65 °C for 1.5 h to aid protein denaturation. This mixture was centrifuged again (13000 rpm for 10 min), the soluble fraction was separated and combined to the first one. Extract was ultrafiltered using PALL Nanosep centrifugal device (3 kDa, 0.5 mL), purified and exchanged to buffer (50 mM ammonium bicarbonate, pH 7.5-8). Proteins were reduced and alkylated and afterwards digested with trypsin as described above. Peptide digests were purified using C18 solid-phase extraction tips, as reported for shell samples above.

MALDI-MS analysis

All of the extracted peptides samples were resuspended in 10 µL TFA solution (0.1%). 0.7 µL of sample 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)) and spotted on a MBT Biotarget 96 MALDI plate (modern shell samples were spotted directly; for the archaeological samples two modes of matrix-mixture deposition were tested: spotting directly as for the modern shell samples and mixing with the matrix separately in an eppendorf and then spotting on the MALDI plate). The samples were analysed on a bench-top Microflex LRF MALDI-TOF mass spectrometer (Bruker Daltonics, Germany) in reflector mode. The parameter settings were as follows: ion source 1 – 18.96 kV; ion source 2 – 16.02 kV; lens voltage – 9.05 kV, reflector – 20.01 kV, laser power 22-24%. 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, was used for external mass calibration to maximize mass accuracy (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). The spectra were exported as text files and further processed using mMass, an open access mass spectrometry interpretation tool (Niedermeyer and Strohalm, 2012). Two spectra replicates were averaged for each sample (three for the archaeological samples) and the spectra were processed by performing

baseline correction and smoothing as indicated in the previous chapter (Sakalauskaite et al., 2020a). Peak picking was performed selecting a S/N threshold ≥ 6 and internal spectrum calibration was carried out using trypsin, keratin and MALDI matrix peaks (*m/z* list used for calibration can be found in (Sakalauskaite et al., 2020a)). *M/z* values that corresponded to common laboratory contaminants (*i.e.* keratin, trypsin, α-cyano MALDI matrix) were filtered out using a reference contaminant database (mass tolerance for peak matching: 0.1 Da). The resulting *m/z* values were used to create the shell PMF library, presented in Table 2. The PMF spectra of archaeological samples were carefully examined and the contaminant peaks excluded using the same reference contaminant database. The resulting peaks, which were interpreted as genuine shell protein matrix components, were compared with the shell PMF library.

Results

Shell peptide mass fingerprint (PMF) library

In this study I have created a reference dataset of shell peptide mass fingerprints (PMFs) by analysing the intracrystalline shell proteins from 34 different shell taxa (27 analysed in this study and 7 added from the previous shell PMF study, presented in Chapter 3 and Sakalauskaite et al., 2020a). Table 2 lists all of the shells that were used to create the reference library, together with their taxonomic classification, shell microstructure information (nacre, prismatic, crossed-lamellar, foliated, homogeneous), mineralogy (aragonite and/or calcite) and the obtained *m/z* values. Table 2 indicates markers that are found to be unique to each studied shell specimen and also several *m/z* values (labelled with sign '#') that were found to be shared by phylogenetically close species and/or species with similar microstructures. Figure 2 shows the PMFs of all the shells analysed in this study. Marker peaks are highlighted with asterisks and shared *m/z* are indicated with sign '#'.

In general, figure 2 shows that the MALDI-TOF spectra obtained from intracrystalline proteins of different shells clearly display the distinctiveness of their PMFs. Most of the shell peptide markers were identified in the 1000-2000 *m/z* range (Table 2, Figure 2). The majority of the taxa were identified by ≥10 peptide markers (22 specimens). Seven taxa were identified with a higher number or peptide markers (≥ 15), *i.e. Glycymeris glycymeris, Pinna nobilis*, *Nautilus macromphalus, Haliotis asina, Haliotis tuberculata, Modiolus modiolus, Littorina littorea*. Seven specimens showed very few PMF markers (≤ 4) - *Ruditapes philippinarum, Mytilus galloprovincialis, Mercenaria mercenaria, Venus verrucosa, Cerastoderma edule*,

Dentalium and *Antalis* sp. and the markers of shell *Cerastoderma edule* were identified only in a single extract.

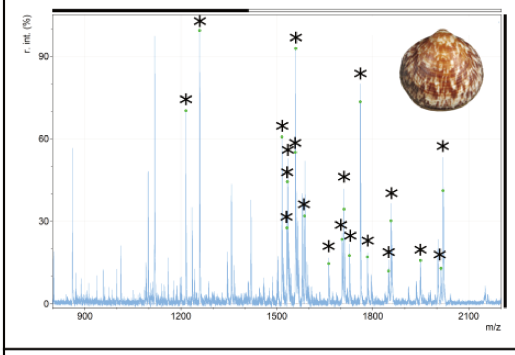
Table 2. Shell PMF library (available as electronic version).

Chapter 4. Table 2. Shell PMF library.xlsx

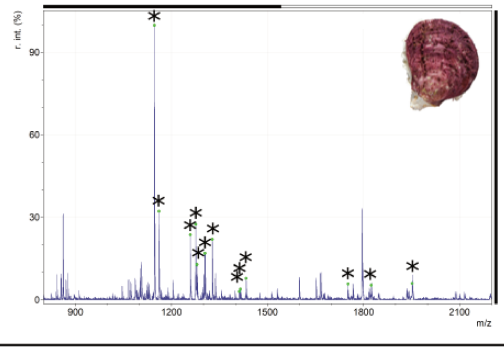
A number of *m/z* values were identified as "shared markers" in different taxa that were phylogenetically close and/or had the same microstructures (Table 2 and Figure 2, shared markers are noted with symbol '**#**'). The shared *m/z* values include:

- 'Unionid' markers – the peak at 1570.8 was found in all of the five Unionida order shells analysed in this study: *Potamida littoralis, Unio crassus, Anodonta cygnea, Margaritifera margaritifera, Unio pictorum*; the marker at *m/z* 2101.0 was shared by *Anodonta cygnea and Potamida litoralis*; 1372.6 was shared by *Margaritifera margaritifera* and *Anodonta cygnea;* 1781.7 was shared by *Margaritifera margaritifera and Unio crassus* and *m/z* 1268.5 was found in *Margaritifera margaritifera* and *Unio pictorum* shells.
- "Veneridae" markers: peptide markers at *m/z* 1435.8, 1436.8 were found in *Mercenaria mercenaria* and *Venus verrucosa* shells.
- "Dentaliidae" marker: the tusk shells, *Dentalium* and *Antalis* sp., share the same peptide marker at *m/z* 1893.0.
- "Nacre markers": several additional markers were also identified as being common in different nacreous shells, including *Haliotis* sp. *Pinna nobilis*, *Modiolus modiolus (*Table 2*)*. However, it is difficult to determine if the observed *m/z* values represent shared markers (*e.g.* due to microstructure) because of lack of comparative molecular data and phylogenetic distance between these shells.
- The nacreous and prismatic layers of *Pinna nobilis* shell were extracted and analysed separately owing to the valves' large size. The PMF spectra obtained from the prismatic layer showed a considerably higher number of marker peptides compared to the spectra obtained from the nacreous layer. However, some shared peptide markers were also identified in both microstructures – *m/z* values 1253.7, 1401.7, 2139.0.

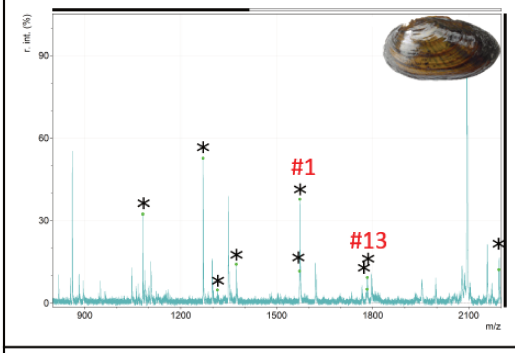
1) *Glycimeris glycimeris*



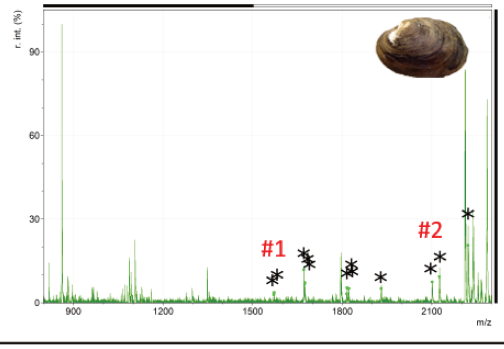
2) *Spondylus gaederopus*



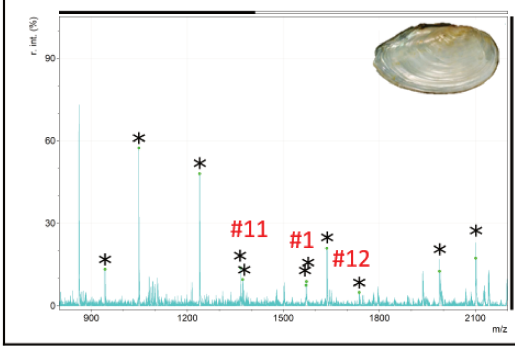
3) *Unio crassus*



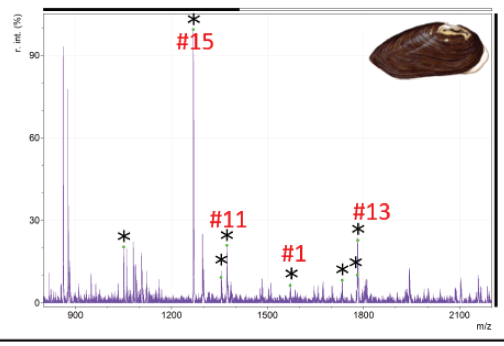
4) *Potamida littoralis*



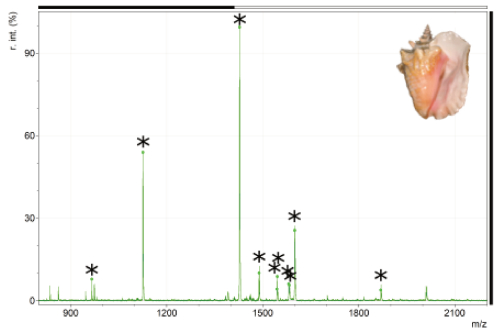
5) *Anodonta cygnea*



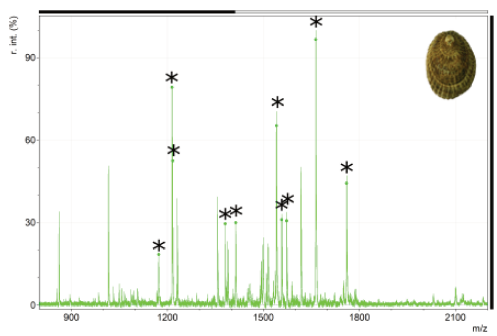
6) *Margaritifera margaritifera*



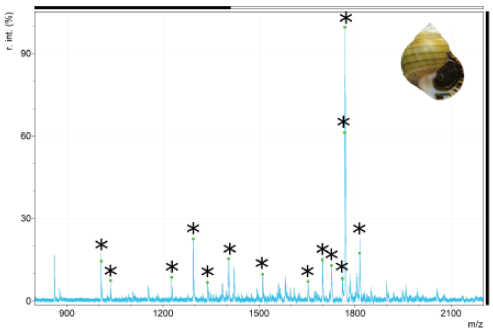
7) *Aliger gigas*



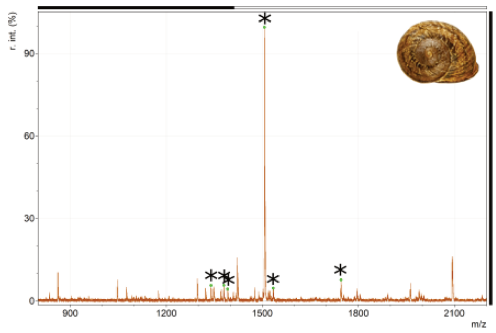
8) *Lottia gigantea*



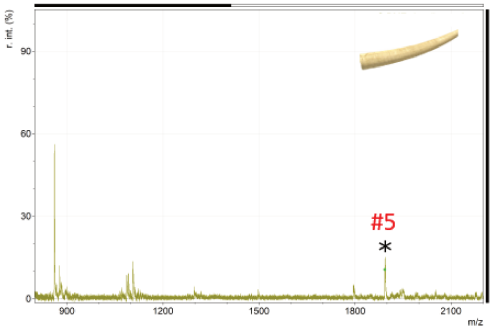
9) *Littorina littorea*



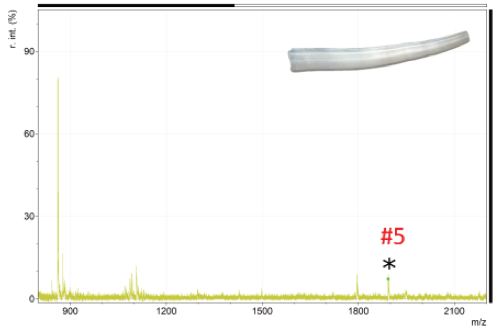
10) *Cornu aspersum*



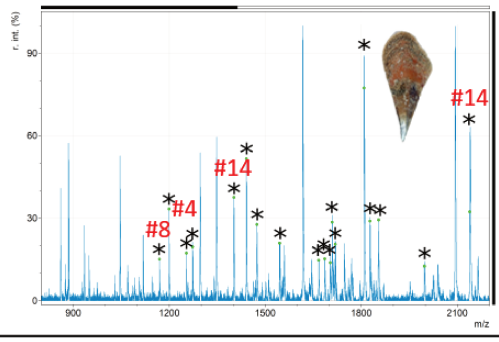
11) *Antalis* sp.



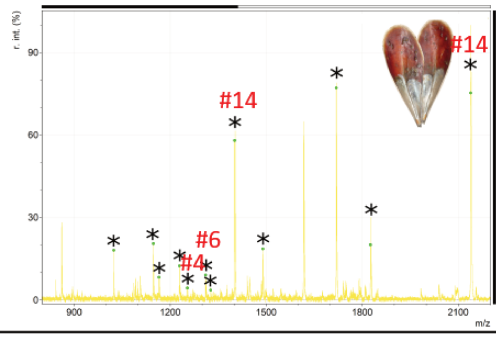
12) *Dentalium* sp.



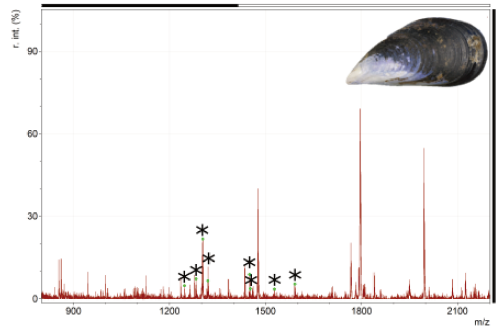
13) *Pinna nobilis* (prismatic layer)



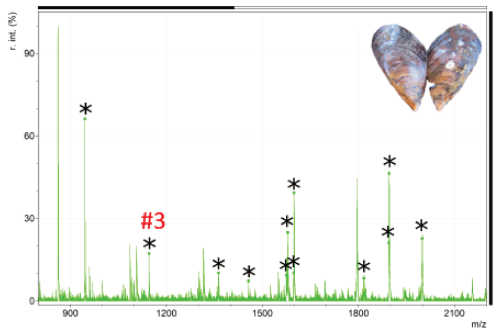
14) *Pinna nobilis* (nacreous layer)



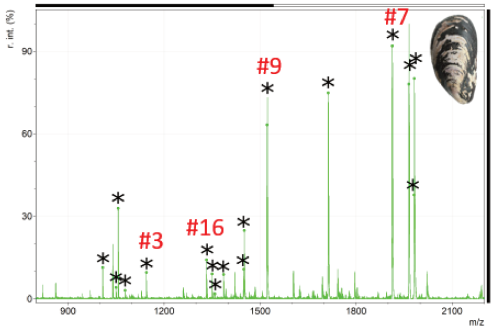
15) *Mytilus galloprovincialis*



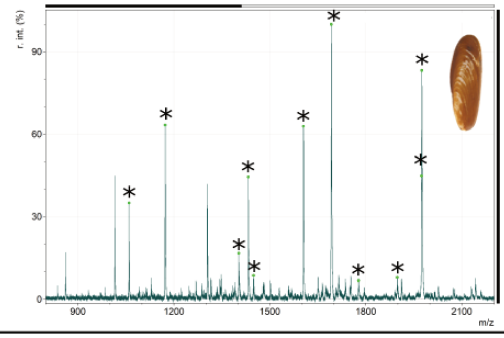
16) *Mytilus californianus*



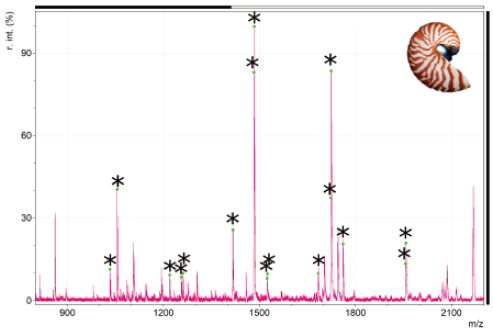
17) *Modiolus modiolus*



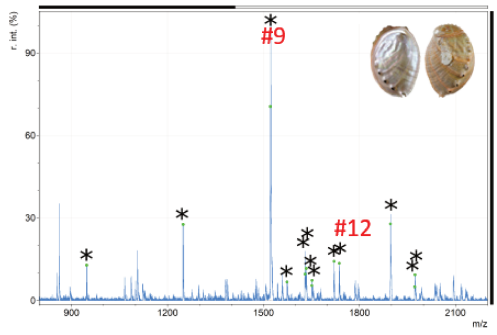
18) *Lithophaga lithophaga*



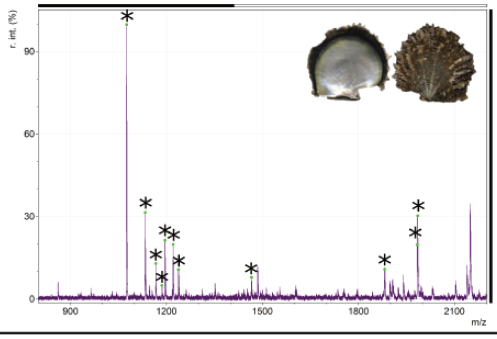
19) *Nautilus macromphalus*



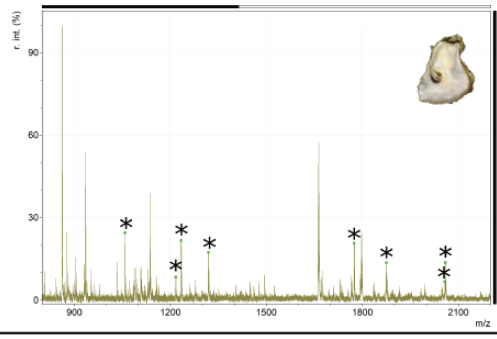
20) *Haliotis discus*



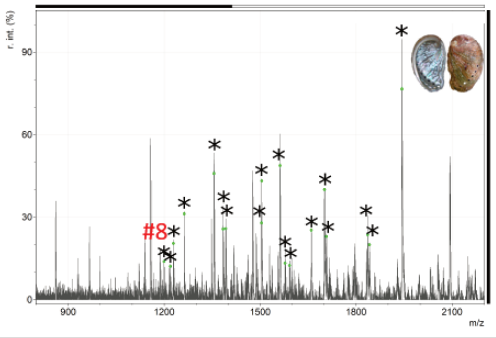
21) *Pinctada margaritifera*



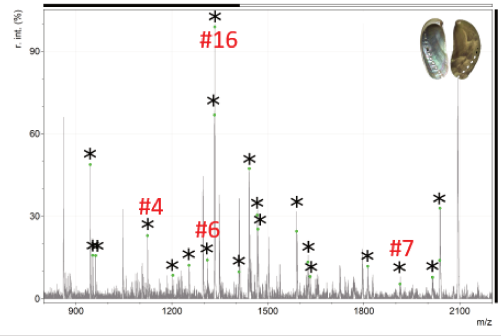
22) *Crassostrea gigas*



23) *Haliotis tuberculata*



24) *Haliotis asinina*



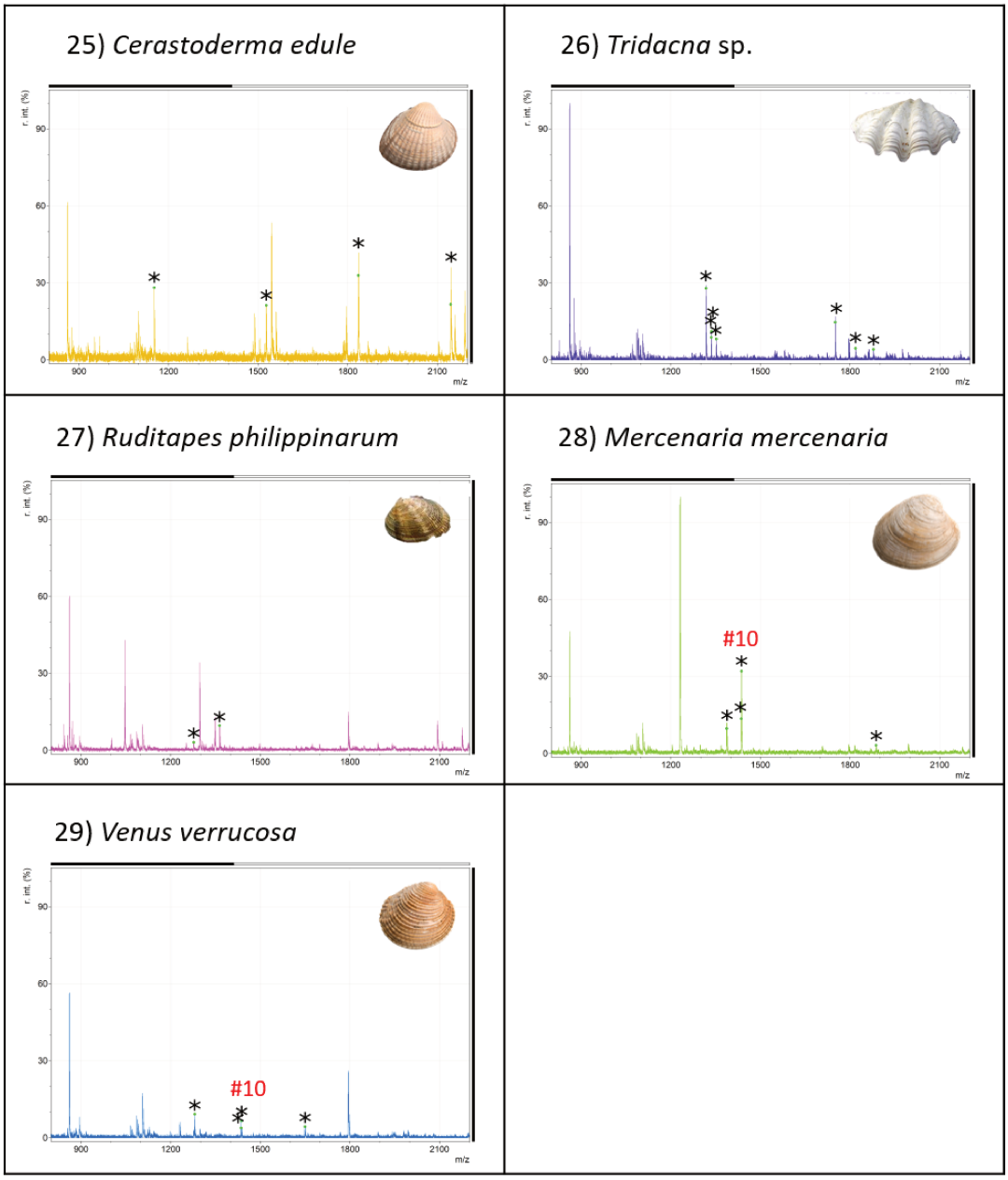


Figure 2. Shell peptide mass fingerprints (PMFs) obtained from the intracrystalline proteins of different modern shells. The data were used to create a reference shell PMF library, presented in Table 2. The asterisks indicate identified shell markers and the '#' symbol (in red) indicates those markers that were found to be shared between different taxa as detailed in Table 2. The shell PMF spectra show distinctive profiles suggesting good potential for molecular barcoding of shells.

Archaeological shell ornaments

Structural and mineralogical characterisation

X-ray micro tomography (micro CT system), scanning electron microscopy (SEM) and infrared spectroscopy in attenuated total reflectance mode (FTIR-ATR) were used to characterise the structure and the mineralogy of all the archaeological samples. The 3D images obtained by X-ray micro tomography (Figure 3) showed that sample P179 has clear mollusc shell growth lines and sample P167 showed contours of lamellae, which are characteristic of crossed-lamellar mollusc shell microstructure (raw 3D files are documented in Appendix 4, SI material 2). The X-ray scan of sample P186 showed that the ornament was originally fragmented and reconstructed, likely with an organic-based glue (as indicated by the "empty space" in the centre between the fragments).

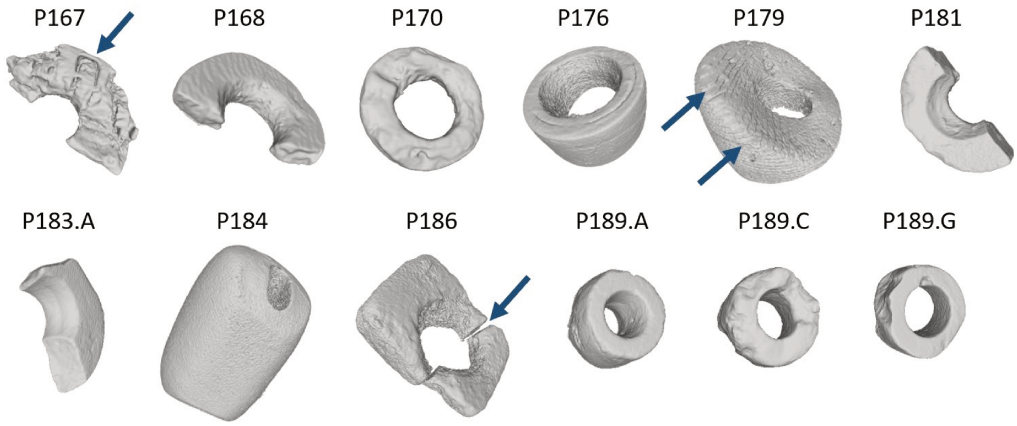


Figure 3. Archaeological ornament samples analysed in this study and their 3D structure representations obtained by X-ray micro tomography. Arrows indicate: P167 – the lamella characteristic to crossed-lamellar shell microstructure; P179 – shell growth lines; P186 – empty spaces which indicate conservation treatment: presence of organics-based glue that was used to reconstruct fragmented ornament. Full 3D MESH files are available in Appendix 4, SI material 2.

Figure 4 shows the combination of SEM images and the corresponding FTIR-ATR spectra of four archaeological samples – P167, P176, P179, P184. Table 3 summarises the obtained structural data for all of the ornaments analysed in this study (all SEM micrographs and FTIR-ATR spectra are documented in Appendix 4, SI material 3). The SEM and FTIR-ATR data showed that aragonitic crossed-lamellar mollusc shell structures were found in samples P167, P168, P170 (all of the three beads from Le Taï archaeological site) and P176, P179, P184 (shell ornaments from archaeological sites in the Yonne valley). FTIR-ATR spectra of

these samples display distinctive aragonite absorption bands: a CO in-plane bending doublet at around 712 and 700 cm^{-1} ($\nu 4$), an out-of-plane bending mid-size peak positioned at 856 cm^{-1} ($\nu 2$) and a weak peak at 1082 cm^{-1}, which is due to symmetrical stretching ($\nu 1$). The FTIR-ATR spectra of sample P175 show aragonite absorption bands and the corresponding SEM micrographs display nacreous and prismatic microstructures which are characteristic features of freshwater Unionida bivalves.

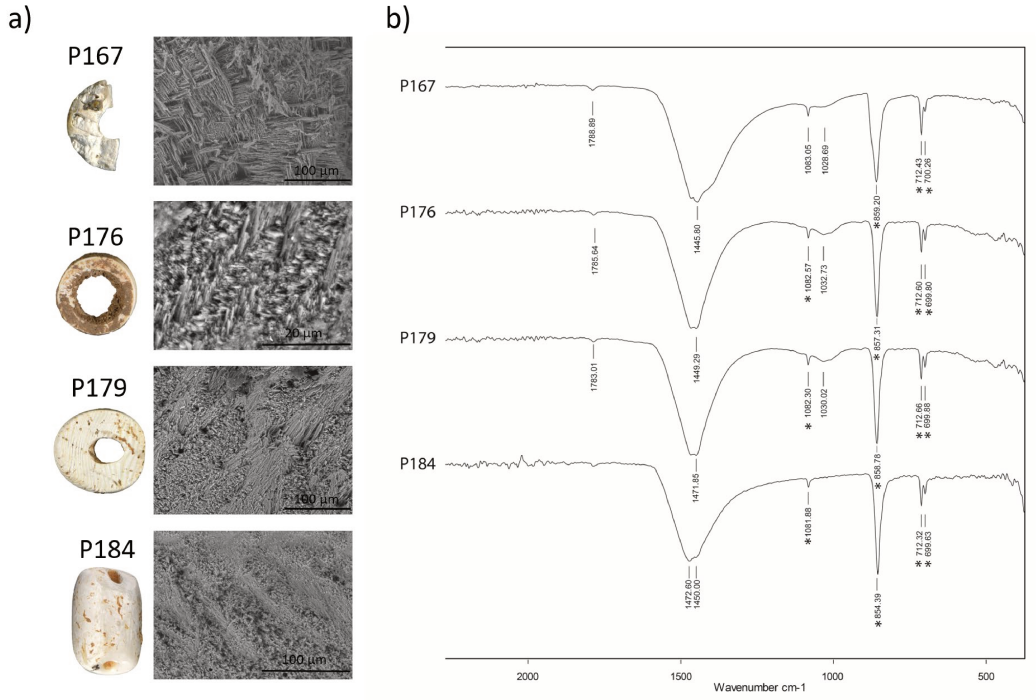


Figure 4. SEM micrographs and corresponding FTIR-ATR spectra of samples P167 (from Le Taï archaeological site), P176, P179, P184 (from Yonne valley archaeological sites): a) crossed-lamellar molluscan shell microstructure was observed in these four archaeological samples by SEM; b) corresponding FTIR-ATR spectra of these samples show distinctive aragonite bands, which are marked with asterisks.

The FTIR-ATR spectra of samples P181, P183.A, P186, P187, P189.A, P189.C, P189.G, P189.J (all from Yonne valley archaeological sites) display distinctive calcite absorption bands: a single $\nu 4$ peak at 711 cm^{-1}, a $\nu 2$ peak which is shifted to 875 cm^{-1} and absence of the $\nu 1$ stretching at 1082 cm^{-1}. The SEM micrographs showed a homogeneous surface and characteristic microstructure of these samples could not be determined. This could be due to the presence of surface contamination, or this may imply that the material is not made of biogenic carbonate, and instead of geological mineral. The difficulty in identifying the

(bio)mineral microstructure is mainly due to the absence of chemical cleaning (*e.g.* etching with mild acid) or mechanical preparation (by abrading). This was a requirement from our collaborators – to keep the archaeological samples as intact as possible prior to proteomics, *i.e.* to perform strictly non-destructive preliminary characterization.

Sample P189.L is a black fragment, a broken piece of a circular-shape ornament. The structural and mineralogical analyses by SEM and FTIR-ATR indicate that the sample is made of bone-like material (Figure 5, Table 3). The peaks at 472.37, 522.35, 1043.04 cm^{-1} can be attributed to the phosphate group vibration modes v2, v4, v3, respectively. We can observe collagen amide II and amide III bands at 1560.55 and 1224.82 cm^{-1} respectively. However, the amide I band, which, in bone appears around 1660 cm^{-1} and usually is the strongest among the three amide bands, in sample P189.L, is absent. A v(CH) bending absorption band at 2921.34 cm^{-1} which is characteristic to lipids and two broad bands at 3131.15 and 3263.88 cm^{-1} that are due to OH stretching (Marques et al., 2018; Festa et al., 2019; Stathopoulou et al., 2019) were also observed.

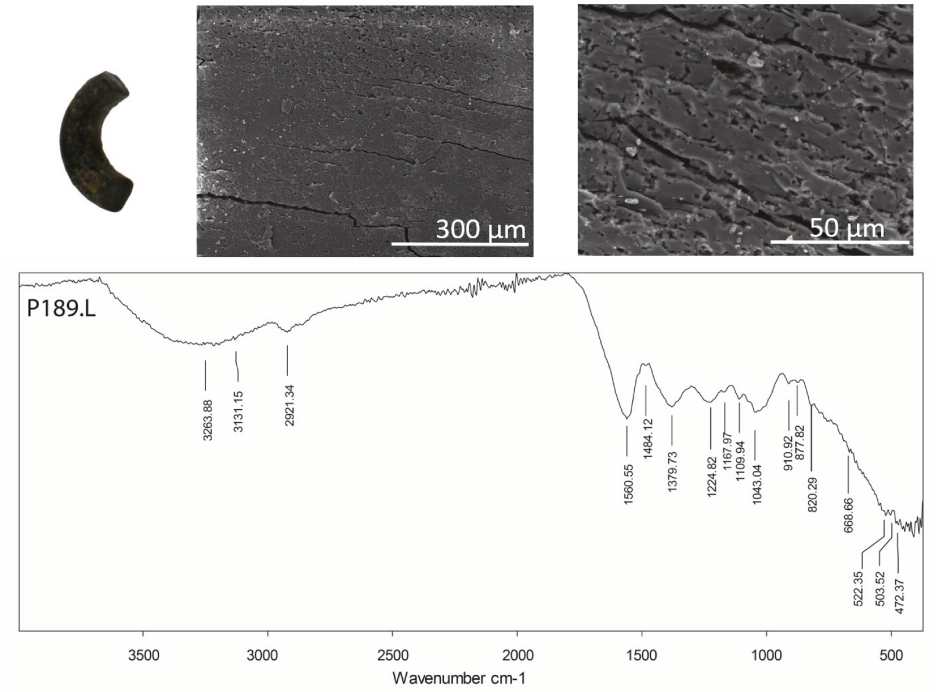


Figure 5. SEM micrographs and FTIR-ATR spectra of sample P189.L, a black fragmented ornament from Yonne valley archaeological site. It shows a bone-like structure and chemistry.

Table 3. Summary of the results obtained via structural and mineralogical analysis of the archaeological ornament samples by SEM and FTIR-ATR. Microstructure abbreviations: CL – crossed-lamellar, N – nacre, P – prismatic, ND – not determined.

Sample	Type	Total Mass (g)	Archaeological site	Micro-structure (SEM)	Mineralogy (FTIR-ATR)	FTIR-ATR marker bands (cm^{-1})	Summary
P167	Circular bead, fragment	0.056	Le Taï, Remoulins, France	CL	Aragonite	700.26, 712.43, 859.2, 1083.05	Crossed-lamellar, aragonitic mollusc shell
P168	Circular bead, fragment	0.056	Le Taï, Remoulins, France	CL	Aragonite	700.04, 712.79, 858.54, 1083.09	Crossed-lamellar, aragonitic mollusc shell
P170	Circular bead	0.023	Le Taï, Remoulins, France	CL	Aragonite	700.12, 713.03, 858.54, 1083.16	Crossed-lamellar, aragonitic mollusc shell
P175	Shell valve, fragment	1.96	Passy, France	N, P	Aragonite	698.73, 711.87, 855.89, 1082.37	Nacro-prismatic, aragonitic Unionid shell

P176	Circular bead	0.73	Villeneuve-la-Guyard, France	CL	Aragonite	699.8, 712.6, 857.31, 1082.57	Crossed-lamellar, aragonitic mollusc shell
P179	Circular/Triangular bead	1.41	Villeneuve-la-Guyard, France	CL	Aragonite	699.88, 712.66, 858.78, 1082.3	Crossed-lamellar, aragonitic mollusc shell
P181	Circular bead, fragment	0.3	Villeneuve-sur-Yonne, France	ND	Calcite	712.05, 872.11	Calcitic material, type not determined
P183.A	Circular bead, fragment	0.19	Villeneuve-sur-Yonne, France	ND	Calcite	712.25, 873.57	Calcitic material, type not determined
P184	Tubular bead	14.4	Villeneuve-la-Guyard, France	CL	Aragonite	699.63, 712.32, 854.39, 1081.88	Crossed-lamellar, aragonitic mollusc shell
P186	Trapezoid bead, glued	0.78	Vinneuf, France	ND	Calcite	712.17, 872.78	Calcitic material, type not determined

P187	Fragment	0.98	Passy, France	ND	Calcite	695.1, 712.45, 873.61, 1078.52	Calcitic material, possibly recrystallized, type not determined
P189.A	Circular bead	0.24	Passy, France	ND	Calcite	712.19, 872.52	Calcitic material, type not determined
P189.C	Circular bead	0.09	Passy, France	ND	Calcite	712.21, 872.57	Calcitic material, type not determined
P189.G	Circular bead	0.18	Passy, France	ND	Calcite	712.2, 872.54	Calcitic material, type not determined
P189.J	Circular bead, fragment	0.02	Passy, France	ND	Calcite	712.08, 871.54	Calcitic material, type not determined
P189.L	Circular bead, fragment	0.02	Passy, France	Bone	Bone-like material	472.37, 503.52, 522.35, 668.66, 820.29, 877.82 910.92, 1043.04, 1109.94,	Possibly bone, burnt

						1167.97, 1224.82, 1379.73, 1484.12, 1560.55, 2921.34 3131.15, 3263.88	

Shell palaeoproteomics

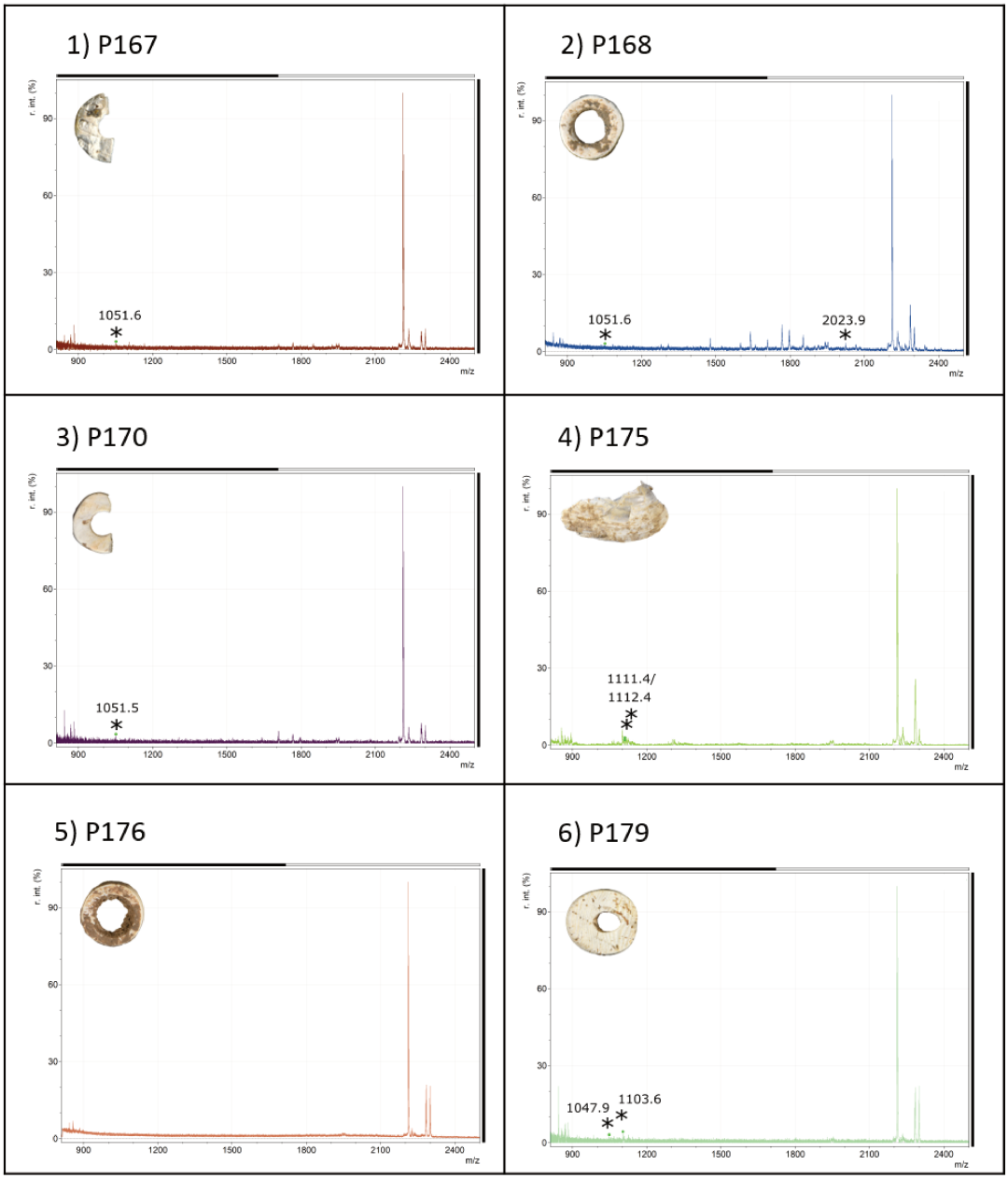
All archaeological samples selected for this study were also analysed by proteomics. Proteins were extracted and the resulting peptide digests were characterised by MALDI-TOF-MS. Figure 6 displays the obtained PMF spectra. Individual *m/z* values of each sample, which were obtained after a careful filtering of common laboratory contaminants (trypsin, keratin) are presented in Table 4.

Table 4. List of *m/z* values obtained from PMFs of archaeological samples, analysed by MALDI-TOF-MS. The third column reports possible matches to reference shell markers, collected in this study. The table lists only those samples for which characteristic *m/z* values were obtained.

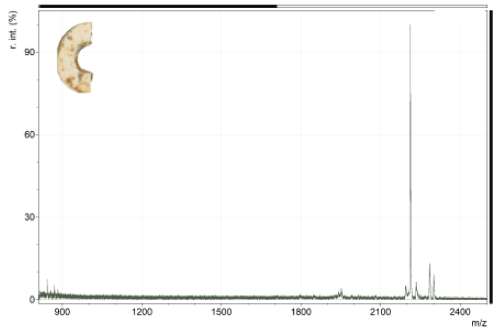
Sample	Peaks	Peptide match – ID?
P167	1051.6 (?)	n.d
P168	1051.6 (?), 2023.9	n.d
P170	1051.5 (?)	n.d
P175	1111.4, 1112.4	*Pseudunio auricularius* (match to marker at 1111.6, error 0.2 Da)
P179	1047.9, 1103.6	n.d
P187	1047.9, 1798.5, 2920.9, 3314,.6, 3767.7	n.d
P189.J	1046.9, 1101.0, 1297.0, 3117.4	n.d
P189.L	1051.8 (?), 1554.7, 1556.0, 1611.8, 1612,9, 2664.9, 2805.9	n.d

In general, the majority of sample spectra showed notable presence of common laboratory contaminants (mostly trypsin and MALDI matrix, which are used for sample preparation, but also some peaks of keratin). Samples P184, a tubular bead with an aragonitic crossed-lamellar type microstructure and P186, a calcitic trapezoid-shape bead (both from archaeological sites in Yonne valley) yielded poor quality spectra. For sample P175 (from Yonne valley archaeological site), a fragmented shell valve with a nacro-prismatic aragonitic microstructure typical of Unionida shells, a peak at *m/z* 1111.4 was identified, which likely corresponds to the shell marker of *Pseudunio auricularius* (m/z 1111.6, error 0.2 Da). For samples P167, P168, P170 (from Le Taï archaeological site), which were characterised by crossed-lamellar shell microstructures, a peak at *m/z* 1051.6 was identified in the spectra of all the three samples. However, it did not correspond to a known shell marker (considering diagenetic effects, it was also assessed with possible mass differences due to oxidation or deamidation modifications). Hence, it remains

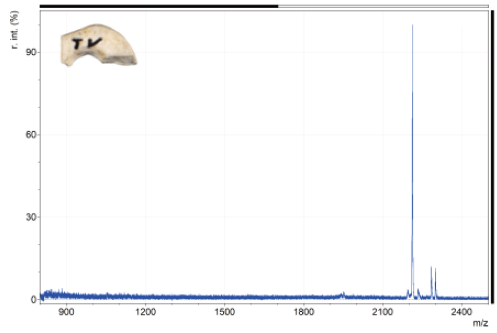
uncertain if it is a true “shared” marker, or could be due to contamination (for this reason, in the table it is noted with a question mark). Unfortunately, none of the *m/z* values obtained for the other samples could be matched to the reference shell markers collected in this study. The identified values likely represent other types of shells/biological materials or may be due to degradation products.



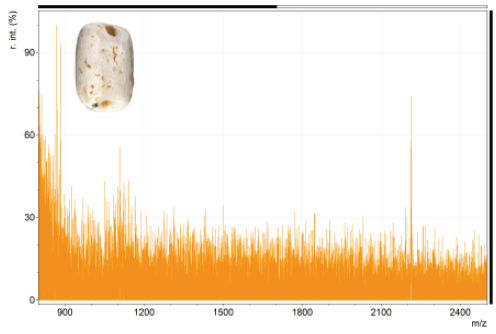
7) P181



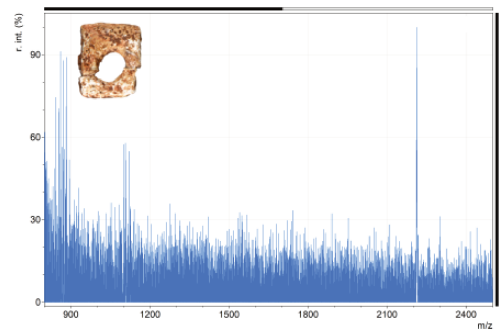
8) P183.A



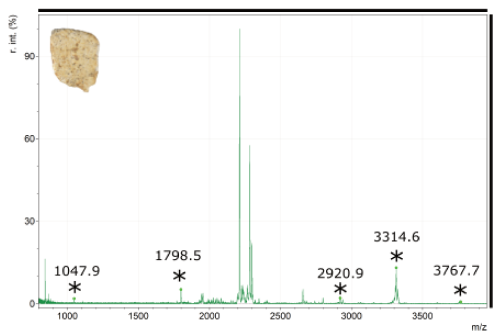
9) P184



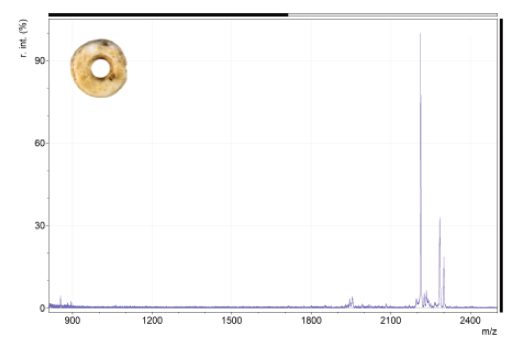
10) P186



11) P187



12) P189.A



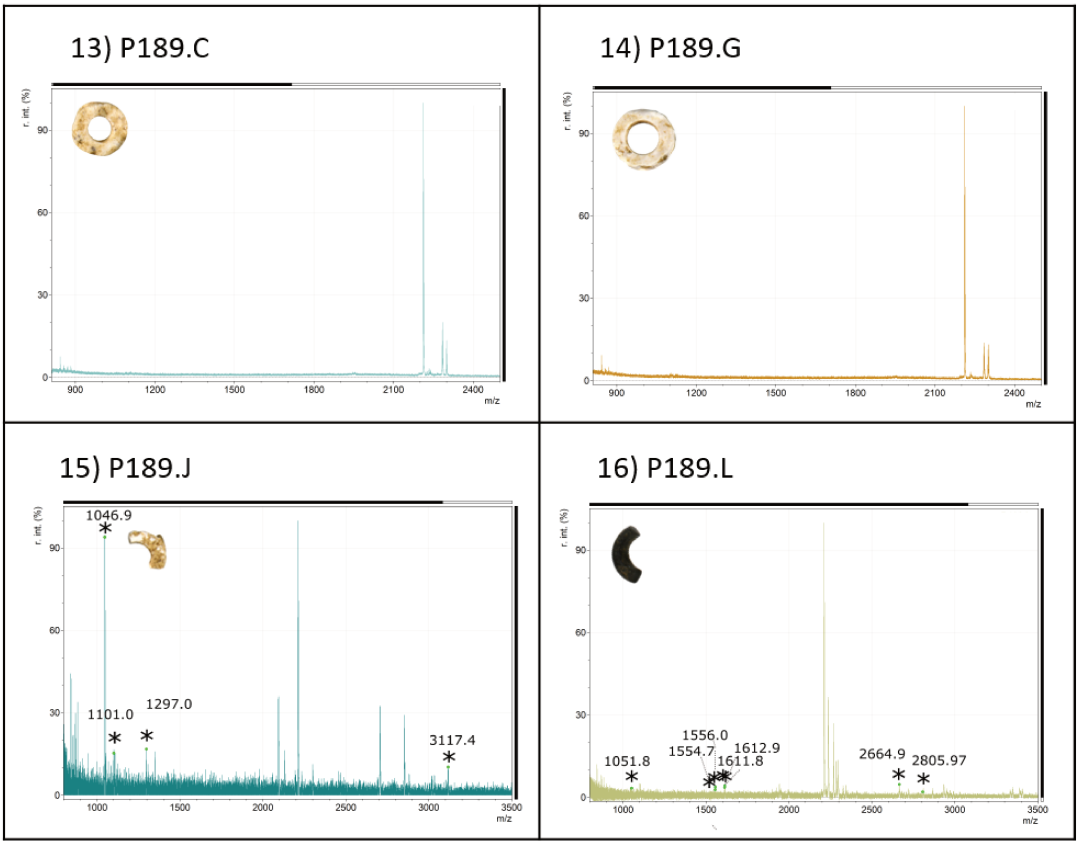


Figure 6. Peptide mass fingerprints (PMFs) obtained from the studied archaeological samples by MALDI-TOF-MS. Characteristic *m/z* values are marked in asterisks and are reported in Table 4. Note that in most of the spectra, the common laboratory contaminants (*e.g.* trypsin) are overrepresented.

Sample P189.L, which exhibits a bone-like structure (as seen by SEM and FTIR-ATR), in the PMF spectra showed presence of several low-intensity peaks, which however could not be matched to known collagen markers published in the literature (Welker et al., 2016).

Discussion

This study presents an investigation of peptide mass fingerprinting (PMF) by MALDI-TOF-MS to discriminate mollusc shells and the first method application to archaeological shell ornaments. It is the second part of the shell palaeoproteomics work (Sakalauskaite et al., 2020a) presented in Chapter 3. The primary objective of this study was to create an in-house shell PMF library by analysing the intracrystalline proteins from a large variety of shell taxa. The second objective was

to test the developed methodology for archaeological ornaments samples of unknown origin, assess protein preservation and the possibility to obtain shell identification.

First of all, the majority of modern shells analysed in this study generated good quality PMFs. For these, the intracrystalline shell proteins resulted in distinct peptide mass fingerprints (PMFs), validating their use for shell discrimination. The in-house shell PMF library reports the data on 34 shell species, 27 of which were analysed in this study and 7 that were included from our previous work (Sakalauskaite et al., 2020a). In general, I find that:

- **Nacro-prismatic shells** yielded good quality PMFs and generated a relatively high number of individual marker peaks, as seen for *Pinna nobilis*, *Haliotis* sp., *Pinctada* shell, the freshwater mother-of-pearl unionoid shells (*Unio* sp., *Margaritifera*, *Potamida*, *Anodonta*), the gastropod *Phorcus turbinatus* and the cephalopod *Nautilus macromphalus*.
- Veneridae, Cardiidae and Dentaliidae shells, all of which have dense skeletons made of **crossed-lamellar (CL) microstructures**, showed poor quality PMFs and only a few markers were identified. However, the same was not observed for other shells with CL microstructures: *Glycymeris glycymeris, Lottia gigantea, Aliger gigas, Littorina littorea, Spondylus gaederopus* which showed a high number of markers.

This suggests that the ability to yield useful PMF spectra depends on a combination of factors, including protein content and sequence characteristics. Compared to nacreous shells, CL microstructures have a smaller fraction of intracrystalline organics (up to two orders of magnitude) (Penkman et al., 2008; Sakalauskaite et al., 2020b). Extremely low quantities of intracrystalline shell proteins could have an effect on protein recovery using the SP3 purification method, because increased proportion between mineral salts and proteins can negatively impact protein binding to magnetic beads (Hughes et al., 2019). Moreover, shell proteins are characterised by peculiar sequence characteristics, most notably with the presence of repetitive, low complexity domains (RLCDs) (Marin et al., 2013). Continuous RLCD regions hamper enzymatic digestion with proteases (*e.g.* commonly used trypsin) and may result in partial extraction and low peptide recovery. In addition, the choice of the right MALDI matrix is key to the success for MALDI-TOF mass spectrometry (O'Rourke et al., 2018). In this study, the analysis was performed with the α-Cyano-4-hydroxycinnamic acid (HCCA) matrix that was previously found suitable for shell proteins of other mollusc taxa (Sakalauskaite et al., 2020a). However, as shell proteins can be very diverse among different species (Marin, 2020) the matrix may not be suitable to all. Besides, specific shell sequence characteristics, such as the lack of charged amino acids or the presence of RLCD domains with hydrophobic

amino acids, could also cause poor sample-matrix crystallization or poor ionisation effect (Dreisewerd, 2003), resulting in low-quality spectra.

A number of peptide markers shared across phylogenetically close species were identified in the studied shells. The most notable example is the peptide at *m/z* 1570.8 which was found in all of the freshwater mother-of pearl (Unionida order) shells (*U. crassus, M. margaritifera, P. litoralis, A. cygnea* from this study and also *U. pictorum, P. auricularius* from the first shell PMF study (Sakalauskaite et al., 2020a)). This marker presumably corresponds to peptide EAD(-18.01)DLALLSLLFGGR from nacre protein Hic74 (Liu et al., 2017; Sakalauskaite et al., 2019), also identified and characterised in the previous study, detailed in Chapter 3 (Sakalauskaite et al., 2020a). This supports that Unionida nacre proteins share similar sequences (Marie et al., 2017) and suggests that such peaks can be used to easily discriminate freshwater mother-of-pearl among other nacreous shells.

Proteins extracted from *Pinna nobilis* aragonitic nacreous and calcitic prismatic microstructural layers, generated considerably different PMFs. Aside from several markers that were identified in both structures, the majority of *m/z* values were distinct. This is in accordance with previous studies of *Pinna* shell proteins (Marin et al., 2000; 2005; Marin and Luquet, 2005) and is also in agreement with other works which showed that different microstructures of the same shell actually have distinct protein repertoires (Marie et al., 2012). More importantly, the data imply that additional analyses are needed to investigate and understand the variability of shell PMFs obtained from different shell layers. This is because many archaeological shell artifacts, as for example ornaments or figurines, were often heavily worked, polished or may have had entire layers removed, which alters the original microstructural proportions in the shell.

Shared peptide markers were also identified between: i) two venus clams *Mercenaria mercenaria* and *Venus verrucosa*, ii) two tusk shells *Dentalium* sp. and *Antalis* sp. iii) several nacro-prismatic shells that belong to different taxonomic groups. However, due to the lack of information on shell protein sequences for these species, it remains difficult to assess if markers represent phylogenetic relationship or shared conserved protein domains (Marie et al., 2009; Jackson et al., 2010). In the future, tandem mass spectrometry by MALDI-TOF/TOF could be used to investigate “shared” peptide markers, for example, by fragmenting the “shared” marker peptides and comparing the tandem spectra obtained for different species. MALDI-TOF/TOF also could be used to attempt a *de novo* sequencing of the shell marker peptides. Moreover, given the large size of phylum Mollusca (GBIF.org, 2020) and the diversity of shell proteins (Marin, 2020), the accuracy and

confidence of shell PMF barcoding may need the application of novel bioinformatic techniques. For example, for collagen-based samples, semi-supervised automated machine learning algorithms showed a greater sensitivity and accuracy for species discrimination (Gu and Buckley, 2018). A similar method could also be applied to shell PMFs and help to better discriminate mollusc shell taxa. In addition, in this study we have not been able to analyse several biological replicates of the same species. In the future we will aim to assess if there is any variability of intracrystalline PMFs within species and so, to securely use this approach for archaeological applications.

Compared to modern shells, the PMF analysis of archaeological samples appeared to be more difficult to interpret. The preliminary structural and mineralogical analysis of archaeological samples showed that:

- Six ornament samples had aragonitic crossed-lamellar structures, typical of mollusc shells: P167, 168, P170 (samples from Le Taï archaeological site), P176, P179, P184 (samples from archaeological sites in the Yonne valley);
- One sample, P175, was a semi-intact bivalve shell, with aragonitic nacreous and prismatic layers, typical of freshwater mother-of-pearl Unionida shell (sample from Yonne valley archaeological site);
- Eight samples showed unidentifiable calcitic microstructures: P181, P183.A, P186, P187, P189.A, P189.C, P189.G, P189.J (all samples from archaeological sites in the Yonne valley);
- One sample showed a bone-like structure (P.189.L, sample from Yonne valley archaeological site).

Considering the seven ornaments that were identified as being made of mollusc shells (P167, P168, P170, P175, P176, P179, P184), PMFs were obtained for five of them (P167, P168, P170, P175, P179, P179). In general, the spectra were dominated by trypsin, MALDI matrix and keratin peaks. The peak intensity of putative shell *m/z* values was notably lower compared to trypsin, which suggests that the protein content in archaeological samples was very low.

- Samples **P167, P168, P170** (all from Le Taï archaeological site) were identified with aragonite crossed-lamellar microstructures and in their PMFs showed presence of same *m/z* (1051.6), which however, could not be securely assigned to known shell markers. Also, the obtained *m/z* values in the PMFs of sample **P179** which was characterised by aragonitic crossed-lamellar microstructures and samples **P187, P189.J** which showed calcitic structures (all from Yonne valley archaeological sites) did not match with known shell markers.
- One marker was identified in **sample P175** (a semi-intact Unionida shell valve from Yonne valley archaeological site, *m/z* 1111.4) which likely

corresponds to shell marker of *Pseudunio auricularius* shell (sub-fossil origin). However, due to low intensity of the peak and the absence of additional markers, it remains difficult to confirm the shell species.

- Sample **P184,** which was identified as an aragonitic crossed-lamellar shell, and sample **P186,** which was characterised by a calcitic type of material (both from Yonne valley), showed poor quality spectra. Notably, both of the archaeological samples were partially coated with a reddish substance (likely ochre, the earth pigment that contains ferric oxides). Iron ions, which could have been dissolved by the chelating agent EDTA, may interfere in MALDI-MS analysis and explain the poor quality of the PMF spectra (Engelke et al., 2004).
- Several individual *m/z* values were identified in the PMF spectrum of sample **189.L** (Yonne valley), which showed a bone-like structure. However, the obtained peaks could not be assigned to known collagen markers of different species (Welker et al., 2016; Buckley, 2018). The black colour of the sample and the absence of amide I band in the FTIR-ATR spectra suggest that the fragment was likely heated or burnt, resulting in at least partial loss of organic fraction and the observed *m/z* values in PMF spectra may represent degraded collagen peptides.

There are several reasons which may explain why archaeological shell ornament samples did not result in sufficient quality PMFs when analysed by MALDI-TOF-MS. First of all, it is possible that the protein content was too low, similarly as was observed for some of the modern shells discussed above. It is also possible that strong bleaching treatment applied to archaeological samples significantly reduced the quantity of proteins and resulted in a fraction that was below the limit of detection. Secondly, the poor quality PMFs may also indicate that archaeological proteins are too degraded. Considering sample P189.L, the ornament fragment made of bone-like material, its black colour and FTIR-ATR spectra indicated that the sample was likely burnt. The data obtained in the PMF spectra is in accordance, because the obtained *m/z* values could not be matched to know collagen marker peaks. They likely represent degraded peptides. Collagen makes up to 30% of the total bone weight (Piez, 1976), but bone is an open system, hence collagen is not protected by the mineral phase (hydroxyapatite) and can easily be degraded by heating (Yates, 2013; Festa et al., 2019).

In the future, the preservation state of shell samples could be determined by additional analyses. For example, cathodoluminescence radiation could be useful to detect if the shell samples with aragonitic microstructures show the presence of diagenetically recrystallized calcite over the whole surface (Barbin, 2000; Bajnóczi et al., 2013). Moreover, amino acid racemization analysis (AAR) could be used to

directly determine protein preservation state from the archaeological samples. In another 'palaeoshellomics' study (Sakalauskaite et al., 2019) (Chapter 5) AAR data of archaeological "double-buttons" showed that one of the samples was actually burnt and indeed, HPLC-MS/MS analyses of this sample did not find any genuine sequences (Chapter 5). Finally, in the future, the extracts obtained from archaeological samples will be re-analysed by high resolution proteomics using tandem mass spectrometry (HPLC-MS/MS). This technique can provide higher sensitivity (Aebersold and Mann, 2003) and is more suited to characterise heavily degraded sequences (Demarchi et al., 2016; Welker et al., 2020). HPLC-MS/MS could help to identify archaeological proteins including peptides that may have been successfully extracted, but were poorly ionised in MALDI-TOF-MS and consequently, were not detected.

Conclusions

This work presents further advancements in developing a palaeoproteomics method for taxonomic shell barcoding based on peptide mass fingerprints (PMFs). Intracrystalline mollusc shell proteins were analysed by MALDI-TOF mass spectrometry after having been extracted from a variety of different molluscan species. The data thus collected were used to build an in-house reference shell PMF library. The obtained PMF dataset represents taxa of the four main molluscan classes, including species with the most common shell microstructures and those frequently encountered in archaeological sites. This work highlighted the distinctiveness of shell PMFs among the different studied species, confirming that taxonomic shell discrimination can be obtained using protein barcodes. A number of shared markers were identified, which were coherent with phylogenetic closeness. The shell PMF method was tested on a set of unidentified archaeological samples. A subset of samples was determined as being made of mollusc shells (using SEM and FTIR-ATR), however the obtained PMFs did not allow a simple identification of the taxa due to low protein content or because proteins were too degraded. In the future I will attempt to re-analyse these samples by high performance liquid chromatography coupled to tandem mass spectrometry (HPLC-MS/MS) which could address some of the issues encountered in this study. The study shows that additional optimization is needed to use the shell PMF method to archaeological shell samples. At the same time, it highlights a promising future potential for the discrimination of modern shells, especially to determine the different types of nacreous shells, as well as thick-valve shells common in coastal areas such as *Glycymeris*, *Spondylus* and *Pecten*.

References

Aebersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. Nature 422, 198–207.

Bajnóczi, B., Schöll-Barna, G., Kalicz, N., Siklósi, Z., Hourmouziadis, G.H., Ifantidis, F., Kyparissi-Apostolika, A., Pappa, M., Veropoulidou, R., Ziota, C., 2013. Tracing the source of Late Neolithic Spondylus shell ornaments by stable isotope geochemistry and cathodoluminescence microscopy. Journal of archaeological science 40, 874–882.

Barbin, V., 2000. Cathodoluminescence of Carbonate Shells: Biochemical vs Diagenetic Process. In: Pagel, M., Barbin, V., Blanc, P., Ohnenstetter, D. (Eds.), Cathodoluminescence in Geosciences. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 303–329.

Bar-Yosef, D., 2005. Archaeomalacology: Molluscs in former environments of human behaviour. Oxbow Books.

Bar-Yosef Mayer, D.E., 2018. Shell ornaments and artifacts in Neolithic Cyprus and correlations with other Mediterranean regions. Quaternary international: the journal of the International Union for Quaternary Research 464, 206–215.

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

Bertin, L., 2015. Exploitation of mother of pearl in the Middle Ages, Clos d'Ugnac archaeological site (Pennautier, Aude, France): Malacological study, consumption, exploitation and utilization of the nacre. Quaternary international: the journal of the International Union for Quaternary Research 375, 145–152.

Boggild, O.B., 1930. The shell structure of the mollusks. A.F. Host, Kjobenhavn.

Bonnardin, S., 2009. La parure funéraire au Néolithique ancien dans les bassins parisien et rhénan: Rubané, Hinkelstein et Villeneuve-Saint-Germain. Société préhistorique française.

Bonnardin, S., 2014. La parure: Nécropoles Préhistoriques et Protohistoriques de Passy (Yonne) N3P (Yonne). Projet Collectif de recherche. CEPAM - Culture et Environnements, Préhistoire, Antiquité, Moyen-Age.

Brandl, B., 1984. The Engraved Tridacna-Shell Discs*. Anatolian Studies 34, 15–41.

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

Buckley, M., Collins, M.J., 2011. Collagen survival and its use for species identification in Holocene-lower Pleistocene bone fragments from British

archaeological and paleontological sites. Antiqua 1, 1.
Buckley, M., Collins, M., Thomas-Oates, J., Wilson, J.C., 2009. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid communications in mass spectrometry: RCM 23, 3843–3854.
Caro, J., Manen, C., Bruxelles, L., Convertini, F., Perrin, T., Sordoillet, D., 2014. Les productions céramiques du Néolithique ancien du Taï (Remoulins, Gard). Approche spatiale, caractérisation typo-technologique et attribution culturelle. In: Sénépart, I., Leandri, F., Cauliez, J., Perrin, T., Thirault, É. (Eds.), Chronologie de La Préhistoire Récente Dans Le Sud de La France. Acquis 1992-2012. Presented at the 10e Rencontres méridionales de Préhistoire récente, Archives d'Écologie Préhistorique, Toulouse, pp. 511–521.
Carter, J.G. (Ed.), 1990. Skeletal Biomineralization: Pattern, Processes, and Evolutionary Trends. Amer Geophysical Union.
Cartwright, C., 2003. Unusual Use of Freshwater Mussel (*Unio* sp.) Shells during the Early Bronze Age at Tell es-Sa'idiyeh, Jordan. Environmental Archaeology.
Chapman, J., Gaydarska, B., 2015. *Spondylus Gaederopus/Glycymeris* Exchange Networks in the European Neolithic and Chalcolithic. In: Fowler, C., Harding, J., Hofmann, D. (Eds.), The Oxford Handbook of Neolithic Europe. Oxford University Press.
Charpentier, V., Phillips, C.S., Méry, S., 2012. Pearl fishing in the ancient world: 7500 BP. Arabian Archaeology and Epigraphy 23, 1–6.
Cleland, T.P., Schroeter, E.R., 2018. A Comparison of Common Mass Spectrometry Approaches for Paleoproteomics. Journal of proteome research 17, 936–945.
Colonese, A.C., Mannino, M.A., Bar-Yosef Mayer, D.E., Fa, D.A., Finlayson, J.C., Lubell, D., Stiner, M.C., 2011. Marine mollusc exploitation in Mediterranean prehistory: An overview. Quaternary international: the journal of the International Union for Quaternary Research 239, 86–103.
Demarchi, B., Hall, S., Roncal-Herrero, T., Freeman, C.L., Woolley, J., Crisp, M.K., Wilson, J., Fotakis, A., Fischer, R., Kessler, B.M., Rakownikow Jersie-Christensen, R., Olsen, J.V., Haile, J., Thomas, J., Marean, C.W., Parkington, J., Presslee, S., Lee-Thorp, J., Ditchfield, P., Hamilton, J.F., Ward, M.W., Wang, C.M., Shaw, M.D., Harrison, T., Domínguez-Rodrigo, M., MacPhee, R.D.E., Kwekason, A., Ecker, M., Kolska Horwitz, L., Chazan, M., Kröger, R., Thomas-Oates, J., Harding, J.H., Cappellini, E., Penkman, K., Collins, M.J., 2016. Protein sequences bound to mineral surfaces persist into deep time. eLife 5. doi:10.7554/eLife.17092
Demarchi, B., Presslee, S., Sakalauskaite, J., Fischer, R., Best, J., 2020. The role of birds at Çatalhöyük revealed by the analysis of eggshell. Quaternary international: the journal of the International Union for Quaternary Research.

doi:10.1016/j.quaint.2020.02.009
Dimitrijević, V., Tripković, B., 2006. *Spondylus* and *Glycymeris* bracelets: trade reflections at Neolithic Vinča-Belo Brdo. Documenta praehistorica 33, 237–252.
Dreisewerd, K., 2003. The Desorption Process in MALDI. Chemical reviews 103, 395–426.
Engelke, C.J., Knicker, H., Kögel-Knabner, I., 2004. Iron interference on matrix-assisted laser desorption/ionisation time-of-flight mass spectra of condensed tannins. Environmental chemistry letters 2, 71–75.
Festa, G., Andreani, C., Baldoni, M., Cipollari, V., Martínez-Labarga, C., Martini, F., Rickards, O., Rolfo, M.F., Sarti, L., Volante, N., Senesi, R., Stasolla, F.R., Parker, S.F., Vassalo, A.R., Mamede, A.P., Batista de Carvalho, L.A.E., Marques, M.P.M., 2019. First analysis of ancient burned human skeletal remains probed by neutron and optical vibrational spectroscopy. Science advances 5, eaaw1292.
GBIF.org, Global Biodiversity Information Facility. Available from: https://www.gbif.org
Gu, M., Buckley, M., 2018. Semi-supervised machine learning for automated species identification by collagen peptide mass fingerprinting. BMC bioinformatics 19, 241.
Henry, D.G., Watson, J.S., John, C.M., 2017. Assessing and calibrating the ATR-FTIR approach as a carbonate rock characterization tool. Sedimentary geology 347, 36–52.
Hoffmann, E. de, Stroobant, V., 2013. Mass Spectrometry: Principles and Applications, 3rd ed. Wiley-Interscience.
Hughes, C.S., Moggridge, S., Müller, T., Sorensen, P.H., Morin, G.B., Krijgsveld, J., 2019. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nature protocols 14, 68–85.
Jackson, D.J., McDougall, C., Woodcroft, B., Moase, P., Rose, R.A., Kube, M., Reinhardt, R., Rokhsar, D.S., Montagnani, C., Joubert, C., Piquemal, D., Degnan, B.M., 2010. Parallel evolution of nacre building gene sets in molluscs. Molecular biology and evolution 27, 591–608.
Jonuks, T., Oras, E., Best, J., Demarchi, B., Mänd, R., Presslee, S., Vahur, S., 2018. Multi-method Analysis of Avian Eggs as Grave Goods: Revealing Symbolism in Conversion Period Burials at Kukruse, NE Estonia. Environmental Archaeology 23, 109–122.
Kirby, J., 2018. Treasures from the Sea: Sea Silk and Shellfish Purple Dye in Antiquity. Ambix 65, 405–407.
Liu, X., Jin, C., Wu, L., Dong, S., Zeng, S., Li, J., 2017. Hic74, a novel alanine and glycine rich matrix protein related to nacreous layer formation in the mollusc *Hyriopsis cumingii*. Aquaculture and Fisheries 2, 119–123.

Loftus, E., Rogers, K., Lee-Thorp, J., 2015. A simple method to establish calcite:aragonite ratios in archaeological mollusc shells. Journal of Quaternary Science 30, 731–735.

Marie, B., Le Roy, N., Marie, A., Dubost, L., Milet, C., Bedouet, L., Becchi, M., Zanella-Cléon, I., Jackson, D., Degnan, B., Luquet, G., Marin, F., 2009. Nacre Evolution: A Proteomic Approach. MRS Proceedings 1187. doi:10.1557/PROC-1187-KK01-03

Marie, B., Joubert, C., Tayalé, A., Zanella-Cléon, I., Belliard, C., Piquemal, D., Cochennec-Laureau, N., Marin, F., Gueguen, Y., Montagnani, C., 2012. Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. Proceedings of the National Academy of Sciences of the United States of America 109, 20986–20991.

Marie, B., Arivalagan, J., Mathéron, L., Bolbach, G., Berland, S., Marie, A., Marin, F., 2017. Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the *Unionoida Elliptio complanata* and *Villosa lienosa*. Journal of the Royal Society, Interface / the Royal Society 14. doi:10.1098/rsif.2016.0846

Marin, F., 2020. Mollusc shellomes: Past, present and future. Journal of structural biology 212, 107583.

Marin, F., Luquet, G., 2005. Molluscan biomineralization: The proteinaceous shell constituents of *Pinna nobilis* L. Materials Science and Engineering: C 25, 105–111.

Marin, F., Corstjens, P., Gaulejac, B. de, Vrind-De Jong, E. de, Westbroek, P., 2000. Mucins and molluscan calcification. Molecular characterization of mucoperlin, a novel mucin-like protein from the nacreous shell layer of the fan mussel *Pinna nobilis* (Bivalvia, pteriomorphia). The Journal of biological chemistry 275, 20667–20675.

Marin, F., Amons, R., Guichard, N., Stigter, M., Hecker, A., Luquet, G., Layrolle, P., Alcaraz, G., Riondet, C., Westbroek, P., 2005. Caspartin and calprismin, two proteins of the shell calcitic prisms of the Mediterranean fan mussel *Pinna nobilis*. The Journal of biological chemistry 280, 33895–33908.

Marin, F., Marie, B., Hamada, S.B., Ramos-Silva, P., Le Roy, N., Guichard, N., Wolf, S.E., Montagnani, C., Joubert, C., Piquemal, D., Saulnier, D., Gueguen, Y., 2013. “Shellome”: Proteins involved in mollusk shell biomineralization-diversity, functions. In: Watabe, S., Maeyama, K., Nagasawa, H. (Eds.), International Symposium on Pearl Research. Terrapub Tokyo, pp. 149–166.

Marques, M.P.M., Mamede, A.P., Vassalo, A.R., Makhoul, C., Cunha, E., Gonçalves, D., Parker, S.F., Batista de Carvalho, L.A.E., 2018. Heat-induced Bone Diagenesis Probed by Vibrational Spectroscopy. Scientific reports 8, 15935.

Mayer, D.E.B.-Y., 2005. The Exploitation of Shells as Beads in the Palaeolithic and

Neolithic of the Levant. Paléorient 31, 176–185.
Minniti, C., 2005. Shells at the Bronze Age settlement of Coppa Nevigata (Apulia, Italy). In: Bar-Yosef Mayer, D.E. (Ed.), Archaeomalacology: Molluscs in Former Environments of Human Behaviour. Presented at the 9th ICAZ Conference, Oxbow Books, Oxford, pp. 71–81.
Niedermeyer, T.H.J., Strohalm, M., 2012. mMass as a software tool for the annotation of cyclic peptide tandem mass spectra. PloS one 7, e44913.
O'Rourke, M.B., Djordjevic, S.P., Padula, M.P., 2018. The quest for improved reproducibility in MALDI mass spectrometry. Mass spectrometry reviews 37, 217–228.
Paulsen, A.C., 1974. The Thorny Oyster and the Voice of God: *Spondylus* and *Strombus* in Andean Prehistory. American antiquity 39, 597–607.
Penkman, K.E.H., Kaufman, D.S., Maddy, D., Collins, M.J., 2008. Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells. Quaternary geochronology 3, 2–25.
Piez, K.A., 1976. Biochemistry of collagen. In: Ramachandran, G.N., Reddi, A.H. (Eds.), Biochemistry of Collagen. Springer, Boston, MA, pp. 1–44.
Pothier Bouchard, G., Riel-Salvatore, J., Negrino, F., Buckley, M., 2020. Archaeozoological, taphonomic and ZooMS insights into The Protoaurignacian faunal record from Riparo Bombrini. Quaternary international: the journal of the International Union for Quaternary Research. doi:10.1016/j.quaint.2020.01.007
Presslee, S., Wilson, J., Woolley, J., Best, J., Russell, D., Radini, A., Fischer, R., Kessler, B., Boano, R., Collins, M., Demarchi, B., 2017. The identification of archaeological eggshell using peptide markers. STAR: Science & Technology of Archaeological Research 3, 89–99.
Prestreau, M., 1992. Le site Néolithique et protohistorique des falaises de prépoux à Villeneuve-La-Guyard (Yonne). Gallia préhistoire 34.
Ridout-Sharpe, J., 2017. Shell ornaments, icons and other artefacts from the eastern Mediterranean and Levant. In: Allen, M.J. (Ed.), Molluscs in Archaeology: Methods, Approaches and Applications. Oxbow Books, pp. 290–307.
Rigaud, S., Errico, F. d', Vanhaeren, M., 2015. Ornaments reveal resistance of North European cultures to the spread of farming. PloS one 10, e0121166.
Sakalauskaite, J., Andersen, S.H., Biagi, P., Borrello, M.A., Cocquerez, T., Colonese, A.C., Dal Bello, F., Girod, A., Heumüller, M., Koon, H., Mandili, G., Medana, C., Penkman, K.E., Plasseraud, L., Schlichtherle, H., Taylor, S., Tokarski, C., Thomas, J., Wilson, J., Marin, F., Demarchi, B., 2019. "Palaeoshellomics" reveals the use of freshwater mother-of-pearl in prehistory. eLife 8. doi:10.7554/eLife.45644
Sakalauskaite, J., Marin, F., Pergolizzi, B., Demarchi, B., 2020a. Shell

palaeoproteomics: First application of peptide mass fingerprinting for the rapid identification of mollusc shells in archaeology. Journal of proteomics 227, 103920.

Sakalauskaite, J., Plasseraud, L., Thomas, J., Albéric, M., Thoury, M., Perrin, J., Jamme, F., Broussard, C., Demarchi, B., Marin, F., 2020b. The shell matrix of the european thorny oyster, *Spondylus gaederopus*: microstructural and molecular characterization. Journal of structural biology 211, 107497.

Stathopoulou, E., Phoca Cosmetatou, N., Theodoropoulou, T., Mallouchou, M., Margariti, E., Psycharis, V., 2019. Origin of archaeological black bones within a waterlogged context: A multidisciplinary approach. Palaeogeography, palaeoclimatology, palaeoecology 534, 109334.

Stewart, J.R.M., Allen, R.B., Jones, A.K.G., Kendall, T., Penkman, K.E.H., Demarchi, B., O'Connor, T., Collins, M.J., 2014. Walking on Eggshells: A Study of Egg Use in Anglo-Scandinavian York Based on Eggshell Identification Using ZooMS. International Journal of Osteoarchaeology 24, 247–255.

Taylor, J.D., Kennedy, W.J., Hall, A., 1969. The shell structure and mineralogy of Bivalvia: introduction, *Nuculacea-Trigonacea*, Bulletin of the British Museum (Natural History), Zoology, Supplement. The British Museum (Natural History), London.

IUCN 2020. The IUCN Red List of Threatened Species. Version 2020-2. https://www.iucnredlist.org. Accessed on 10 October 2020.

Vanhaeren, M., Errico, F. d', 2006. Aurignacian ethno-linguistic geography of Europe revealed by personal ornaments. Journal of archaeological science 33, 1105–1128.

Welker, F., 2018. Palaeoproteomics for human evolution studies. Quaternary science reviews 190, 137–147.

Welker, F., Hajdinjak, M., Talamo, S., Jaouen, K., Dannemann, M., David, F., Julien, M., Meyer, M., Kelso, J., Barnes, I., Brace, S., Kamminga, P., Fischer, R., Kessler, B.M., Stewart, J.R., Pääbo, S., Collins, M.J., Hublin, J.-J., 2016. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. Proceedings of the National Academy of Sciences of the United States of America 113, 11162–11167.

Welker, F., Ramos-Madrigal, J., Gutenbrunner, P., Mackie, M., Tiwary, S., Rakownikow Jersie-Christensen, R., Chiva, C., Dickinson, M.R., Kuhlwilm, M., Manuel, M. de, Gelabert, P., Martinón-Torres, M., Margvelashvili, A., Arsuaga, J.L., Carbonell, E., Marques-Bonet, T., Penkman, K., Sabidó, E., Cox, J., Olsen, J.V., Lordkipanidze, D., Racimo, F., Lalueza-Fox, C., Bermúdez de Castro, J.M., Willerslev, E., Cappellini, E., 2020. The dental proteome of Homo antecessor. Nature 580, 235–238.

WoRMS Editorial Board (2020). World Register of Marine Species. Available from https://www.marinespecies.org at VLIZ. Accessed 2020-10-14.

doi:10.14284/170.
Yates, J.F., 2013. It will not be possible to use Zooarchaeology by Mass Spectrometry (ZooMS) to identify species in samples of cremated bone that have been burnt higher than 155 ºC (BSc Bioarchaeology). University of York.
Zilhão, J., Angelucci, D.E., Badal-García, E., Errico, F. d', Daniel, F., Dayet, L., Douka, K., Higham, T.F.G., Martínez-Sánchez, M.J., Montes-Bernárdez, R., Murcia-Mascarós, S., Pérez-Sirvent, C., Roldán-García, C., Vanhaeren, M., Villaverde, V., Wood, R., Zapata, J., 2010. Symbolic use of marine shells and mineral pigments by Iberian Neandertals. Proceedings of the National Academy of Sciences of the United States of America 107, 1023–1028.

Chapter 5

'Palaeoshellomics' reveals the use of freshwater mother-of-pearl in prehistory

Digest

Since prehistoric times, mollusc shells have been used as raw materials for making ornaments, for example crafted into beads, pendants, bracelets, and other elaborate shapes. Shell ornaments were used for symbolic purposes, to display status, wealth and prestige. Tracking the biological origin of the shells used for ornament-making provides some keys to interpret their cultural significance, including cross-cultural interactions, past trading exchanges and, ultimately, how prehistoric people exploited natural resources. However, due to the small sample size and the loss of morphological features, the identification of shell types used for crafting small ornaments is difficult, and may introduce a bias in archaeological interpretations.

Miniature white mother-of-pearl "double-button" ornaments ("Doppelknöpfe"), analysed in this study, are a typical example of heavily worked raw materials that have lost all traces of their taxonomic origin, and consequently, of their geographical origin. They were found between 4200–3800 BCE in different archaeological sites spanning all across Europe – a Neolithic village in Germany, a coastal shell midden site in Denmark and a cave site in Transylvania, Romania. All of them were highly standardised. Marine shells found in inland archaeological sites are frequently considered as the exotic raw materials of choice, typically viewed as more "prestigious". Hence, one hypothesis suggested that "double-buttons" may have been made of marine shells, abundant in coastal Danish sites, and transported further south. However, the theory remained debated (and untested) because accurate identification of shells was simply not possible based on their morphology and structure.

In the recent decade, palaeoproteomics, the analysis of ancient proteins, has enabled the characterization and identification of numerous archaeological and palaeontological artifacts of animal origin. The use of high performance liquid chromatography coupled to tandem mass spectrometry (HPLC-MS/MS) provides a clear advantage in characterising complex proteomes. Moreover, the technique is extremely sensitive even to extremely small protein quantities, such as those encountered in degraded mollusc shells.

In this study, I employed an integrated palaeoproteomics approach combined with structural, mineralogical, and geochemical analyses to study the “double-button” ornaments of the three localities mentioned above. The main goal was to find out if the shell proteins preserved in the “double-buttons” allowed us to identify their biological (taxonomic) origin and therefore to address the longstanding issue of their provenance. The microstructural analysis showed that all three sets of “double-buttons” were composed of aragonitic prismatic and nacreous layers. Carbon and oxygen stable isotope data suggested freshwater origin of the shell raw materials and amino acid racemization data confirmed that shells were broadly contemporary to the makers (*i.e*. not of fossil origin). However, it was shell palaeoproteomics that unequivocally demonstrated that ornaments were made of freshwater mother-of-pearl shells. Multiple bioinformatic data analyses were employed to compare ancient protein sequences from the ornaments and a set of reference shells, all of which consistently assigned “double-button” proteins to those of Unionoida bivalve shells.

This work is immensely significant because it marks the first palaeoproteomics identification of archaeological mollusc shells, and indeed, of any invertebrate system. Shell palaeoproteomics allowed to identify the biological (taxonomic) origin of miniature “double-button” pearls and showed that biomolecular approaches are key in order to address similar archaeological questions. I also note that further identification of molluscan taxa at genus or species levels, for the moment, remains out of reach, mainly because of limited molecular data for different species. However, in the future, this may be achieved via alternative data analyses routes, for example, by developing database-independent bioinformatic approaches. Finally, our findings revealed that freshwater mother-of-pearl shells were prized and constituted a raw material of high quality, highlighting the importance of unionoid shells in Europe’s prehistory.

eLIFE elifesciences.org

RESEARCH ARTICLE

'Palaeoshellomics' reveals the use of freshwater mother-of-pearl in prehistory

Jorune Sakalauskaite[1,2], Søren H Andersen[3], Paolo Biagi[4], Maria A Borrello[5], Théophile Cocquerez[2], André Carlo Colonese[6], Federica Dal Bello[7], Alberto Girod[8], Marion Heumüller[9], Hannah Koon[10], Giorgia Mandili[7,11], Claudio Medana[7], Kirsty EH Penkman[12], Laurent Plasseraud[13], Helmut Schlichtherle[14], Sheila Taylor[12], Caroline Tokarski[15†], Jérôme Thomas[2], Julie Wilson[16], Frédéric Marin[2], Beatrice Demarchi[1,6*]

[1]Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy; [2]UMR CNRS 6282 Biogéosciences, University of Burgundy-Franche-Comté, Dijon, France; [3]Moesgaard Museum, Højbjerg, Denmark; [4]Department of Asian and North African Studies, University of Ca' Foscari, Venice, Italy; [5]Independent researcher, Genève, Switzerland; [6]Department of Archaeology, University of York, Heslington, United Kingdom; [7]Department of Molecular Biotechnology and Health Sciences, University of Turin, Turin, Italy; [8]Italian Malacological Society, Sorengo, Switzerland; [9]Niedersächsisches Landesamt für Denkmalpflege, Hannover, Germany; [10]School of Archaeological and Forensic Sciences, University of Bradford, Bradford, United Kingdom; [11]Centre for Experimental and Clinical Studies, University of Turin, Turin, Italy; [12]Department of Chemistry, University of York, Heslington, United Kingdom; [13]Institute of Molecular Chemistry, ICMUB UMR CNRS 6302, University of Burgundy-Franche-Comté, Dijon, France; [14]Landesamt für Denkmalpflege im Regierungspräsidium Stuttgart, Gaienhofen, Germany; [15]Miniaturization for Synthesis, Analysis & Proteomics, USR CNRS 3290, University of Lille, Lille, France; [16]Department of Mathematics, University of York, Heslington, United Kingdom

***For correspondence:** beatrice@palaeo.eu

Present address: †Institute of Chemistry & Biology of Membranes & Nano-objects, UMR CNRS 5248, Proteome Platform, University of Bordeaux, Bordeaux, France

Competing interests: The authors declare that no competing interests exist.

Funding: See page 20

Received: 30 January 2019
Accepted: 20 April 2019
Published: 07 May 2019

Reviewing editor: Jessica C Thompson, Yale University, United States

Abstract The extensive use of mollusc shell as a versatile raw material is testament to its importance in prehistoric times. The consistent choice of certain species for different purposes, including the making of ornaments, is a direct representation of how humans viewed and exploited their environment. The necessary taxonomic information, however, is often impossible to obtain from objects that are small, heavily worked or degraded. Here we propose a novel biogeochemical approach to track the biological origin of prehistoric mollusc shell. We conducted an in-depth study of archaeological ornaments using microstructural, geochemical and biomolecular analyses, including 'palaeoshellomics', the first application of palaeoproteomics to mollusc shells (and indeed to any invertebrate calcified tissue). We reveal the consistent use of locally-sourced freshwater mother-of-pearl for the standardized manufacture of 'double-buttons'. This craft is found throughout Europe between 4200–3800 BCE, highlighting the ornament-makers' profound knowledge of the biogeosphere and the existence of cross-cultural traditions.
DOI: https://doi.org/10.7554/eLife.45644.001

Introduction

The selection of shell as a raw material by prehistoric populations implies that it possesses an inherent attractiveness that makes it suitable for displaying social connections, wealth and prestige (***Bar-***

eLife digest Just like people do today, prehistoric humans liked to adorn themselves with beautiful objects. Shells, from creatures like clams and snails, were used to decorate clothing or worn as jewelry at least as far back as 100,000 years ago. Later people used shells as the raw materials to make beads or bracelets. Learning where the shells came from may help scientists understand why prehistoric people chose certain shells and not others. It may also offer clues about how they used natural resources and the cultural significance of these objects. But identifying the shells is difficult because they lose many of their original distinctive features when worked into ornaments.

New tools that use DNA or proteins to identify the raw materials used to craft ancient artifacts have emerged that may help. So far, scientists have mostly used these genomic and proteomic tools to identify the source of materials made from animal hide, ivory or bone – where collagen is the most abundant protein molecule. Yet it is more challenging to extract and characterize proteins or genetic material from mollusc shells. This is partly because the amount of proteins in shells is at least 300 times lower than in bone, and also because the makeup of proteins in shells is not as well-known as in collagen.

Sakalauskaite et al. have now overcome these issues by combining the analytical tools used to study the proteins and mineral content of modern shells with those of ancient protein research. They then used this approach, which they named palaeoshellomics, to extract proteins from seven "double-buttons" – pearl-like ornaments crafted by prehistoric people in Europe. The double-buttons were made between 4200 and 3800 BC and found at archeological sites in Denmark, Germany and Romania. Comparing the extracted proteins to those from various mollusc shells showed that the double-buttons were made from freshwater mussels belonging to a group known as the Unionoida.

The discovery helps settle a decade-long debate in archeology about the origin of the shells used to make double-buttons in prehistoric Europe. Ancient people often crafted ornaments from marine shells, because they were exotic and considered more prestigious. But the results on the double-buttons suggest instead that mother-of-pearl from fresh water shells was valued and used by groups throughout Europe, even those living in coastal areas. The palaeoshellomics technique used by Sakalauskaite et al. may now help identify the origins of shells from archeological and palaeontological sites.

DOI: https://doi.org/10.7554/eLife.45644.002

Yosef Mayer et al., 2009; ***d'Errico et al., 2005***; ***Giacobini, 2007***; ***Kuhn et al., 2001***; ***Saunders, 1999***; ***Zilhão et al., 2010***). Unravelling these nuances is especially important in times of life-style transformation (e.g. from mobile hunting and gathering to sedentism), when the way people perceived their 'homeland' may also have shifted: personal adornments thus help track cultural continuity or discontinuity (***Rigaud et al., 2018***; ***Rigaud et al., 2015***; ***Stiner, 2014***; ***Stiner et al., 2013***; ***Taborin, 1974***; ***Vanhaeren and d'Errico, 2006***). *Exotic* marine shells, transported to inland sites, are thought to have acquired special value and are interpreted as a marker of status and a proxy for long-distance exchange and trade (***Alarashi et al., 2018***; ***Bajnóczi et al., 2013***; ***Borrello and Micheli, 2011***; ***Taborin, 1974***; ***Trubitt, 2003***). The exploitation of *local* shells is instead usually viewed in utilitarian terms (e.g. ***Colonese et al., 2011***), as an easily accessible, convenient, but inherently less prestigious resource. However, the collection of certain local marine shells may have a deeper meaning; for example, creating a feeling of familiarity with a new environment for seafarers and colonisers (***Bar-Yosef Mayer, 2018***). The use of *freshwater* shells as raw material, although often documented in archaeological sites, has been somewhat overlooked (***Borrello and Girod, 2008***), and the relative importance of freshwater over marine shells has never been systematically addressed, introducing a bias in archaeological interpretations.

A typical example of such bias is the interpretation as *exotica* of the findings of prehistoric mother-of-pearl (shell) miniature double-buttons (*doppelknöpfe*), worked in a way to look like 'true' pearls (***Figure 1***). Experimental archaeological work has shown that they are excellent as ornaments pressed into thin leather, for example armbands or belts (***Kannegaard, 2013***). The raw materials

Figure 1. The double-buttons analysed in this study.
DOI: https://doi.org/10.7554/eLife.45644.003

actually vary from shell to (possibly) copies made of stone, teeth, bone and ceramic, but they are all consistently white, and their appearance standardized (***Heumüller, 2009***). Double-buttons have been reported primarily from a range of Neolithic Central European sites on the Danube Valley, but occasional findings occur from the Ligurian coast to Northern and Southeastern Europe (***Girod, 2010a***; ***Heumüller, 2012***). Among these sites, the shell double-buttons from the Danish Ertebølle/Early Funnel Beaker shell midden at Havnø (***Andersen, 2008***) and the hundreds of examples from the Neolithic submerged pile dwelling settlement of Hornstaad-Hörnle IA on Lake Constance (***Heumüller, 2009***) are especially significant, as they could represent an instance of exchange of materials or ideas between hunter-gatherers (Havnø) and farmers (Hornstaad-Hörnle IA; ***Figures 1*** and ***2***, ***Table 1***). Furthermore, if the double-buttons were made of marine shell (in particular of oyster shells, abundant at Havnø [***Andersen, 2008***]), then a case could be made for exotic materials from the coastal Ertebølle site of Havnø being imported (because highly desirable) to the inland Neolithic settlement of Hornstaad-Hörnle IA (***Rowley-Conwy, 2014***). However, the edible flat oyster (*Ostrea edulis*), so abundant at Havnø, can hardly be the raw material of the double-buttons: visual inspection (later confirmed by our analytical data) revealed that these are made of mother-of-pearl (nacre) and *O. edulis* does not form this microstructure. Furthermore, the 'true' pearl oysters (marine genus *Pinctada*) are not found in North Atlantic cold waters.

Despite being the basis for archaeological inference, the knowledge of the biological origin of molluscan mother-of-pearl in prehistoric Europe is only vague (***Taborin, 1974***). This is usually based on macroscopic and microscopic observations of heavily worked and degraded objects, which have thus lost any diagnostic feature that might have been useful for identification (***Demarchi et al., 2014***). Using analytical methodologies that are the basis of shell biomineralization studies (***Marin et al., 2016***; ***Marin et al., 2013***), including the first application of proteomics to archaeological shells ('palaeoshellomics'), we investigated seven archaeological double-buttons, dated

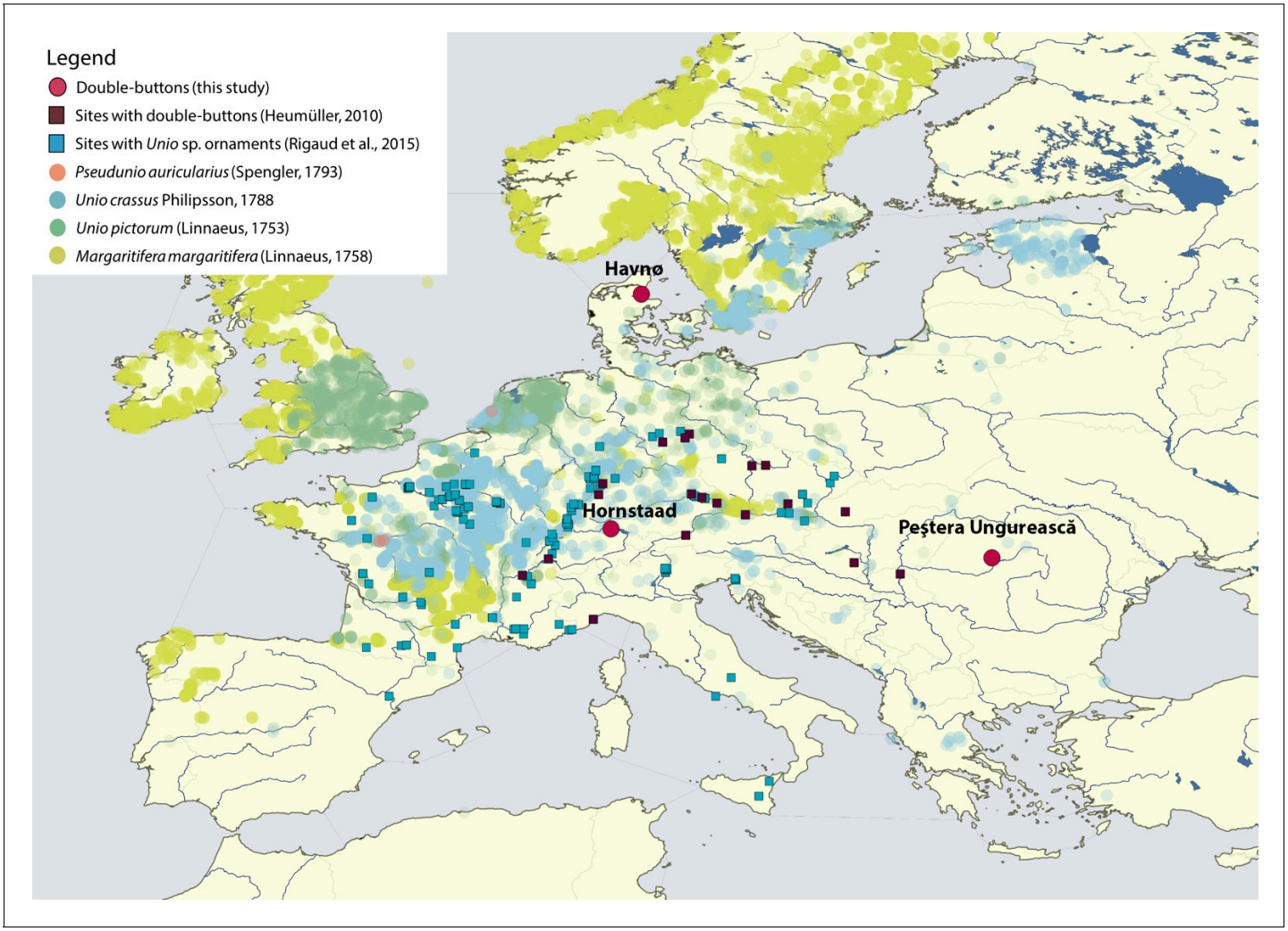


Figure 2. Map displaying the location of Havnø, Hornstaad-Hörnle IA and Peştera Ungurească, together with other archaeological sites from which double-buttons (***Heumüller, 2012***) and a variety of ornaments made with *Unio* sp. shells (***Rigaud et al., 2015***) have been reported.
DOI: https://doi.org/10.7554/eLife.45644.004

between ~4200 and ~3800 BCE and recovered from a wide geographic area (***Figure 2***): three (HorA, HorB, HorC) from Hornstaad-Hörnle IA (***Borrello and Girod, 2008***; ***Heumüller, 2012***), three (HavA, HavB, HavC) from Havnø (***Andersen, 2008***), and one (PesB) from Peştera Ungurească, a cave site in Transylvania (***Biagi and Voytek, 2006***; ***Girod, 2010a***). The richest record comes from Hornstaad-Hörnle IA, where more than 564 double-buttons were recovered, although many were partially fragmented and affected by fire. The origin of the raw material had been debated (***Table 1***): either edible oyster (marine) or freshwater pearl mussel (***Borrello and Girod, 2008***; ***Heumüller, 2012***). Around 40 double-buttons were excavated from the shell midden in Havnø - the raw material used to make these was undetermined. A handful of double-buttons were found at the cave site of Peştera Ungurească, together with *Unio crassus* valves, and therefore this freshwater mollusc was assumed to be the raw material used to make the double-buttons at Peştera Ungurească (***Girod, 2010a***).

Our work had two main aims:

1. to develop 'palaeoshellomics', a new molecular approach to characterize ancient proteomes preserved in mollusc shells. The recovery and identification of these proteomes is challenging, because shell contains only a small fraction of proteins embedded in the mineral skeleton (approximately 0.1–1% vs 30% in bone) and there is a lack of molecular sequence data, which

Table 1. Summary of the materials analysed in this study (archaeological ornaments and reference shells) and information on their context, chronology and taxonomic determination.

Sample type	Site	Cultural group	Time span	Taxonomic determination before this study	Other molluscan fauna present at the site
Double-buttons	Havnø	Ertebølle	4200–4000 cal BCE (radiocarbon dating of the Late Mesolithic horizon; ***Andersen, 2000***; ***Andersen, 2008***)	Unknown, presumed marine shell	Abundant edible marine shells: *Ostrea edulis*, *Littorina* sp., *Mytilus edulis*, *Cerastoderma edule* (***Andersen, 2000***; ***Andersen, 2008***)
	Hornstaad-Hörnle IA	Hornstaad group (early phase of the regional Late Neolithic)	3917–3902 BCE (dendrochronology; ***Billamboz, 2006***)	Debated: marine (*Ostrea edulis*) vs freshwater (*Margaritifera margaritifera*; ***Heumüller, 2010***; ***Borrello and Girod, 2008***)	Mediterranean marine shells (exotic, non edible): *Columbella rustica* (***Borrello and Girod, 2008***), *Callista chione*, *Astarte borealis*, *Dentalium vulgare* (***Heumüller, 2010***)
	Peştera Ungurească	Toarte Pastilate and transition to Coţofeni	4260–3820 cal BCE (range of radiocarbon dates, at 1σ, of layers 2B, 2A3 and 2A, Toarte Pastilate) (***Biagi and Voytek, 2006***)	*Unio* cf. *crassus* (***Girod, 2010a***, ***Figure 6B***)	Abundant terrestrial taxa (naturally occurring). Occasional presence of freshwater species (*Anisus spirorbis*, *Pisidium milium*, *Lithoglyphus naticoides*, *Lymnaea truncatula*, *Planorbis* cf. *carinatus*; ***Girod, 2010a***)
Reference shell	Limfjord (Northern Jutland)		Modern	*Ostrea edulis*	
	Limfjord (Northern Jutland)		Modern	*Modiolus modiolus*	
	Limfjord (Northern Jutland)		Modern	*Margaritifera margaritifera* (determined by F.M.)	
	France (Izeure)		Modern	*Unio pictorum*	
	Peştera Ungurească	Toarte Pastilate and transition to Coţofeni	4260–3820 cal BCE (range of radiocarbon dates, at 1σ, of layers 2B, 2A3 and 2A, Toarte Pastilate) (***Biagi and Voytek, 2006***)	*Unio* cf. *crassus* (***Girod, 2010a***)	Abundant terrestrial taxa (environmental signal). Occasional occurrence of freshwater species (*Anisus spirorbis*, *Pisidium milium*, *Lithoglyphus naticoides*, *Lymnaea truncatula*, *Planorbis* cf. *carinatus*). Not suitable as raw material for the double-buttons (***Girod, 2010a***)
	Isorella	Vhò	5226–5023 cal BCE at 2σ (***Starnini et al., 2018***)	*Pseudunio auricularius* (***Biddittu and Girod, 2003***; ***Girod, 2010b***)	Marine taxa (typically used as ornaments): *C. rustica*, *Spondylus* (fragment of a bracelet; ***Girod, 2010b***)

DOI: https://doi.org/10.7554/eLife.45644.005

are needed for comparing taxa within the same clade. When studying ancient materials, these issues are further compounded by the effects of diagenesis and by the small sample sizes available. Therefore, developing palaeoshellomics has an impact on understanding the fundamental mechanisms of biomineralization (protein-mineral interactions) in molluscan shells and other invertebrates, the extent of proteome variability within the same molluscan clade, as well as the diagenetic pathways of degradation and preservation.

2. to integrate 'palaeoshellomics' within a set of well-established analytical techniques, in order to determine the origin of the shell ornaments from Hornstaad-Hörnle IA and Havnø, and to confirm that of the double-buttons from Peştera Ungurească. This is archaeologically significant, as similar ornaments were recovered from three geographically distant sites belonging to three different and broadly contemporary cultural groups (***Table 1***): Late Mesolithic (Ertebølle), early Late Neolithic (Hornstaad Group), and Copper Age (Toarte Pastilate/Coţofeni).

To achieve both of these goals, the archaeological double-buttons were studied alongside a selected set of reference shells (***Table 1***). These include two marine species (*Ostrea edulis*, *Modiolus modiolus*), which commonly occur in northwestern European seas, and four freshwater mussels,

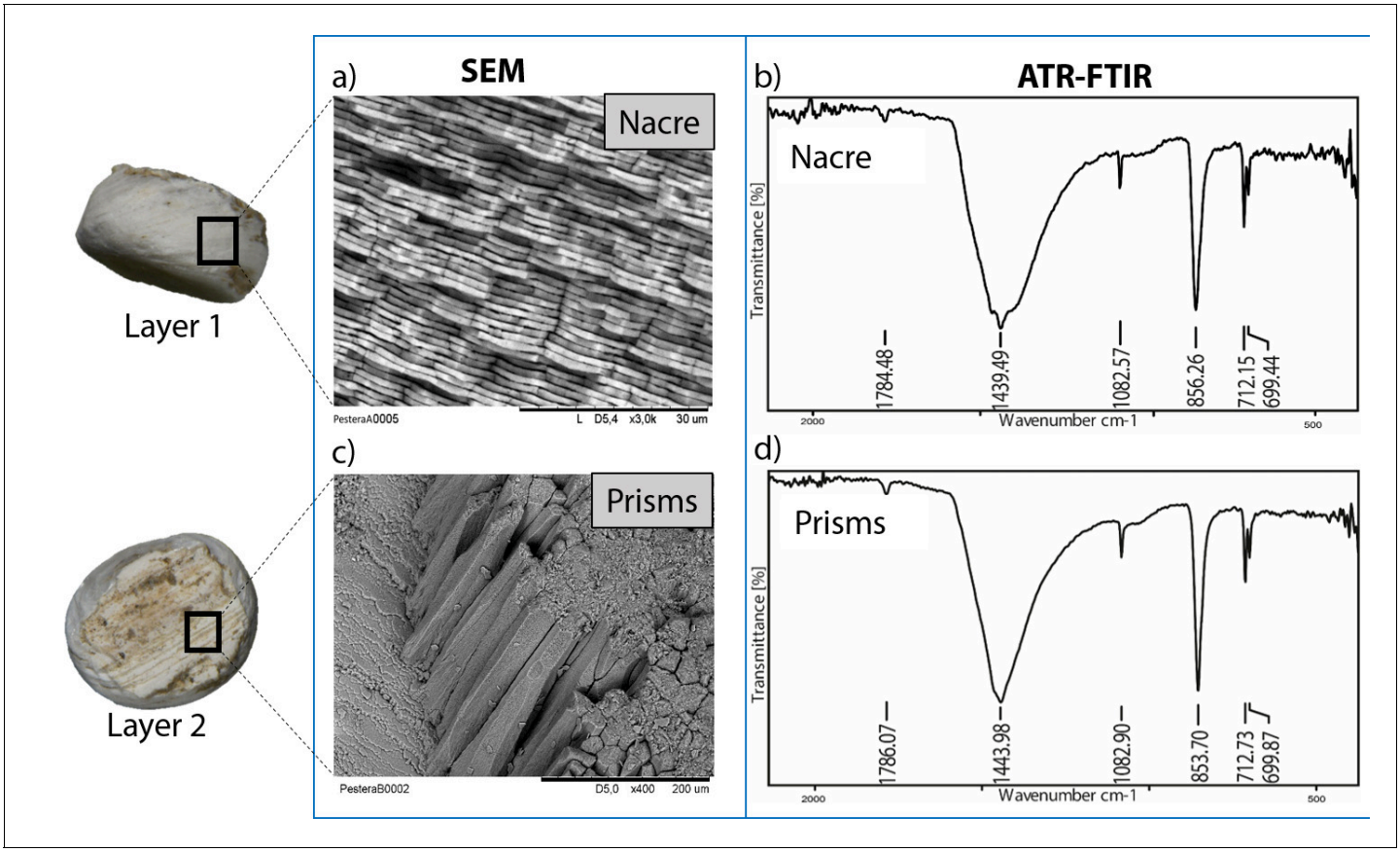


Figure 3. Microstructure (SEM) and mineralogy (FTIR-ATR) of double-buttons, showing shiny nacreous (a) and matte prismatic (c) layers, both aragonitic (b, d).

DOI: https://doi.org/10.7554/eLife.45644.006

which are characterized by a thick layer of mother-of-pearl, that is *Unio pictorum*, *Unio crassus*, *Margaritifera margaritifera*, *Pseudunio auricularius*.

Sample selection

Two modern marine mollusc shells, *O. edulis* and *M. modiolus*, were collected in northern Jutland (Denmark) by Søren H. Andersen and were selected for the following reasons: *O. edulis* shells had been suggested as the potential raw material for the Hornstaad-Hörnle IA assemblage (***Heumüller, 2010***) and are very abundant at the shell midden site of Havnø; *M. modiolus* is a thick-shelled mussel with a nacreous layer, therefore a suitable raw material for the Havnø ornaments (Appendix 1, section 2). Furthermore, close relatives of both species are present in public sequence databases, which is important for palaeoproteomics: *O. edulis* belongs to family Ostreidae (genomes available for *Crassostrea gigas and C. virginica*) and *M. modiolus* to family Mytilidae (genome available for *Mytilus galloprovincialis*).

With regard to the freshwater species (order Unionoida), *U. pictorum* and *U. crassus* belong to family Unionidae, *P. auricularius* and *M. margaritifera* to family Margaritiferidae. Modern *U. pictorum* shells were collected in a stream close to Izeure (Burgundy) by Frédéric Marin and modern *M. margaritifera* was collected in northern Jutland by Søren H. Andersen. The morphological determination of both taxa was carried out by Frédéric Marin. *U. crassus* and *P. auricularius* are archaeological shell specimens from the sites of Peştera Ungurească and Isorella (Neolithic, Po Plain, Italy [***Starnini et al., 2018***]). The determination of *U.* cf. *crassus* had been carried out by Alberto Girod on the basis of morphological observations of the whole shell valves and comparison to extant specimens from the area (***Girod, 2010a***). Archaeological *P. auricularius* was used as this species is critically endangered (***Altaba, 1990***) and extant populations rare (***Appendix 1—figure 1***). The morphological

determination of the species had been carried out by Alberto Girod using comparative specimens from museum collections (**Biddittu and Girod, 2003**). An advantage of including archaeological shells as reference materials is that we were able to assess the extent of molecular preservation in shells that are contemporary to the double-buttons. Furthermore, whole and fragmented valves of *U.* cf. *crassus* had been recovered from all the archaeological layers that yielded the double-buttons at Peştera Ungurească (**Girod, 2010a**). Therefore, in this case, potential raw material and finished product have experienced the same post-depositional conditions.

None of the Unionoida species is well-represented in public sequence databases, especially with regard to proteins related to shell biomineralization (see Methods section).

Results and discussion

Morphological analysis

The double-buttons described here are circular, with a groove in the middle of the body and no perforation. The main body is shiny but all of them have a thin matte layer on one of the two surfaces (**Figures 1** and **3**). These two layers, both aragonitic (as shown by infrared spectroscopy, see Appendix 1, section 3.2), are the nacre and prisms of a mollusc shell. Scanning electron microscopy showed the presence of the 'brickwall' microstructure of nacre (sheet nacre) juxtaposed with the thin layer of prisms, the latter having elongation axes perpendicular to the nacre plane (**Figure 3**). No secondary calcite was observed and there was no sign of the occurrence of diagenetic recrystallization. The overall nacre appearance is typical of bivalves and not of gastropods. The combination of nacroprismatic microstructure and aragonitic mineralogy is observed in freshwater unionoid mussels but also in marine trigonioids and anomalodesmatans (Appendix 1, section 2).

Stable isotopes of carbon and oxygen

Stable isotope analyses for all of the samples yielded average $\delta^{18}O$ and $\delta^{13}C$ values of -5.3 ± 0.4 and -11.1 ± 0.6 ‰ for Havnø, -6.1 ± 1.0 and -11.9 ± 1.7 ‰ for Peştera Ungurească and -9.3 ± 0.5 and -10.6 ± 1.6 ‰ for Hornstaad, respectively (Appendix 1, section 3.3). The consistently low $\delta^{18}O$ and $\delta^{13}C$ values of shells from Peştera Ungurească and Hornstaad indicate a local freshwater origin for the shells (**Keith et al., 1964**; **Leng and Lewis, 2016**), whereas the $\delta^{18}O$ values at Havnø suggest some mixing of marine water or changes in the atmospheric circulation, with precipitations slightly enriched in ^{18}O compared to present day over the region. Our interpretations are broadly supported by the average annual $\delta^{18}O$ values of modern local precipitations for the sites (**WaterIsotopes.org, 2018**). The isotope data therefore suggest that the shells were locally sourced (**Keith et al., 1964**).

Amino acid analysis

The absence of recrystallization observed by SEM was consistent with the concentration, composition and relatively low D/L values for all the amino acids analysed (e.g. alanine D/L ~0.1 for Peştera Ungureaşca, ~0.2 for Havnø and Hornstaad), except for sample HorA. This supports a non-fossil origin for the shells used to make the double-buttons, that is the makers used 'fresh' or recently dead mollusc shells. However, the extent of degradation for HorA was significantly higher and both D/L and concentration values showed a clear 'burning' signal (**Crisp, 2013**; **Demarchi et al., 2011**). The amino acid composition was similar to that of freshwater bivalves (*Unio*, *Margaritifera*) present in the reference database of **Demarchi et al. (2014)**, although HorA and PesB appeared to be rather different from the other double-buttons (Appendix 1, section 3.4).

Palaeoproteomics

We characterized the proteomes preserved within the seven double-buttons and performed bioinformatic searches (PEAKS 8.5, Bioinformatics Solutions Inc [**Ma et al., 2003**]) of the product ion spectra against both a Protein sequences and an Expressed Sequence Tags (ESTs) database, restricting the taxonomy to Mollusca (see Methods section). This resulted in the identification of 1973 and 3233 peptide sequences, respectively, which represent the 3.5% and 5.1% of the total number of sequences generated by the de novo algorithm of the software (excluding contaminant sequences). For comparison purposes, we also performed shotgun proteomics on the shell matrices extracted

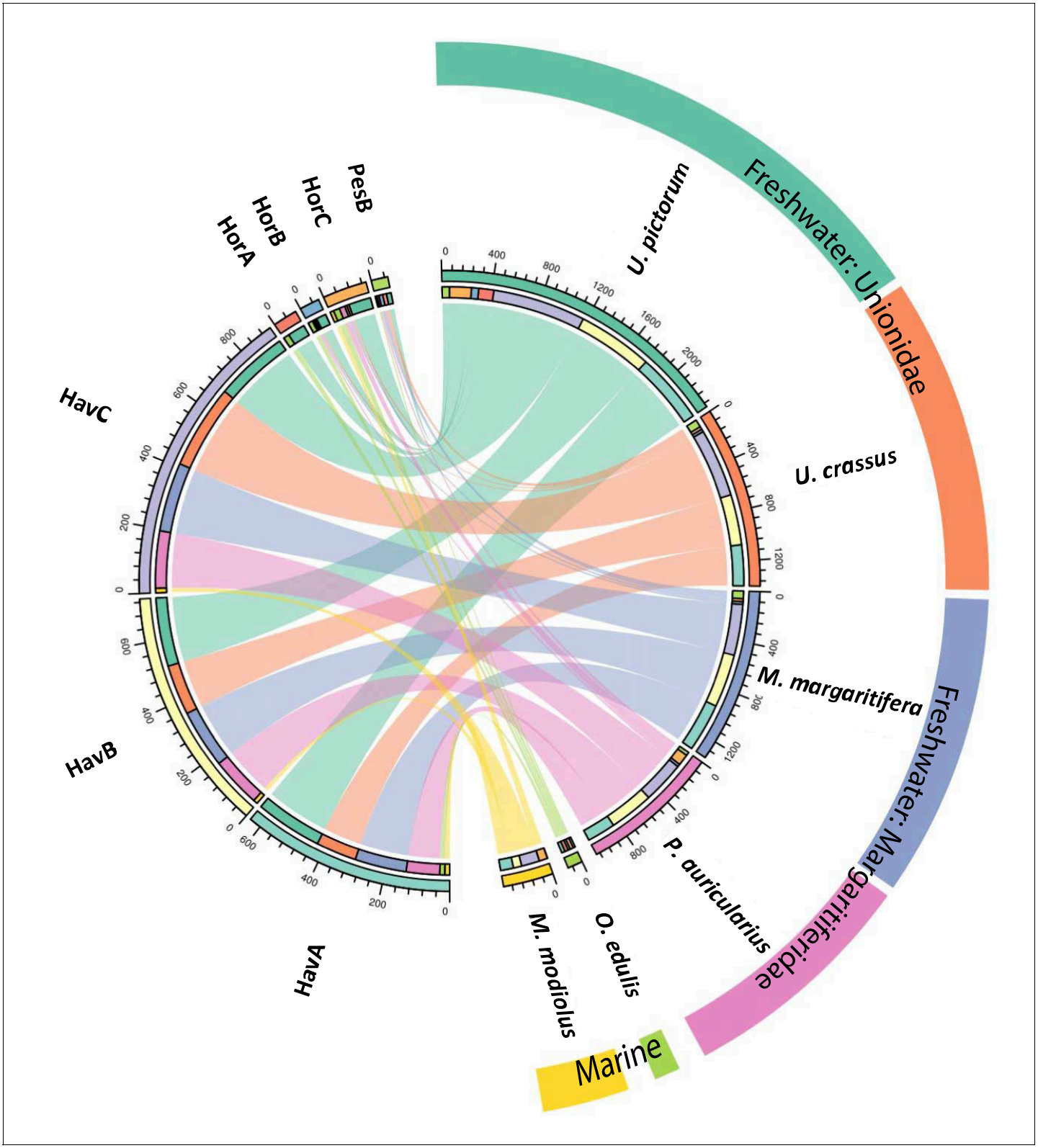


Figure 4. Circular diagram representing the similarity between the proteomes of seven double-buttons (left) and six mollusc shell taxa (right).
DOI: https://doi.org/10.7554/eLife.45644.009
The following source data is available for figure 4:

Source data 1. R code and data files for ***Figure 4***.
DOI: https://doi.org/10.7554/eLife.45644.010

Table 2. Main protein sequences identified in the double-buttons from Havnø, Hornstaad-Hörnle IA and Peştera Ungurească and their presence/absence in the analysed set of reference freshwater and marine shells (black dots).

Numbers indicate total number of peptide sequences identified and the cell colour is proportional to the coverage of the sequence itself. Threshold values for peptide and protein identification: false discovery rate (protein FDR) = 0.5%, protein score −10lgP ≥ 40, unique peptides ≥ 2, de novo sequences scores (ALC%) ≥ 50. Asterisks (*) indicate proteins identified only when using less stringent parametres: protein score −10lgP ≥ 20; unique peptides ≥ 1. Note that molecular sequence databases for molluscan species are incomplete and biased towards well-studied model organisms. The peptide sequences recovered in our study were identified using sequence homologies with proteins originally described from *Hyriopsis cumingii*, *Crassostrea* sp., *Pinctada* sp., *Mytilus* sp. and several others. As a result of database insufficiency, the bioinformatic search of these 'shellomes' could not identify the exact taxon of our samples, but provided a strong indication of the fact that the closest taxon to that of the ornaments (and of the freshwater reference shells) is the pearl-producing triangle sail mussel *Hyriopsis cumingii* (Unionoida).

Proteins present in database from			Double-buttons							Freshwater				Marine	
Order	Genus	Identified proteins	HavA	HavB	HavC	HorA	HorB	HorC	PesB	U.p	U.c	M.m	P.a	Mo. M	O. e
Unionoida	*Hyriopsis*	Hic74 [*Hyriopsis cumingii*]	132	158	260		6	11	21	•	•	•	•		
		Hic52 nacreous layer matrix protein [*Hyriopsis cumingii*]	1*	1*	2*					•	•	•	•		
		Silkmapin (isoforms: nasilin 1 and nasilin 2) [*Hyriopsis cumingii*]	1*	3	3		5			•	•	•*	•*		
Ostreida	*Pinctada*	MSI60-related protein [*Pinctada fucata*]	6		27			12		•	•		•	•	
		Insoluble matrix protein [*Pinctada fucata*]	4		33					•	•		•		
	Crassostrea	Glycine-rich cell wall structural protein-like [*Crassostrea virginica/gigas*]	17	11	12				14		•				
		Glycine-rich protein 23-like [*Crassostrea virginica*]	8	11	6					•		•			
		Antifreeze protein Maxi-like [*Crassostrea virginica*]		4	4								•		
Mytilida	*Bathymodiolus*	MSI60-related protein partial [*Bathymodiolus platifrons*]		6	11					•			•	•	
	Mytilus	Precollagen D [*Mytilus edulis*]	16	26	23				9	•	•	•			
		Nongradient byssal precursor [*Mytilus edulis*]	10	10					25		•				
Other	*Other*	Predicted: transcription factor hamlet-like partial [*Octopus bimaculoides*]	5	6	11					•		•			
		Hypothetical protein OCBIM_22008720 mg partial [*Octopus bimaculoides*]	6	11											
							Coverage			≥55%	≥35%	≥15%	≥10%	≥1%	

Presence •

DOI: https://doi.org/10.7554/eLife.45644.007

The following source data is available for Table 2:

Source data 1. Palaeoshellomics.

The complete proteomics dataset obtained on reference shells and archaeological ornaments

DOI: https://doi.org/10.7554/eLife.45644.008

from the six reference mollusc species (Appendix 1, section 3.5) and analysed the data using the same databases and parametres (see ***Table 2—source data 1*** for full results of the palaeoproteomic analyses).

Table 2 shows the top-scoring proteins from the seven double-buttons: the numbers indicate the peptides supporting each protein identification, while protein coverage (i.e. the percentage of

sequence for which we could detect peptides) is represented by different colours. Additionally, on the right hand side of the table we indicate if each of the double-button proteins also occurred in the reference shell proteomes (the list of all shell proteins identified in both the double-buttons and the reference shells can be found in ***Table 2—source data 1***).

Shellomes: significant protein hits

The main protein sequences from the double-buttons were identified as belonging to the pearl-producing triangle sail mussel *Hyriopsis cumingii* (***Bai et al., 2013***).

Protein Hic74 (GenBank: ARG42316.1) was found in all of the archaeological samples, except HorA. The percentage coverage for the Hic74 sequence was highest for the Havnø beads (35–55%), where it was supported by 132, 153, 255 unique peptides in HavA, HavB and HavC, respectively (Appendix 1, section 3.5). This protein was also securely identified in all of the freshwater unionoid reference shells (coverage varying from 34% in *M. margaritifera*, supported by 67 peptides, to 50% in *U. crassus,* supported by 203 peptides). Hic74 is an acidic, Ala- and Gly-rich shell matrix protein (***Liu et al., 2017a***). Consisting of 19 poly-A blocks, GA repeats, short acidic motifs (that probably bind to the mineral) and a GS-rich domain at the C-terminus (which resembles that of lustrin-A), this silk fibroin-like protein is likely to play a structural role in nacre formation and in enhancing its mechanical properties (***Liu et al., 2017a***).

Protein Hic52 (GenBank: ARH52598.1) was identified in all the reference unionoid shells and in the Havnø samples, but only when less stringent parameters were used for the identification (i.e. number of unique peptides ≥ 1 (instead of 2) and protein score $-10\lg P \geq 20$ (instead of 40)). Hic52 is a very basic (theoretical pI > 10), Gly- and Gln-rich protein, with few poly-Q and poly-G blocks and several degenerate G-rich repeats of different lengths along the sequence. It possesses a collagen-like structure which suggests a structural role in nacre formation (***Liu et al., 2017b***). Silkmapin (GenBank: AIZ03589.1, and its isoforms nasilin 1 and 2) are Gly-rich non-acidic proteins with a structural function, probably related to the formation of both nacreous and prismatic layers (***Liu et al., 2015***; ***Marie et al., 2017***). Present in the shell matrix of all the unionoids, these proteins were also detected in the Havnø samples and in one of the Hornstaad beads (HorB). Finally, we also identified protein sequences from marine mollusc genera (mainly *Pinctada*, *Crassostrea* and *Mytilus*), but all these 'marine' sequences only displayed repeated low-complexity (RLC) domains (typically consisting of Ala and Gly-rich repeats and/or poly-Ala blocks). RLC-containing peptides are not sufficient for distinguishing between freshwater and marine shells. On the contrary, in double-buttons and in unionoid reference shells, the top-scoring protein Hic74 was supported by remarkably high (for shell proteins) coverages, and, together with Hic52 and silkmapin/nasilin, showed a number of specific peptides that do not exhibit RLC domains. These proteins showed no homologues with any other shell proteins of marine origin currently present in the NCBI database (BLASTp search), being unique to *H. cumingii* and suggesting their specificity to freshwater unionoid shells. We argue that their presence (where identified as the major shell matrix proteins, supported not only by RLC domains) is specific to Unionoida, freshwater mother-of-pearl shells, which in combination with the isotopic data, and supported by the microstructural and amino acid results, excludes a marine origin for the raw material used to make the double-buttons.

Comparison with reference shells

We performed a search of the product ion spectra from the double-buttons and the six reference shells against the redundant EST database, so that we could recover complementary information from non-annotated sequences. For example, a search of *Hyriopsis cumingii* on NCBI will retrieve 246 protein sequences but 10156 EST sequences. The dataset was used to explore the similarities between double-buttons and molluscan shell proteomes, presented as a circular plot in ***Figure 4***. This output was derived from an adjacency matrix, showing which proteins (EST sequence identifiers) occurred in two or more samples (R code and data files can be found in ***Figure 4—source data 1***). From this, the subset of unique identifiers, present both in the double-buttons and in any of the reference shells, was associated with its sequence coverage (%) in the archaeological samples. This information is represented in the right-hand side of ***Figure 4***: the length of the circular segment for each molluscan species is proportional to the number of sequences that the shell shares with the archaeological samples, scored on the basis of the coverage. On the left-hand side of the graph,

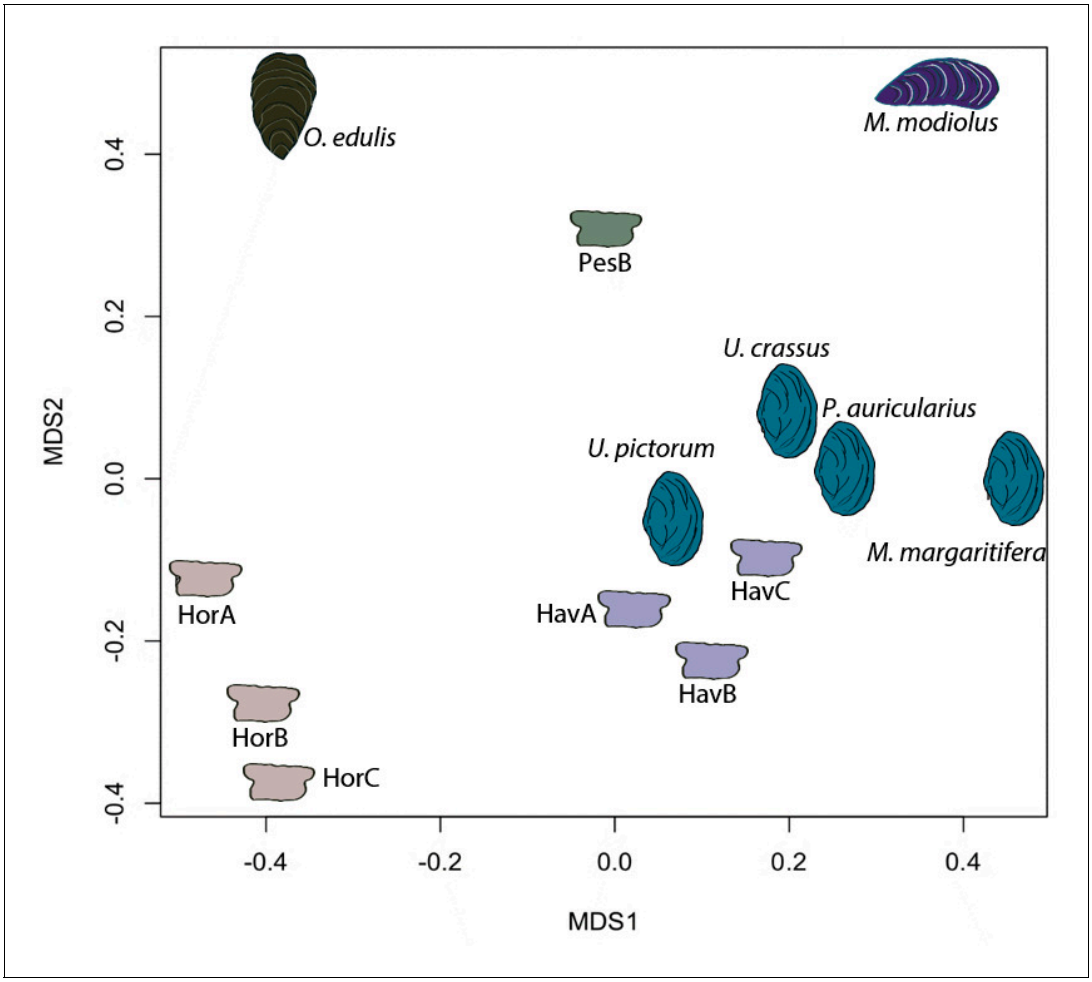


Figure 5. Proteome comparison based on peptide sequence similarity, represented by multi-dimensional scaling (MDS).

DOI: https://doi.org/10.7554/eLife.45644.011

The following source code is available for figure 5:

Source code 1. Pepmatch code Code (developed in C language) for ***Figure 5***.

DOI: https://doi.org/10.7554/eLife.45644.012

each double-button is represented by a circular segment, which is proportional to the number of unique peptides that supports each shared protein sequence. The similarity between double-buttons and mollusc taxa can be visualised through the thickness of the connecting bands. Overall, the data showed that the EST sequences shared between the ornaments and the shells were mainly from the unionoids, consistent with the results obtained by searching the annotated protein database (***Table 2***). From all of the archaeological samples, the Havnø set showed the best match to the freshwater unionoids, owing to better-preserved proteins, that is with high coverage and number of supporting peptides. The LC-MS/MS analysis of PesB produced a number of tandem mass spectra comparable to the other samples, but a lower number of sequences were identified. Sample HorA (burnt), from which no proteins had been identified using the annotated protein database, yielded some matches to EST sequences, most of which shared with *U. pictorum*.

Database-independent comparison

In order to provide further, independent, evidence for the origin of the raw material, we developed an in-house tool (in C language, available in ***Figure 5—source code 1***) for 'proteome comparison', using all the peptide sequences generated by the de novo algorithm of the software PEAKS, that is before performing any database search (Appendix 1, section 3.6). The tool was able to provide a

Table 3. Potential amino acid substitutions detected in the samples analysed in this study, compared to the reference Hic74 sequence [*Hyriopsis cumingii*].

Positions are derived from the sequence alignment shown in ***Figure 6***. Dashes indicate that the position was not covered for that sample; question mark symbols indicate ambiguous substitutions. Hic74 coverages for each sample and supporting product ion spectra are presented in ***Figure 6—figure supplement 1***, ***Figure 6—figure supplement 2***, ***Figure 6—figure supplement 3***, ***Figure 6—figure supplement 4***, ***Figure 6—figure supplement 5***, ***Figure 6—figure supplement 6***, ***Figure 6—figure supplement 7***, ***Figure 6—figure supplement 8***, ***Figure 6—figure supplement 9***, ***Figure 6—figure supplement 10***, as well as in ***Figure 6—source data 1***.

Hic74 [*Hyriopsis cumingii*]	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**G**	**D**	**G**	**S**	**E**	**G**	**A**	**A**	**L**	**V**	**G**	**L**	**I**	**A**	**G**	**A**	**Q**	**R**	**E**
AA position	83	85	91	93	106	108	110	111	151	152	163	172	175	282	283	284	289	292	306	310	403	801	804	821	822	827
U. crassus	-	-	A	A	A	A	A	G	D	G	-	Q	G	A	?	L	?	G	L	F	A	-	-	Q	R	E
U. pictorum	-	-	?	?	?	-	-	-	-	-	-	Q	G	-	-	-	-	-	L	F	?	S	S	Q	R	E
M. margaritifera	-	-	-	A	-	-	-	-	-	-	-	Q	G	-	A	L	L	G	L	F	A	-	-	Q	R	E
P. auricularius	-	A	A	A	A	A	A	-	?	?	D	E	L	-	A	F	V	E	F	I	-	-	-	H	H	D
HavA	A	S	A	A	A	?	?	A	-	-	-	Q	G	A	A	L	L	G	L	F	-	-	-	Q	G	E
HavB	?	S	A	A	A	V	-	-	-	-	-	-	-	-	A	L	L	G	L	F	-	-	-	Q	G	E
HavC	?	S	A	A	A	V	?	A	D	G	G	-	G	?	A	L	L	G	L	F	-	G	S	Q	G	E

DOI: https://doi.org/10.7554/eLife.45644.025

score for the sequence similarity between two lists of peptides and to generate a similarity matrix from all pairwise comparisons, which was then converted to a distance matrix. Multidimensional scaling (MDS; ***Gower, 1966***) was used to visualise the similarity of observations (***Figure 5***) and confirmed that the Havnø set and the freshwater reference unionoids display the higher degree of proteome similarity, while the samples from Hornstaad and PesB fall in a different area of the plot from each other and from the marine reference shells. The results were also in accordance with those obtained from another database-independent approach ***Appendix 1—figure 25***, based on direct product ion spectra comparison (***Rieder et al., 2017***), which was adapted for this study (Appendix 1, section 3.6). Overall, our study, which represents one of the few that attempts to compare molluscan proteomes within the same clade (genus, family or order), shows that unionoid shells exhibit very similar proteome profiles, sharing many sequences between species ***Appendix 1—figure 23***. This may suggest that this group has a rather conserved, homogeneous and recognisable proteomic signature, a conclusion that is completely congruent with earlier findings (***Marie et al., 2017***). Furthermore, all analyses showed that the three sets of archaeological ornaments have similar proteome profiles (***Table 2***, ***Figures 4*** and ***5***, ***Appendix 1—figure 24***, but do not exhibit a simple correspondence to a molluscan species, at least among the Unionoida considered here, further highlighting the complexity of molluscan shell proteomes.

Analysis of the Hic74 sequence

We examined the sequence of the top-scoring protein, Hic74, recovered from the reference shells and ornaments, with the aim of assessing the presence and frequency of any amino acid substitutions, which could potentially yield taxonomic resolution within Unionoida. ***Figure 6*** shows the alignment (performed using the software Geneious Prime 2019.1.1) of these incomplete sequences to the reference (Hic74 from *Hyriopsis cumingii*). The sequence coverage of each sample was obtained from the '*Spider*' output of PEAKS 8.5. The *Spider* algorithm takes into account potential amino acid substitutions, as well as a large number of in vivo, laboratory-induced (e.g. carbamidomethylation), and diagenetically-relevant (e.g. deamidation) modifications, therefore it is especially useful in highlighting possible mutation sites. In our sequence reconstruction we only considered peptides displaying typical sample preparation-induced or diagenesis-induced modifications (***Figure 6—figure supplement 1***, ***Figure 6—figure supplement 2***, ***Figure 6—figure supplement 3***, ***Figure 6—figure supplement 4***, ***Figure 6—figure supplement 5***, ***Figure 6—figure supplement 6***, ***Figure 6—figure supplement 7***, ***Figure 6—figure supplement 8***, ***Figure 6—figure supplement 9***, ***Figure 6—figure supplement 10***). The potential amino acid substitutions and their positions are summarised in ***Table 3*** (supporting product ion spectra can be found in ***Figure 6—source data 1***).

In the best-case scenario, only around half of the Hic74 (*Hyriopsis cummingii*) reference sequence was covered (**Table 2**, **Figure 6**). This may be due to: genuine sequence differences between *Hyriopsis* and all the other Unionoida examined here; low susceptibility of low-complexity domains to enzymatic cleavage; selective *post mortem* (or laboratory-induced) degradation of half of the sequence; errors in the transcriptome assembly of the protein. It is likely that a combination of all of these factors is responsible for this, particularly as the Hic74 regions not covered in the samples are mainly low-complexity domains, sometimes highly polar and thus prone to hydrolysis (Ser and Asp-rich). Within the limits due to the incomplete coverage of the mutation sites in reference shells, it is interesting to note that the Hic74 sequence of *P. auricularius* diverges significantly from those of *U. pictorum*, *U. crassus* and *M. margaritifera*, which appear to have a higher degree of sequence similarity (**Table 3**). While this is in contrast with a recent taxonomy reassessment of Unionida based on mitochondrial DNA (**Lopes-Lima et al., 2018**), it has been noted before that shell proteomes do not follow a simple phylogenetic signal (**Jackson et al., 2010**). The sequence coverage was insufficient to attempt any further consideration for the Hornstaad and PesB samples, but the better-preserved Havnø double-buttons shared the same amino acid substitutions, supporting the hypothesis that the same species was used to make these three ornaments (**Table 3**). Furthermore, **Table 3** shows that this taxon was unlikely to be *P. auricularius*, and more likely to be *Unio* or *Margaritifera* sp.

Raw material identification

The raw material used to manufacture the seven double-buttons can be firmly and consistently identified as Unionoida, freshwater shells with a thick mother-of-pearl layer, on the basis of morphological, microstructural, mineralogical, geochemical and biomolecular data. From a microstructural viewpoint alone, the combination of aragonitic prisms and sheet nacre ('brickwall type') structures is restricted to three bivalvian orders: the Unionoida, the Trigonioida and the Anomalodesmata, the first two belonging to sub-class Palaeoheterodonta (**Taylor et al., 1973**). The Trigonioida relic order could be ruled out, since it is represented nowadays by a single genus, *Neotrigonia*, with very small shells and living exclusively on the Australian and Tasmanian coasts. Geochemical data, that is stable isotope values of carbon and oxygen, overall indicated that all biominerals studied here were formed in freshwater environments, and $\delta^{18}O$ values in double-buttons tracked the average annual $\delta^{18}O$ values of local precipitations. This excluded Anomalodesmata as potential candidates, since this order of enigmatic, rare and specialised bivalves are strictly marine (**Taylor et al., 1973**). This finally left only one possibility, Unionoida, the representatives of which are all freshwater bivalves.

Biomolecular analyses showed that the proteome similarity is highest between the double-buttons and the unionoid reference shells. The identification of proteins Hic74, Hic52 and silkmapin in almost all of the archaeological samples confirms the freshwater nacre (Unionoida) origins of the double-buttons. With our current knowledge on shell proteins, these sequences probably represent taxon-specific adaptations for the biomineralization of nacroprismatic structures in unionoid shells: they do not bear any homologues with other shell proteins and are not found in the proteomes of other non-nacreous shell structures characterized here (*O. edulis*). Furthermore, the analysis of the amino acid substitutions on the Hic74 sequence, recovered from both the Unionoida shells and the ornaments, indicates that *Unio* or *M. margaritifera* (not *Pseudunio*) had been used for making the Havnø ornaments.

Sources of bias in 'palaeoshellomics'

Except technical bias, inherent to standard proteomics per se and discussed elsewhere (**Marin et al., 2016**), we identified three potential sources of bias that may hamper, limit or confound the current use of 'palaeoshellomics': 1) the intrinsic peculiarities of several shell matrix proteins; 2) the completeness of the dataset used for identification searches; 3) the diagenetic degradation of shell proteins.

1. Sequence analysis of shell matrix proteins has revealed that a large proportion exhibits unconventional primary structures, with abundant long stretches of repeated residues, the RLCDs/LCDs (repetitive low-complexity domains; **Marie et al., 2013**). The sequences of such domains are neither taxon-specific nor do they carry phylogenetic information. Because they play a structural role, they are ubiquitous and may be detected in phylogenetically-distant lineages. For example, our proteomic analysis identified many sequences known in marine bivalves such as glycine-rich, insoluble matrix and MSI60-related proteins, but these findings are only

supported by RLC domains that can be detected in very different molluscs. More generally, silk fibroin-like domains (poly-Ala), acidic (D-rich or poly-D) or collagen-like repeats (G-X-A triplets) are widespread, ubiquitous and cannot be simply assigned to one given shell protein and/or mollusc genus.

2. As stated before, protein identification is database-dependent and the quality of the interpretation is proportional to the size of the data set. The list of known shell proteins has dramatically expanded in the past few years, with the use of high-throughput screening (***Marin et al., 2016***). However, the 'shellomes' of only about thirty mollusc genera have been identified, in a phylum that comprises between 80000 and 100000 species. We do not know yet whether the set of sequences at our disposal is a representative sample of the sequence diversity for the whole phylum. In other words, the taxonomic origin of a given archaeological shell sample (such as the double-buttons analysed here) may be indeterminable because extant representatives of the corresponding genus (or of closely related genera) are not yet registered in molecular databases via their genomes, mantle transcriptomes or shellomes. As a consequence, palaeoshellomics, for the time being, will provide information mainly at intermediate taxonomic levels (order, family, sub-family). Nevertheless, the continuous expansion of the dataset of shell protein sequences will increase the power of palaeoshellomics: archaeological shell samples, which are currently unidentified, can be revisited and re-interpreted in the future via in silico investigations alone (i.e. without any further proteomic analysis). Furthermore, the detection of possible amino acid mutations in the sequence of different reference shells, such as Hic74 in this study (***Figure 6***, ***Table 3***), can be helpful in narrowing the range of candidates and in obtaining more precise taxonomic information.
3. Diagenesis, that is the slow transformation of sediments and their contents with time, also affects shell biominerals, whether from continental, lacustrine or marine environments. As shell biominerals are organo-mineral composites, diagenesis alters shell proteins too. The diagenetic stability of one given shell protein is a complex phenomenon that depends on several parameters, including its primary structure, its conformation, its localization within the biomineral, the availability of water molecules at the vicinity of the protein, the presence of saccharide moieties. Some shell proteins that could be important for taxonomic determination may be diagenetically unstable, that is easily degraded. As a result, the information they carry would be lost. In the present case, we were fortunate to obtain a consistent set of peptides that do not correspond to LC/RLC domains, i.e. peptides that carry relevant information for protein/clade identification. These unambiguously target Hic74, Hic52 and silkmapin proteins that correspond to freshwater unionoid bivalves. Beside taxonomic determination per se, our data suggest that these proteins (especially Hic74) are diagenetically stable and that they may constitute accurate markers for further studies.

While diagenesis may represent a true source of bias, we were able to accurately evaluate its effects, and we found that the extent of protein degradation (racemization, deamidation) was consistent with the age and burial history of the samples. More specifically, we observed that samples from Havnø were the best preserved - the coverage of the main proteins was high, especially for Hic74 (up to 55%). Surprisingly, we found that the coverage of this protein in Havnø samples is similar to that of modern *U. pictorum* and Neolithic *U. crassus* and *P. auricularius* - indeed, the number of Hic74 peptides in Havnø even surpasses that of the reference unionoid shells, despite the fact that the sample size for the archaeological double-buttons was at least 100 times smaller (Appendix 1, section 3). We assume that this effect is due to early diagenetic changes (such as protein unfolding, loss of linked sugars) that render the protein backbone more accessible to proteolytic enzymes, thus increasing the chance of releasing and identifying peptides. Interestingly, we observed a similar phenomenon in other mineralised systems, for example ostrich eggshell (***Demarchi et al., 2016***).

Samples from Hornstaad yielded significantly lower coverages (7% and 12% respectively for HorB and HorC) with fewer supporting peptides (6 and 11; ***Table 2***). In one instance (sample HorA), no proteins were identified. This is consistent with the results of the chiral amino acid analysis, which had flagged this sample as compromised and probably burnt, as well as with archaeological evidence for widespread fire destruction of the settlement (***Heumüller, 2012***). However, the same sample yielded a high number of unidentified peptide sequences (~6000), some of which appeared to be highly acidic and reminiscent of biogenic carbonate-associated proteins (***Marin and Luquet, 2008***). Furthermore, neither the microstructure nor the mineralogy of the double-buttons from Hornstaad showed any apparent sign of recrystallization to secondary calcite. We therefore hypothesize that the exposure to high temperatures had been relatively moderate, sufficient for inducing protein

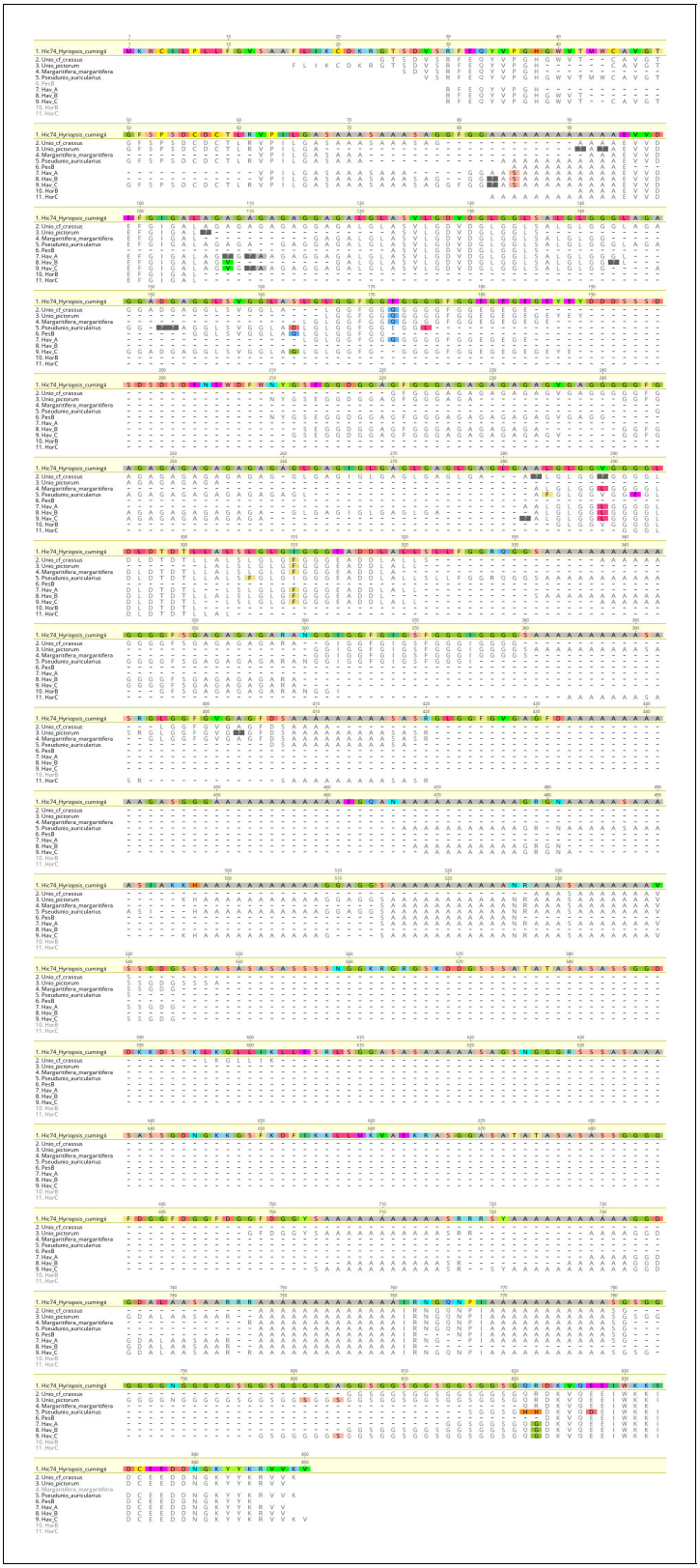


Figure 6. Alignment of Hic74 sequences recovered from the Unionoida reference shells and the ornaments. The reference Hic74 [*Hyriopsis cumingii*] is shown at the top of the alignment and is highlighted in yellow. Dashes indicate where the sequence was not covered in the samples analysed in this study; amino acid residues highlighted in colour show all disagreements with the reference Hic74 [*Hyriopsis cumingii*].

DOI: https://doi.org/10.7554/eLife.45644.013

Figure 6 continued on next page

Figure 6 continued

The following source data and figure supplements are available for figure 6:

Source data 1. Product ion spectra supporting the amino acid mutations shown in ***Figure 6***.

DOI: https://doi.org/10.7554/eLife.45644.024

Figure supplement 1. Unio crassus - Hic74 coverage.

DOI: https://doi.org/10.7554/eLife.45644.014

Figure supplement 2. Unio pictorum - Hic74 coverage.

DOI: https://doi.org/10.7554/eLife.45644.015

Figure supplement 3. Margaritifera margaritifera - Hic74 coverage.

DOI: https://doi.org/10.7554/eLife.45644.016

Figure supplement 4. Pseudunio auricularius - Hic74 coverage.

DOI: https://doi.org/10.7554/eLife.45644.017

Figure supplement 5. PesB - Hic74 coverage.

DOI: https://doi.org/10.7554/eLife.45644.018

Figure supplement 6. HavA - Hic74 coverage.

DOI: https://doi.org/10.7554/eLife.45644.019

Figure supplement 7. HavB - Hic74 coverage.

DOI: https://doi.org/10.7554/eLife.45644.020

Figure supplement 8. HavC - Hic74 coverage.

DOI: https://doi.org/10.7554/eLife.45644.021

Figure supplement 9. HorB - Hic74 coverage.

DOI: https://doi.org/10.7554/eLife.45644.022

Figure supplement 10. HorC - Hic74 coverage.

DOI: https://doi.org/10.7554/eLife.45644.023

degradation and/or modification (Asp and Ser decomposition, amino acid racemization, ***Appendix 1—figure 9***) but not high enough to induce mineral conversion, which starts to occur around 300 ˚C (***Yoshioka and Kitano, 1985***). Double-button PesB yielded low D/L values but only a modest number of peptides were identified (~100, much less compared to the Havnø samples, where at least 200–400 peptides were matched to known shell protein sequences); this suggests that the sample had not been diagenetically compromised (also supported by the amino acid data, ***Appendix 1—figure 7***, ***Appendix 1—figure 9***).

Freshwater mother-of-pearl: archaeological significance

We found that mother-of-pearl of freshwater origin (Unionoida) was used in three European sites over a wide geographical range but relatively short time span (~4200–3800 BCE). Crucially, the crafters manufacturing such highly-standardized ornaments belong to different cultural groups: Late Mesolithic, Neolithic and Copper Age. Our results settle the 'marine vs freshwater' debate (***Heumüller, 2012***) for the double-buttons from Hornstaad (***Borrello and Girod, 2008***), and confirm previous identifications for the Peştera Ungurească examples (***Girod, 2010a***). The use of freshwater nacre (*Unio* or *Margaritifera*) comes as a surprise for the Havnø material, a coastal shell midden with a dominance of marine resource exploitation and rich in marine shells perfectly suitable for the purpose of making beads, including the horsemussel *M. modiolus*. Therefore, this finding suggests that the importance of freshwater mother-of-pearl be re-evaluated.

Unio sp. shells were probably selected to make 'disc beads' in the Epipaleolithic of the Levant, at Eynan (Natufian, 10,000–8,000 BCE; ***Bar-Yosef Mayer, 2013***), and in Europe the presence of *Unio* sp. beads has been recorded 259 times according to the dataset gathered from the literature by ***Rigaud et al. (2015)***, mainly from Neolithic sites. Despite this relative frequency of (presumed) freshwater mollusc ornaments in prehistoric Europe, a systematic study of their exploitation as raw materials is almost completely lacking. This is especially surprising since it is known that unionoid shells were exploited for mother-of-pearl until the Middle Ages (***Bertin, 2015***). Indeed, North American freshwater mussels were the basis for the 'pearl rush' during the 19th century, and their overexploitation for pearl harvesting, for making nuclei to be inserted in *Pinctada* pearl oysters as well as button-making on an industrial scale, almost drove a high number of species to extinction (***Haag, 2012***).

The lack of comprehensive archaeological studies on freshwater molluscs can be explained by two main factors. The first is methodological: the typical *chaîne opératoire* of bead-making involves several steps that obliterate most of the anatomical features (e.g. hinge apophysis) that are usable for taxonomic identification. These include: cutting and abrading small pieces until they take a circular shape; perforating the disc (**Gurova and Bonsall, 2017**) or, in the case of the double-buttons, working the side (with an abrasive wire?) to shape the central groove (**Bertin, 2015**; **Borrello and Girod, 2008**). Our work provides a series of analytical tools for overcoming this issue and for determining the biological origin of the raw material.

A second, perhaps more relevant, factor is the long-standing perception that freshwater molluscs are inherently less 'prestigious' than marine species, because of their presumed local origin. However, marine and freshwater molluscs are used side-by-side in a number of instances, for example the high-status burials at Mulhouse-Est (**Bonnardin, 2009**), or complex *parures* from the Swiss Early Bronze Age (**Borrello and Girod, 2008**). This clearly demonstrates that both were held in the same 'esteem' by the craftsman and that her/his choice was dictated by reasons other than the 'exoticism' of the material. The use of freshwater mother-of-pearl at Hornstaad and Peştera Ungurească, two sites with large procurement networks of exotic raw materials and at which there are clear signs of specialised production of ornaments (including gold at Peştera Ungurească; **Biagi and Voytek, 2006**; **Heumüller, 2012**), also confirms that freshwater pearl mussels were seen as prized materials, locally available. Furthermore, the use of freshwater molluscs for the manufacture of the *doppelknöpfe* recovered from Havnø (together with unworked fragments of the shells, **Appendix 1—figure 2**) shows that the manufacture of these ornaments was consistently associated with the use of freshwater mother-of-pearl, even in marine settings. Therefore, the Late Mesolithic people of Jutland and the Neolithic people of central Europe were either exchanging the finished products/raw materials, or the knowledge that the manufacture of the double-buttons required the use of unionoid shells.

Why freshwater nacre?

It is clear that mother-of-pearl (nacre) from freshwater molluscs was a prime material of choice for the manufacture of shell double-buttons. Further investigation of other types of shell ornaments may reveal that this raw material was more frequently selected than previously thought, but in the meanwhile it is necessary to consider the reasons behind this choice.

Unionoids inhabit clean flowing waters (they are occasionally also found in lakes) and are dependent upon the presence of sufficient salmonid fish to carry the larval glochidial stage of the pearl mussel life cycle (**IUCN: International Union for Conservation of Nature, 2019**). It is highly likely that freshwater mussels were collected near the site (as supported by the 'local' isotope signatures in this study), and that the procurement of the mussels was not especially difficult (for example, *M. margaritifera* lives at depths of up to two meters) nor too time-consuming. Therefore, the choice of this material must have been linked to reasons other than its long-distance provenance, the skills involved in procurement, or its rarity; rather, it is more likely a result of the characteristics of the raw material per se (mechanical properties and aesthetic qualities) and its connection to other *things*, be these in the sensory world (the river and its water, the landscape) or in the symbolic.

Mother-of-pearl is exceptionally hard - a thousand times more resistant to fractures than its mineral alone (**Currey, 1977**; **Jackson et al., 1990**) - and unionoid shells (*Unio* and *Margaritifera* sp., but not *Anodonta*) typically have a rather thick layer of nacre, while in some of the marine molluscs (particularly those occurring in European waters, such as *Modiolus* sp.), the ratio between nacre and prisms shifts, favouring the latter, where the nacre only partly covers the inner surface of the shell. The preservation of the prismatic matte layer may indicate that the coloured periostracum, which can give an appealing effect of chromatic contrast, was deliberately kept, for aesthetic reasons. Alternatively, if the periostracum was removed by mild abrasion, this would have resulted in fully white ornaments, showing both the brilliance of nacre and the dullness of prisms. The white colour of the ornaments may have been associated to wellbeing, peace and fertility (**Trubitt, 2003**). White was certainly a sought-after effect, so much so that red-purple *Spondylus* shells were often worked in order to remove the striking hue and reveal the white underneath (**Borrello and Micheli, 2011**). At the same time, the gloss of mother-of-pearl has been linked, in historical periods, with spirituality, life, royalty, and pearl fishing is a tradition that dates back to the same period considered here, around or a few centuries before 5500 BCE, in the Arabian Peninsula (**Charpentier et al., 2012**).

The choice of the raw material could also be a reflection of the role of freshwater environments: the Neolithic is the period in which water, together with plants and animals, is 'domesticated' (**Garfinkel et al., 2006**; **Mithen, 2010**). Rivers provided fast access routes to Central, Western and Northern Europe for hunter-gatherers during the Palaeolithic and Mesolithic and, later on, for agriculturalists coming from the East (**Rowley-Conwy, 2011**). Despite their 'fluidity', rivers and lakes were meaningful and persistent places in the prehistoric landscape.

In summary, the streams, rivers and lakes near occupation sites were inhabited by organisms that provided the crafters with exceptional-quality raw material, easy to procure and which could be worked following a well-established *chaîne opératoire* in order to obtain a standardized result. The small white double-buttons could then be threaded using the central groove or pressed into the fabric or leather (**Kannegaard, 2013**). Our work thus highlights an interpretative bias whereby exoticism is considered the primary reason for choice of raw materials, and suggest that local environments held an equally important place in the mind of prehistoric people.

Conclusions

The first application of 'palaeoshellomics' has demonstrated that it is possible to recover and identify ancient proteins sequences from mollusc shell, despite significant analytical challenges due to the combined effects of several factors, including low protein concentrations, small samples sizes, diagenesis and database insufficiency (**Table 2**).

Our molecular data showed that molluscan proteins are similar across the four freshwater taxa we examined (*U. pictorum*, *U. crassus*, *M. margaritifera*, *P. auricularius*) and differ significantly from the two marine species (*O. edulis*, *M. modiolus*; **Table 2**, **Figures 4** and **5**). We confirmed that freshwater molluscan matrix proteins are characterized by highly repetitive low complexity domains (RLCs). This is consistent with results obtained on other shell taxa, and improves our understanding of the biomineralization mechanisms within these invertebrate systems.

The archaeological double-buttons examined here were all confidently identified as Unionoida, freshwater shells with a thick layer of mother-of-pearl, using a combination of mineralogical, geochemical and biochemical techniques (SEM, FTIR-ATR, oxygen and carbon isotopes, chiral amino acid analyses, palaeoshellomics). The analysis of the sequence of the shell matrix protein Hic74 supports the use of *Unio* or *Margaritifera* as the raw material for the three Havnø ornaments (excluding *Pseudunio*), but lack of coverage of most amino acid modification sites in the reference samples hampered identification to a lower taxonomic level (**Figure 6**, **Table 3**).

The high degree of standardization of the ornaments (**Figures 1** and **2**), as well as the consistent choice of freshwater mother-of-pearl as raw material indicate that, in Europe, between ~ 4200 and~3800 BCE, there was a common notion of the manufacture of the *doppelknöpfe*, which was shared by different cultural groups: Late Mesolithic (Ertebølle), early Late Neolithic (Hornstaad Group), and Copper Age (Toarte Pastilate/Coţofeni). Our in-depth study therefore puts into question the most commonly accepted interpretations, which privilege the preponderant use of exotic marine shells as prestigious raw materials for the manufacture of prehistoric shell ornaments.

Methods

Non-destructive characterization

Whole beads and fragments of the reference shells were observed using an environmental Scanning Electron Microscope (Hitachi TM1000 Tabletop Microscope). The mineralogy of the beads was identified by infrared spectroscopy in attenuated total reflectance (ATR) mode (FTIR-ATR) (Appendix 1, sections 3.1 and 3.2).

Biogenic carbonate isotopic analyses

Isotopic analysis was carried out on biogenic carbonate to obtain bulk $\delta^{13}C$ and $\delta^{18}O$ values for the double-buttons. Small amounts of cleaned samples (bleached using concentrated NaOCl (12% w/v) for 48 hr) were analysed using a Delta V Plus mass spectrometer coupled with a Kiel IV carbonate device (ThermoFisher). All steps are detailed in Appendix 1, section 3.3.

Protein analysis

All reference shells and beads were powdered using a clean mortar and pestle and accurately weighed.

Amino acid racemization

~2 mg of powder were selected for each double-button and were bleached for 48 hr using concentrated NaOCl (12% w/v) in order to isolate the intra-crystalline amino acids. The analysis of total hydrolysable amino acids (THAA) was carried out as detailed in **Demarchi et al. (2014)** and Appendix 1 (section 3.4).

Proteomics (Appendix 1, sections 3.5, 3.6)

Powdered reference shells and double-buttons were bleached using diluted NaOCl (2.6%) for 48 and 3 hr respectively. Demineralization was carried out using cold diluted acetic acid, the resulting solutions were thoroughly desalted, concentrated and lyophilized. All samples were digested using two proteolytic enzymes (trypsin and elastase) in order to maximise sequence coverage. LC-MS/MS analyses were carried out using a nanoflow HPLC instrument (U3000 RSLC Thermo Fisher Scientific) coupled to a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) for *M. modiolus*, *O. edulis*, *U. pictorum*, *M. margaritifera* and the Havnø and Hornstaad-Hörnle IA ornaments (MSAP CNRS laboratory, University of Lille); an Ultimate 3000 Dionex nanoHPLC instrument coupled with an Orbitrap Fusion (Thermo Fisher Scientific) mass analyzer was used for the analysis of *U. crassus*, *P. auricularius* and for sample PesB (Mass Spectrometry Biomolecules core facility, University of Turin). No systematic difference was detected between the data obtained at the two facilities, which could have potentially affected the identification of the samples.

Bioinformatic analysis was carried out using PEAKS Studio 8.5 (Bioinformatics Solutions Inc, **Ma et al., 2003**). The thresholds for peptide and protein identification were set as follows: protein false discovery rate (FDR) = 0.5%, protein score $-10\lg P \geq 40$, unique peptides ≥ 2, de novo sequences scores (ALC%) ≥ 50. The FDR is calculated by the software PEAKS using an approach called decoy fusion (**Zhang et al., 2012**), whereby target and decoy are concatenated for each protein, rather than searching a target database and a decoy database separately (which can result in FDR underestimation); the effect is that of improving accuracy without impacting on sensitivity. The de novo algorithm derives the peptide sequences from the tandem mass spectra without using a database, and it is therefore suitable for the study of organisms where molecular reference sequences are scarce, such as in the present case. Each amino acid residue in the sequence is given a score (0–99%), indicating how confident the software is for that *local* identification. The overall confidence of the de novo sequence is calculated as the Average of Local Confidence (ALC) score.

The Molluscan Protein Database used in this study comprised 633061 protein sequences, that is all sequences available on the National Centre for Biotechnology Information (NCBI) repository restricting the taxonomy to Mollusca (fasta database downloaded on 15/02/2018), excluding all common contaminants (cRAP; common Repository of Adventitious Proteins: http://www.thegpm.org/crap/). The ESTs database included 1149,723 expressed sequence tags, also restricting the taxonomy to Mollusca (fasta database downloaded on 15/02/2018) and including cRAP sequences. Unionoida molluscs are poorly represented in these databases: a search for Unionoidae on the NCBI Identical Protein Database retrieves 4562 entries, almost exclusively belonging to soft tissue proteins, for example cytochrome oxidase, NADH dehydrogenase and ATP synthase. Shell matrix-related sequences are few (<35), almost all from the transcriptome of the triangle sail mussel, *Hyriopsis cumingii* (first released in 2013 [**Bai et al., 2013**]), including proteins related to shell biomineralization (i.e. upsalin, Pif, calmodulin, hicsilin, hichin, Hic74, Hic52, nasilin, Ca-binding P-glycoprotein, perlucin). The same search for Margaritiferidae yields 207 entries, none of which related to shell matrix proteins.

Data deposition

All the mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (**Vizcaíno et al., 2013**) with the data set identifier PXD011985.

Acknowledgements

The authors are sincerely grateful to Peter Rowley-Conwy and two anonymous reviewers, as well as the Reviewing and Senior Editors, for exceptionally clear, valuable and insightful comments on this manuscript.

BD and JS would like to thank Rosa Boano, Cristina Giacoma, Massimo Maffei, Nadia Barbero and their teams at the University of Turin. JS and FM would like to thank Justine Briard (uB-FC) for help with sample preparation and Emmanuel Fara (uB-FC) for valuable insights and comments. BD, FM and JS are grateful to Matthew Collins (Universities of Copenhagen and Cambridge) and Harry Robson (University of York) for useful discussion.

BD is supported by the Italian 'Ministero dell'Istruzione dell'Università e della Ricerca' (Programma per Giovani Ricercatori 'Rita Levi Montalcini') and wishes to acknowledge the Department of Archaeology at the University of York (UK) and the EU FP7 Re(In)tegration grant PERG07-GA-2010–268429 (project: mAARiTIME) for funding pilot research.

JS, BD and FM are also supported by the PHC Galilée programme, Italo-French University (UIF/UFI; project G18−464/39612 SB) and JS acknowledges the support of MIUR and of a BGF funded by the French embassy in Italy.

The work performed at UMR CNRS 6282 Biogeosciences (JS, BD, FM), Dijon was financed via the annual recurrent budget of this research unit.

KP acknowledges the support of the Leverhulme Trust, PLP-2012–116.

LP thanks the CNRS (France) and the University of Bourgogne Franche-Comté (France).

GM is supported by 'Ricerca Locale Università di Torino'.

CT would like to thank Stéphanie Devassine, Fabrice Bray and Christian Rolando (MSAP, University of Lille) for their technical contribution for MS operation. CT also thanks the Institut Universitaire de France, the European community (FEDER), the Région Hauts de France, the IBISA network, the CNRS, the University of Lille and the FR 3624 FT-ICR.

Additional information

Funding

Funder	Grant reference number	Author
Ministry of Education, Universities and Research	Young Researchers	Beatrice Demarchi
European Commission	PERG-GA-2010-268429	Kirsty EH Penkman Beatrice Demarchi
Leverhulme Trust		Kirsty EH Penkman
Centre National de la Recherche Scientifique		Théophile Cocquerez Laurent Plasseraud Caroline Tokarski Jérôme Thomas Frédéric Marin
Campus France, Università Italo-Francese	PHC Galilée programme	Jorune Sakalauskaite Frédéric Marin Beatrice Demarchi

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Jorune Sakalauskaite, Resources, Data curation, Software, Formal analysis, Validation, Visualization, Methodology, Writing—original draft; Søren H Andersen, Paolo Biagi, Maria A Borrello, Alberto Girod, Resources, Writing—review and editing; Théophile Cocquerez, Laurent Plasseraud, Formal analysis, Investigation; André Carlo Colonese, Formal analysis, Writing—original draft; Federica Dal Bello, Sheila Taylor, Investigation; Marion Heumüller, Helmut Schlichtherle, Resources, Writing—original draft; Hannah Koon, Investigation, Writing—review and editing; Giorgia Mandili, Claudio Medana, Jérôme Thomas, Resources; Kirsty EH Penkman, Resources, Investigation, Writing—review

and editing; Caroline Tokarski, Resources, Formal analysis, Methodology, Writing—review and editing; Julie Wilson, Software, Formal analysis, Validation, Visualization, Writing—original draft; Frédéric Marin, Conceptualization, Resources, Formal analysis, Supervision, Validation, Investigation, Methodology, Writing—original draft; Beatrice Demarchi, Conceptualization, Resources, Data curation, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing—original draft, Project administration

Author ORCIDs
Jorune Sakalauskaite https://orcid.org/0000-0002-8029-8120
Federica Dal Bello http://orcid.org/0000-0003-0726-3025
Kirsty EH Penkman https://orcid.org/0000-0002-6226-9799
Jérôme Thomas http://orcid.org/0000-0002-1602-4416
Beatrice Demarchi https://orcid.org/0000-0002-8398-4409

Decision letter and Author response
Decision letter https://doi.org/10.7554/eLife.45644.061
Author response https://doi.org/10.7554/eLife.45644.062

Additional files

Supplementary files
• Transparent reporting form
DOI: https://doi.org/10.7554/eLife.45644.026

Data availability
All data generated or analysed during this study are included in the manuscript and supporting files. Source data files have been provided for Figures 4 and 5. All the mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the data set identifier PXD011985.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Demarchi B	2019	Palaeoshellomics	https://www.ebi.ac.uk/pride/archive/projects/PXD011985	ProteomeXchange, PXD011985

References
Alarashi H, Ortiz A, Molist M. 2018. Sea shells on the riverside: cowrie ornaments from the PPNB site of tell halula (Euphrates, northern Syria). *Quaternary International* **490**:98–112. DOI: https://doi.org/10.1016/j.quaint.2018.05.004

Albeck S, Aizenberg J, Addadi L, Weiner S. 1993. Interactions of various skeletal intracrystalline components with calcite crystals. *Journal of the American Chemical Society* **115**:11691–11697. DOI: https://doi.org/10.1021/ja00078a005

Altaba CR. 1990. The last known population of the freshwater mussel margaritifera auricularia (Bivalvia, unionoida): A conservation priority. *Biological Conservation* **52**:271–286. DOI: https://doi.org/10.1016/0006-3207(90)90072-W

Andersen SH. 2000. 'Køkkenmøddinger' (Shell Middens) in Denmark: a Survey. *Proceedings of the Prehistoric Society* **66**:361–384. DOI: https://doi.org/10.1017/S0079497X00001857

Andersen SH. 2008. A report on recent excavations at the shell midden of havnø in Denmark. *Mesolithic Miscellany* **19**:3–6.

Arivalagan J, Yarra T, Marie B, Sleight VA, Duvernois-Berthet E, Clark MS, Marie A, Berland S. 2017. Insights from the shell proteome: biomineralization to adaptation. *Molecular Biology and Evolution* **34**:66–77. DOI: https://doi.org/10.1093/molbev/msw219, PMID: 27744410

Aucour A-M, Sheppard SMF, Savoye R. 2003. δ^{13}C of fluvial mollusk shells (Rhône river): A proxy for dissolved inorganic carbon? *Limnology and Oceanography* **48**:2186–2193. DOI: https://doi.org/10.4319/lo.2003.48.6.2186

Bai Z, Zheng H, Lin J, Wang G, Li J. 2013. Comparative analysis of the transcriptome in tissues secreting purple and white nacre in the pearl mussel hyriopsis cumingii. *PLOS ONE* **8**:e53617. DOI: https://doi.org/10.1371/journal.pone.0053617, PMID: 23341956

Bajnóczi B, Schöll-Barna G, Kalicz N, Siklósi Z, Hourmouziadis GH, Ifantidis F, Kyparissi-Apostolika A, Pappa M, Veropoulidou R, Ziota C. 2013. Tracing the source of late neolithic spondylus shell ornaments by stable isotope geochemistry and cathodoluminescence microscopy. *Journal of Archaeological Science* **40**:874–882. DOI: https://doi.org/10.1016/j.jas.2012.09.022

Bar-Yosef Mayer DE, Vandermeersch B, Bar-Yosef O. 2009. Shells and ochre in middle paleolithic qafzeh cave, Israel: indications for modern behavior. *Journal of Human Evolution* **56**:307–314. DOI: https://doi.org/10.1016/j.jhevol.2008.10.005, PMID: 19285591

Bar-Yosef Mayer DE. 2013. Towards a typology of stone beads in the neolithic levant. *Journal of Field Archaeology* **38**:129–142. DOI: https://doi.org/10.1179/0093469013Z.00000000043

Bar-Yosef Mayer DE. 2018. Shell ornaments and artifacts in neolithic Cyprus and correlations with other mediterranean regions. *Quaternary International* **464**:206–215. DOI: https://doi.org/10.1016/j.quaint.2017.06.034

Bartolomei G. 2013. I micromammiferi della grotta peştera ungurească nella gola di cheile turzii. *Atti Della Società Per La Preistoria E Protostoria Della Regione Friuli-Venezia Giulia* **2011**:107–117.

Bertin L. 2015. Exploitation of mother of pearl in the middle ages, clos d'ugnac archaeological site (Pennautier, aude, France): Malacological study, consumption, exploitation and utilization of the nacre. *Quaternary International* **375**:145–152. DOI: https://doi.org/10.1016/j.quaint.2015.04.042

Biagi P, Voytek B. 2006. Excavations at Pestera Ungureasca (Caprelor) (Cheile turzii, petresti de jos, Transylvania) 2003-2004: a preliminary report on the chipped stone assemblages from the chalcolithic toarte pastilate (Bodrogkeresztur) layers. *Analele Banatului* **14**:177–202.

Biddittu I, Girod A. 2003. La diffusione pleistocenica e olocenica di pseudunio auricularius (Mollusca, bivalvia) in italia. Atti Del 4° Convegno Nazionale Di Archeozoologia, Quaderni Del Museo Archeologico Del Friuli Occidentale , Rovereto 127–137.

Billamboz A. 2006. Dendroarchäologische Untersuchungen in den neolithischen Ufersiedlungen von Hornstaad-HörnleHornstaad-Hörnle 1A. Die Befunde Einer Jungneolithischen Pfahlbausiedlung Am Westlichen Bodensee, Vol IX. Siedlungarchologie Im Alpenworland. Stuttgart:Theiss.

Boggild OB. 1930. *The shell structure of the mollusks*. Kjobenhavn: A.F. Host.

Bolotov IN, Vikhrev IV, Bespalaya YV, Gofarov MY, Kondakov AV, Konopleva ES, Bolotov NN, Lyubas AA. 2016. Multi-locus fossil-calibrated phylogeny, biogeography and a subgeneric revision of the margaritiferidae (Mollusca: bivalvia: unionoida). *Molecular Phylogenetics and Evolution* **103**:104–121. DOI: https://doi.org/10.1016/j.ympev.2016.07.020, PMID: 27444710

Bonnardin S. 2009. *La Parure Funéraire Au Néolithique Ancien Dans Les Bassins Parisien Et Rhénan: Rubané, Hinkelstein Et Villeneuve-Saint-Germain*. Société Préhistorique Française.

Borrello M, Girod A. 2008. Bivalves d'eau douce, matière première pour la fabrication de parures en Suisse, du Néolithique à l'âge du Bronze. *Annuaire d'Archéologie Suisse*:108–115.

Borrello MA, Micheli R. 2011. *Spondylus gaederopus* in Prehistoric Italy: Jewels from Neolithic and Copper Age Sites. In: Ifantidis F, Nikolaïdou M (Eds). *Spondylus in Prehistory: New Data and Approaches. Contributions to the Archaeology of Shell Technologies*. Archaeopress. p. 25–37.

Bosch MD, Mannino MA, Prendergast AL, O'Connell TC, Demarchi B, Taylor SM, Niven L, van der Plicht J, Hublin JJ. 2015. New chronology for ksâr 'Akil (Lebanon) supports Levantine route of modern human dispersal into Europe. *PNAS* **112**:7683–7688. DOI: https://doi.org/10.1073/pnas.1501529112, PMID: 26034284

Boschian G. 2010. Soil micromorphological observations on the anthropogenous deposits of peştera Ungurească. *Atti Della Società Per La Preistoria E Protostoria Della Regione Friuli-Venezia Giulia* **2009**:157–169.

Carter JG. 1990. *Skeletal Biomineralization: Pattern, Processes, and Evolutionary Trends*. London: Springer.

Charpentier V, Phillips CS, Méry S. 2012. Pearl fishing in the ancient world: 7500 BP. *Arabian Archaeology and Epigraphy* **23**:1–6. DOI: https://doi.org/10.1111/j.1600-0471.2011.00351.x

Colonese AC, Mannino MA, Bar-Yosef Mayer DE, Fa DA, Finlayson JC, Lubell D, Stiner MC. 2011. Marine mollusc exploitation in Mediterranean prehistory: an overview. *Quaternary International* **239**:86–103. DOI: https://doi.org/10.1016/j.quaint.2010.09.001

Crisp MK. 2013. *Amino Acid Racemization Dating: Method Development Using African Ostrich (Struthio Camelus) Eggshell*. University of York.

Currey JD. 1977. Mechanical properties of mother of pearl in tension. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **196**:443–463. DOI: https://doi.org/10.1098/rspb.1977.0050

d'Errico F, Henshilwood C, Vanhaeren M, van Niekerk K. 2005. *Nassarius kraussianus* shell beads from blombos cave: evidence for symbolic behaviour in the middle stone age. *Journal of Human Evolution* **48**:3–24. DOI: https://doi.org/10.1016/j.jhevol.2004.09.002, PMID: 15656934

Demarchi B, Williams MG, Milner N, Russell N, Bailey G, Penkman K. 2011. Amino acid racemization dating of marine shells: a mound of possibilities. *Quaternary International* **239**:114–124. DOI: https://doi.org/10.1016/j.quaint.2010.05.029, PMID: 21776187

Demarchi B, O'Connor S, de Lima Ponzoni A, de Almeida Rocha Ponzoni R, Sheridan A, Penkman K, Hancock Y, Wilson J. 2014. An integrated approach to the taxonomic identification of prehistoric shell ornaments. *PLOS ONE* **9**:e99839. DOI: https://doi.org/10.1371/journal.pone.0099839

Demarchi B, Clements E, Coltorti M, van de Locht R, Kröger R, Penkman K, Rose J. 2015. Testing the effect of bleaching on the bivalve *glycymeris*: A case study of amino acid geochronology on key Mediterranean raised beach deposits. *Quaternary Geochronology* **25**:49–65. DOI: https://doi.org/10.1016/j.quageo.2014.09.003

Demarchi B, Hall S, Roncal-Herrero T, Freeman CL, Woolley J, Crisp MK, Wilson J, Fotakis A, Fischer R, Kessler BM, Rakownikow Jersie-Christensen R, Olsen JV, Haile J, Thomas J, Marean CW, Parkington J, Presslee S, Lee-Thorp J, Ditchfield P, Hamilton JF, et al. 2016. Protein sequences bound to mineral surfaces persist into deep time. *eLife* **5**:e17092. DOI: https://doi.org/10.7554/eLife.17092, PMID: 27668515

Dieckmann B, Harwath A, Hoffstadt J. 2006. *HornstaadHörnle 1A. Die Befunde Einer Jungneolithischen Pfahlbausiedlung Am Westlichen Bodensee*. In: Baden-Württemberg LFD (Ed). *Siedlungsarch Im Alpenvorland 9. K. Theiss*. Globesystems. p. 11–275.

Dieckmann B, Harwath A, Heumüller M, Hoffstadt J, Maier U, Matuschik I, Schwoerbel A, Stephan E, Schweizer-Strobel P, Styring A, Schlichtherle H, Theune-Großkopf B. 2016. Pfahlbauten J (Ed). *Eine Kurze Dorfgeschichte. Hornstaad-Hörnle I Am Bodensee*. J Thorbecke.

Eerkens JW, Herbert GS, Rosenthal JS, Spero HJ. 2005. Provenance analysis of olivella biplicata shell beads from the California and Oregon coast by stable isotope fingerprinting. *Journal of Archaeological Science* **32**:1501–1514. DOI: https://doi.org/10.1016/j.jas.2005.04.005

Gao P, Liao Z, Wang XX, Bao LF, Fan MH, Li XM, Wu CW, Xia SW. 2015. Layer-by-Layer proteomic analysis of *Mytilus galloprovincialis* Shell. *PLOS ONE* **10**:e0133913. DOI: https://doi.org/10.1371/journal.pone.0133913, PMID: 26218932

Garfinkel Y, Vered A, Bar-Yosef O. 2006. The domestication of water: the neolithic well at Sha'ar Hagolan, Jordan Valley, Israel. *Antiquity* **80**:686–696. DOI: https://doi.org/10.1017/S0003598X00094138

GBIF.org. 2018. Free and open access to biodiversity data. https://www.gbif.org [Accessed November 20, 2018].

Giacobini G. 2007. Richness and diversity of burial rituals in the upper paleolithic. *Diogenes* **54**:19–39. DOI: https://doi.org/10.1177/0392192107077649

Gillikin DP, Hutchinson KA, Kumai Y. 2009. Ontogenic increase of metabolic carbon in freshwater mussel shells (*Pyganodon cataracta*) . *Journal of Geophysical Research* **114**:2186. DOI: https://doi.org/10.1029/2008JG000829

Girod A. 2010a. Land and freshwater snails from peştera Ungurească (Cheile turzii gorge, Transylvania, Romania). *Atti Della Società Per La Preistoria E Protostoria Della Regione Friuli-Venezia Giulia* **2008-2009**:187–199.

Girod A. 2010b. Lavorazioni su conchiglia di *pseudunio auricularius* (Mollusca, bivalvia) nel neolitico antico a isorella (BS), Italia settentrionale. *Atti Della Società Italiana Di Scienze Naturali E Del Museo Civico Di Storia Naturale Di Milano* **151**:89–98.

Gower JC. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**:325–338. DOI: https://doi.org/10.1093/biomet/53.3-4.325

Gurova M, Bonsall C. 2017. Experimental Replication of Stone, Bone and Shell Beads from Early Neolithic Sites in Southeast Europe. In: Bar-Yosef Mayer D, Bonsall C, Choyke A. M (Eds). *Not Just for Show: The Archaeology of Beads, Beadwork and Personal Ornaments*. Oxbow Books. p. 159–168.

Haag WR. 2012. *North American Freshwater Mussels: Natural History, Ecology, and Conservation*. Cambridge University Press. DOI: https://doi.org/10.1017/cbo9781139048217

Henry DG, Watson JS, John CM. 2017. Assessing and calibrating the ATR-FTIR approach as a carbonate rock characterization tool. *Sedimentary Geology* **347**:36–52. DOI: https://doi.org/10.1016/j.sedgeo.2016.07.003

Heumüller M. 2009. Der Schmuck der jungneolithischen Seeufersiedlung Hornstaad-Hörnle IA im Rahmen des mitteleuropäischen Mittel- und Jungneolithikums. K. Theiss.

Heumüller M. 2010. Schmuck als Zeichen weit gespannter und lang andauernder Kommunikationsräume im 5. und frühen 4. Jahrtausend. In: Ralf Gleser V. B (Ed). *Neolithikumund ältere Metallzeiten: Studien und Materialien*. Universität Münster .

Heumüller M. 2012. Schmuck als Zeichen weit gespannter und lang andauernder Kommunikationsräume im 5. und frühen 4. Jahrtausend. In: Ralf G, Valeska B (Eds). *Neolithikum und ältere Metallzeiten: Studien und Materialien*. Universität Münster. p. 359–387.

IUCN: International Union for Conservation of Nature. 2019. Red list. http://www.iucnredlist.org/searchNovember 12, 2018].

Jackson AP, Vincent JFV, Turner RM. 1990. Comparison of nacre with other ceramic composites. *Journal of Materials Science* **25**:3173–3178. DOI: https://doi.org/10.1007/BF00587670

Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R, Rokhsar DS, Montagnani C, Joubert C, Piquemal D, Degnan BM. 2010. Parallel evolution of nacre building gene sets in molluscs. *Molecular Biology and Evolution* **27**:591–608. DOI: https://doi.org/10.1093/molbev/msp278, PMID: 19915030

Kannegaard E. 2013. Småting fra nederst. *Årbog Fra Museum Østjylland*:207–208.

Keith ML, Anderson GM, Eichler R. 1964. Carbon and oxygen isotopic composition of mollusk shells from marine and fresh-water environments. *Geochimica Et Cosmochimica Acta* **28**:1757–1786. DOI: https://doi.org/10.1016/0016-7037(64)90021-3

Kuhn SL, Stiner MC, Reese DS, Güleç E. 2001. Ornaments of the earliest upper paleolithic: new insights from the levant. *PNAS* **98**:7641–7646. DOI: https://doi.org/10.1073/pnas.121590798, PMID: 11390976

Launey S, Ledu C, Boudry P, Bonhomme F, Naciri-Graven Y. 2002. Geographic structure in the european flat oyster (*Ostrea edulis L.*) as revealed by microsatellite polymorphism. *Journal of Heredity* **93**:331–351. DOI: https://doi.org/10.1093/jhered/93.5.331, PMID: 12547921

Lazarovici G, Lazarovici CM. 2013. Etapa veche din atelierul de bijuterii de la cheile turzii - Peştera Ungurească. *Archeovest I*.

Lazarovici CM, Lazarovici G. 2016. The jewelry workshop at Cheile turzii: cave Binder, copper age. *Annales Universitatis Apulensis Series Historica* **20**:335–357.
Le Roy N, Jackson DJ, Marie B, Ramos-Silva P, Marin F. 2014. The evolution of metazoan α-carbonic anhydrases and their roles in calcium carbonate biomineralization. *Frontiers in Zoology* **11**:75. DOI: https://doi.org/10.1186/s12983-014-0075-8
Leng MJ, Lewis JP. 2016. Oxygen isotopes in Molluscan shell: applications in environmental archaeology. *Environmental Archaeology* **21**:295–306. DOI: https://doi.org/10.1179/1749631414Y.0000000048
Liu X, Dong S, Jin C, Bai Z, Wang G, Li J. 2015. Silkmapin of *hyriopsis cumingii*, a novel silk-like shell matrix protein involved in nacre formation. *Gene* **555**:217–222. DOI: https://doi.org/10.1016/j.gene.2014.11.006, PMID: 25447895
Liu X, Jin C, Wu L, Dong S, Zeng S, Li J. 2017a. Hic74, a novel alanine and glycine rich matrix protein related to nacreous layer formation in the mollusc *hyriopsis cumingii*. *Aquaculture and Fisheries* **2**:119–123. DOI: https://doi.org/10.1016/j.aaf.2017.03.005
Liu X, Pu J, Zeng S, Jin C, Dong S, Li J. 2017b. *Hyriopsis cumingii* Hic52-A novel nacreous layer matrix protein with a collagen-like structure. *International Journal of Biological Macromolecules* **102**:667–673. DOI: https://doi.org/10.1016/j.ijbiomac.2017.03.154, PMID: 28392384
Loftus E, Rogers K, Lee-Thorp J. 2015. A simple method to establish calcite:aragonite ratios in archaeological mollusc shells. *Journal of Quaternary Science* **30**:731–735. DOI: https://doi.org/10.1002/jqs.2819
Lopes-Lima M, Sousa R, Geist J, Aldridge DC, Araujo R, Bergengren J, Bespalaya Y, Bódis E, Burlakova L, Van Damme D, Douda K, Froufe E, Georgiev D, Gumpinger C, Karatayev A, Ü K, Killeen I, Lajtner J, Larsen BM, Lauceri R, et al. 2017. Conservation status of freshwater mussels in Europe: state of the art and future challenges: conservation of european freshwater mussels. *Biol Rev, Invertebrates* **92**:572–607. DOI: https://doi.org/10.1111/brv.12244
Lopes-Lima M, Bolotov IN, Do VT, Aldridge DC, Fonseca MM, Gan HM, Gofarov MY, Kondakov AV, Prié V, Sousa R, Varandas S, Vikhrev IV, Teixeira A, Wu RW, Wu X, Zieritz A, Froufe E, Bogan AE. 2018. Expansion and systematics redefinition of the most threatened freshwater mussel family, the margaritiferidae. *Molecular Phylogenetics and Evolution* **127**:98–118. DOI: https://doi.org/10.1016/j.ympev.2018.04.041, PMID: 29729933
Lydeard C, Cowie RH, Ponder WF, Bogan AE, Bouchet P, Clark SA, Cummings KS, Frest TJ, Gargominy O, HERBERT DAIG, Hershler R, Perez KE, Roth B, Seddon M, Strong EE, Thompson FG. 2004. The global decline of nonmarine mollusks. *BioScience* **54**:321–330. DOI: https://doi.org/10.1641/0006-3568(2004)054[0321:TGDONM]2.0.CO;2
Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G. 2003. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* **17**:2337–2342. DOI: https://doi.org/10.1002/rcm.1196, PMID: 14558135
Marie B, Marin F, Marie A, Bédouet L, Dubost L, Alcaraz G, Milet C, Luquet G. 2009. Evolution of nacre: biochemistry and proteomics of the shell organic matrix of the cephalopod *nautilus macromphalus*. *ChemBioChem* **10**:1495–1506. DOI: https://doi.org/10.1002/cbic.200900009, PMID: 19472248
Marie B, Jackson DJ, Ramos-Silva P, Zanella-Cléon I, Guichard N, Marin F. 2013. The shell-forming proteome of *lottia gigantea* reveals both deep conservations and lineage-specific novelties. *FEBS Journal* **280**:214–232. DOI: https://doi.org/10.1111/febs.12062, PMID: 23145877
Marie B, Arivalagan J, Mathéron L, Bolbach G, Berland S, Marie A, Marin F. 2017. Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the Unionoida *Elliptio complanata* and *Villosa lienosa* . *Journal of the Royal Society Interface* **14**. DOI: https://doi.org/10.1098/rsif.2016.0846
Marin F, Pokroy B, Luquet G, Layrolle P, De Groot K. 2007. Protein mapping of calcium carbonate biominerals by immunogold. *Biomaterials* **28**:2368–2377. DOI: https://doi.org/10.1016/j.biomaterials.2007.01.029, PMID: 17306364
Marin F, Luquet G. 2008. *Unusually Acidic Proteins in Biomineralization*. In: Bäuerlein E (Ed). *Handbook of Biomineralization: Biological Aspects and Structure Formation*. John Wiley & Sons. p. 273–290.
Marin F, Le Roy N, Marie B.2012. The formation and mineralization of mollusk shell. *Frontiers in Bioscience* **S4**: 1099–1125. DOI: https://doi.org/10.2741/s321
Marin F, Marie B, Hamada SB, Ramos-Silva P, Le Roy N, Guichard N, Wolf SE, Montagnani C, Joubert C, Piquemal D, Saulnier D, Gueguen Y. 2013. 'Shellome': Proteins Involved in Mollusk Shell Biomineralization-Diversity, Functions. International Symposium on Pearl Research 149–166.
Marin F, Bundeleva I, Takeuchi T, Immel F, Medakovic D. 2016. Organic matrices in metazoan calcium carbonate skeletons: composition, functions, evolution. *Journal of Structural Biology* **196**:98–106. DOI: https://doi.org/10.1016/j.jsb.2016.04.006, PMID: 27085423
Marin F, Luquet G. 2004. Molluscan shell proteins. *Comptes Rendus Palevol* **3**:469–492. DOI: https://doi.org/10.1016/j.crpv.2004.07.009
McConnaughey TA, Gillikin DP. 2008. Carbon isotopes in mollusk shell carbonates. *Geo-Marine Letters* **28**:287–299. DOI: https://doi.org/10.1007/s00367-008-0116-4
Mithen S. 2010. The domestication of water: water management in the ancient world and its prehistoric origins in the jordan valley. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **368**:5249–5274. DOI: https://doi.org/10.1098/rsta.2010.0191
Nisbet R. 2010. Copper age wild and cultivated plants from peştera ungurească cave (Transylvania, Romania). *Atti Della Società Per La Preistoria E Protostoria Della Regione Friuli-Venezia Giulia* **2008-2009**:171–185.
Ortiz JE, Sánchez-Palencia Y, Gutiérrez-Zugasti I, Torres T, González-Morales M. 2018. Protein diagenesis in archaeological gastropod shells and the suitability of this material for amino acid racemisation dating: *phorcus*

lineatus (da Costa, 1778). *Quaternary Geochronology* **46**:16–27. DOI: https://doi.org/10.1016/j.quageo.2018.02.002

Palmblad M, Deelder AM. 2012. Molecular phylogenetics by direct comparison of tandem mass spectra. *Rapid Communications in Mass Spectrometry* **26**:728–732. DOI: https://doi.org/10.1002/rcm.6162, PMID: 22368051

Penkman KE, Preece RC, Keen DH, Maddy D, Schreve DC, Collins MJ. 2007. Testing the aminostratigraphy of fluvial archives: the evidence from intra-crystalline proteins within freshwater shells. *Quaternary Science Reviews* **26**:2958–2969. DOI: https://doi.org/10.1016/j.quascirev.2007.06.034, PMID: 19684880

Penkman KE, Kaufman DS, Maddy D, Collins MJ. 2008. Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells. *Quaternary Geochronology* **3**:2–25. DOI: https://doi.org/10.1016/j.quageo.2007.07.001, PMID: 19684879

Penkman KE, Preece RC, Bridgland DR, Keen DH, Meijer T, Parfitt SA, White TS, Collins MJ. 2011. A chronological framework for the british quaternary based on *Bithynia opercula*. *Nature* **476**:446–449. DOI: https://doi.org/10.1038/nature10305, PMID: 21804567

Pierini F, Demarchi B, Turner J, Penkman K. 2016. Pecten as a new substrate for IcPD dating: the quaternary raised beaches in the gulf of Corinth, Greece. *Quaternary Geochronology* **31**:40–52. DOI: https://doi.org/10.1016/j.quageo.2015.10.006

Ramos-Silva P, Benhamada S, Le Roy N, Marie B, Guichard N, Zanella-Cléon I, Plasseraud L, Corneillat M, Alcaraz G, Kaandorp J, Marin F. 2012. Novel molluskan biomineralization proteins retrieved from proteomics: a case study with upsalin. *ChemBioChem* **13**:1067–1078. DOI: https://doi.org/10.1002/cbic.201100708, PMID: 22474002

Rieder V, Blank-Landeshammer B, Stuhr M, Schell T, Biß K, Kollipara L, Meyer A, Pfenninger M, Westphal H, Sickmann A, Rahnenführer J. 2017. DISMS2: a flexible algorithm for direct proteome- wide distance calculation of LC-MS/MS runs. *BMC Bioinformatics* **18**:148. DOI: https://doi.org/10.1186/s12859-017-1514-2, PMID: 28253837

Rigaud S, d'Errico F, Vanhaeren M. 2015. Ornaments reveal resistance of north european cultures to the spread of farming. *PLOS ONE* **10**:e0121166. DOI: https://doi.org/10.1371/journal.pone.0121166, PMID: 25853888

Rigaud S, Manen C, García-Martínez de Lagrán I. 2018. Symbols in motion: flexible cultural boundaries and the fast spread of the neolithic in the western mediterranean. *PLOS ONE* **13**:e0196488. DOI: https://doi.org/10.1371/journal.pone.0196488, PMID: 29715284

Rowley-Conwy P. 2011. Westward ho! the spread of agriculturalism from central europe to the Atlantic. *Current Anthropology* **52**:S431–S451. DOI: https://doi.org/10.1086/658368

Rowley-Conwy P. 2014. Foragers and farmers in Mesolithic/Neolithic Europe, 5500-3900 cal BC: beyond the anthropological comfort zone. In: Foulds F, Drinkall H, Perri A, Clinnick D, Walker J (Eds). *Wild Things: Recent Advances in Palaeolithic and Mesolithic Research*. Oxford: Oxbow Books. p. 185–201.

Saunders NJ. 1999. Biographies of brilliance: pearls, transformations of matter and being, c. AD 1492. *World Archaeology* **31**:243–257. DOI: https://doi.org/10.1080/00438243.1999.9980444

Schlichtherle H. 1990. Die Sondagen 1973–1978 in den Ufersiedlungen Hornstaad-Hörnle ISiedlungsarchäologie Im Alpenvorland I. Forsch. U. Ber. Zur Vor- U. Frühgesch. Baden-Württemberg 36. K Theiss.

Starnini E, Biagi P, Mazzucco N. 2018. The beginning of the neolithic in the po plain (northern Italy): Problems and perspectives. *Quaternary International* **470**:301–317. DOI: https://doi.org/10.1016/j.quaint.2017.05.059

Stiner MC, Kuhn SL, Güleç E. 2013. Early upper paleolithic shell beads at Üçağızlı cave I (Turkey): technology and the socioeconomic context of ornament life-histories. *Journal of Human Evolution* **64**:380–398. DOI: https://doi.org/10.1016/j.jhevol.2013.01.008, PMID: 23481346

Stiner MC. 2014. Finding a common bandwidth: causes of convergence and diversity in paleolithic beads. *Biological Theory* **9**:51–64. DOI: https://doi.org/10.1007/s13752-013-0157-4

Sudo S, Fujikawa T, Nagakura T, Ohkubo T, Sakaguchi K, Tanaka M, Nakashima K, Takahashi T. 1997. Structures of mollusc shell framework proteins. *Nature* **387**:563–564. DOI: https://doi.org/10.1038/42391, PMID: 9177341

Taborin Y. 1974. La parure en coquillage de l'Épipaléolilhique au bronze ancien en France (suite). *Gallia Préhistoire* **17**:307–417. DOI: https://doi.org/10.3406/galip.1974.1470

Taylor JD, Kennedy WJ, Hall A. 1969. The shell structure and mineralogy of bivalvia: introduction, Nuculacea-Trigonacea, bulletin of the british museum (Natural history). *Zoology*.

Taylor JD, Kennedy WJ, Hall A. 1973. The shell structure and mineralogy of the bivalvia. *II. Lucinacea-Clavagellacea, Conclusions, 9. London: The British Museum*.

Trubitt MBD. 2003. The production and exchange of marine shell prestige goods. *Journal of Archaeological Research* **11**:243–277.

Vanhaeren M, d'Errico F, Billy I, Grousset F. 2004. Tracing the source of upper palaeolithic shell beads by strontium isotope dating. *Journal of Archaeological Science* **31**:1481–1488. DOI: https://doi.org/10.1016/j.jas.2004.03.011

Vanhaeren M, d'Errico F. 2006. Aurignacian ethno-linguistic geography of europe revealed by personal ornaments. *Journal of Archaeological Science* **33**:1105–1128. DOI: https://doi.org/10.1016/j.jas.2005.11.017

Verdegaal S, Troelstra SR, Beets CJ, Vonhof HB. 2005. Stable isotopic records in unionid shells as a paleoenvironmental tool. *Netherlands Journal of Geosciences* **84**:403–408. DOI: https://doi.org/10.1017/S0016774600021211

Versteegh EAA, Vonhof HB, Troelstra SR, Kaandorp RJG, Kroon D. 2010. Seasonally resolved growth of freshwater bivalves determined by oxygen and carbon isotope shell chemistry. *Geochemistry, Geophysics, Geosystems : G3* **11**. DOI: https://doi.org/10.1029/2009gc002961

Vizcaíno JA, Côté RG, Csordas A, Dianes JA, Fabregat A, Foster JM, Griss J, Alpi E, Birim M, Contell J, O'Kelly G, Schoenegger A, Ovelleiro D, Pérez-Riverol Y, Reisinger F, Ríos D, Wang R, Hermjakob H. 2013. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Research* **41**: D1063–D1069. DOI: https://doi.org/10.1093/nar/gks1262, PMID: 23203882

WaterIsotopes.org. 2018. Gridded maps of the isotopic composition of meteoric waters. http://wateriso.utah.edu/waterisotopes/index.html [Accessed November 7, 2018].

Westbroek P, Marin F. 1998. A marriage of bone and nacre. *Nature* **392**:861–862. DOI: https://doi.org/10.1038/31798, PMID: 9582064

Wustman BA, Morse DE, Evans JS. 2002. Structural analyses of polyelectrolyte sequence domains within the adhesive elastomeric biomineralization protein lustrin A. *Langmuir* **18**:9901–9906. DOI: https://doi.org/10.1021/la025927m

Yoshioka S, Kitano Y. 1985. Transformation of aragonite to calcite through heating. *Geochemical Journal* **19**: 245–249. DOI: https://doi.org/10.2343/geochemj.19.245

Yılmaz Ş, Victor B, Hulstaert N, Vandermarliere E, Barsnes H, Degroeve S, Gupta S, Sticker A, Gabriël S, Dorny P, Palmblad M, Martens L. 2016. A pipeline for differential proteomics in unsequenced species. *Journal of Proteome Research* **15**:1963–1970. DOI: https://doi.org/10.1021/acs.jproteome.6b00140, PMID: 27089233

Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie GA, Ma B. 2012. PEAKS DB: *De Novo* Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. *Molecular & Cellular Proteomics* **11**:M111.010587. DOI: https://doi.org/10.1074/mcp.M111.010587

Zilhão J, Angelucci DE, Badal-García E, d'Errico F, Daniel F, Dayet L, Douka K, Higham TF, Martínez-Sánchez MJ, Montes-Bernárdez R, Murcia-Mascarós S, Pérez-Sirvent C, Roldán-García C, Vanhaeren M, Villaverde V, Wood R, Zapata J. 2010. Symbolic use of marine shells and mineral pigments by iberian neandertals. *PNAS* **107**:1023–1028. DOI: https://doi.org/10.1073/pnas.0914088107, PMID: 20080653

Discussion

This PhD thesis comprises five research articles which develop topics within the subdisciplines of biomineralization research, palaeoproteomics and archaeology. The overarching aim of the work was to develop and apply proteomics approaches to the study of prehistoric mollusc shells.

As presented in the introduction, mollusc shell remains and shell-made objects are abundant in archaeological sites and represent an important source of information for studying past cultures. However, in archaeological research, the knowledge about shell use is often incomplete, due to the usual limitations of standard analytical techniques for identifying shell sources. This is particularly the case when archaeological shell objects are small, heavily worked, or are found fragmented and/or degraded.

The aim of this thesis was to develop a proteomics-based approach to characterise and identify sub-fossil mollusc shells from the archaeological/paleontological record. The approach developed here is therefore grounded in biomineralization research and represents the first application of palaeoproteomics to calcified molluscan skeletons, and in general to any invertebrate system.

Three main research lines were developed in this thesis:

- An in-depth assessment of shell proteins and their use as molecular archives for taxonomic and phylogenetic information;
- The study of shell protein diagenesis patterns in order to assess the potential of protein preservation over geological timescales;
- The development and application of shell palaeoproteomics to study archaeological mollusc shells.

This discussion summarises the main findings, highlighting the key advances in shell palaeoproteomics research.

Shell proteins – molecular archives of phylogenetic and taxonomic information

The first key question addressed in this thesis was whether shell proteins carry phylogenetic/taxonomic information that could be decoded and used as molecular barcodes for taxonomic determination. Indeed, an in-depth biomolecular and proteomics study of *Spondylus gaederopus* shell presented in Chapter 1 suggests that *Spondylus* proteome represents lineage-specific sequences. The study supported previous “shellomics” research on different shell models, which had revealed that shell protein sequences can diverge drastically or be completely

different between species (Jackson et al., 2010; Kocot et al., 2016; Marin et al., 2016; McDougall and Degnan, 2018; Marin, 2020).

The proteome of *Spondylus* was characterised by several protein sequences, almost all belonging to a pectinoid shell reference, the Yesso scallop *Mizuhopecten yessoensis* (Chapter 1). From a phylogenetic point of view, the two specimens belong to the same Pectinida order, although some molecular studies consider *Spondylus* to be a sister group of the Pectinidae family (Waller, 2006; Smedley et al., 2019). Interestingly, the shells are also characterised by different structures, as *Pectinidae* shells are calcitic foliated while *Spondylus* (and other members of the Spondylidae family) exhibit a complex arrangement of different microstructures, dominated by aragonitic crossed-lamellar (Carter, 1990). Hence, the exact evolutionary relationship between the specimens is unclear. The reason why the *Spondylus* proteome contains sequences from *Mizuhopecten* is because it is the phylogenetically closest species which has been extensively studied by molecular approaches (Wang et al., 2017) and has a full genome sequenced (Mao et al., 2018). This supports the idea that shell proteins do carry a taxonomic – even phylogenetic – signal, although this one may be blurred. For instance, considering the variety of shells that have been studied by proteomics so far, in *Spondylus*, we do not detect sequences from oysters (Marie et al., 2011a; Yarra et al., 2016) (even though *Spondylus* is often named as the thorny or spiny oyster, it is not a true oyster, *i.e.* of order Ostreida). We have also not found sequences of other crossed-lamellar shells currently available in the database *e.g. Lymnaea stagnalis* (Herlitze et al., 2018), which could possibly suggest a correlation between shell proteins and microstructural information. The Spondylid family diverged from Pectinoidea around 154 million years ago (Waller, 2006; Smedley et al., 2019), and among other shell specimens, it is considered to be of relatively recent evolution. Taking into account other shells that have evolved recently, as for example the venerid bivalves (Carter, 1990), homologous proteins from these shells are not found in the proteome of *Spondylus* (Marie et al., 2011b; Arivalagan et al., 2016). This indicates that protein similarities between *Spondylus* and *Pecten* shells are likely due to phylogenetic relatedness of these species.

However, it is important to note that *Spondylus* shell proteins (*i.e.* from *Mizuhopecten yessoensis*) which were identified by mass spectrometry analysis and PEAKS bioinformatics software, displayed a low peptidic coverage and their identification was supported only by few peptide sequences. Moreover, the data analysis by PEAKS showed that there was a considerably large number of the *de novo* only peptide sequences, which represented more than 90% of the total information generated by bioinformatic analysis of proteomics data. These sequences are obtained by reconstructing the product ion spectra based on

mathematical calculations and are noted as "*de novo* only", because they were not further matched to proteins contained in the database used for searching (*i.e.* molluscan protein database). In simple terms, the '*de novo* only' peptides can be considered as 'unexplained' data highlighting that a large part of *Spondylus* proteome is different to that of *Pecten* shell and for the moment, remains unknown. After completion of the first study (Chapter 1), the *Spondylus* proteomics dataset was reanalysed using an updated molluscan protein database (downloaded from the NCBI protein repository in early 2020, study presented in Chapter 2). In addition to updated proteins, it also included sequences from another pectinid bivalve *Pecten maximus*, with a full genome sequenced in 2020 as part of Wellcome Sanger 25 Genomes Project (Kenny et al., 2020). The protein make-up was found to be very similar to the results of the first study, but this time the best protein matches were with *Pecten maximus* shell proteins. The two specimens, *M. yessoensis* and *P. maximus* belong to the same family (Pectinidae), but to two different subfamilies, Pedinae and Pectininae, respectively. Based on molecular data of different Pectinida order specimens (which includes both Pectinidae and Spondylidae families), the time of evolutionary divergence of *Spondylus* is closer to *P. maximus* compared to *M. yessoensis* (Smedley et al., 2019). This again suggests that mollusc shell proteins may carry, to some extent, a phylogenetic signal and also evolutionary information (Kocot et al., 2016; McDougall and Degnan, 2018). In the future, this can be further investigated by using more direct comparison, *i.e.* by analysing the full set of 'omics data (proteomics, transcriptomics, genomics) of *Spondylus* and *Pecten* specimens. The data for *Pecten* specimens are available and for *Spondylus*, transcriptomics data will be available in the near future. We are currently directing a parallel project to sequence *Spondylus* mantle transcriptome, funded by Assemble+ program and University of Copenhagen. We have successfully collected live *Spondylus* specimens from the Mediterranean Sea near the island of Crete and the sequencing part is currently ongoing. Hence the obtained transcriptome will allow us to directly compare molecular datasets of different Pectinidae shells and will also be used as a reference 'scaffold' to map *Spondylus* proteomics data obtained in this thesis in order to better characterise *Spondylus* shellome.

A body of biomineralization research suggests that mollusc shell protein sequences had rapid evolution which explains the shell protein diversity and their distinctiveness between different species. As presented in the introduction, shell proteins are characterised by the presence of repetitive low complexity domains (RLCDs) and intrinsically disordered regions (IDRs). Their precise role in biomineralization is not well understood, but it is supposed that RLCDs and IDRs may combine a variety of different functions for matrix assembly and structuring the biomineral framework (Boskey and Villarreal-Ramirez, 2016; Evans, 2019; Marin,

2020). The evolution of RLCDs and IDRs is fast, which may explain the huge diversity of lineage-specific shell proteins among different mollusc types (Kocot et al., 2016). These domains and lineage-specific sequences are of particular interest in shell palaeoproteomics research and may represent one of the best sources for taxonomic information. The biomolecular and proteomics study of *Spondylus* shell (Chapter 1) showed that the majority of identified sequences were characterised by intrinsically disordered structures, some of which represented more than half of total sequence length. Two homologous proteins which were identified in *Spondylus* matrices with top scores, the Uncharacterized protein LOC110461617 from *Mizuhopecten yessoensis* (Chapter 1) and Uncharacterized protein LOC117318053 from *Pecten maximus* (Chapter 2) were almost identical, except for the intrinsically disordered parts, which were almost completely "uncovered". This means that the evolutionary differences lie precisely in the IDR regions (Zarin et al., 2019) and the sequence homology of *Spondylus* and *Pecten* shell proteins are due to conserved protein regions. The fast-evolving and lineage-specific IDR regions may comprise evolutionarily meaningful signatures and, in the future, could be the target when studying phylogenetic relationships of different shell proteins. In the future, this could be achieved by developing novel biomolecular tools specifically adapted to characterise and discriminate low complexity or disordered regions. For instance, this could be done by analysing the specific patterns of amino acids that occur in RLCD domains. These types of profiles obtained for different shell species could help to better discriminate shell proteins and may lead to a better understanding of the evolution of RLCD domains. The sequence homology analysis of selected IDR regions between different shell proteins would also help to better estimate the extent of their evolutionary plasticity and ultimately, may aid to study phylogenetic relationship of mollusc shell proteins. These examples also highlight a clear fact that the advancements in shell palaeoproteomics research greatly depends on the availability of bioinformatic tools for in-depth shell sequence analysis.

However, the LCD/RLCD-type domains and IDR regions are challenging to analyse and characterise mainly because of methodological limitations. First of all, reference molecular data are lacking for many shell species, including *Spondylus*, as seen in Chapter 1, and this hampers our ability to identify the fast-evolving protein regions and fully characterise IDR sequences. Secondly, proteins which possess LCD/RLCD and IDR domains are difficult to characterise with currently available bioinformatic tools. For example, BLASTp, which is typically used to assess protein sequence similarities, is poorly suited for homology search of shell sequences that have long RLCD domains (Marie et al., 2017). This may explain the fact why so many shell proteins are found to be unique and are noted as lineage-specific. Homology search can also be performed by masking the low complexity

regions (Frith, 2011), but in shellomic studies, as discussed above, this would remove an important part of taxonomic information. However, recently published bioinformatic tools that were specifically adapted to shell protein analysis showed good potential in treating different types of RLCDs (Jarnot et al., 2020) and in the future, could help to better investigate the functionality of these domains in biomineralization (Skeffington and Donath, 2020). The combination of bioinformatic studies with experimental analysis using modern microscopy techniques, as for example electron cryo-microscopy (cryo-EM) (Yip et al., 2020) or atomic force microscopy (AFM) (Ando, 2017), will also help to better understand the function of low complexity domains, IDRs regions and how they interact with the mineral crystals.

It is important to note that RLCDs sequences are a *confusing* source of phylogenetic information. LCD peptides with biased amino acid composition are ubiquitous in many shell proteins and, for example, motifs with poly-Ala, poly-Gly, poly-Ser are particularly common among shells with nacreous microstructures (Marin et al., 2016; Marie et al., 2017; Marin, 2020). It is difficult to securely assign LCD peptides to specific proteins. In shell palaeoproteomics, this could mislead shell species identification. This issue was discussed in the palaeoproteomics study of “double-buttons” (Chapter 5) where LCD-type peptides were identified in the archaeological ornament samples and in many other shell proteomes, including freshwater mother-of-pearl mussels, marine mussels (*Mytilus* sp.) and true pearl oysters (*Pinctada* sp.). In this case, the LCD peptides were not considered in further data interpretation. The study highlighted that a critical mind-set is important for shell analysis by palaeoproteomics.

The distinctiveness of shell proteins was once more confirmed in the two shell peptide mass fingerprint (PMF) studies, in which, a number of mollusc shell proteins were characterised and showed unique PMF patterns (Chapter 3, 4). The PMFs of *Spondylus* and *Pecten* intracrystalline proteins obtained by MALDI-TOF mass spectrometry displayed clearly different spectra, despite the fact that *Pecten* proteins were previously identified in *Spondylus* matrix by tandem mass spectrometry (HPLC-MS/MS), as discussed above. Moreover, the analysis of 34 different shell species revealed a great variety of shell PMFs, even for molluscs that belong to the same family. However, protein and peptide differences were less marked for some shells that were of the same order. In particular, we observed this in freshwater mother-of-pearl bivalves (Unionida order) *e.g. Unio pictorum*, *Margaritifera margaritifera*, *Unio crassus*, *Anodonta cygnea* and others. A number of peptide markers shared by all of these shells were found in their PMFs, which suggests that Unionida may have similar or homogeneous shell proteomes or that proteins share similar sequence domains that are ionised and detected by MALDI-

TOF-MS. Comparable findings were observed in the proteomics data of several freshwater mother-of-pearl shells and archaeological “double-button” ornaments obtained by tandem mass spectrometry analysis (HPLC-MS/MS) and presented in Chapter 5. In this study, four Unionida shells that belong to Unionidae and Margaritiferidae families were analysed in order to build a “reference protein collection”. All identified proteomes were characterised by sequences from *Hyriopsis cumingii* specimen, which belongs to order Unionida. These proteomes were very similar and did not allow us to simply discriminate shell proteins at family level. On the one hand, this can be explained by the fact that ‘omics’ data are lacking for more closely related species. On the other hand, it may also indicate that unionoid bivalves form a very homogeneous clade, which representatives share a large number of conserved protein sequences (Marie et al., 2017). However, we were able to obtain a higher taxonomic resolution by carrying out an in-depth sequence analysis of the top-scoring nacre protein Hic74, which was found in all Unionida bivalves. By manually reassessing identified peptides and using sequence alignment tools, we were able to demonstrate that shell proteins of *P. auricularius* (family Margaritiferidae) diverged more, when compared to the other three freshwater nacreous bivalve – *M. margaritifera*, *U. pictorum* and *U. crassus*. The study highlights the potential of shell proteins as carriers of valuable biological information, which can be accessed by using different types of approaches for data analysis.

Shell palaeoproteomics: methodological approaches

In shell palaeoproteomics, the choice of proteomics approach significantly influences the obtained taxonomic resolution. This is well demonstrated in the different studies carried out in this thesis in which two mass spectrometry approaches were investigated:

- high performance liquid chromatography tandem mass spectrometry (HPLC-MS/MS) was used in the two proteomic studies of *Spondylus* shell (Chapter 1, Chapter 2) and the palaeoproteomics study of archaeological “double-buttons” (Chapter 5);
- matrix-assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-TOF-MS) was used to develop the shell peptide mass fingerprinting (PMF) approach (Chapter 2, 3).

The individual advantages and disadvantages of these techniques for shell protein studies are highlighted below.

Shell proteomics by HPLC-MS/MS

Shell protein analysis by HPLC-MS/MS provides a great amount of data and allows to perform a detailed investigation of protein sequences. For example, the use of HPLC-MS/MS allowed an in-depth characterisation of shell protein sequences in the proteomics study of *Spondylus* shell (Chapter 1, 2) and the "palaeoshellomics" study of modern and archaeological shell samples (Chapter 5). It was possible to identify the presence of functional domains, disordered regions (IDRs), repetitive low complexity domains (RLCDs) and also perform sequence similarity, *i.e.* protein homology search. It also allowed us to identify the presence of post-translational modifications (PTMs), as for example the diagenetically-induced deamidation of Gln and Asn residues. The observed patterns of deamidation were particularly useful to investigate protein degradation mechanisms in *Spondylus* shell (Chapter 2). In palaeoproteomic research, deamidated peptides are commonly used as markers to verify the authenticity of ancient sequences (Wilson et al., 2012; Schroeter and Cleland, 2016; Simpson et al., 2016; Hendy et al., 2018). In the shell palaeoproteomics study of archaeological "double-buttons" (Chapter 5) a number of deamidated peptides were identified in the ancient samples, supporting that the extracted proteins were genuine. Moreover, in the same study, the proteomic data obtained by HPLC-MS/MS enabled an in-depth sequence analysis of nacre proteins. It was possible to identify protein regions that were covered with peptides with slightly different amino acid composition, helping to assess the relatedness of these nacre proteins between four studied Unionida shells.

The drawback of proteomic analysis by tandem mass spectrometry is that the success of protein characterisation and thus, shell type identification, is directly dependent on the availability of reference molecular data. Public sequence databases (*e.g*. NCBI) currently host complete molecular data of only around ~30 shell genera (Marin, 2020) and many species, especially those that are not of high commercial importance, as for example *Spondylus,* are underrepresented.

However, the data obtained by tandem mass spectrometry may also be used in order to develop database-independent bioinformatic approaches for characterising shell proteins. For example, raw MS/MS data can be directly used to investigate proteome similarities and differences based on ion fragmentation patterns (Rieder et al., 2017). This method was applied to study the proteomes of reference shells and archaeological "double-buttons" in the shell palaeoproteomics study (Chapter 5), though this method showed some limitations in the case when different mass spectrometry instruments or analytical parameters were used. However, in the future, similar approaches could be applied to study a larger

amount of shell proteomics data, including that which has already been published and so, is publicly available. Database-independent shell protein analysis could also be carried out using the information provided by the *de novo* peptides, *i.e.* sequences that are mathematically constructed from the product ion spectra after bioinformatic analysis and search by PEAKS software. A prototype of this in-house algorithm was applied in the shell palaeoproteomics study of archaeological "double-buttons" and several reference shells (Chapter 5). The obtained *de novo* peptidomes of each sample were searched against characteristic motifs and grouped according to the recurring patterns. The obtained results, combined together with database search, helped to identify the material of the archaeological ornaments, supporting that they were made of freshwater mother-of-pearl. In the future, the use of database-independent methods, as for example development of peptidome libraries, could provide a way to overcome challenges in shell protein studies. In the late 2000s, there were attempts to create a public peptidome database (Slotta et al., 2009) which was unfortunately discontinued due to lack of funding. However, a database restricted to shell proteins may be more easily set up, used and curated. Shell peptidome libraries could be easily created *in silico*, *i.e.* just by reanalysing already existing shell proteomics data, or, in publications, provided as peptidome datasets, similarly as for biomolecular *Spondylus* shell study (Chapter 1). In general, the *de novo* only peptide datasets are rarely exploited in palaeoproteomic studies, but, in my opinion, in the future, they will provide an important source to investigate rare shell models, extinct species as well as archaeological shell artifacts.

Shell PMF by MALDI-TOF-MS

Considering the discussion above, one of the main advantages of peptide mass fingerprinting (PMF) analysis by MALDI-TOF mass spectrometry is that the characterisation of shell proteins (peptides) can be easily completed in database-independent mode and can be easily enlarged to include diversity of species. The shell PMF library created as part of this PhD project and presented in Chapter 4 hosts the proteomics data of a variety of molluscan shells and it was built in order to represent species that are important in archaeological research, as for example *Unio pictorum, Margaritifera margaritifera*, *Spondylus gaederopus*, *Glycymeris glycymeris*, *Pinna nobilis*, *Pecten maximus*. Moreover, the shell PMF dataset also included mollusc species which have not been previously characterised by proteomic approaches, as for example *Phorcus turbinatus*, *Patella vulgata*, *Littorina littorea.* Hence, the shell PMF method provides a simple, cost- and time-effective tool to create an in-house shell protein library without the need of genomics/transcriptomics data. It can also be easily adapted to archaeological projects, by including species that are relevant in specific archaeological

sites/contexts. For instance, considering that 34 different reference shells and 16 archaeological samples were analysed using MALDI-TOF-MS in the two shell PMF studies (Chapter 3 and 4), if tandem mass spectrometry (HPLC-MS/MS) was used, the cost of proteomic analyses would have been about 3 orders of magnitude higher (also considering the analysis of replicates). In palaeoproteomics, the shell PMF studies detailed in Chapters 3 and 4 are significant because they represent the first attempt to characterise intracrystalline mollusc shell protein PMFs. Together with the PMF analysis of avian eggshells (Stewart et al., 2013; Presslee et al., 2017), they are truly the first of that type for $CaCO_3$ biominerals, in contrast to the many proteomic studies for collagen/calcium phosphate based (mainly bone) samples.

The drawback of PMF by MALDI-TOF-MS approach, as noted in the discussion above, is that the level of information compared to shotgun proteomics (HPLC-MS/MS) is considerably lower. The peptide mass fingerprints do not provide sequence information neither of the identified peptides, nor of post-translational modification (PTMs), as well as not guaranteeing a secure characterisation of shell markers. It is also more complicated to assess sequences that are shared by closely related species. In the future, the shell PMF data of reference shells obtained in this study could be reanalysed in tandem mass spectrometry mode (*e.g.* using MALDI-TOF/TOF system) (Gogichaeva et al., 2007). The obtained fragmentation patterns of marker peptides would enable the identification of homologous peptides, which is what happens with collagen-based samples (Buckley et al., 2014). The approach could also be used for *de novo* sequencing in order to characterise the amino acid sequences of marker peptides. In the future, the expansion of mollusc shell PMF dataset would help to better understand the diversity of shell proteins among different species and may provide a new way to perform comparative analyses to study shell protein relatedness.

Intracrystalline shell proteins

The research presented in this thesis highlighted some interesting features of intracrystalline mollusc shell proteins. First of all, intracrystalline shell proteins are isolated using a strong bleaching step and, in shell palaeoproteomics, this provides an important advantage by eliminating the impact of environmental contamination, a recurrent issue in ancient protein studies of other types of archaeological substrates (Hendy et al., 2018). Different studies presented in this thesis demonstrated that the intracrystalline proteins isolated from a variety of mollusc shells, can be characterised by both MALDI-TOF-MS and HPLC-MS/MS even when small-size samples are used (20-35 mg) (Chapters 2-4). It is worth to note that while the intracrystalline proteins of different mollusc shells have been characterised by their amino acid composition (as for example in Demarchi et al.,

2013b; 2014; Pierini et al., 2016; Ortiz et al., 2018) and used for AAR dating of sub-fossil shells (Penkman et al., 2008; 2011; Demarchi et al., 2011), their sequences have never been investigated by proteomics. In mollusc shells, the protein content is usually very low (as presented in the literature review, it can be around 0.001-0.01% of total shell weight) and the majority of proteomic studies of shells typically use weak bleaching treatments to isolate a semi-intracrystalline fraction so as to retrieve more matrix proteins (Marie et al., 2013). However, the studies carried out in this thesis showed that, contrary to initial expectations, the more intensive is the bleaching used, the better is the quality of the proteomics data obtained. Bleaching significantly reduces the fraction of organics that is being analysed and this "matrix simplification" may result in an easier identification of peptides and proteins. This effect was observed in the methodological shell PMF study presented in Chapter 3. Comparing three different bleaching approaches, I found that the quality of PMF spectra is significantly better for the intracrystalline protein fraction. Similarly, in the proteomics study of *Spondylus gaederopus* shell by HPLC-MS/MS, I also observed that protein coverages and number of identified peptides were higher in samples that were more thoroughly bleached, compared to just mild oxidation treatment (Chapter 1). In addition, immunochemistry analyses that were used to characterise intracrystalline shell matrix fraction (Chapter 2) showed that a strong bleaching treatment reduced the signal of binding carbohydrate groups, but increased the signal of reactive protein epitopes. Prolonged oxidative treatments may also help to partially degrade some of the crossed-linked intercrystalline shell matrix structures, therefore helping to better solubilise matrix proteins.

Intracrystalline proteins are typically characterised by sequences of acidic amino acids that may play a role in mineral binding activity (Albeck et al., 1993; 1996b; Marin et al., 2005; Demarchi et al., 2016). Some studies suggest that intracrystalline proteins may differ significantly from the sequences inside the intercrystalline matrix, based on considerably different amino acid compositions found in the two fractions (Penkman et al., 2008; Demarchi et al., 2013b; Pierini et al., 2016). However, in *Spondylus*, the opposite is true. The proteomic studies of *Spondylus* (Chapter 1 and 2) showed that the intracrystalline shell sequences were very similar to its (inter+intra)crystalline fraction. In fact, a similar finding was also documented for another crossed-lamellar shell, *Glycymeris glycymeris*, based on its amino acid composition (Demarchi et al., 2015). The shell organic matrix, when extracted with an organic acid, typically exists in two fractions – the acid soluble (ASM) and acid insoluble (AIM) one. Despite the considerably different solubility, the protein set in both ASM and AIM fractions, was found to be very similar in *Spondylus* (Chapter 1), and the same effect has been observed in other shell models (Marie et al., 2009; Marin et al., 2013). These observations indicate that the solubility of the matrix (AIM *vs* ASM) or the fraction type, *i.e.* inter- vs intracrystalline is not necessarily due to

different protein sets, but may be because of protein cross-linking with other shell macromolecules, such as carbohydrates.

Moreover, the top scoring *Spondylus* protein from the intracrystalline fraction was actually basic. Proteins with high pI value have also been identified in other shell models, as for example nacro-prismatic freshwater mother-of-pearl shells (Ramos-Silva et al., 2012; Liu et al., 2017; Jin et al., 2020), however the way these proteins interact with the mineral crystals and their function in shell biomineralization is poorly investigated. The protein-mineral interaction may occur via different protein regions or are likely mediated by additional functional groups from post-translational modifications (PTMs), such as phosphorylation (Borbas et al., 1991; Hecker et al., 2003) and glycosylation (Albeck et al., 1996a; Lang et al., 2020). These, however, have not been observed in *Spondylus* matrices. In the future, the identification of glycosylation and phosphorylation PTMs could also be carried out *in silico* by using adapted data search algorithms (Mann and Edsinger, 2014) and could help us to better understand their possible role in mineral binding activity.

Protein diagenesis

Shell protein diagenesis is a complex process and the degradation pathways are, as yet, poorly understood. However, protein diagenesis can be modelled and studied using accelerated aging experiments, as presented for *Spondylus* (Chapter 2). In this study immunochemical assays and quantitative proteomics were combined to simultaneously track the loss of molecular structure and peptide bond hydrolysis. As expected, the results exposed complex patterns of shell protein diagenesis. The degradation of secondary matrix structure, as investigated by immunochemistry, could be explained by kinetic models and it was found to follow an apparent first order reaction. The observed kinetic rates were found to be similar to the rates of hydrolysis as investigated for a variety of amino acids from other shell systems (Demarchi et al., 2013a; Pierini et al., 2016), implying that structural loss could overlap with peptide bond hydrolysis. Interestingly but not unexpectedly, the proteomics data obtained here showed that the stability of peptide bonds could not be explained simply by using theoretical “models” and experimental values which are based on the behaviour of peptides in solution (Hill, 1965; White, 1984; Demarchi et al., 2013c). The study highlighted that protein structural features, mineral-binding properties and the whole microchemical environment of the shell biomineral system significantly impacts protein stability and the pathways of protein degradation (Marin et al., 2007; Parker et al., 2015; Demarchi et al., 2016; Demarchi, 2020). In the future, shell protein degradation will be further investigated by proteomics, analysing all of the samples sets that were artificially aged as part of this study (heated at different temperatures). This may provide new data on

protein breakdown patterns and will hopefully allow us to directly retrieve kinetic information on shell protein hydrolysis for *Spondylus*. Moreover, as part of an ongoing project, the samples will also be analysed for amino acid racemisation and, together with proteomics and immunochemistry data, will provide the first extensive dataset to characterise shell protein diagenesis.

Another intriguing finding was that, during the initial phase of accelerated aging (heating the samples at high temperatures), proteins do not undergo instant degradation, but in contrast, appear to be more abundant, *i.e.* are more easily detected. In simple terms, I observed that mild degradation leads to higher protein recovery. Some of the shell proteins may be crossed-linked with other macromolecules, possibly carbohydrates, or may exist as glycoproteins (Albeck et al., 1996b; Agbaje et al., 2019; Marin, 2020). Hence, heating could help to uncoil/denature some of these structures, making them more soluble. Quantitative proteomics showed that in samples that were heated for short periods, the relative peptide abundances was considerably higher compared to untreated shells. Immunochemical assays also showed that carbohydrate structures were stable in the first phase of aging and short-term heating also increased their relative cross-reactivity (*i.e.* detectability). Protein glycosylation also improves its molecular stability (Hanson et al., 2009; Solá and Griebenow, 2009) which could also explain the stability of shell proteins and carbohydrate groups when heated at high temperatures.

This interesting observation was already noted in other sub-fossil shell samples but, so far, has never been thoroughly investigated. For example, in the palaeoshellomics study of archaeological “double-buttons” (Chapter 5), we found that the 6000-year-old sub-fossil reference shells and archaeological samples, when analysed by tandem mass spectrometry and compared to modern shells, showed higher protein coverages and number of supporting peptides. Similar findings were also observed in the proteomics study of sub-fossil *Tridacna* shells from Polynesia (F. Marin, unpublished data) (Marin, 2020) and in sub-fossil as well as heated *Pecten* shells that were characterised by their amino acid composition (Pierini et al., 2016).

Interestingly, a reverse outcome was observed in the peptide mass fingerprinting (PMF) analysis of archaeological shell samples by MALDI-TOF mass spectrometry (Chapter 4). The majority of modern shells yielded high quality PMFs, whereas the archaeological shell samples resulted in poor-quality spectra. On the one hand, this may be explained by low abundance of intracrystalline proteins (considering that ~20 mg samples were used) or that the preservation state was poor. However, the archaeological shell samples, more precisely the aragonitic shells, did not show

any clues of recrystallization to calcite, a transformation that we interpret as a potentially poor preservation state of the sample (Loftus et al., 2015). Moreover, two sets of archaeological “double-buttons” which were analysed by tandem mass spectrometry in Chapter 5 were of similar age but showed good protein preservation, that even surpassed modern shells. It is important to note that the “double-buttons” were recovered from sites with considerably different annual mean temperatures and soil chemistries (*i.e.*, site conditions from the North of Denmark cannot be simply compared to the conditions in the South of France), however, in the shell PMF study, proteins were also not detected from archaeological samples that were recovered from archaeological sites in the North of France. Therefore, we cannot exclude that the poor quality PMFs may also be due to some degradation products, which may negatively impact protein extraction or their detection by MALDI-TOF-MS. It is possible that the studied archaeological samples are just partially degraded but the protein extracts may be dominated by glycoproteins, glycosylated peptides or longer carbohydrate units, which, in modern shells may not have been solubilised at all. Glycosides are highly hydrophilic in nature and due to their physico chemical properties they do not ionise well. This means that the MALDI-TOF-MS analysis of peptide mixture with glycoside moieties or carbohydrates could result in poor quality spectra (Shinohara et al., 2004). In the future, protein preservation state could be assessed by amino acid racemisation analysis. Moreover, shell proteins could be identified by high resolution tandem mass spectrometry (HPLC-MS/MS) which, in previous studies, showed good ability to characterise archaeological and degraded shell samples.

Overall, our findings highlighted the biological complexity of mollusc shell systems, which require a holistic and multidisciplinary approach for an in-depth investigation. Nevertheless, the studies presented in this PhD thesis also demonstrate that palaeoproteomics has an excellent potential for the study of ancient shell artifacts.

References

Agbaje, O.B.A., Thomas, D.E., Dominguez, J.G., Mclnerney, B.V., Kosnik, M.A., Jacob, D.E., 2019. Biomacromolecules in bivalve shells with crossed lamellar architecture. Journal of Materials Science 54, 4952–4969.

Albeck, S., Aizenberg, J., Addadi, L., Weiner, S., 1993. Interactions of various skeletal intracrystalline components with calcite crystals. Journal of the American Chemical Society 115, 11691–11697.

Albeck, S., Weiner, S., Addadi, L., 1996a. Polysaccharides of Intracrystalline Glycoproteins Modulate Calcite Crystal Growth in Vitro. Chemistry A European Journal, ACS Symp. Ser 2, 278–284.

Albeck, S., Addadi, I., Weiner, S., 1996b. Regulation of calcite crystal morphology by intracrystalline acidic proteins and glycoproteins. Connective tissue research 35, 365–370.

Ando, T., 2017. Directly watching biomolecules in action by high-speed atomic force microscopy. Biophysical reviews 9, 421–429.

Arivalagan, J., Marie, B., Sleight, V.A., Clark, M.S., Berland, S., Marie, A., 2016. Shell matrix proteins of the clam, *Mya truncata*: Roles beyond shell formation through proteomic study. Marine genomics 27, 69–74.

Borbas, J.E., Wheeler, A.P., Sikes, C.S., 1991. Molluscan shell matrix phosphoproteins: Correlation of degree of phosphorylation to shell mineral microstructure and to in vitro regulation of mineralization. The Journal of experimental zoology 258, 1–13.

Boskey, A.L., Villarreal-Ramirez, E., 2016. Intrinsically disordered proteins and biomineralization. Matrix biology: journal of the International Society for Matrix Biology 52-54, 43–59.

Buckley, M., Fraser, S., Herman, J., Melton, N.D., Mulville, J., Pálsdóttir, A.H., 2014. Species identification of archaeological marine mammals using collagen fingerprinting. Journal of archaeological science 41, 631–641.

Carter, J.G. (Ed.), 1990. Skeletal Biomineralization: Pattern, Processes, and Evolutionary Trends. Amer Geophysical Union.

Cleland, T.P., Schroeter, E.R., 2018. A Comparison of Common Mass Spectrometry Approaches for Paleoproteomics. Journal of proteome research 17, 936–945.

Demarchi, B., 2020. Amino Acids and Proteins in Fossil Biominerals: An Introduction for Archaeologists and Palaeontologists. John Wiley & Sons.

Demarchi, B., Williams, M.G., Milner, N., Russell, N., Bailey, G., Penkman, K., 2011. Amino acid racemization dating of marine shells: A mound of possibilities. Quaternary international: the journal of the International Union for Quaternary Research 239, 114–124.

Demarchi, B., Collins, M.J., Tomiak, P.J., Davies, B.J., Penkman, K.E.H., 2013a. Intra-crystalline protein diagenesis (IcPD) in *Patella vulgata*. Part II: Breakdown and temperature sensitivity. Quaternary geochronology 16, 158–172.

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

Demarchi, B., Collins, M., Bergström, E., Dowle, A., Penkman, K., Thomas-Oates, J., Wilson, J., 2013c. New experimental evidence for in-chain amino acid racemization of serine in a model peptide. Analytical chemistry 85, 5835–5842.

Demarchi, B., O'Connor, S., Lima Ponzoni, A. de, Almeida Rocha Ponzoni, R. de, Sheridan, A., Penkman, K., Hancock, Y., Wilson, J., 2014. An integrated approach to the taxonomic identification of prehistoric shell ornaments. PloS one 9, e99839.

Demarchi, B., Clements, E., Coltorti, M., Locht, R. van de, Kröger, R., Penkman, K., Rose, J., 2015. Testing the effect of bleaching on the bivalve *Glycymeris*: A case study of amino acid geochronology on key Mediterranean raised beach deposits. Quaternary geochronology 25, 49–65.

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

Evans, J.S., 2019. The Biomineralization Proteome: Protein Complexity for a Complex Bioceramic Assembly Process. Proteomics 19, e1900036.

Frith, M.C., 2011. Gentle masking of low-complexity sequences improves homology search. PloS one 6, e28819.

Gogichaeva, N.V., Williams, T., Alterman, M.A., 2007. MALDI TOF/TOF tandem mass spectrometry as a new tool for amino acid analysis. Journal of the American Society for Mass Spectrometry 18, 279–284.

Hanson, S.R., Culyba, E.K., Hsu, T.-L., Wong, C.-H., Kelly, J.W., Powers, E.T., 2009. The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. Proceedings of the National Academy of Sciences of the United States of America 106, 3131–3136.

Hecker, A., Testenière, O., Marin, F., Luquet, G., 2003. Phosphorylation of serine residues is fundamental for the calcium-binding ability of Orchestin, a soluble

matrix protein from crustacean calcium storage structures. FEBS letters 535, 49–54.
Hendy, J., Welker, F., Demarchi, B., Speller, C., Warinner, C., Collins, M.J., 2018. A guide to ancient protein studies. Nature ecology & evolution 2, 791–799.
Herlitze, I., Marie, B., Marin, F., Jackson, D.J., 2018. Molecular modularity and asymmetry of the molluscan mantle revealed by a gene expression atlas. GigaScience 7. doi:10.1093/gigascience/giy056
Hill, R.L., 1965. Hydrolysis of proteins. Advances in protein chemistry 20, 37–107.
Jackson, D.J., McDougall, C., Woodcroft, B., Moase, P., Rose, R.A., Kube, M., Reinhardt, R., Rokhsar, D.S., Montagnani, C., Joubert, C., Piquemal, D., Degnan, B.M., 2010. Parallel evolution of nacre building gene sets in molluscs. Molecular biology and evolution 27, 591–608.
Jarnot, P., Ziemska-Legiecka, J., Dobson, L., Merski, M., Mier, P., Andrade-Navarro, M.A., Hancock, J.M., Dosztányi, Z., Paladin, L., Necci, M., Piovesan, D., Tosatto, S.C.E., Promponas, V.J., Grynberg, M., Gruca, A., 2020. PlaToLoCo: the first web meta-server for visualization and annotation of low complexity regions in proteins. Nucleic acids research 48, W77–W84.
Jin, C., Li, J.-L., Liu, X.-J., 2020. Teosin, a novel basic shell matrix protein from *Hyriopsis cumingii* induces calcium carbonate polycrystal formation. International journal of biological macromolecules 150, 1229–1237.
Kenny, N.J., McCarthy, S.A., Dudchenko, O., James, K., Betteridge, E., Corton, C., Dolucan, J., Mead, D., Oliver, K., Omer, A.D., Pelan, S., Ryan, Y., Sims, Y., Skelton, J., Smith, M., Torrance, J., Weisz, D., Wipat, A., Aiden, E.L., Howe, K., Williams, S.T., 2020. The gene-rich genome of the scallop *Pecten maximus*. GigaScience 9. doi:10.1093/gigascience/giaa037
Kocot, K.M., Aguilera, F., McDougall, C., Jackson, D.J., Degnan, B.M., 2016. Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. Frontiers in zoology 13, 23.
Lang, A., Mijowska, S., Polishchuk, I., Fermani, S., Falini, G., Katsman, A., Marin, F., Pokroy, B., 2020. Acidic monosaccharides become incorporated into calcite single crystals. Chemistry A European Journal 2020.08.03.234310.
Liu, X., Pu, J., Zeng, S., Jin, C., Dong, S., Li, J., 2017. *Hyriopsis cumingii* Hic52-A novel nacreous layer matrix protein with a collagen-like structure. International journal of biological macromolecules 102, 667–673.
Loftus, E., Rogers, K., Lee-Thorp, J., 2015. A simple method to establish calcite:aragonite ratios in archaeological mollusc shells. Journal of Quaternary Science 30, 731–735.
Mann, K., Edsinger, E., 2014. The *Lottia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. Proteome science 12, 28.
Mao, J., Zhang, W., Zhang, X., Tian, Y., Wang, X., Hao, Z., Chang, Y., 2018.

Transcriptional changes in the Japanese scallop (*Mizuhopecten yessoensis*) shellinfested by Polydora provide insights into the molecular mechanism of shell formation and immunomodulation. Scientific reports 8, 17664.

Marie, B., Marin, F., Marie, A., Bédouet, L., Dubost, L., Alcaraz, G., Milet, C., Luquet, G., 2009. Evolution of nacre: biochemistry and proteomics of the shell organic matrix of the cephalopod *Nautilus macromphalus*. Chembiochem: a European journal of chemical biology 10, 1495–1506.

Marie, B., Zanella-Cléon, I., Guichard, N., Becchi, M., Marin, F., 2011a. Novel proteins from the calcifying shell matrix of the Pacific oyster *Crassostrea gigas*. Marine biotechnology 13, 1159–1168.

Marie, B., Trinkler, N., Zanella-Cleon, I., Guichard, N., Becchi, M., Paillard, C., Marin, F., 2011b. Proteomic identification of novel proteins from the calcifying shell matrix of the Manila clam *Venerupis philippinarum*. Marine biotechnology 13, 955–962.

Marie, B., Ramos-Silva, P., Marin, F., Marie, A., 2013. Proteomics of CaCO3 biomineral-associated proteins: how to properly address their analysis. Proteomics 13, 3109–3116.

Marie, B., Arivalagan, J., Mathéron, L., Bolbach, G., Berland, S., Marie, A., Marin, F., 2017. Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the *Unionoida Elliptio complanata* and *Villosa lienosa*. Journal of the Royal Society, Interface / the Royal Society 14. doi:10.1098/rsif.2016.0846

Marin, F., 2020. Mollusc shellomes: Past, present and future. Journal of structural biology 212, 107583.

Marin, F., Amons, R., Guichard, N., Stigter, M., Hecker, A., Luquet, G., Layrolle, P., Alcaraz, G., Riondet, C., Westbroek, P., 2005. Caspartin and calprismin, two proteins of the shell calcitic prisms of the Mediterranean fan mussel *Pinna nobilis*. The Journal of biological chemistry 280, 33895–33908.

Marin, F., Morin, V., Knap, F., Guichard, N., 2007. Caspartin: Thermal stability and occurrence in mollusk calcified tissues. In: Arias, J.L.A., Fernández, M.S. (Eds.), Biomineralization: From Paleontology to Materials Science. Presented at the 9th International Symposium on Biomineralization, researchgate.net, pp. 281–288.

Marin, F., Marie, B., Hamada, S.B., Ramos-Silva, P., Le Roy, N., Guichard, N., Wolf, S.E., Montagnani, C., Joubert, C., Piquemal, D., Saulnier, D., Gueguen, Y., 2013. “Shellome”: Proteins involved in mollusk shell biomineralization-diversity, functions. In: Watabe, S., Maeyama, K., Nagasawa, H. (Eds.), International Symposium on Pearl Research. Terrapub Tokyo, pp. 149–166.

Marin, F., Bundeleva, I., Takeuchi, T., Immel, F., Medakovic, D., 2016. Organic matrices in metazoan calcium carbonate skeletons: Composition, functions, evolution. Journal of structural biology 196, 98–106.

McDougall, C., Degnan, B.M., 2018. The evolution of mollusc shells. Wiley interdisciplinary reviews. Developmental biology 7, e313.

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

Parker, A., Immel, F., Guichard, N., Broussard, C., Marin, F., 2015. Thermal stability of nacre proteins of the Polynesian pearl oyster: a proteomic study. In: Marin, F., Brümmer, F., Checa, A., Furtos, G., Lesci, I.G., Šiller, L. (Eds.), Key Engineering Materials. Trans Tech Publ, pp. 222–231.

Penkman, K.E.H., Kaufman, D.S., Maddy, D., Collins, M.J., 2008. Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells. Quaternary geochronology 3, 2–25.

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

Pierini, F., Demarchi, B., Turner, J., Penkman, K., 2016. *Pecten* as a new substrate for IcPD dating: The quaternary raised beaches in the Gulf of Corinth, Greece. Quaternary geochronology 31, 40–52.

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

Ramos-Silva, P., Benhamada, S., Le Roy, N., Marie, B., Guichard, N., Zanella-Cléon, I., Plasseraud, L., Corneillat, M., Alcaraz, G., Kaandorp, J., Marin, F., 2012. Novel molluskan biomineralization proteins retrieved from proteomics: a case study with Upsalin. Chembiochem: a European journal of chemical biology 13, 1067–1078.

Rieder, V., Blank-Landeshammer, B., Stuhr, M., Schell, T., Biß, K., Kollipara, L., Meyer, A., Pfenninger, M., Westphal, H., Sickmann, A., Rahnenführer, J., 2017. DISMS2: A flexible algorithm for direct proteome- wide distance calculation of LC-MS/MS runs. BMC bioinformatics 18, 148.

Schroeter, E.R., Cleland, T.P., 2016. Glutamine deamidation: an indicator of antiquity, or preservational quality? Rapid communications in mass spectrometry: RCM 30, 251–255.

Shinohara, Y., Furukawa, J.-I., Niikura, K., Miura, N., Nishimura, S.-I., 2004. Direct N-glycan profiling in the presence of tryptic peptides on MALDI-TOF by controlled ion enhancement and suppression upon glycan-selective derivatization. Analytical chemistry 76, 6989–6997.

Simpson, J.P., Penkman, K.E.H., Demarchi, B., Koon, H., Collins, M.J., Thomas-Oates, J., Shapiro, B., Stark, M., Wilson, J., 2016. The effects of

demineralisation and sampling point variability on the measurement of glutamine deamidation in type I collagen extracted from bone. Journal of archaeological science 69, 29–38.

Skeffington, A.W., Donath, A., 2020. ProminTools: Shedding light on proteins of unknown function in biomineralization with user friendly tools illustrated using mollusc shell matrix protein sequences. Cold Spring Harbor Laboratory. doi:10.1101/2020.03.05.978718

Slotta, D.J., Barrett, T., Edgar, R., 2009. NCBI Peptidome: a new public repository for mass spectrometry peptide identifications. Nature biotechnology 27, 600–601.

Smedley, G.D., Audino, J.A., Grula, C., Porath-Krause, A., Pairett, A.N., Alejandrino, A., Lacey, L., Masters, F., Duncan, P.F., Strong, E.E., Serb, J.M., 2019. Molecular phylogeny of the Pectinoidea (Bivalvia) indicates Propeamussiidae to be a non-monophyletic family with one clade sister to the scallops (Pectinidae). Molecular phylogenetics and evolution 137, 293–299.

Solá, R.J., Griebenow, K., 2009. Effects of glycosylation on the stability of protein pharmaceuticals. Journal of pharmaceutical sciences 98, 1223–1245.

Stewart, J.R.M., Allen, R.B., Jones, A.K.G., Penkman, K.E.H., Collins, M.J., 2013. ZooMS: making eggshell visible in the archaeological record. Journal of archaeological science 40, 1797–1804.

Waller, T.R., 2006. Phylogeny of families in the Pectinoidea (Mollusca: Bivalvia): importance of the fossil record. Zoological journal of the Linnean Society 148, 313–342.

Wang, S., Zhang, J., Jiao, W., Li, J., Xun, X., Sun, Y., Guo, X., Huan, P., Dong, B., Zhang, L., Hu, X., Sun, X., Wang, J., Zhao, C., Wang, Y., Wang, D., Huang, X., Wang, R., Lv, J., Li, Y., Zhang, Z., Liu, B., Lu, W., Hui, Y., Liang, J., Zhou, Z., Hou, R., Li, X., Liu, Y., Li, H., Ning, X., Lin, Y., Zhao, L., Xing, Q., Dou, J., Li, Y., Mao, J., Guo, H., Dou, H., Li, T., Mu, C., Jiang, W., Fu, Q., Fu, X., Miao, Y., Liu, J., Yu, Q., Li, R., Liao, H., Li, X., Kong, Y., Jiang, Z., Chourrout, D., Li, R., Bao, Z., 2017. Scallop genome provides insights into evolution of bilaterian karyotype and development. Nature ecology & evolution 1, 120.

White, R.H., 1984. Hydrolytic stability of biomolecules at high temperatures and its implication for life at 250 degrees C. Nature 310, 430–432.

Wilson, J., Doorn, N.L. van, Collins, M.J., 2012. Assessing the extent of bone degradation using glutamine deamidation in collagen. Analytical chemistry 84, 9041–9048.

Yarra, T., Gharbi, K., Blaxter, M., Peck, L.S., Clark, M.S., 2016. Characterization of the mantle transcriptome in bivalves: *Pecten maximus*, *Mytilus edulis* and *Crassostrea gigas*. Marine genomics 27, 9–15.

Yip, K.M., Fischer, N., Paknia, E., Chari, A., Stark, H., 2020. Atomic-resolution protein structure determination by cryo-EM. Nature 587, 157–161.

Zarin, T., Strome, B., Nguyen Ba, A.N., Alberti, S., Forman-Kay, J.D., Moses, A.M., 2019. Proteome-wide signatures of function in highly diverged intrinsically disordered regions. eLife 8. doi:10.7554/eLife.46883

Conclusions

My doctoral work lies at the interface of biomineralization and archaeological sciences, two disciplines that have rarely interacted so far. The three-years research presented in this thesis consisted of extensive experimental work in order to develop proteomic-based approaches to study archaeological mollusc shell artifacts.

The outcomes of my research highlight that:

1. Mollusc shell proteins are complex biological molecules, which however carry taxonomic information usable for clade identification.
2. *Spondylus gaederopus* shell represents an adequate system for protein preservation. However, the diagenesis of intracrystalline proteins involves complex chemical processes that we are just beginning to understand using proteomics.
3. The analysis of intracrystalline shell proteins by MALDI-TOF-MS results in characteristic PMFs which could be used as molecular barcodes for taxonomic discrimination. Additional future optimization is required for the application to archaeological shell samples.
4. Palaeoproteomic analysis of archaeological “double-buttons” by tandem mass spectrometry (HPLC-MS/MS) enabled us to identify mother-of-pearl from freshwater unionoid bivalves as the source of raw material for making these ornaments.

The comprehensive biochemical, structural and proteomics-based investigation of the chosen model taxon, *Spondylus gaederopus,* demonstrated that shell proteins entrapped in mineral crystals preserve biological information (Chapter 1). The data are among the few which have been obtained from shells with complex crossed-lamellar microstructures and therefore provide an important baseline to better understand the biochemical characteristics of this microstructural type, as well as the evolution of shell-associated proteins. I have determined that *Spondylus* shell proteome represents lineage-specific sequences which potentially encode taxonomic information. Furthermore, I have used MALDI-TOF mass spectrometry to characterise intracrystalline shell proteins of different mollusc taxa, including *Spondylus* (Chapters 3 and 4). These results also support the hypothesis that shell proteins can be used as molecular barcodes for taxonomic identification. These datasets therefore provide the first solid basis for the scientific identification of prehistoric shell artifacts.

Finding *Spondylus* proteins in ancient shells should be possible. I have demonstrated that it represents a suitable system for protein preservation (Chapter

2). This was shown by artificial aging experiments where I studied the degradation, *i.e.*, diagenesis, of intracrystalline shell proteins using immunochemistry and, for the first time, quantitative proteomics, which allowed me to simultaneously track both the structural loss of the organic matrix and the hydrolysis of peptide bonds. The study showed that shell proteins follow complex degradation patterns, which are dependent on the overall microchemical environment. The study also highlighted that early diagenesis, even at high temperature, aids protein solubilisation and denaturation by uncoiling cross-linked matrix structures. This feature has been empirically observed in other sub-fossil shells and archaeological "double-buttons" (Chapter 5).

In this thesis, I present a novel tool to characterise and discriminate mollusc shells based on their peptide mass fingerprints (PMFs) (Chapters 3 and 4). First of all, an extraction protocol was optimised to isolate and analyse intracrystalline shell proteins by MALDI-TOF-MS (Chapter 3). Secondly, I created an in-house reference library of shell PMFs by analysing intracrystalline proteins from 34 different molluscan taxa. The method was also used to characterise a set of archaeological samples from Neolithic sites in France but the results showed that future optimization is needed. In particular, it will be necessary to overcome limitations that are either due to characteristics of archaeological samples (degradation) or methodology (*e.g.* sensitivity of the technique). Nevertheless, the study represents the first attempt to discriminate shell proteins based on their PMFs and provides a significant advancement in palaeoproteomics research.

The potential of shell palaeoproteomics in archaeological applications was fully revealed by the study of ~6000 years old archaeological "double-buttons", which were recovered from three archaeological sites in Denmark, Germany and Romania (all dated between 4200–3800 BCE). The archaeological samples were investigated using structural, biomolecular and geochemical analyses coupled with high resolution HPLC-MS/MS, which enabled to unequivocally identify that these artifacts were made of mother-of-pearl shells from freshwater unionoid bivalves (Chapter 5). Contrary to what was initially thought, shells were likely sourced locally and not subject to long distance exchanges, providing the answer to a years-long archaeological debate over their provenance. Overall, this study marked the first application of palaeoproteomics to mollusc shells, and more generally, to any invertebrate system.

Limitations and future perspectives

At the same time, this PhD project has also exposed current challenges and open questions in the field of shell palaeoproteomics. These are mainly linked to limited

molecular reference data for molluscs and the complexity of shell matrices. However, future analyses and development of novel approaches may provide avenues to overcome current limitations. More specifically:

- *Spondylus* proteins identified in this thesis, in reality, represent only a partial picture of the full shell proteome which is due to the absence of reference transcriptome/genome (Chapter 1).

In the future, this will be addressed by additional molecular analyses. I am currently leading a project to sequence *Spondylus* mantle transcriptome which will provide a so-called "reference scaffold" to map proteomics data of this shell and characterise its full proteome. This, in turn, will enable us to study sequence characteristics associated with crossed-lamellar shells and investigate the evolution of shell proteins. Moreover, outside proteomics, future analyses will be required to characterise the full biomolecular composition of the shell, in particular, the sugar moieties, that are important constituents involved in biomineralization, as well in the diagenesis of shell organic matrix.

- Due to the same reasons (*i.e.* 'partial picture of the matrix'), protein hydrolysis patterns, identified in the experimental diagenetic study of *Spondylus* shell matrix, represent only a fraction of all occurring reactions (Chapter 2).

In the future, as proteome data for *Spondylus* becomes available, it will be possible to re-investigate proteomics data of artificially aged samples and identify additional degradation patterns. Moreover, I will aim to analyse amino acid racemisation reactions and further investigate peptide bond hydrolysis patterns that occur at different temperatures.

- In the palaeoproteomics study of archaeological "double-buttons", the partial knowledge on shell proteomes hampered the achievement of a more accurate identification, *i.e.,* at family or genus levels (Chapter 5).

In the future, this obstacle could be bypassed by developing database-independent bioinformatic algorithms for data analysis, for example by creating peptidome libraries for mollusc genera and species that are of interest in archaeological research.

- Finally, while the shell PMF method was successfully applied to the characterisation of the intracrystalline proteins from a variety of modern shells, it did not work properly for the selected set of archaeological samples (Chapter 4). This issue may be due to methodological approach, sample characteristics, low protein content (*e.g.* if samples were too degraded).

To overcome these issues, amino acid racemisation analysis can be used to verify protein preservation state. I will also aim to re-analyse the samples by high resolution tandem mass spectrometry (HPLC-MS/MS) which could identify sequences, even partially degraded ones. Considering future advancements of shell PMF method, the obtained fingerprints can be re-examined by tandem mass

spectrometry (MALDI-TOF/TOF) to identify (*de novo*) the peptide sequences and assess the presence of shared markers. The library discussed in this thesis should also be further extended by including more species of interest for archaeology and palaeontology as well as more biological replicates.

To conclude, the research carried out as part of this doctoral thesis provides a solid foundation for advancing shell biomineralization studies and palaeoproteomics, an emerging discipline. It demonstrates that proteomic-based approaches provide an undoubtedly powerful tool to investigate mollusc shell remains from archaeological and palaeontological sites.

Appendices

Appendix 1. Materials and methods

1. Materials

In this thesis work, 34 different mollusc shells specimens and 23 archaeological shell samples were analysed, all of which are summarised in Table 1 and Table 2. The majority of the studied mollusc shells (32) refer to modern specimens and two shell samples were of subfossil origin, recovered from archaeological excavations (Table 1, M7, M9). Mollusc shell samples (M1-34, Table 1) were obtained from personal collections of laboratory directors and collaborators: Frédéric Marin (Biogéosciences, UMR CNRS 6282, University of Burgundy-Franche-Comté, Dijon, France), Beatrice Demarchi (Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy), Søren H Andersen (Moesgaard Museum, Højbjerg, Denmark), Alberto Girod (Italian Malacological Society, Sorengo, Switzerland). Archaeological samples (A1-23, Table 2) were obtained from museum collections via our collaborators: Søren H Andersen (Moesgaard Museum, Højbjerg, Denmark), Helmut Schlichtherle (Landesamt für Denkmalpflege im Regierungspräsidium Stuttgart, Gaienhofen, Germany), Alberto Girod (Italian Malacological Society, Sorengo, Switzerland), Solange Rigaud and Claire Manen from PACEA institute (UMR 5199 CNRS, University of Bordeaux, France), Monique de Cargouët and Nicolas Potier from CEREP institute, Archaeological museum of Sens (Société archéologique de Sens et Musées de Sens, France).

The material, physical and chemical properties of the shell samples were characterised by different techniques that were chosen according to the type of sample and the aim of the study (summarised in Table 1 and Table 2). These mainly included:

- Structural analyses (SEM, CT-scan, DUV photoluminescence imaging);
- Mineralogical analysis (FTIR-ATR);
- Geochemical analysis (stable isotopes - $\delta^{13}C_{carb}$ and $\delta^{18}O_{carb}$);
- Biochemical analyses (SDS-PAGE, Western-blot, ELLA/ELISA, FTIR-ATR, AAR);
- Proteomics (HPLC-MS/MS and PMF by MALDI-TOF-MS).

The shells of taxon *Spondylus gaederopus* (M1, Table 1) were characterised by multiple structural, biochemical and proteomics techniques and represented an in-depth biomineralization study of a model system.

The methodologies of the techniques are presented below. In this section, for simplicity reasons, the analysed mollusc shells and archaeological samples are

referred with their numbers as in Table 1 and 2 (M1-34 for mollusc shells and A1-26 for archaeological samples).

Table 1. Shell samples analysed in this thesis work. The first column reports sample number and the last column refers to the study chapter where the data are presented. Abbreviations: SEM – scanning electron microscopy; DUV – Deep-Ultraviolet photoluminescence imaging; FTIR-ATR – infrared spectroscopy analysis in attenuated total reflectance mode; ELLA – enzyme-linked lectin assay; ELISA – enzyme-linked immunosorbent assay; HPLC-MS/MS - high resolution liquid chromatography-tandem mass spectrometry; MALDI-TOF-MS – matrix assisted laser desorption-ionisation, time-of-flight mass spectrometry; TMT – tandem mass tag. Samples obtained from personal collections of lab directors and collaborators: F.M. – Frédéric Marin (Biogéosciences, UMR CNRS 6282, University of Burgundy-Franche-Comté, Dijon, France), B.D. – Beatrice Demarchi (Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy), S.H.A. – Søren H Andersen (Moesgaard Museum, Højbjerg, Denmark), A.G. – Alberto Girod (Italian Malacological Society, Sorengo, Switzerland)

No.	Species	Order, family, genus	Sample type, source	Age	Shell provenance	Analysis	Chapters
M1	*Spondylus gaederopus* Linnaeus, 1758	Pectinida, Spondylidae, *Spondylus*	Intact shells (full valves), purchased from Conchology, Inc (*Conchology, Inc*, 2020).	Modern	Saronikos, Greece	**Characterization:** Microstructural (SEM, DUV); Biochemical (FTIR-ATR, ELLA/ ELISA). **Proteomics:** Tandem mass spectrometry by HPLC-MS/MS; Shell PMF by MALDI-TOF-MS; Quantitative proteomics by HPLC-MS/MS using TMT labelling.	C1, C2, C3

M2	*Unio pictorum* (Linnaeus, 1758)	Unionida, Unionidae, *Unio*	Intact shells (full valves), obtained from reference collection of F.M.	Modern	Izeure, Côte d'Or, Burgundy, France	**Characterization:** Microstructural (SEM); Mineralogical (FTIR). **Proteomics:** Tandem mass spectrometry by HPLC-MS/MS; Shell PMF by MALDI-TOF-MS.	C3, C5
M3	*Ostrea edulis*	Ostreida, Ostreidae, *Ostrea*	Intact shells (full valves), obtained from reference collection of S.H.A.	Modern	Limfjord, Northern Jutland, Denmark	**Characterization:** Microstructural (SEM); Mineralogical (FTIR). **Proteomics:** Tandem mass spectrometry by HPLC-MS/MS; Shell PMF by MALDI-TOF-MS.	C3, C5
M4	*Patella vulgata* Linnaeus, 1758	Patellidae, *Patella*	Shell powder, obtained from reference collection of B.D.	Modern	St Marys Lighthouse, Newcastle, UK	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C3

			(Demarchi et al., 2014)				
M5	*Phorcus turbinatus* (Born, 1778)	Trochida, Trochidae, *Phorcus*	Shell powder, obtained from reference collection of B.D. (Demarchi et al., 2014)	Modern	Lipari, Sicily, Italy	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C3
M6	*Pecten maximus* (Linnaeus, 1758)	Pectinida, Pectinidae, *Pecten*	Shell powder, obtained from reference collection of B.D. (Demarchi et al., 2014)	Modern	Purchased at local supermarket, York, UK	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C3
M7	*Pseudunio auricularius* (Spengler, 1793)	Unionida, Margaritiferidae, *Pseudunio*	Intact shells (full valves and also fragments), obtained from research collection of A.G.	Archaeological. 5226-5023 cal BCE at 2σ (Starnini et al., 2018)	Isorella, Po plain, Italy	**Characterization:** Microstructural (SEM); Mineralogical (FTIR). **Proteomics:** Tandem mass spectrometry by HPLC-MS/MS; Shell PMF by MALDI-TOF-MS.	C3, C5
M8	*Modiolus modiolus*	Mytilida, Mytilidae, *Modiolus*	Intact shells (full valves),	Modern	Limfjord, Northern	**Characterization:** Microstructural	C4, C5

	(Linnaeus, 1758*)*		obtained from reference collection of S.H.A.		Jutland, Denmark	(SEM); Mineralogical (FTIR). **Proteomics:** Tandem mass spectrometry by HPLC-MS/MS; Shell PMF by MALDI-TOF-MS.	
M9	*Unio crassus (*Philipsson, 1788)	Unionida, Unionidae, *Unio*	Intact shells (full valves and also fragments), obtained from research collection of A.G.	Archaeological. 4260-3820 cal BCE (range of radiocarbon dates, at 1σ, of layers 2B, 2A3 and 2A, Toarte Pastilate (Biagi and Voytek, 2006)	Peştera Ungurească, Romania	**Characterization:** Microstructural (SEM); Mineralogical (FTIR). **Proteomics:** Tandem mass spectrometry by HPLC-MS/MS;	C5
M9a			Shell powder, obtained from reference collection of F.M.	Modern	South of Dijon, Côte d'Or, Burgundy, France	**Proteomics:** Shell PMF by MALDI-TOF-MS	C4
M10	*Margaritifera margaritifera* (Linnaeus, 1758)	Unionida, Margaritiferidae, *Margaritifera*	Intact shells (full valves), obtained from reference	Modern	Limfjord, Northern Jutland, Denmark	**Characterization:** Microstructural (SEM); Mineralogical	C4, C5

			collection of S.H.A.			(FTIR). **Proteomics:** Tandem mass spectrometry by HPLC-MS/MS; Shell PMF by MALDI-TOF-MS.	
M11	*Ruditapes philippinarum*	Venerida, Veneridae, *Ruditapes*	Shell powder, obtained from reference collection of F.M	Modern	Ile aux Oiseaux, Morbihan Gulf, Brittany, France	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M12	*Glycimeris glycimers*	Arcida, Glycymerididae, *Glycymeris*	Shell powder, obtained from reference collection of B.D (Demarchi et al., 2014)	Modern	Modern Glycymeris samples were collected from the Ebro Valley, Spain (collected dead on the beach, courtesy A.C. Colonese)	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M13	*Lithophaga lithophaga*	Mytilida, Mytilidae, *Lithophaga*	Shell powder, obtained from reference collection of F.M.	Modern	Côte d'azur, Mediterranean coast, France	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M14	*Lottia gigantea*	Lottiidae*, Lottia*	Shell powder, obtained from reference	Modern	California coast, USA	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4

			collection of F.M.				
M15	*Nautilus macromphalus*	Nautilida, Nautilidae, *Nautilus*	Shell powder, obtained from reference collection of F.M.	Modern	Noumea, New Caledonia, France	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M16	*Crassostrea gigas*	Ostreida, Ostreidae, *Crassostrea*	Shell powder, obtained from reference collection of F.M.	Modern	Brittany coast, France	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M17	*Mytilus galloprovincialis*	Mytilida, Mytilidae, *Mytilus*	Shell powder, obtained from reference collection of F.M.	Modern	Rovinj, Adriatic Sea Croatia coast, Croatia	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M18	*Mytilus californianus*	Mytilida, Mytilidae, Mytilus	Shell powder, obtained from reference collection of F.M.	Modern	California coasts, USA	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M19	*Aliger gigas* (known also as *Strombus gigas*)	Littorinimorpha, Strombidae, *Strombus*	Shell powder, obtained from reference collection of F.M.	Modern	French Antilles	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4

M20	*Potamida littoralis*	Unionida, Unionidae, *Potomida*	Shell powder, obtained from reference collection of F.M.	Modern	Côte d'Or, Burgundy, France	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M21	*Cornu aspersum* (known also as *Helix aspersa maxima*)	Stylommatophora, Helicidae, *Cornu*	Shell powder, obtained from reference collection of F.M.	Modern	Bouzy, Champaign, France (snail farm)	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M22	*Pinctada margaritifera*	Ostreida, Margaritidae, *Pinctada*	Shell powder, obtained from reference collection of F.M.	Modern	French Polynesia	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M23	*Haliotis asinina*	Lepetellida, Haliotidae, *Haliotis*	Shell powder, obtained from reference collection of F.M.	Modern	Bribie Island, Queensland, Australia (aquaculture research facility)	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M24	*Haliotis tuberculata* (known also as *Haliotis lamellosa*)	Lepetellida, Haliotidae, *Haliotis*	Shell powder, obtained from reference collection of F.M.	Modern	Cherbourg, Normandy	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4

M25	*Haliotis discus*	Lepetellida, Haliotidae , *Haliotis*	Shell powder, obtained from reference collection of F.M.	Modern	Japan	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M26	*Mercenaria mercenaria*	Venerida, Veneridae, *Mercenaria*	Shell powder, obtained from reference collection of F.M.	Modern	Fisheries in Scheveningen, The Netherlands	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M27	*Venus verrucosa*	Venerida, Veneridae, *Venus*	Shell powder, obtained from reference collection of F.M.	Modern	Brittany coasts, France	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M28	*Anodonta cygnea*	Unionida, Unionidae, *Anodonta*	Shell powder, obtained from reference collection of F.M.	Modern	Côte d'Or, Burgundy, France	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M29	*Pinna nobilis*	Ostreida, Pinnidae, *Pinna*	Shell powder, obtained from reference collection of F.M.	Modern	Adriatic sea, Croatia and Villefranche / mer, France	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M30	*Cerastoderma edule*	Cardiida, Cardiidae, *Cerastoderma*	Shell powder, obtained from reference collection of	Modern	Brittany coasts, France	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4

			F.M.				
M31	*Tridacna* sp.	Cardiida, Cardiidae, *Tridacna*	Shell powder, obtained from reference collection of F.M.	Modern	French Polynesia	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M32	*Littorina littorea*	Littorinimorpha, Littorinidae, *Littorina*	Shell powder, obtained from reference collection of B.D. (Demarchi et al., 2014)	Modern	Lipari, Sicily, Italy	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M33	*Dentalium* sp.	Dentaliida, Dentaliidae, *Dentalium*	Shell powder, obtained from reference collection of B.D. (Demarchi et al., 2014)	Modern	Lipari, Sicily, Italy	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4
M34	*Antalis* sp.	Dentaliida, Dentaliidae, *Antalis*	Shell powder, obtained from reference collection of B.D. (Demarchi et al., 2014)	Modern	Lipari, Sicily, Italy	**Proteomics:** Shell PMF by MALDI-TOF-MS.	C4

Table 2. Archaeological samples analysed in this thesis work. The first column reports sample number and the last column refers to the study chapter where the data are presented. Abbreviations: SEM – scanning electron microscopy; FTIR-ATR – infrared spectroscopy analysis in attenuated total reflectance mode; AAR – amino acid racemisation analysis; HPLC-MS/MS – high resolution liquid chromatography-tandem mass spectrometry; MALDI-TOF-MS – matrix assisted laser desorption-ionisation, time-of-flight mass spectrometry.

No.	Names	Archaeological site	Age	Source	Type	Analyses	Identification	Chapter
A1	HavA	Havnø (Denmark)	4200-4000 cal BCE (radiocarbon dating of the Late Mesolithic horizon) (Andersen, 2000; 2008)	Søren H Andersen, Moesgaard Museum, Højbjerg, Denmark	Double-button	**Characterization:** Microstructural (SEM); Mineralogical (FTIR), Molecular (AAR), Geochemical (stable isotopes) **Proteomics:** Tandem mass spectrometry by HPLC-MS/MS;	Freshwater mother-of-pearl Unionoida	C5
A2	HavB	Havnø (Denmark)			Double-button		Freshwater mother-of-pearl Unionoida	C5
A3	HavC	Havnø (Denmark)			Double-button		Freshwater mother-of-pearl Unionoida	C5
A4	HorA	Hornstaad-Hörnle IA (Germany)	3917- 3902 BCE (dendrochronology) (Billamboz, 2006)	Obtained from Helmut Schlichtherle, Landesamt für Denkmalpflege im Regierungspräsidium Stuttgart, Gaienhofen, Germany	Double-button	**Characterization:** Microstructural (SEM); Mineralogical (FTIR), Molecular (AAR), Geochemical (stable isotopes) **Proteomics:** Tandem mass spectrometry by HPLC-MS/MS;	Freshwater mother-of-pearl Unionoida	C5
A5	HorB	Hornstaad-Hörnle IA (Germany)			Double-button		Freshwater mother-of-pearl Unionoida	C5
A6	HorC	Hornstaad-Hörnle IA (Germany)			Double-button		Freshwater mother-of-pearl Unionoida	C5

A7	PesB	Peştera Ungurească (Romania)	4260-3820 cal BCE (range of radiocarbon dates, at 1σ, of layers 2B, 2A3 and 2A, Toarte Pastilate) (Biagi and Voytek, 2006)	Alberto Girod, Italian Malacological Society, Sorengo, Switzerland	Double-button		Freshwater mother-of-pearl Unionoida	C5
A8	P167	Le Taï, (Remoulins, France)	Neolithic	Solange Rigaud and Claire Manen from PACEA institute, UMR 5199 CNRS, University of Bordeaux, France	Circular bead (fragmented)	**Characterization:** Microstructural (SEM, CT-scan); Mineralogical (FTIR). **Proteomics:** Shell PMF by MALDI-TOF-MS.	Crossed-lamellar, aragonitic mollusc shell taxon n/d	C4
A9	P168	Le Taï, (Remoulins, France)	Neolithic		Circular bead (fragmented)			
A10	P170	Le Taï, (Remoulins, France)	Neolithic		Circular bead			
A11	P175	Passy (Yonne, France)	Neolithic	CEREP institute, Archaeological museum of Sens (Société archéologique de Sens et Musées de Sens), France	Shell valve	**Characterization:** Microstructural (SEM); Mineralogical (FTIR). **Proteomics:** Shell PMF by MALDI-TOF-MS.	Unionid shell	C4
A12	P176	Villeneuve-la-Guyard (Yonne)	Neolithic		Circular bead	**Characterization:** Microstructural (SEM, CT-scan); Mineralogical (FTIR). **Proteomics:** Shell PMF by MALDI-TOF-MS.	Crossed-lamellar, aragonitic mollusc shell taxon n/d	C4
A13	P179	Villeneuve-la-Guyard (Yonne)	Neolithic	CEREP institute, Archaeological museum of Sens (Société archéologique de	Tubular triangular bead	**Characterization:** Microstructural (SEM); Mineralogical (FTIR).	Crossed-lamellar, aragonitic mollusc shell taxon n/d	C4

A14	P181	Villeneuve-sur-Yonne (Yonne)	Neolithic	Sens et Musées de Sens), France	Circular bead (fragmented)	**Proteomics:** Shell PMF by MALDI-TOF-MS.	Calcitic material, type n/d	C4
A15	P183.A	Villeneuve-sur-Yonne (Yonne)	Neolithic	CEREP institute, Archaeological museum of Sens (Société archéologique de Sens et Musées de Sens), France	Circular bead (?) (fragmented)	**Characterization:** Microstructural (SEM); Mineralogical (FTIR). **Proteomics:** Shell PMF by MALDI-TOF-MS.	Calcitic material, type n/d	C4
A16	P184	Villeneuve-la-Guyard (Yonne)	Neolithic		Tubular bead		Crossed-lamellar, aragonitic mollusc shell taxon n/d	C4
A17	P186	Vinneuf (Yonne)	Neolithic	CEREP institute, Archaeological museum of Sens (Société archéologique de Sens et Musées de Sens), France	Trapezoid bead	**Characterization:** Microstructural (SEM); Mineralogical (FTIR). **Proteomics:** Shell PMF by MALDI-TOF-MS.	Calcitic material, type n/d	C4
A18	P187	Passy (Yonne, France)	Neolithic		Flat fragment		Calcitic material, type n/d	C4
A19	P189.A	Passy (Yonne, France)	Neolithic		Circular bead	**Characterization:** Microstructural (SEM, CT-scan); Mineralogical (FTIR). **Proteomics:** Shell PMF by MALDI-TOF-MS.	Calcitic material, type n/d	C4
A20	P189.C	Passy (Yonne, France)	Neolithic	CEREP institute, Archaeological museum of Sens (Société archéologique de Sens et Musées de Sens), France	Circular bead		Calcitic material, type n/d	C4
A21	P189.G	Passy (Yonne, France)	Neolithic		Circular bead		Calcitic material, type n/d	C4

A22	P189.J	Passy (Yonne, France)	Neolithic	CEREP institute, Archaeological museum of Sens (Société archéologique de Sens et Musées de Sens), France	Circular bead (fragmented)	**Characterization:** Microstructural (SEM); Mineralogical (FTIR). **Proteomics:** Shell PMF by MALDI-TOF-MS.	Calcitic material, type n/d	C4
A23	P189.L	Passy (Yonne, France)	Neolithic		Circular bead (fragmented)		Bone-like material, burnt	C4

2. Analytical techniques

2.1 Structural and mineralogical characterisation

2.1.1 Scanning Electron Microscopy (SEM)

Scanning electron microscopy (SEM) was used to study the microstructural organization of recent mollusc shells and archaeological samples. The main aims of SEM analyses were:

1) The characterisation of the microstructure of several reference modern and archaeological shells (M2, M3, M7, M8, M9, M10, Chapter 5). Preparation: small transverse fragments of the shell were obtained using a Dremel drill equipped with a diamond blade.
2) The characterisation of the microstructure of archaeological samples (A1-23, Chapter 4, 5). Preparation: non-invasive sampling was performed for the archaeological samples – intact fragments were characterised.
3) Detailed investigation of microstructural layers in *Spondylus gaederopus* shell (M1, Chapter 1). Preparation: a transverse section was obtained from the upper valve of the shell and it was further cut to have a small squared-shaped fragment. It was polished using grinding papers with progressively higher grit sizes (P200, P400, P600, P800, P1200) and using a thin-section polishing machine. Lastly, the fragment was mirror-polished using an aluminium oxide suspension (0.05 µm particle size) and rinsed with water.

The samples were etched with 1% (w/v) EDTA solution in a sonication bath in order to provide a better resolution of the surface topography and of shell microstructure patterns, when observed by SEM (with the exception for archaeological samples A8-23, which had to be observed without any further pre-treatment). Modern mollusc shells were etched for 2-3 min and the archaeological samples (A1-7) and subfossil shells (M7, M9) were treated for 1 minute. No carbon coating was applied. The samples were observed using a Hitachi TM1000 Tabletop Microscope in low vacuum mode (UMR CNRS 6282 Biogéosciences, UBFC, Dijon, France).

2. 1.2 Infrared spectroscopy (FTIR)

Fourier Transform Infrared spectroscopy (FTIR) in attenuated total reflectance (ATR) was used to characterise the mineral phase of the archaeological samples (A1-23, Chapter 4,5) and several reference modern shells (M2, M3, M7, M8, M9, M10, Chapter 5). The FTIR samples were obtained by scraping carefully – in particular in the case of archaeological objects – tiny fragments of the shell. The ATR spectra were recorded with a Bruker Vector 22 instrument (BrukerOptics Sarl, France, Marne la Vallée) fitted with a GoldenGate attenuated total reflectance

device (SpecacLtd, Orpington, UK) in the 4000–500 cm^{-1} range (twelve scans at a spectral resolution of 4 cm^{-1}). Spectral analyses were performed with the OPUS software provided by the instrument manufacturer (BrukerOptics Sarl). The analyses were carried out at the Institute of Molecular Chemistry, ICMUB UMR CNRS 6302, UBFC, Dijon, France. The assignment of the different absorption bands was obtained by comparison with previous spectra descriptions available in the bibliography.

2.1.3 X-ray microtomography

CT-scan, non-invasive X-ray detection analysis was used to obtain a 3D structural representation of the archaeological samples (A8-10, A12-17, A19-21, Chapter 4). The samples were scanned with Micro CT scanning system (SkyScan 1174, Bruker; UMR CNRS 6282 Biogéosciences, UBFC, Dijon, France). The data were exported as 3Dmodel files and software MeshLab was used to investigate and obtain the images from the 3D scans.

2.1.4 Bio-structural characterization by luminescence imaging using DUV synchrotron photoemission beamline

The Deep-Ultraviolet (DUV-visible) photoluminescence imaging was used to investigate the distribution of organics at the microstructural and mineralogical transition in *Spondylus* shell (M1, Chapter 1). The fluorescence maps were recorded on the DISCO beamline at the SOLEIL Synchrotron (Jamme et al., 2013) using the setup TELEMOS build around an Olympus IX71 inverted microscope stand and a Peltier iDus CCD detector with 26x26 µm pixel size. Sample preparation: a small fragment was cut from the upper valve of the *Spondylus gaederopus* shell and was mirror-polished. The area was scanned using a x40 objective and the luminescence maps were reconstructed and corrected using an in-house script in Matlab, developed at DISCO beamline by one of the collaborators (Frédéric Jamme).

2.2 Geochemical analysis. Stable isotopes

The stable isotopes of the shell biogenic carbonate, i.e. the $\delta^{13}C_{carb}$ and $\delta^{18}O_{carb}$, were studied to determine the environmental conditions at the time the organisms were actively biomineralizing their shell. The analysis of $^{18}O/^{16}O$ ($\delta^{18}O$), was carried out to discriminate freshwater *vs* marine carbonates of archaeological shell samples (A1-7, Chapter 5).

Sample preparation: small amounts of shell powder (about several milligrams) were sampled in bulk and in the case of sample A7, which showed a good state of preservation of its nacreous and prismatic shell layers, the different microstructures were sampled separately. The shell powders were bleached for 48 hours using

NaOCl (12% w/v) to preclude any contamination. The stable isotope analyses were carried out at two laboratories: samples A1-6 were analysed at 1) the Light Stable Isotope Laboratory (School of Archaeological Sciences, University of Bradford, UK) and sample A7 was analysed at 2) GISMO platform (UMR CNRS 6282 Biogéosciences, UBFC, Dijon, France).

1) Analytical procedure at Light Stable Isotope Laboratory: between 100 and 300 µg of the bleached biogenic calcium carbonate powders were loaded into 12 ml Exetainer ® tubes. Carbon and oxygen isotope values were determined by online phosphoric acid dissolution at 70°C using a Thermo GasBench 2 preparation system coupled to a Thermo Delta V Advantage Isotope-Ratio mass spectrometer. Standardisation of $\delta^{18}O$ values against the V-PDB reference was undertaken using repeated measurements of international standards IAEA NBS-19, IAEA-CO-8 and IAEA-CO-1, as well as internal laboratory standards (Merck $CaCO_3$ and OES). The analytical precision of the instrument was better than ± 0.1 ‰.
2) Analytical procedure at GISMO platform: between 35 and 40 µg of the bleached biogenic calcium carbonate were loaded into glass vials. Carbon and oxygen isotope values were determined by online 100% phosphoric acid dissolution (250 µl) at 70°C for 12 minutes using a Delta V Plus mass spectrometer coupled with a Kiel IV carbonate device (ThermoFisher). All isotopic values are reported in the standard δ-notation in permil (‰) *vs* VPDB. The reproducibility (2σ) of the IAEA NBS19 used as an external standard is better than 0.04 ‰ for the $\delta^{13}C$ and 0.08 ‰ for the $\delta^{18}O$.

2.3 Shell organic matrix

The shell organic matrix was extracted from all of the samples studied in this thesis work. The matrices were characterised by different biochemical approaches and proteomic techniques. The shell preparation, matrix extraction and further sample processing followed several main steps:

1) Preparation: shell cleaning, crushing and powder forming;
2) Bleaching of the shell powder in order to isolate mineral-bound shell organics;
3) Demineralisation of the shell;
4) Purification of shell matrix/proteins;
5) Digestion of proteins and purification of peptides for proteomic analysis.

The steps are detailed below.

2.3.1 Cleaning and powdering the shell

Reference shells M1, M2, M3, M7, M8, M9, M10 (Table 1) were obtained as intact shell valves and were carefully prepared to have cleaned powdered samples. First of all, the shells were cleaned using abrasive tools and immersed in diluted NaOCl to remove surface contamination from the valves (~1–1.5% active chlorine) for 2 h. The valves were well rinsed with water, air-dried and crushed into fragments (~2 mm) with a Jaw-crusher (Retsch BB200; UMR CNRS 6282 Biogéosciences, UBFC, Dijon, France). The coarsely crushed shells were further powdered with a mortar grinder (Frisch Pulverisette 2) and sieved to a particle size < 200 µm. For sample M1, intermediate bleaching step was also applied for coarsely crushed powder as described in section 2.3.2. The rest of modern reference shells (M4-6; M11-34) from Table 1 were available as cleaned shell powders (grain size 200-500 µm). Archaeological samples (A1-23, Table 2) were placed directly into eppendorfs and grinded into powder manually using micro-pestles.

2.3.2 Bleaching

All of the mollusc shells and archaeological shell samples analysed in this study were thoroughly cleaned by bleaching the shell powders with sodium hypochlorite (NaOCl). For modern shells, the valves were also cleaned by mechanical abrasion before they were grinded into powder. NaOCl is used to isolate mineral-bound shell matrix fraction and removes environmental contamination, intercrystalline as well as extracellular shell organics.

In this thesis work, several different bleaching approaches were used, which were selected on the basis of the samples and techniques by which shell matrices were analysed. The different bleaching methods used in the study are presented in Table 3, grouped according to the principal aims of the study. After bleaching, the shell powders were well rinsed with ultrapure water, dried under air and processed for matrix extraction.

Table 3. Summary of the different bleaching approaches used in the different studies of the thesis.

Aim	Chapter	Study	Sample preparation	Bleaching approach	Notes and details
Compare shell matrix fractions obtained via different bleaching approaches.	1	The shell matrix of the european thorny oyster, *Spondylus gaederopus*: microstructural and molecular characterisation.	Intact *Spondylus* shell valves were cleaned with diluted NaOCl (~1-1.5% active chlorine) for 2h (step noted as the 1st bleaching treatment - 1BL). Cleaned shell valves were crushed into ~2 mm fragments with a Jaw-crusher (Retsch BB200) to obtain a coarsely grinded powder.	Two types of bleaching approaches were tested: **Fraction 2BL**: coarsely ground powder was bleached with ~1-1.5% NaOCl for 4 h. **Fraction 3BL**: the 2BL shell powder was further grinded with a mortar grinder (Frisch Pulverisette 2) to 200 µm particle size and bleached with ~1-1.5% NaOCl for 14 h.	The 2BL matrix represents a close approximation of (inter+intra)crystalline fraction and the 3BL matrix – intracrystalline organics.
	3	Shell palaeoproteomics: First application of peptide mass fingerprinting for	Shell samples were available as fine-grained powders (particle size: 200-500	Three bleaching approaches were tested – “mild”, “intermediate” and “strong”. **“Mild”**:	The matrix obtained via “strong” bleaching treatment (48 h with 10-15% NaOCl) refers to

		the rapid identification of mollusc shells in archaeology.	µm).	shell powders were bleached using diluted NaOCl (1.0-1.5%) for 4 h. **“Intermediate”**: shell powders were bleached using diluted NaOCl (1.0-1.5%) for 24 h. **“Strong”**: shell powders were bleached using concentrated NaOCl (10-15%) for 48 h.	operational intracrystalline shell matrix fraction.
Proteomic characterisation of reference mollusc shells by HPLC-MS/MS.	5	'Palaeoshellomics' reveals the use of freshwater mother-of-pearl in prehistory.	Reference modern and subfossil shells were grinded into fine powder.	The powdered shell samples were bleached using NaOCl (2.8%) for 48 h.	
The study of archaeological shell double-buttons by HPLC-MS/MS.			Fragments of archaeological shell samples were crushed with a micro pestle and carefully powdered directly in an eppendorf.	The samples were bleached for 3 h using NaOCl (2.8%).	The archaeological samples were treated using a mild bleaching approach, considering the small size of samples and the unknown state of protein preservation.

Proteomics characterisation of intracrystalline shell proteins from modern shells and archaeological samples.	2	Proteomic study of intracrystalline *Spondylus* shell protein diagenesis.	Shell samples were available as fine-grained powders (particle size: 200 µm).	Shell powders were bleached using concentrated NaOCl (10-15% active chlorine) for 48 h (with the exception of sample A23 (Table 2) which was characterised by bone-like structure).	In all of the studies the targeted fraction was the operational intracrystalline shell matrix.
	3	Shell palaeoproteomics: First application of peptide mass fingerprinting for the rapid identification of mollusc shells in archaeology.	Reference shell samples were available as fine-grained powders (particle size: 200 µm). The archaeological samples were sampled by grating with a scalpel and further grinded into fine powder directly in an eppendorf.		
	4	Shell palaeoproteomics II: peptide mass fingerprinting of mollusc shells in archaeology.			

2.3.3 Shell demineralization

In the studies presented in this thesis, two types of reagents were used to demineralize shell powders – acetic acid (AcOH) and ethylenediaminetetraacetic acid (EDTA).

1) **Acetic acid:** cold diluted AcOH (10% v/v) was added to shell powders (in suspension in Milli-Q water) in small aliquots (for sample M1 as presented in Table 1, automatic burette was used to keep the rate of 100 µL every 5 sec.). The solution was constantly mixed and kept in cold conditions (4°C). The demineralised extracts were centrifuged, resulting in two separate fractions – acid soluble (ASM) and acid insoluble matrices (AIM). This approach was used in these studies: the biomolecular and proteomics characterisation of *Spondylus* shell (Chapter 1), shell protein diagenesis (Chapter 2), analysis of shell peptide mass fingerprints (Chapter 3) and palaeoshellomics study of modern and archaeological shell samples (Chapter 5).
2) **EDTA**: The EDTA solution (pH 8, ~0.5 M) was added at room temperature and shell samples were thoroughly mixed for 4-16 h until completely demineralised. This demineralisation approach was used for samples analysed by immunochemistry in shell protein diagenesis study (Chapter 2) and also for reference and archaeological samples studied by shell PMF approach (samples M1-34 and A1-22 as in Table 1 and 2; data presented in Chapters 3, 4). Note, archaeological sample A23 (bone-like structure, Table 2) was demineralised using cold 0.6M hydrochloric acid.

2.3.4 Extraction and purification

Several different shell matrix extraction approaches were used in this thesis which were chosen according to sample size, protein amount and type of analysis.

Matrix extraction from large-size samples. This approach was used to extract the matrix from modern and sub-fossil reference shells (Chapter 1, 5).

1) For modern shells (sample weight ≥10 g), the extraction followed a routine protocol developed at the Biomineralization laboratory (UMR CNRS 6282 Biogéosciences, University of Burgundy-Franche-Comté) (Ramos-Silva et al., 2012; Takeuchi et al., 2018). The ASM and AIM extracts obtained after demineralisation with acetic acid (as described in section 2.3.3), were purified as follows: after centrifugation (4500 rpm, 30 min., 5°C), the ASM was filtered (5 µm) and the volume of the solution was reduced by ultrafiltration using 10 kDa cut-off membrane (Millipore, Ref. PLGC07610), dialysed against 1L of Mill-Q water, with a minimum of 6 changes over 2 days, and freeze-dried. The AIM matrix was rinsed (5x) with ultra-pure water

and freeze-dried. This approach was used for the following shell samples (for names, see Table 1): M1 (Chapter 1), M2, M3, M8, M10 (Chapter 5).

2) For sub-fossil shells (sampling size ≥3 g) M7, M9 (Chapter 5), the purification procedure was adapted with minimal changes. The ASMs obtained from the extraction with acetic acid were ultrafiltered using 10 kDa cut-off tubes (Amicon Ultra-15, Ultracel-PL PLGC; centrifuged at 4500 RPM to reduce the volume to 0.75 mL) and washed with water (10 times). The AIM matrices were rinsed with ultrapure water as above (5 times). All matrices were lyophilised.

Matrix extraction from small-size samples. FASP and SP3 methods were used to extract shell proteins from modern and archaeological shells (Chapter 2, 3, 4).

1) **FASP** – **Filter Aided Sample Preparation method. Samples:** it was used to purify proteins from shell sample M1 in intracrystalline protein diagenesis study (Chapter 2) and samples M1, M2, M3, M7 in shell PMF study (Chapter 3); (for shell names, see Table 1). It was also used to purify proteins extracted from archaeological samples A1-7 in the palaeoshellomics study (Chapter 5, for sample names, see Table 2). **Procedure:** the acid soluble shell extracts (ASMs) were loaded into 3 kDa, 0.5 mL centrifugal devices (PALL Nanosep/Amicon Ultra). The samples were concentrated and desalted washing five times with HPLC-grade water (0.5 mL, centrifuging at 11000 rpm, at room temperature) and afterwards were exchanged to buffer (50 mM ammonium bicarbonate, pH 7.5–8). The AIMs were carefully rinsed with water (HPLC grade, 5x cycles of washes). After this step the samples were ready for digestion. Note: in shell PMF study (Chapter 3) for shells M1, M2, M3, M7, due to the small sample size used for extraction (20 mg), the ASMs and AIMs were loaded and purified together.

   *Note: FASP approach was also used to extract proteins from archaeological sample A23 (Table 2) that showed a bone-like structure. Full extraction procedure: the sample was demineralised with cold 0.6 M hydrochloric acid overnight at 4°C. The sample was centrifuged (13000 rpm for 10 min) and the solubilised fraction was separated. The insoluble fraction was further treated with guanidine·HCl 6M solution at 65 °C for 1.5 h to, centrifuged again (13000 rpm for 10 min) and the second solubilised fraction was separated and combined to the first one. Extracts were ultrafiltered using PALL Nanosep centrifugal devices (3 kDa, 0.5 mL), purified and exchanged to buffer (50 mM ammonium bicarbonate, pH 7.5-8).

2) **SP3** – **Solid-Phase, Single-Pot Sample Preparation method. Samples:** it was used to purify proteins from shell samples M1-34 and archaeological samples A8-22 (for names see Table 1 and Table 2 respectively) in shell

PMF studies (Chapter 3, 4). The extraction followed a published protocol (Hughes et al., 2019). 8 µL of Sera-Mag SpeedBeads (1:1 mixture of hydrophobic and hydrophilic) were added to the demineralised shell extracts. In order to induce binding, 100% EtOH (HPLC-grade) was added to have EtOH concentration of 50%. The mixtures were incubated at 24 °C for 5 min at ~1000 rpm. The tubes were then placed on a magnetic rack for separation. The supernatant was removed and discarded. The proteins bound to the beads were cleaned with 80% EtOH (performing 3x cycles), exchanged to buffer (50 mM ammonium bicarbonate, pH 7.5–8) and the mixture sonicated for 30 s. After this step the samples were ready for digestion. Important note: for extracts in acetic acid, the SP3 purification was performed directly after demineralisation; for extracts in EDTA, the proteins were first reduced and alkylated (as detailed in the next section) and the SP3 extraction was done afterwards. This is because the binding between proteins and beads is more efficient after reduction of disulfide bonds, according to published protocol (Hughes et al., 2019). However, the reduction and alkylation need a neutral or basic pH, hence, for extracts in acetic acid, this step could not have been performed and was done afterwards.

2.4 Proteomics

2.4.1 Sample preparation for proteomic analysis by mass spectrometry

General procedure: purified protein extracts were exchanged or resuspended in a buffer (50 mM ammonium bicarbonate, pH 7.5-8). The disulphide bonds were reduced and alkylated by using M DL-Dithiothreitol (Sigma, Canada) for 1 h at 65 °C and 0.5 M iodoacetamide (Sigma, USA) for 45 min at room temperature in the dark. The enzymatic digestion was carried out using protease trypsin (Promega, 2800 Woods Hollow Road Madison, WI 53,711 USA). Enzymatic digestion with elastase (Worthington, Lakewood, NJ, USA) was also used for samples M2, M3, M7-10; A1-7 (Table 1, 2, Chapter 5). Digestion was carried out overnight at 37 °C. It was stopped with 10% TFA (to have a final concentration of 0.1% TFA). Digested peptides were purified using C18 solid-phase extraction tips (Pierce zip-tip; Thermo-Fisher) and evaporated to dryness.

Note 1: for samples which were prepared by SP3 protein purification method (Chapter 3, 4), the enzyme trypsin was added in the sample mixture together with paramagnetic beads. The digestion was carried out at 37 °C, maintaining constant and light shaking (~1000 rpm). After the digestion, the mixtures were centrifuged at 12.9 rpm for 1 minute and the peptide digests were separated from the beads on a magnetic rack. Peptide purification was carried out as above.

Note 2: Extracts analysed in the study presented in Chapter 1, were prepared by in gel-digestion. The purified shell extracts were first cleaned by running a short separating 8% acrylamide gel stained with Coomassie blue. After the gel runs, samples were de-stained twice with a mixture of 100 mM ammonium bicarbonate (ABC) and 50% (v/v) acetonitrile (ACN) for 30 min at 22 °C and then dehydrated using 100% ACN for 15 min. They were then reduced with 25 mM ABC containing 10 mM DTT for 1 h at 56 °C and alkylated with 55 mM iodoacetamide in 25 mM ABC for 30 min in the dark at 22 °C. Gel pieces were washed twice with 25 mM ABC and dehydrated (twice, 20 min) with 100% ACN. Gel cubes were incubated with sequencing grade modified trypsin (Promega, USA; 12.5 ng/µl in 40 mM ABC with 10% ACN, pH 8.0) overnight at 37 °C. After digestion, peptides were extracted twice with a mixture of 50% ACN-5% formic acid (FA) and then with 100% ACN. Extracts were dried using a vacuum centrifuge Concentrator plus. Resulting peptides were resuspended in 10 µL of 10% ACN-0.1% trifluoroacetic acid (TFA) and were ready for mass spectrometry analysis.

2.4.2 Mass spectrometry analysis

HPLC-MS/MS

Proteomic analyses by tandem mass spectrometry were carried out in the studies presented in Chapters 1, 2 and 5. Four HPLC-MS/MS systems were used:

1. **HPLC-MS/MS system at 3P5 Proteomic Platform** (University of Paris, Cochin Institute, INSERM, U1016, CNRS, UMR8104, F-75014 Paris). Ultimate 3000 Rapid Separation Liquid Chromatographic (RSLC) system (Thermo Fisher Scientific) coupled online with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) was used to analyse protein extracts of *Spondylus* shell (sample M1 as in Table 1, Chapter 1). **Procedure:** 1 µL of peptides was loaded, concentrated and washed on a C18 reverse phase pre-column (3 µm particle size, 100 Å pore size, 75 µm i.d., 2 cm length, Thermo Fisher Scientific). The loading buffer contained 98% H_2O, 2% ACN and 0.1% TFA. Peptides were then separated on a C18 reverse phase resin (2 µm particle size, 100 Å pore size, 75 µm i.d., 25 cm length, Thermo Fisher Scientific) with a 1 h gradient from 99% solvent A (0.1% FA and 100% H_2O) to 90% solvent B (80% ACN, 0.085% FA and 20% H_2O). The mass spectrometer acquired data throughout the elution process and operated in a data dependent scheme with full MS scans acquired with the Orbitrap, followed by MS/MS HCD fragmentations acquired with the Ion Trap on the most abundant ions detected. Mass spectrometer settings were: full MS (AGC: 2x10e5, resolution: 6x10e4, m/z

range 350–1500, maximum ion injection time: 60 ms) and MS/MS (HCD collision energy: 30%, AGC: 2x10e4, resolution: 3x10e4, maximum injection time: 100 ms, isolation windows: 1.6 m/z Da, dynamic exclusion time setting: 30 s). The fragmentation was permitted for precursors with a charge state of 2, 3, 4 and up.

2. **HPLC-MS/MS system at The Globe Institute** (University of Copenhagen, Copenhagen, Denmark). EASY-nLC 1200 (Thermo Fischer Scientific, Bremen, Germany) connected to a Q-Exactive HF-X orbitrap mass spectrometer (Thermo Fischer Scientific) was used for quantitative proteomics using TMT labelling to analyse *Spondylus* extracts (sample M1 as in Table 1, Chapter 2). **TMT labelling:** Dried peptides were sent to The Globe Institute, University of Copenhagen, Denmark for TMT labelling and tandem mass spectrometry analysis. The samples were resuspended in 50 µL 50% acetonitrile (ACN). Due to the error caused by low amounts of protein detectable by spectrophotometer analysis (at 280 nm and 205 nm), the samples were normalised by the starting weight of the shell used. Therefore, a portion of the sample (34-40 µl) was taken forward for TMT labelling. HEPES buffer and ACN were then added for a total concentration of 50% ACN and 30 mM HEPES, with the pH checked to be around 8. Thermo-Scientific TMT labels (11-plex) were prepared by resuspension in anhydrous ACN and then 0.02 mg of label was added to each sample, which was then vortexed and incubated at room temperature (covered) for 1 hour. The reaction was quenched by adding 1% hydroxylamine and incubated at room temperature for a further 15 min. The samples were pooled and then cleaned using an in-house made C18 StageTip. **Mass Spectrometry (MS) analysis:** The StageTip was eluted using 20 µl each 40% then 60% ACN and vacuum centrifuged at 40 °C until approximately 3 µl remained. It was then resuspended with 10 µL 0.1% trifluoroacetic acid (TFA) and 5% ACN solution. 5 µL was then analysed by an EASY-nLC 1200 (Thermo Fisher Scientific, Bremen, Germany) coupled to a Q-Exactive HF-X orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) on a 77 min gradient. Chromatographic and MS parameters were then adjusted based on previously published methods for ancient and degraded samples (Mackie et al., 2018), adjusted for TMT analysis. Therefore, the isolation window was narrowed to 0.8 m/z and the normalised collision energy raised to 33. In short, the other MS parameters were set as follows: MS1- 120,000 resolution at m/z 200 over the m/z range 350–1400, target of 3e6, maximum injection time (IT) of 25 ms; MS2- top 10 mode, 60,000 resolution, target of 2e5, maximum IT of 118 ms, and dynamic exclusion of 20 s.
3. **HPLC-MS/MS system at MSAP CNRS laboratory** (University of Lille1, France). A nanoflow HPLC instrument (U3000 RSLC Thermo Fisher

Scientific) coupled to a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) was used to analyse shell samples M2, M3, M8, M10 (Table 1) and archaeological samples A1-6 (Table 2), presented in Chapter 5. **Procedure:** 1 µL of peptide mixture was loaded onto the preconcentration trap (Thermo Scientific, Acclaim PepMap100 C18, 5 µm, 300 µm i.d x 5 mm) using a partial loop injection and a flow rate of 10 µL/min (duration 5 min) with buffer A (5% ACN and 0.1% FA). The peptide mixture was then separated using a nanocolumn (Acclaim PepMap100 C18, 3 µm, 75 mm i.d. × 500 mm) and a linear gradient of 5–40% buffer B (75% ACN and 0.1% FA) at a flow rate of 250 nL/min and a temperature of 45°C. The total duration of an LC-MS/MS run was 120 min. MS data were acquired using a data-dependent top 20 method that selects the 20 most abundant precursor ions from the survey scan (400–1600 m/z range) for HCD fragmentation. The dynamic exclusion duration was 60 s. The isolation of precursors was performed with a 1.6 *m/z* window and MS/MS scans were acquired with a starting mass of 80 *m/z*. Survey scans were acquired at a resolution of 70,000 at *m/z* 400 (AGC set to 106 ions with a maximum fill time of 180 ms). Resolution for HCD spectra was set to 35,500 at *m/z* 200 (AGC set to 105 ions with a maximum fill time of 120 ms). Normalized collision energy was 28 eV. The underfill ratio, which specifies the minimum percentage of the target value likely to be reached at maximum fill time, was defined as 0.3%. The instrument was run with the peptide recognition mode (i.e. from 2 to 8 charges) and exclusions of singly charged ions and unassigned precursor ions. Fifteen blanks of 1 h and one reference sample (80 fmol of Cytochrome C digest; Thermo Fisher Scientific) were injected and controlled (e.g. retention time, sensitivity, carry over, contaminations) between each sample injection.

4. **HPLC-MS/MS system at Biomolecules core facility** (Molecular Biotechnology and Health Sciences Department, University of Turin, Italy). An Ultimate 3000 Dionex nanoHPLC instrument coupled with an Orbitrap Fusion (Thermo Scientific, Milan, Italy) mass analyser was used to analyse shell samples M7, M9 (Table 1) and archaeological sample M7, presented in Chapter 5. **Procedure:** The separation of peptide mixture was achieved using a PepMap RSLC C18, 2 µm, 100 Å, 75 µm × 50 cm column (Thermo Scientific) and a PepMap C18, 5 µm × 5 mm, 100 Å preconcentration column (Thermo Scientific). The eluent used for the preconcentration step was 0.05 % TFA in H_2O/ ACN 98/2 and the flowrate was 5 µL/min. The eluents used for chromatographic separation were 0.1% FA in H_2O (solvent A) and 0.1% FA in ACN/H_2O at ratio 8/2 (solvent B). The program was initially isocratic at 5:95 (A:B %) for 5 minutes, then increased to 75:25 in 55 minutes, ran up to 60:40 in 6 minutes, and then to 10:90 in 5 minutes.

Recondition time was 20 minutes. The injection volume was 1 µL and the flow rate 300 nL/min. The nanocolumn was provided with the ESI source. The mass spectrometry parameters were: positive spray voltage 2300 (V), sweep gas 1 (Arb) and ion transfer tube temperature 275 ºC. Full scan spectra were acquired in the range of *m/z* 375-1500 (resolution 120000 @ *m/z* 200). MS^n spectra in data dependent analysis were acquired in the range between the ion trap cut-off and precursor ion *m/z* values. HCD collision energy was fixed at 28%, orbitrap resolution 50000 and the isolation window was 1.6 *m/z* units.

PMF analysis by MALDI-TOF-MS.

A bench-top Microflex LRF MALDI-TOF mass spectrometer (Bruker Daltonics, Germany) was used at the Department of Clinical and Biological Sciences of University of Turin University of Turin (AOU S. Luigi, 10043 Orbassano, Italy). The PMF by MALDI-TOF-MS approach was used to analyse shell samples M1-34 (Table 1) and archaeological samples A8-23 (Table 2) presented in Chapter 3, 4. **Procedure:** The peptide extracts 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 (modern shell samples were spotted directly; for the archaeological samples two modes of matrix-mixture deposition were tested: spotting directly as for the modern shell samples and mixing with the matrix separately in an eppendorf and then spotting on the MALDI plate). 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%. 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.

2.4.3 Bioinformatics. Data treatment.

Tandem mass spectrometry data analysis (HPLC-MS/MS)

The obtained product ion spectra were analysed using PEAKS Studio (v. 8.5, Bioinformatics Solutions Inc. (BSI) (Ma et al., 2003). The data were searched against a "Mollusca protein" database which was created by downloading all sequences from NCBI protein repository, restricting the taxonomy to Mollusca (the dates of protein database creation are indicated in material and methods section for each of the studies presented in Chapter 1, 2 and 5). All of the searches also included a database of common laboratory contaminants (cRAP; common Repository of Adventitious Proteins, downloaded from: http://www.thegpm.org/crap/). Search parameters were defined assuming no enzyme digestion (in order to also detect naturally cleaved peptides, if any), fragment ion mass tolerance of 0.05 Da and a parent ion tolerance of 10 ppm. Results obtained by SPIDER searches (*i.e*. including all possible modifications) were used for peptide identification and protein characterization. The threshold values for acceptance of high-quality peptides were as follows:

- false discovery rate (FDR) threshold 0.5%;
- protein scores (parameters varied according to the type of samples analysed):
    - −10lgP ≥ 20 (Chapter 1);
    - −10lgP ≥ 40 (Chapter 5);
    - −10lgP ≥ 30 (Chapter 2);
- unique peptides ≥ 2;
- *de novo* sequences scores (ALC %) ≥ 50.

The peptide sequences identified in shell proteins were individually checked using the BLASTp tool (https://blast.ncbi.nlm.nih.gov/), and any sequences that were homologous to common laboratory contaminants were excluded from any further analysis.

Note 1: Several variations of these parameters were also applied to obtain additional data and are indicated separately in each of the chapter, section material and method.

Note 2: Protein quantification for *Spondylus* protein extracts (shell sample M1, Table 1, the study presented in Chapter 2) was performed by PeaksX software. Selected parameters: quantification mass tolerance: 0.2 Da; FDR threshold (%): 1.0; spectrum filter ≥31.9, quality ≥0, reporter ion intensity ≥0E0, Protein significance ≥0, significance method – PEAKSQ, Unique peptides ≥1. The relative quantification of peptides and proteins in each of the samples was obtained by calculating the relative reporter ion intensity in respect to values found in blank samples. The data were used to create relative abundance profiles, represented by bar plots and line graphs.

PMF data analysis (MALDI-TOF-MS)

The spectra were exported in the format of text files and were further processed using mMass (Niedermeyer and Strohalm, 2012), an open access mass spectrometry interpretation tool. Two spectra were obtained and averaged for each sample. All 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. *M/z* values corresponding to common laboratory contaminants (*i.e*. keratin, trypsin, α-cyano MALDI matrix) were filtered out (mass tolerance for peak matching: 0.1 Da). In addition, an in-house contaminant database was created by collecting together all *m/z* values identified in blank samples and was used to filter the spectra from recurring contaminant peaks.

Proteomics data deposition

Published mass spectrometry proteomics data (obtained by HPLC-MS/MS) were deposited to the ProteomeXchange Consortium via the PRIDE partner repository.
The proteomics dataset of *Spondylus gaederopus* shell (Chapter 1) is deposited with the dataset identifier PXD016760.
The proteomics dataset of several reference modern and subfossil shells (M2, M3, M7, M8, M9, M10) and archaeological samples (A1-7, Table 2) analysed in the study presented in Chapter 5, is deposited with the dataset identified PXD011985.

2.4.4 Bioinformatics. Protein characterisation

The physical and chemical parameters of identified proteins, i.e. MW, theoretical PI, amino acid composition, were assessed using the ProtParam tool on the EXPASY server. Protein classification, prediction of domains and motifs was carried using Interpro database (Mitchell et al., 2019) as well as Motif scan tool on the EXPASY server. The prediction of intrinsically disordered regions was carried out using IUPred2A (Mészáros et al., 2018). The BLASTp tool was used to identify homologous proteins.

2.5 Biochemical characterisation

2.5.1 SDS-PAGE gel electrophoresis

The gel electrophoresis was used to characterise *Spondylus* shell matrices (sample M1 as in Table 1, study in Chapter 1). The ASM extracts were dissolved in ultrapure water to have a final concentration of 4 μg/mL and 4x Laemmli buffer was added (ratio 3:1). The AIM extracts were resuspended and treated similarly. AIM has

limited solubility thus it refers to only the Laemmli-soluble supernatants – the LS-AIM. Denaturation was carried out at 99 °C for 5 min, then samples were cooled on ice. Proteins were run on precast (4–15%) gradient precast gels (Bio-Rad) or handcasted 12% gels, then stained with silver nitrate and Stains-all respectively (Morrissey, 1981; Campbell et al., 1983).

2.5.2 Immunochemistry assays

Enzyme-linked lectin assay (ELLA) and Enzyme-linked immunosorbent assay (ELISA)

ELLA and ELISA assays were performed to characterise *Spondylus* shell matrices (sample M1 as in Table 1). Analyses were carried out in *Spondylus* biomolecular and protein diagenesis studies presented in Chapter 1 and 2 respectively.

In the biomolecular *Spondylus* study (Chapter 1), ELLA was carried out to test three sets of 7 biotinylated lectins, total of 21 lectins (Vector Laboratories, Peterborough, UK, Ref. BK-1000, BK-2000, BK-3000). Lectins were diluted in TBS/Tween-20 with ratio 1:200 for kits 1, 2 and 1:100 for kit 3. The assays were directly tested on the obtained ASM fractions, obtained via acetic acid extraction (following the methodology described in sections 2.3.3 and 2.3.4).

In the *Spondylus* shell protein diagenesis study (Chapter 2), the ELLA assay was carried out using lectin jacalin and the ELISA assay was carried out with shell antibody K5090, raised against the ASM of the nacreous layer of the bivalve, *P. nobilis (Marin et al., 1994)*. Both assays were tested on EDTA soluble fraction, obtained via demineralisation with EDTA (section 2.3.3).

Procedure: The shell matrix solutions (100 µL per well, approx. ~200 ng/ well) were deposited in 96-well microplates (MaxiSorp, Nunc/Thermo Scientific, Nunc A/S, Roskilde, Denmark), incubated for 90 min at 37 °C, washed with TBS/Tween20 and blocked with Carbon-Free solution (for ELLA) or gelatin 1% solution (for ELISA). Lectin or antibody solutions were added to the wells and incubated for 90 min at 37 °C, washed (3 times TBS/Tween20), and a solution of alkaline phosphatase-conjugated avidin (A7294 Sigma, 1:70000, for ELLA) or alternately, GARP-AB (for ELISA) was added for another 90 min at 37 °C. Microplates were washed and incubated with the ELISA (aqueous diethanolamine solution, pH 9.8) substrate solution, containing phosphatase substrate (p-nitrophenylphosphate) at 37 °C. Optical density was read using a BioRad Model 680 microplate reader at 405 nm; background values (blank tests) were subtracted and the values were converted to percentage of reactivity. A total of 4 replicates were performed.

Data representation: the optical density (OD) values of reference samples were given 100% and were used to calculate the cross-reactivity intensities. In the biomolecular study of *Spondylus* (Chapter 1), 100% were given to values with the highest OD; in the study of *Spondylus* shell protein diagenesis (Chapter 2), 100%

values were given to OD values of non-heated shell samples (for more details see material and methods in Chapter 2).

2.5.3 Western-blot

Western-blot analyses were used for the characterisation of *Spondylus* shell matrices (shell sample M1 as in Table1) to investigate components that bind to lectin jacalin and shell antibody K5090. The analysis was carried out in two studies – the biomolecular characterisation of *Spondylus* (Chapter 1) and shell protein diagenesis (Chapter 2). **Procedure:** ASM and AIM shell extracts were run on a 12% SDS-PAGE gel as reported above (section 2.5.1), then electro-transferred on a PVDF Immobilon P membrane (Millipore), for 90 min. at constant intensity (100 mA). One membrane was incubated in Carbo-Free blocking solution (Vector Laboratories, Peterborough, UK, ref. SP-5040) then in a solution that contained the biotinylated lectin jacalin (Vector Labs. Ref. B-1155, diluted in TBS/Tween20, ratio 1:100) and finally, after rinsing, in a solution containing avidin-AP conjugate (Sigma A7294, dilution 1/ 70000). The second membrane was incubated in 1% gelatin blocking solution then in a solution that contained the shell antibody K5090 (diluted in ratio 1:100) and finally, after rinsing, in a solution containing the secondary antibody (GAR-AP, Sigma A7294, dilution 1/70000). Thorough rinsing steps with TBS/Tween20 were performed. In both cases, the membrane was stained with NBT/BCIP (Sigma B5655).

2.5.4 In situ lectin-gold assay

In situ lectin-gold localisation was performed in the biomolecular study of *Spondylus* shell (Chapter 1) using lectin jacalin. **Sample preparation:** a transverse section perpendicular to the growth line was cut from the upper valve of the shell (area closer to the margin) using a diamond saw. The section was polished with fine-grain sandpaper of decreasing grain size, then with alumina suspension (0.05 μm) and cut into small square fragments. These were subsequently cleaned by sonication, then bleached (for 10 min with NaOCl ~1–1.5% active chlorine) to clean the surface, and finally rinsed with ultrapure water. Fragments were etched with 1% EDTA (v/v) in the ultrasound bath for 5 min and rinsed again with ultrapure water. **Procedure:** the step followed a published procedure (Takeuchi et al., 2018). First of all, the fragments were incubated in a Carbo-Free blocking solution for 1 h and then incubated with a jacalin solution (prepared in TBS/Tween-20, diluted 1:10; plus, NaN_3, 1/10000 to prevent bacterial growth) overnight. Following this, the fragments were rinsed with TBS/Tween-20 (6x times, for the duration of 10 min each) and then incubated with goat anti-biotin antibody conjugated to ultra-small gold particles (0.8 nm) (GABio, Ultra Small, Ref. 800.088; Aurion, Wareningen, The Netherlands) for 2 h. After this, the fragments were well rinsed, dried gently and

incubated in the silver enhancement solution (British Biocell International, Ref. SEKL15) for about 10–15 min. Finally, they were rinsed again with milli-Q water and dried at 37 °C. The incubation steps were all performed at room temperature and with gentle shaking. Samples were observed with SEM (Hitachi, TM1000) under back-scattered electron mode. To preclude false positive signals, 3 negative controls were performed in parallel using the same conditions but omitting one or two of the following steps: 1) incubation with jacalin (negative control 1); 2) incubation with gold-coupled anti-biotin (negative control 2); 3) both incubations with jacalin and the gold-coupled anti-biotin (negative control 3). 4 replicas of the experiments were carried out to validate the observations.

2.5.5 In vitro calcium carbonate crystallisation study

In vitro $CaCO_3$ crystallisation test was performed on *Spondylus* shell matrices (shell sample M1 as in Table 1, Chapter 1) and the ASM extracts were analysed. **Procedure:** 200 µL of 10 mM $CaCl_2$ solutions containing ASMs at increasing concentrations (0.25 – 16 µg/ml) were placed in 16-well culture slides (Lab-Tek, Nunc/ Thermo Scientific, Rochester, NY, USA). The plastic cover was pierced beforehand to allow the reaction with ammonium bicarbonate vapours, placed on top of the slide and sealed with parafilm. It was placed in a desiccator under vacuum which contained ammonium bicarbonate crystals and incubated at 4 °C for 72 h. Afterwards, the solution was carefully removed with a blunt-end needle and the slide was air-dried. The 16-well culture slide was disassembled and $CaCO_3$ crystals were observed directly (without any carbon sputtering) using a tabletop scanning electron microscope (TM 1000, Hitachi).

References

Andersen, S.H., 2000. “Køkkenmøddinger” (shell middens) in Denmark: a survey. In: Proceedings of the Prehistoric Society. Cambridge University Press, pp. 361–384.

Andersen, S.H., 2008. A report on recent excavations at the shell midden of Havnø in Denmark. Mesolithic Miscellany 19, 3–6.

Biagi, P., Voytek, B., 2006. Excavations at Pestera Ungureasca (Caprelor) (Cheile Turzii, Petresti de Jos, Transylvania) 2003-2004: A preliminary report on the chipped stone assemblages from the Chalcolithic Toarte Pastilate (Bodrogkeresztur) layers. Analele Banatului 14, 177–202.

Billamboz, A., 2006. Dendroarchäologische Untersuchungen in den neolithischen Ufersiedlungen von Hornstaad-HörnleHornstaad-Hörnle 1A. Die Befunde Einer Jungneolithischen Pfahlbausiedlung Am Westlichen Bodensee. Stuttgart: Theiss 9, 23–232.

Campbell, K.P., MacLennan, D.H., Jorgensen, A.O., 1983. Staining of the Ca2+-binding proteins, calsequestrin, calmodulin, troponin C, and S-100, with the cationic carbocyanine dye “Stains-all.” The Journal of biological chemistry 258, 11267–11273.

Conchology, Inc [WWW Document], n.d. URL https://www.conchology.be/?t=1 (accessed 11.22.19).

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

Hughes, C.S., Moggridge, S., Müller, T., Sorensen, P.H., Morin, G.B., Krijgsveld, J., 2019. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nature protocols 14, 68–85.

Jamme, F., Kascakova, S., Villette, S., Allouche, F., Pallu, S., Rouam, V., Réfrégiers, M., 2013. Deep UV autofluorescence microscopy for cell biology and tissue histology. Biology of the cell / under the auspices of the European Cell Biology Organization 105, 277–288.

Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., Lajoie, G., 2003. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid communications in mass spectrometry: RCM 17, 2337–2342.

Mackie, M., Rüther, P., Samodova, D., Di Gianvincenzo, F., Granzotto, C., Lyon, D., Peggie, D.A., Howard, H., Harrison, L., Jensen, L.J., Olsen, J.V., Cappellini, E., 2018. Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. Angewandte Chemie 57, 7369–7374.

Marin, F., Muyzer, G., Dauphin, Y., 1994. Caractérisations électrophorétique et immunologique des matrices organiques solubles des tests de deux Bivalves Ptériomorphes actuels, *Pinna nobilis* L. et *Pinctada margaritifera* (L.). Comptes rendus de l’Academie des Sciences. Serie 2. Fascicule a. Sciences de la terre et des planetes/Earth and planetary sciences. Montrouge 318, 1653–1659.

Mészáros, B., Erdos, G., Dosztányi, Z., 2018. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic acids research 46, W329–W337.

Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.-Y., El-Gebali, S., Fraser, M.I., Gough, J., Haft, D.R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A.P., Paysan-Lafosse, T., Pesseat, S., Potter, S.C., Qureshi, M.A., Rawlings, N.D., Redaschi, N., Richardson, L.J., Rivoire, C., Salazar, G.A., Sangrador-Vegas, A., Sigrist, C.J.A., Sillitoe, I., Sutton, G.G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Yong, S.-Y., Finn, R.D., 2019. InterPro in 2019: improving coverage,

classification and access to protein sequence annotations. Nucleic acids research 47, D351–D360.

Morrissey, J.H., 1981. Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. Analytical biochemistry 117, 307–310.

Niedermeyer, T.H.J., Strohalm, M., 2012. mMass as a software tool for the annotation of cyclic peptide tandem mass spectra. PloS one 7, e44913.

Ramos-Silva, P., Benhamada, S., Le Roy, N., Marie, B., Guichard, N., Zanella-Cléon, I., Plasseraud, L., Corneillat, M., Alcaraz, G., Kaandorp, J., Marin, F., 2012. Novel molluskan biomineralization proteins retrieved from proteomics: a case study with Upsalin. Chembiochem: a European journal of chemical biology 13, 1067–1078.

Starnini, E., Biagi, P., Mazzucco, N., 2018. The beginning of the Neolithic in the Po Plain (northern Italy): Problems and perspectives. Quaternary international: the journal of the International Union for Quaternary Research 470, 301–317.

Takeuchi, T., Plasseraud, L., Ziegler-Devin, I., Brosse, N., Shinzato, C., Satoh, N., Marin, F., 2018. Biochemical characterization of the skeletal matrix of the massive coral, *Porites australiensis* - The saccharide moieties and their localization. Journal of structural biology 203, 219–229.

Appendix 2. Supplementary information.

The shell matrix of the European thorny oyster, *Spondylus gaederopus*: microstructural and molecular characterisation (Chapter 1)

SI.1. Matrix extraction

Shell matrix extraction was carried out from the upper red and lower white shell valves. Two different cleaning approaches, using ~1-1.5% NaOCl and named as 2BL and 3BL, were applied, as detailed in Figure 3 of the main manuscript. The obtained samples and their quantities are reported in Table SI.1.
After the "mild" 2BL treatment, we find that the organic matrix (ASM + AIM) constitutes about 1.43 - 2.14‰ of the shell powder. The ASM represents between 26% and 30% of the total mass of this matrix. The "harsh" 3BL treatment results in a more drastic reduction of the organics: after the extra bleaching treatment the organic matrix (ASM+AIM) constitutes about 0.24 - 0.035‰, with a pronounced reduction of organics observed in the AIM of the lower valve (sample S8).

Table SI.1: Shell organic matrix extracts and comparative yields (‰) using different pretreatments. All samples were bleached once to remove superficial contaminants, then bleached a second time (2BL) or twice (3BL). Bleaching procedure: 2BL – coarse powder bleached for 4 h; 3BL – fine-sieved powder bleached for additional 14 h. UV – upper valve, LV – lower valve. ASM – acid soluble matrix, AIM – acid insoluble matrix.

Sample name	Cleaning method	Shell part	Amount of shell powder used for extraction (g)	Organic fraction	Obtained organic matrix (mg)	Organics yield (‰)
S1	2BL	UV	15.89	ASM	8.83	0.56
S2				AIM	25.02	1.58
S3		LV	20.00	ASM	8.66	0.43
S4				AIM	20.21	1.0
S5	3BL	UV	17.17	ASM	0.64	0.04
S6				AIM	3.63	0.2
S7		LV	20.00	ASM	0.60	0.03
S8				AIM	0.09	0.005

SI. 2. ImG in situ localization of *Spondylus* shell proteins

Antibody selection

Immuno-gold localization (ImG) was used to localise shell proteins *in situ* in *Spondylus* shell sections. Antibody K5090, elicited against the soluble matrix of the nacreous layer of *Pinna nobilis,* was found to be the most cross-reactive with the *Spondylus* shell ASM matrix out of a set of 11 shell antibodies, which were available in the Biogéosciences lab and for which reactivity was assessed by enzyme-linked immunosorbent assay (ELISA) (Fig. SI.1). The graph in Fig. SI.1 shows the relative cross-reactivity of these 11 antibodies, expressed as a percentage (most reactive antibody = 100%). The highest cross-reactivity was observed with antibody K5090, and the second highest result was obtained with antibody α-Mytilus edulis, which was raised against the soluble matrix of *Mytilus edulis* (AB: α-Mytilus edulis).
Antibody K5090 was selected for further experiments, *i.e*. for staining *Spondylus* proteins on shell cross sections by immunogold (ImG) (Marin et al., 2007).

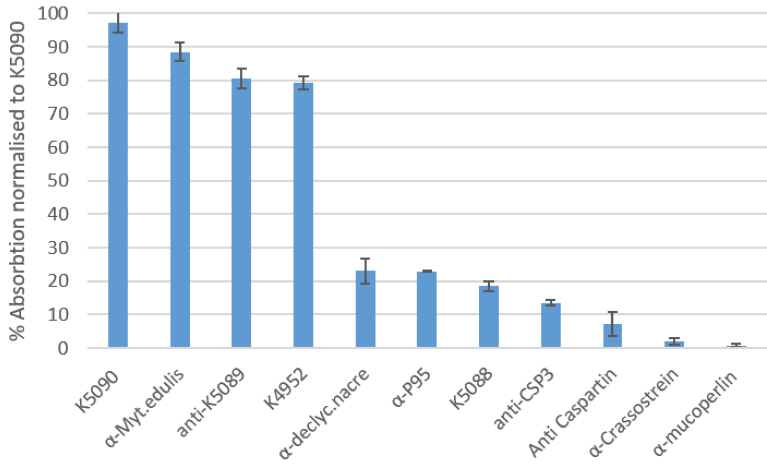


Figure SI.1: Cross-reactivity of the *Spondylus* ASM matrix and 11 available shell matrix antibodies, assessed by enzyme-linked immunosorbent assay (ELISA). Absorbance values at 405 nm were normalized to the highest value (K5090) corresponding to 100% reactivity (n = 4, means ± S.D).

Experimental procedure

Small fragments were sampled from the upper valve of the shell (closer to the margin). The preparations were carried out in two ways in order to obtain (i) fragments with exposed fresh fractures and (ii) small polished sections. For the first preparation (i) small cuts, parallel to the shell growth line, were made using a

diamond saw and tiny fragments were broken off using pliers in order to obtain a natural fracture, perpendicular to the shell growth line. For polished sections (ii), a transverse section, perpendicular to the growth line, was cut using a diamond saw; the section was polished using fine-grain sandpaper and alumina suspension (0.05 μm), cleaned rigorously and cut into small square fragments.

Both types of fragments were cleaned with H_2O in an ultrasound bath (several minutes); the polished sections were also bleached for 10 min (using NaOCl ~1 - 1.5% active chlorine) and rinsed with ultrapure H_2O. All fragments were etched with 1% EDTA in an ultrasound bath (5 min) and rinsed again with ultrapure H_2O. The rest of the analytical procedure follows previously published methods (Marin et al., 2007). Briefly, the fragments were incubated in blocking solution (1% gelatin in TBS/Tween-20) for 1 hour following an overnight incubation with K5090 solution (prepared in a 0.1% gelatin TBS/Tween-20 solution, diluted to a ratio 1:500; a $NaNa_3$ solution was also added at ratio 1/10000 to prevent bacterial growth). All incubations were performed at room temperature and while maintaining gentle shaking. The fragments were cleaned with TBS/Tween-20 (incubating for 10 min, with 6 changes) and incubated with the secondary antibody Goat-Anti Rabbit Immunogold ImG coupled with 5 nm gold nanoparticles (Sigma) for 2 h. Fragments were cleaned again as previously, dried and enhanced with silver solution (British Biocell International, Ref. SEKL15) for about 10-15 min. Finally, they were rinsed with water and dried at 37 °C prior to SEM observation.

In parallel, 3 negative controls were prepared using the same conditions but omitting some of the steps: 1) incubation with K5090; 2) incubation with secondary antibody; 3) both incubations with K5090 and the secondary antibody. A total of 4 replicas of the experiment were carried out in order to validate the observations.

Results

Figure SI.2 shows the SEM images of the *Spondylus* shell cross sections stained with shell antibody K5090. Cross-reactive *Spondylus* shell matrix proteins are present in the upper foliated calcitic layer and are absent in the aragonitic crossed-lamellar. A thicker layer of organics is observed at the structural transition, between the foliated calcitic and aragonitic crossed-lamellar layers. This finding, which has also been confirmed by in situ localization of sugar moieties (Fig. 9, main manuscript), suggests that glycosides and proteins likely co-exist in the same layers and may have a mutual role in the biomineralization of these structures.

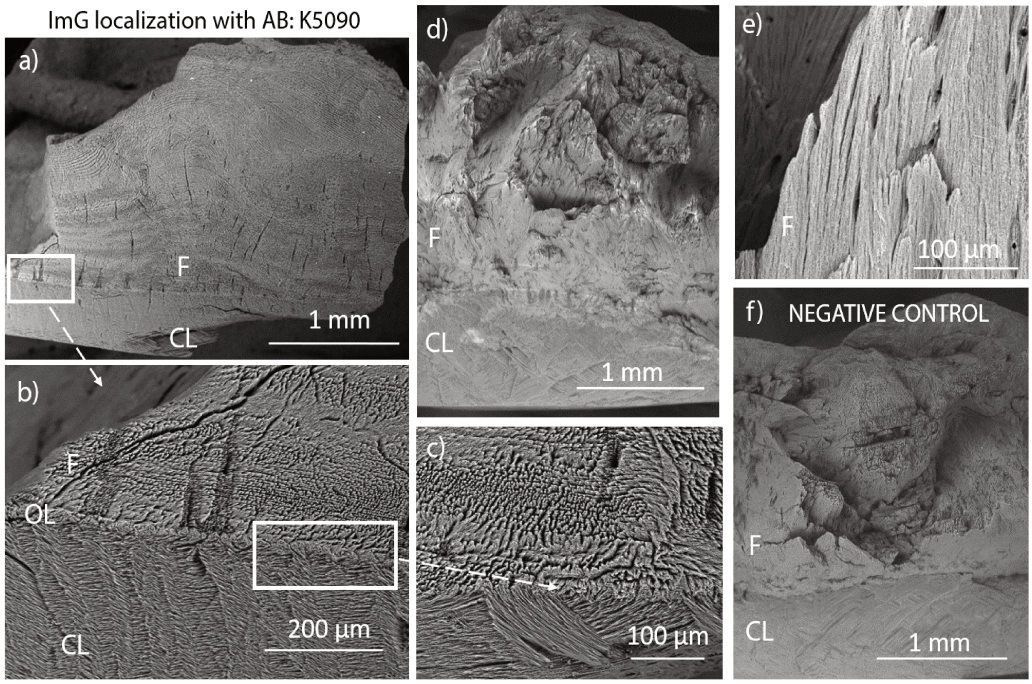


Figure SI.2: In situ ImG localization of *Spondylus* shell matrix proteins with antibody K5090. The test was carried out on fresh fractures and polished sections, sampled in the near margin region of the shell. Images a-e) correspond to positive controls; f) negative control. F - foliated calcite layer; CL - aragonitic crossed-lamellar; OL - organic layer.

SI. 3. Proteomics

Protein identification

The most abundant protein identified in *Spondylus* shell organic matrices was the uncharacterized protein LOC110461617 [*Mizuhopecten yessoensis*]. It was identified using PEAKS Studio 8.5 software and searching product ion spectra against a database of molluscan proteins obtained from the NCBI repository. In sample S8, this protein attained a maximum of 9% of coverage (highest among all of the samples) but the identification is also supported by a high number of *de novo* tags (243) that were sharing at least 6 amino acids (i.e. with sequences partially matching that of the reference *Mizuhopecten* protein). The full protein sequence and its coverage in sample S8 is presented in figure SI.3

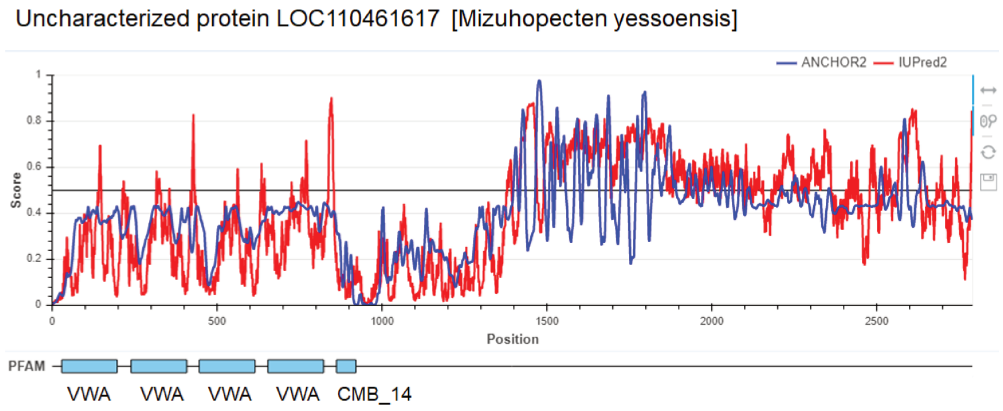

Acetylation (K) (+42.01)
Carbamidomethylation (+57.02)
Deamidation (NQ) (+0.98)
Formylation (+27.99)
Hexose (NSY) (+162.05)
Methylation(KR) (+14.02)
Oxidation (M) (+15.99)
Pyro-glu from Q (-17.03)
Formylation (+27.99), Deamidation (NQ) (+0.98)
K->N
S->A
S->G
1 MKLPISLVLV ALMAHNAYQQ VLNCYVPADI MILIDGSDSI QDQNWREIKN FVSQLVTNFD
61 VGRDAIHVGF VVYSSDVGDH IGLQPFKPKN VLRTLSGILR QPKASTNTAK GIAYVRNEFK
121 TRGRAGVPKI LIVITDGSSD NPRETRKQAN IAKIEGIRVI AVGIGQTFRD ELRQIASRPE
181 KVYTAASFAT LQTLVFEIQR MVCTVITTTT TTTTAIPIPT KPPVVIPVPS DKICDVPGDI
241 VFVMDGSDSI DDADFIRQKQ FVANLIDNFE ISTDAIHVGL IVYSTIIGDT VGLQPPKNKE
301 LLKILARNLR HPKVGTNTAR GIERARDMIR KEGRAMAPKM IVVITDGRSS SPKLTVAQAN
361 MAKVEGITLV SVGVGTQIFS DELNQIASSR RKVFEVSDFR SLELIITSMR DLLCQAITTT
421 TSTTTTTQKP TYIPPPDDLF CNVPADVGIL LDGSDSIADA DWTKQKYFIA SLINNLAVGR
481 DTIHVGVVVF STIIGETVGL TPFKPKELLM ILSNNLKQPK VGTNTALGIQ RMRQMLASQG
541 RANAPKVMII VTDGKSSSPK RTVAQASLAK KEGITVIAVG VGSQLFREEL SQIATNDRKL
601 FTVSDFQGLQ QIIVTLRNLI CQVITTTSTT QLVTTPIPPP VTLQPYPGCE VPAEIMFMMH
661 GSDDIKDDNW GMNKKFVSSL VRNLYVNRGA HHVGMVVYSS SVGDHMSLQP FKEVFDVQNH
721 INSFSHPRGG ANTAAALAKI REAFRTHARG GAPKIGVLIT DQKSTTPADT KLQAELAKNE
781 GVQFVTLGIG SKVNIGELAN IATDSQKLFR AMNFKRLESM MPRIRDMICE IVTTPVTTTT
841 TTTTTTPPPP VSTPAPDFTQ LCAGCLMNSD SGFNPYPGDC TRYVQCWRDG GRIIGAVKVC
901 PYGQFWDEDA ITCRPSMNVM CPQDPCINAP DGFTLGIADH GCRSHWVCVK GYSVASCCPE
961 GTYYVEGMGC MRGMACKMPC PPDGGLIISP NCDKEVHWDG RFYLENVPGK GKMVRRCAPG
1021 TIFDKGYCTC VIATSNYVPV TPTDKCMESV HLNFDVNMKD KSGKSTQINN FAVRKTKYGT
1081 AHFNGNSYLN LWNLGNIDYT DKFALKFKFK FDRWENRFDM GGYPVDAGWK GDFLWKWDSQ
1141 TGWNRNSTVG GSWKVVTLMN MFKRMFAATS ENEMETIMVE MATSPDFQIL LYHLGFMPQS
1201 TNAIQILRQL MSGNGAVFIR QIIRQIGNMQ GGVKTRRLLL QLLMMREVDD WVRKHSGIHV

Figure SI.3: Uncharacterized protein LOC110461617 [Mizuhopecten yessoensis] sequence identified in *Spondylus* shell matrix sample S8. Blue bars indicate the identified peptides fully matching the protein sequence. Grey bars correspond to the identified *de novo* tags sharing at least 6 amino acids. Where detected, post translational modifications (PTMs) and substitutions are indicated. Sequences reconstructed by assisted *de novo* on the basis of mainly mono-charged ions (spectra were acquired on the 400-1600 m/z range and multiply-charged ions were detected); *de novo* sequences scores (ALC%) > 50.

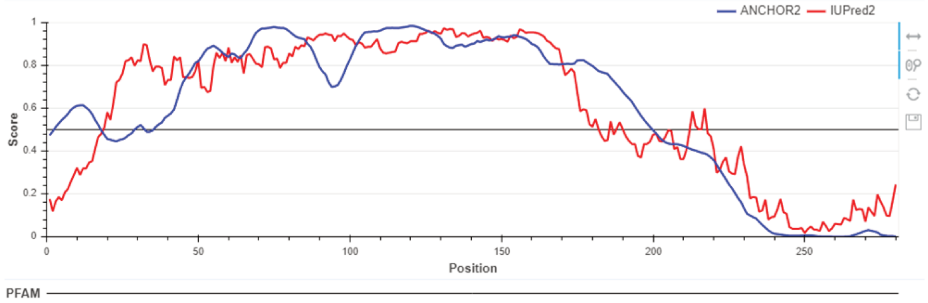
Protein characterization. Intrinsically disordered proteins (IDP)

A number of studies have reported the abundance of intrinsically disordered proteins, which are related to biomineralization (Evans, 2012; Kalmar et al., 2012; Boskey and Villarreal-Ramirez, 2016; Marin et al., 2016) in that they play an important role in mediating the crystal nucleation of calcium carbonate, predominantly due to their flexibility and great binding capacity of calcium ions and other organic/inorganic components.

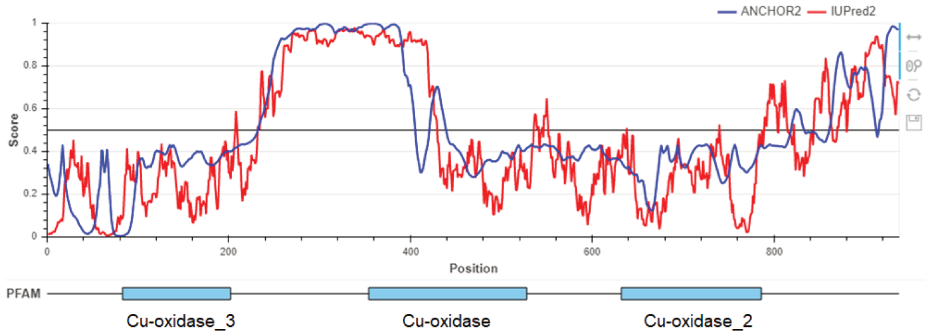
The IUPred2A tool (Mészáros et al., 2018) enables to identify intrinsically disordered protein regions (IDPRs) and was used to assess the presence of these regions in proteins identified in *Spondylus* shell matrices. The IUPred2A interface integrates two methods that are based on energy estimation: the predictions by IUPred2 for ordered and disordered residues and predictions by ANCHOR2 for disordered binding regions. The analysis has been carried out using the default parameters (prediction type: IUPred2 for long disorder and context-dependent prediction ANCHOR2) and the results for all six shell proteins are presented in Figure SI.4. The graph shows the disorder tendency of each residue in all of the analysed proteins, where higher score values (y axes) correspond to a higher probability of disorder (in a scale from 0 to 1). In the case of ANCHOR2, the blue line indicates the probability of a binding region. On the bottom of the graph, PFAM regions (protein family regions, domains), if present, are displayed in all of the analysed sequences. All of the 6 proteins identified in *Spondylus* shell matrix have multiple regions with a very high probability of IDRs.



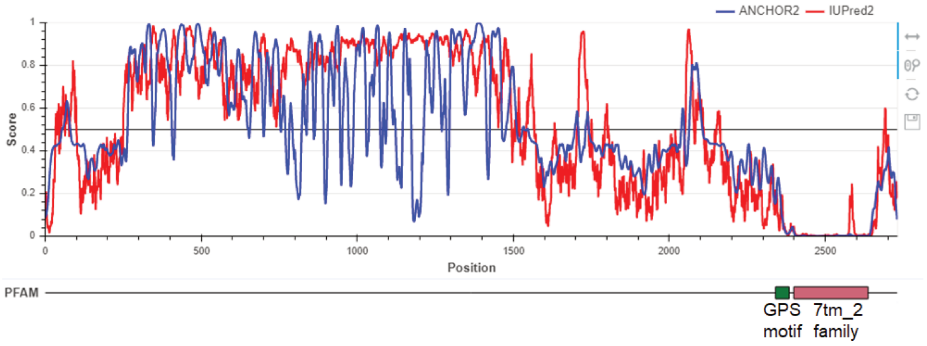
Uncharacterized protein LOC110452388 [Mizuhopecten yessoensis]



Laccase [Mizuhopecten yessoensis]



G-protein coupled receptor 112 [Mizuhopecten yessoensis]



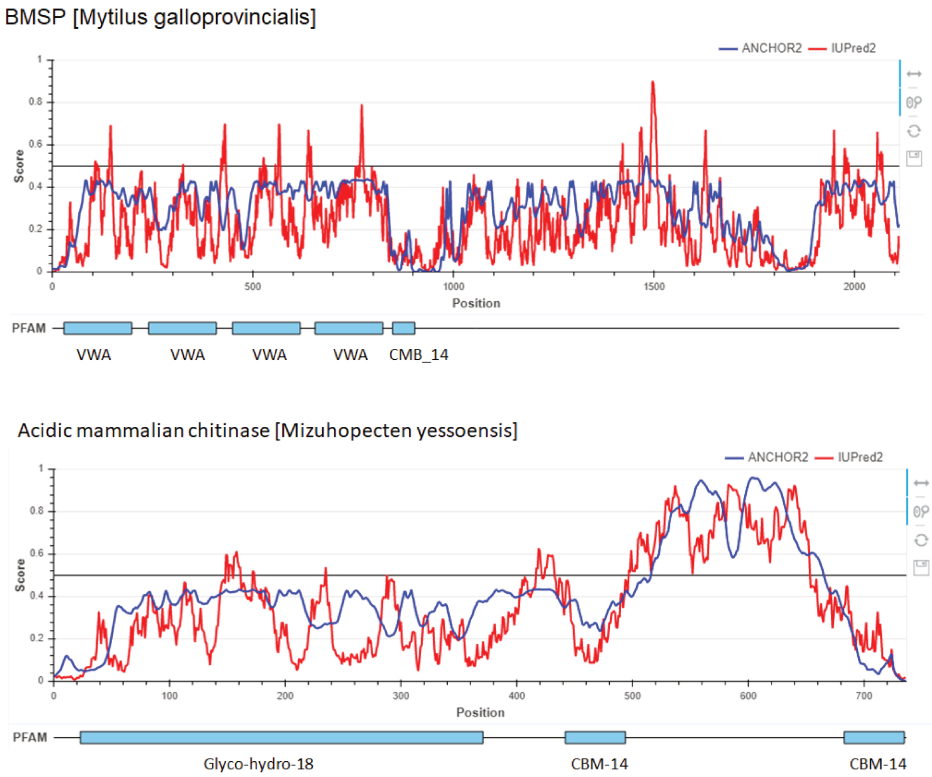


Figure SI.4: Prediction of intrinsically disordered structures in the six shell proteins that have been identified in *Spondylus* shell skeletal matrices. Analysis has been carried out using IUPred2A software which includes the predictions made by IUPred2 and ANCHOR2 tools. Graph shows the disorder tendency of the residues where higher score (y axis) corresponds to a higher probability of disorder. The graph of each protein also indicates PFAM (protein families) if those are identified.

SI.4. The intra- and intercrystalline *Spondylus* shell matrices

The isolation of an "operational" intracrystalline fraction of organics in shells is usually carried out using concentrated NaOCl (12% w/v) and exposure times of 48 hours or more (Penkman et al., 2008). Thus, we aimed to verify the 'level of intra- or inter- crystallinity' of the shell matrices that were obtained using the 2BL bleaching approach (see Fig. 3, main manuscript).

Experimental procedure

Enzyme-linked immunoassay (ELLA) was carried out to compare the signal intensity of organic extracts (carbohydrates and glycoproteins) from shell powders which had been bleached at different levels. Thus, two sets of shell powders were used for matrix extraction:

1) powders treated with the 2BL bleaching approach, as indicated in the matrix extraction section (4 h bleach on coarsely crushed sample with ~1-1.5% (w/v) NaOCl);
2) powders treated by the “classic” bleaching approach (48 h with concentrated NaOCl (12% w/v).
The shell organic matrices were extracted with EDTA (20%) and the solutions were directly deposited on 96-well microplates; the assay was carried out as previously described (section Material and Methods). The ELLA analysis was carried out with the lectin jacalin, which was found to be the most reactive lectin towards *Spondylus* organic shell matrices (carbohydrates and glycoproteins). Figure SI.5 shows the relative intensity of cross-reactivity of the two shell extracts and the values are expressed as a percentage, the most reactive extract giving 100%.

Results

We observe that the ELLA signal of the “classic” intracrystalline shell matrix (IntraC_BL) is only slightly lower than the one observed from the 2BL matrix extracts (by ～10%). This indicates that most of the organics which recognise the lectin jacalin and which are observed after the 2BL bleaching treatment (a rather “mild” cleaning approach), remain also stable even after the strong “classic” bleaching treatment. This also suggests that the other cleaning approach used in our study, the intermediate 3BL bleaching, is able to isolate a fraction of organics that is in all likelihood intracrystalline.

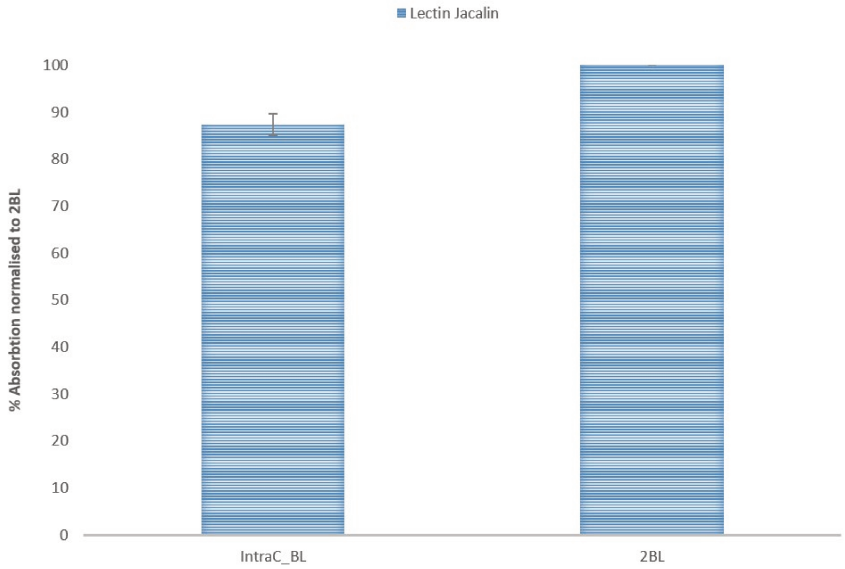


Figure SI.5.: Enzyme-linked lectin assay (ELLA) on the different *Spondylus* shell organic fractions using lectin jacalin: comparison of the signal intensity between the “classic”

intracrystalline fraction (IntraC_BL) and the shell organic matrix extracted by the mild 2BL treatment. The 2BL fraction refers to the isolation of organics by bleaching coarse shell powder with diluted NaOCl (~1-1.5%) for 4 h and the IntraC_BL by exposing to NaOCl (12 % w/v) for 48 h. All powders were demineralised with EDTA (20%). Absorbance values at 405 nm were normalized to the highest value (2BL) corresponding to 100% reactivity (n = 4, means ± S.D).

Table SI.1. Full list of identified protein in *Spondylus* shell. The presented sequences are identified with a minimum of 1 unique peptide in at least one of the eight matrices (S1-S8).

Appendix 2. Table SI.1_Spondylus protei

Table SI.1. *Spondylus* shell peptidomes. The list includes all the *de novo* sequences identified in the eight shell matrices (S1-S8). The d*e novo* peptides are constructed from product ion spectra by software PEAKS.

Appendix 2. Table SI.2_Peptidomes.xlsx

References

Boskey, A.L., Villarreal-Ramirez, E., 2016. Intrinsically disordered proteins and biomineralization. Matrix biology: journal of the International Society for Matrix Biology 52-54, 43–59.

Evans, J.S., 2012. Aragonite-associated biomineralization proteins are disordered and contain interactive motifs. Bioinformatics 28, 3182–3185.

Kalmar, L., Homola, D., Varga, G., Tompa, P., 2012. Structural disorder in proteins brings order to crystal growth in biomineralization. Bone 51, 528–534.

Marin, F., Pokroy, B., Luquet, G., Layrolle, P., De Groot, K., 2007. Protein mapping of calcium carbonate biominerals by immunogold. Biomaterials 28, 2368–2377.

Marin, F., Bundeleva, I., Takeuchi, T., Immel, F., Medakovic, D., 2016. Organic matrices in metazoan calcium carbonate skeletons: Composition, functions, evolution. Journal of structural biology 196, 98–106.

Mészáros, B., Erdos, G., Dosztányi, Z., 2018. IUPred2A: context-dependent

prediction of protein disorder as a function of redox state and protein binding. Nucleic acids research 46, W329–W337.
Penkman, K.E.H., Kaufman, D.S., Maddy, D., Collins, M.J., 2008. Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells. Quaternary geochronology 3, 2–25.

Appendix 3. Supplementary information.

The diagenesis of intracrystalline *Spondylus* shell proteins: a proteomics study

SI.1 *Spondylus* shell matrix

Spondylus shell matrix was tested on Western blots with lectin jacalin and shell antibody K5090 to determine the cross-reactive groups. In this experiment, we used the (inter+intra)crystalline (IIF) *Spondylus* matrix because the quantity of the intracrystalline matrix (IcF) was too low and because ELLA and ELISA analyses confirmed the cross-reactivity of the IcF *Spondylus* matrix being very similar to that of IIF. Acid soluble (ASM) and acid insoluble matrices (AIM) were analysed separately. We note that in western blots, due to poor solubility of AIMs, this fraction actually corresponds to the Laemmli-soluble fraction (LS-AIM).

Westernblots showed that lectin jacalin reacts with a thick discrete band located around 30-34 kDa in the ASM extract and also gave a strong, slightly polydisperse signal at very high molecular weight with bands in the LS-AIMs. This indicates that lectin jacalin recognises acid soluble glycoproteins and high molecular weight, insoluble carbohydrate structures.

Antibody K5090, elicited against the soluble matrix of *Pinna nobilis* shell, reacted with a protein band located around ~40 kDa in the ASM and also showed some cross reactivity signal, albeit weak, with high molecular weight components in the extracts of LC-AIMs. This implies that shell antibody K5090, in *Spondylus* extracts, recognises acid soluble proteins as well as insoluble, high molecular weight proteinaceous compounds.

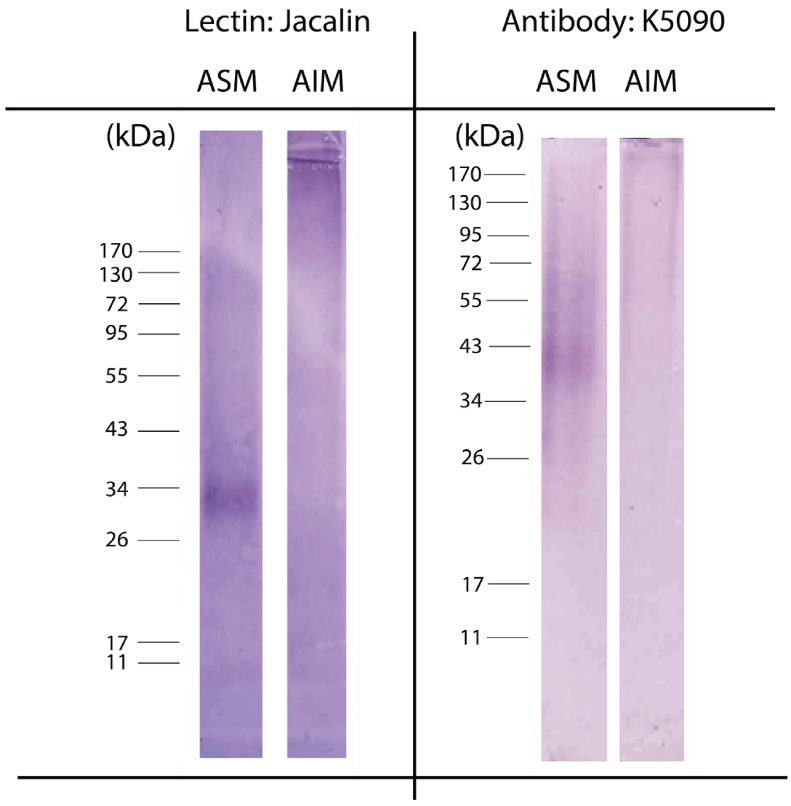


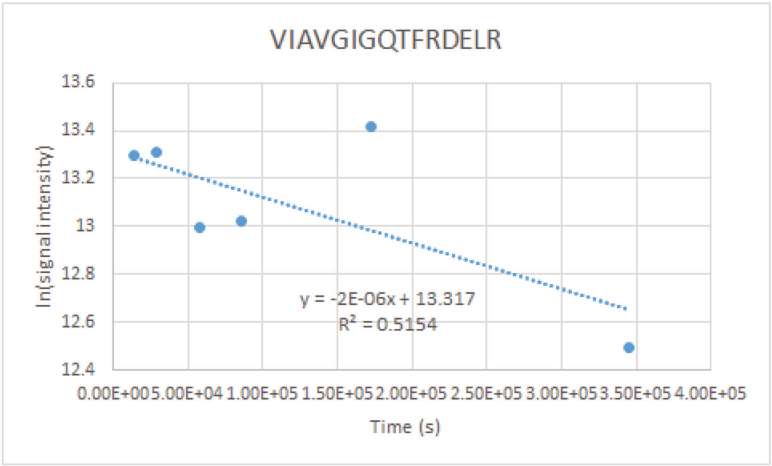
Figure SI.1 Western blots on *Spondylus* shell (inter+intra)crystalline organic matrix (IIF) with lectin jacalin and shell antibody K5090. ASM corresponds to acid soluble matrix and AIM to acid insoluble matrix.

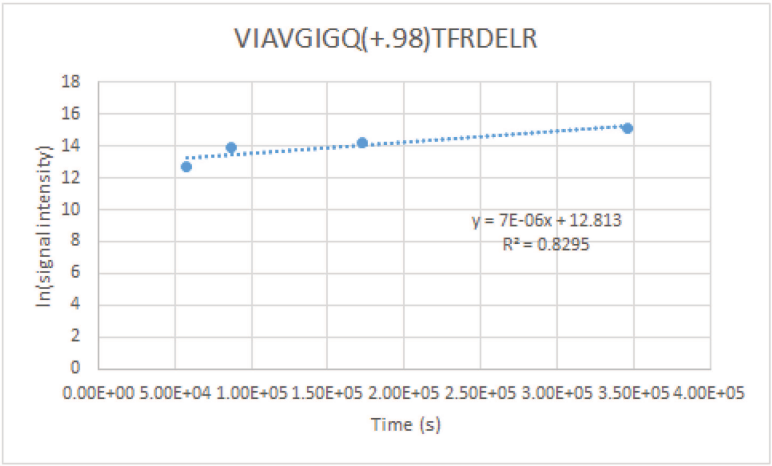
SI2. Protein degradation. Kinetics

Deamidation kinetics

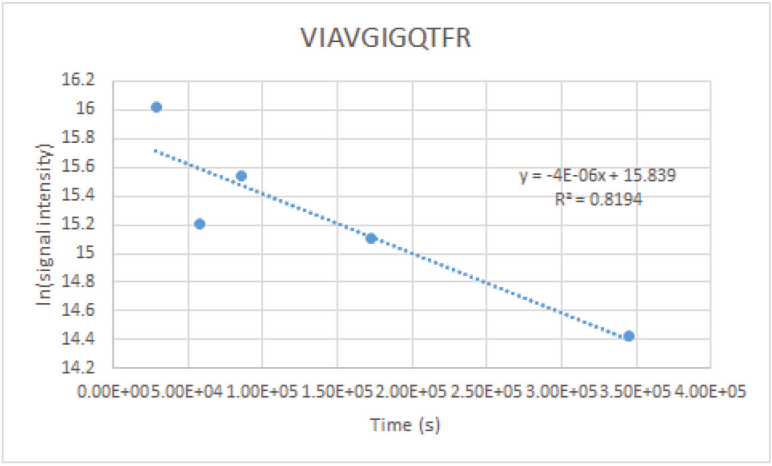
As detailed in the main manuscript, several peptides belonging to uncharacterized protein LOC117318053 [*Pecten maximus*] were identified with either N/Q deamidated or unmodified residues. Hence, we explored the possibility to apply a first order reaction rate kinetics model to find out if linear regression can be fitted to the observed deamidation reactions and obtain their kinetic rates. The fittings are displayed in the graphs below. Peptides are numbered as in Table 4 of the main manuscript. The reactions which showed linear regression fitting higher than $R^2 > 0.9$ were used to determine apparent reaction rate constants and are presented in the main manuscript. The majority of observed reactions did not show a sufficient linear regression fit.

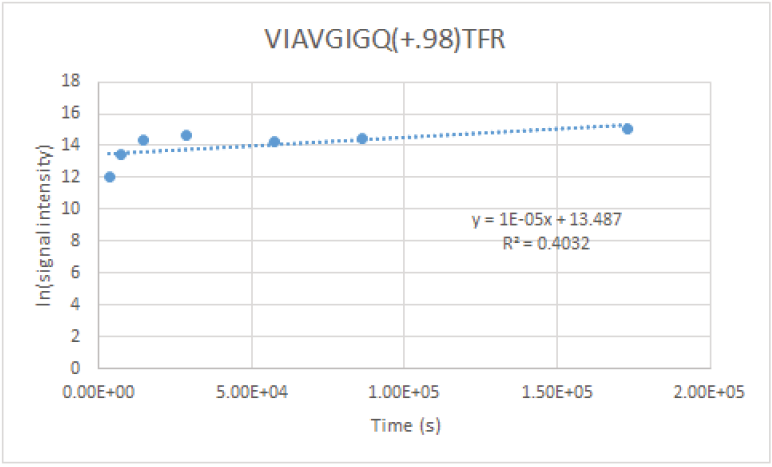
Peptides 1, 2. Glutamine (Q) deamidation.



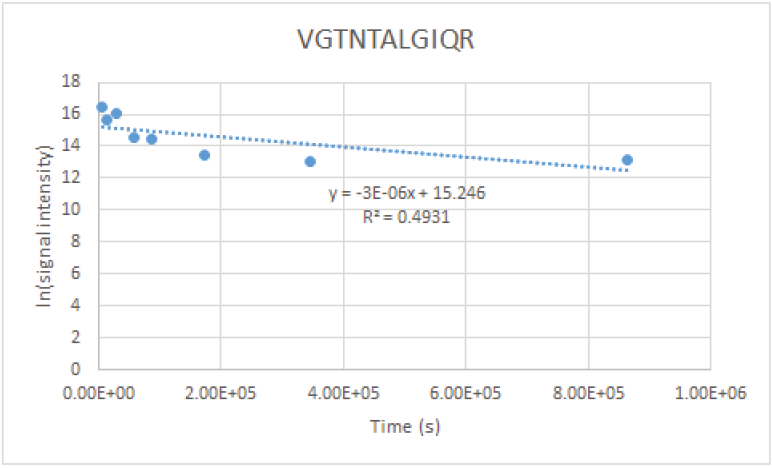


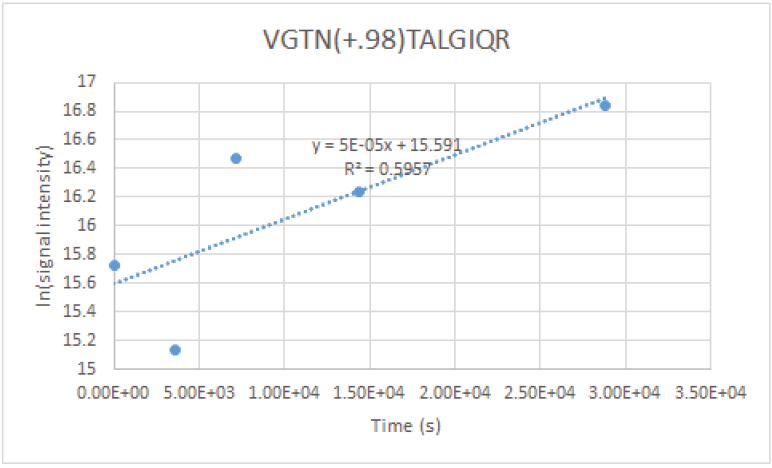
Peptides 3, 4. Glutamine (Q) deamidation.



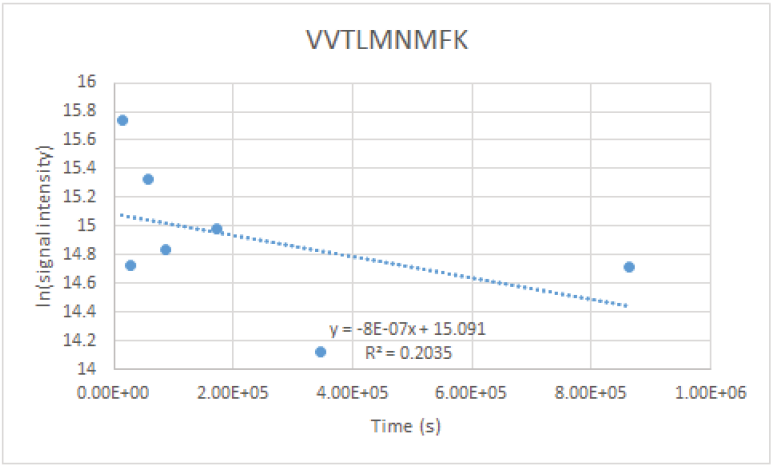


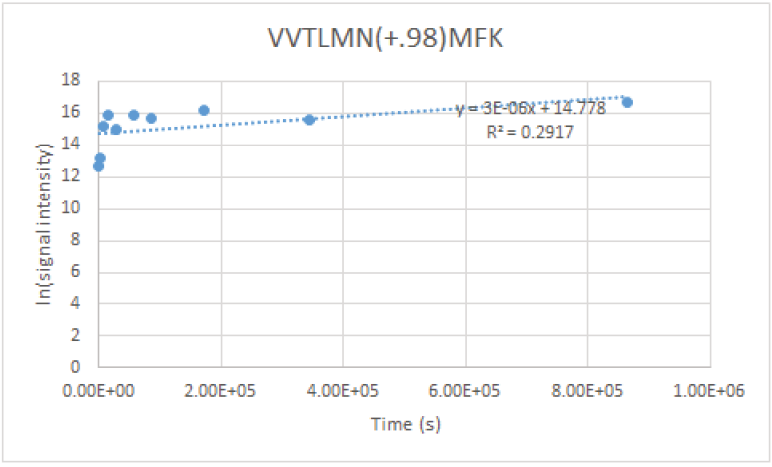
Peptides 15, 16. Asparagine (N) deamidation.





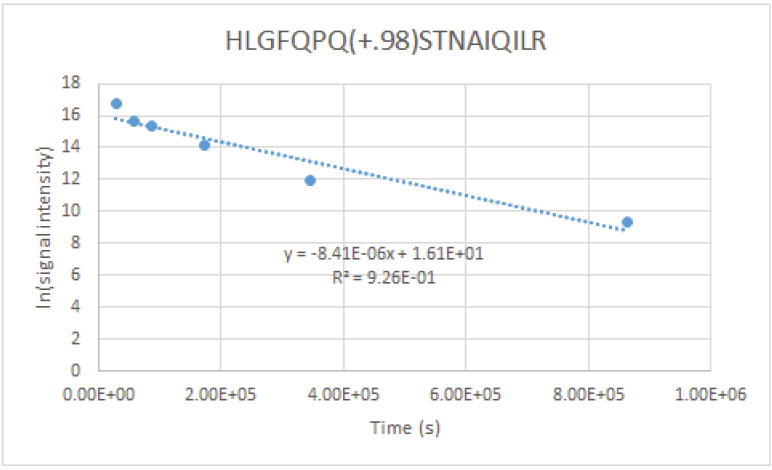
Peptides 19, 20. Asparagine (N) deamidation.

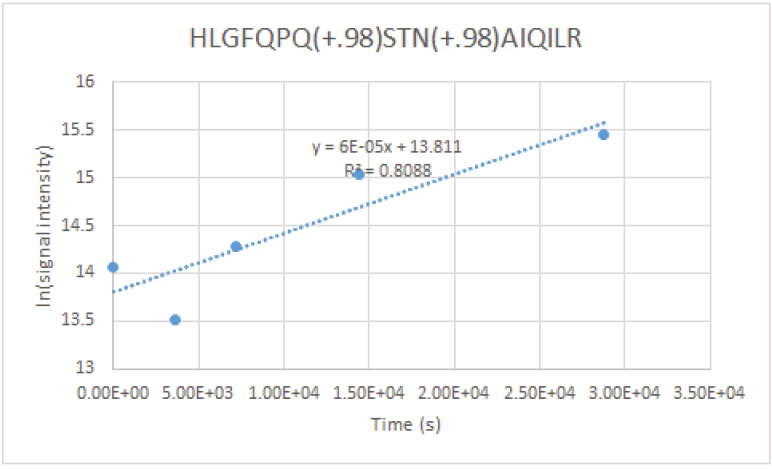




Peptides 24, 25. Asparagine (N) deamidation* (*alternative/additional reactions may be involved).

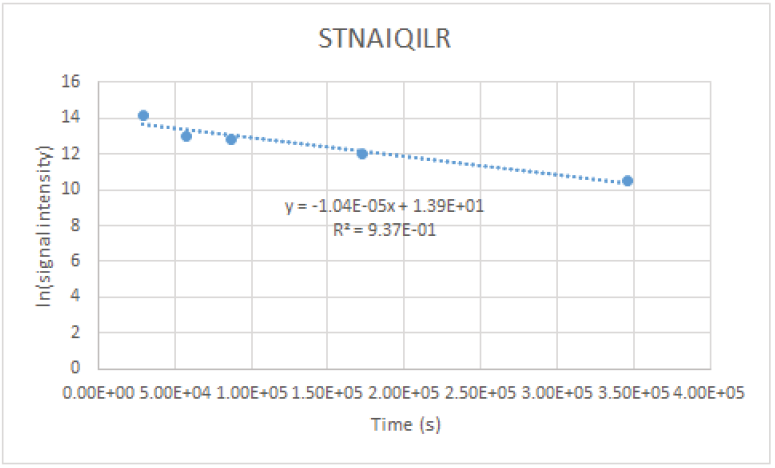
Peptide 24.
Reaction: deamidation of N or hydrolysis Q(+.98)-S.
Linear fitting: R^2 = 0.9257
k_{obs} = 8.41E-06

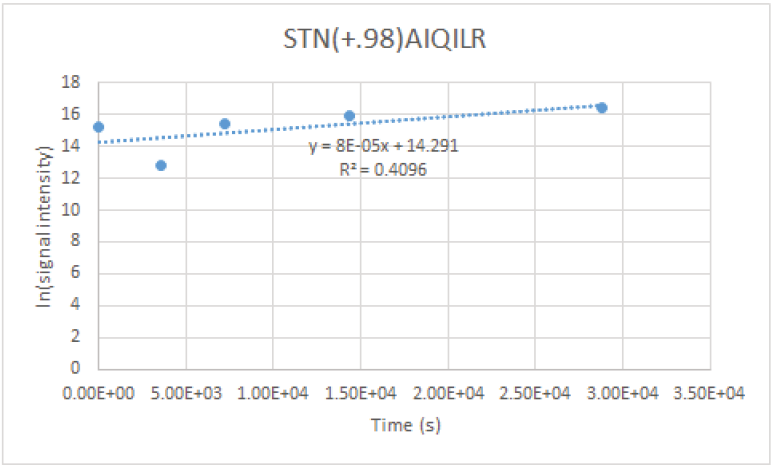


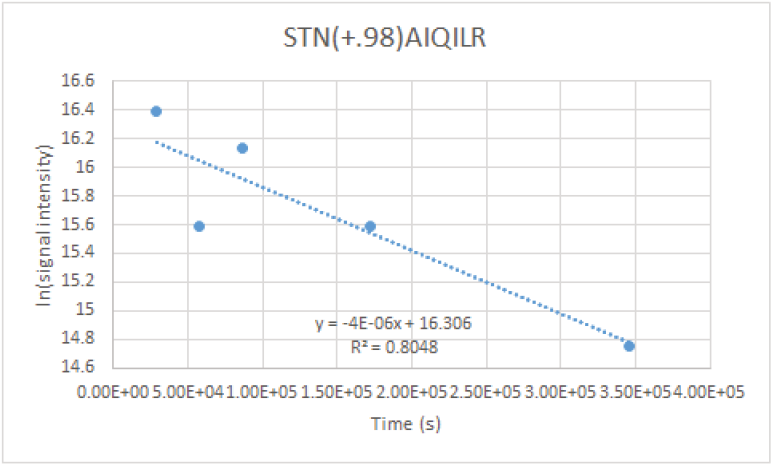


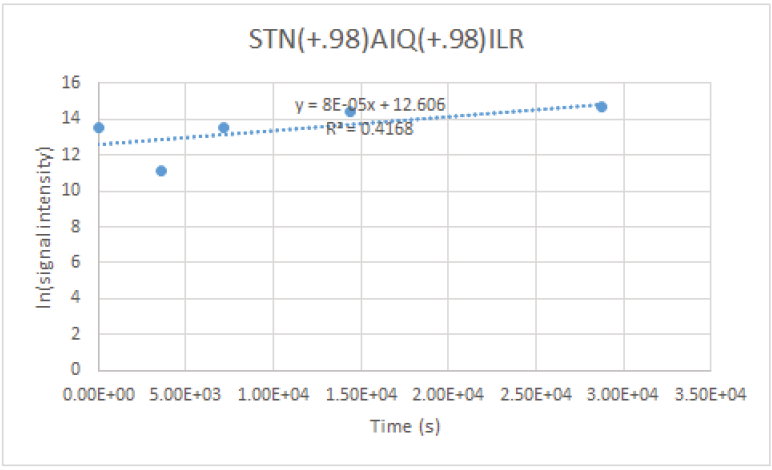
Peptides 26, 27, 28. Glutamine (Q) and asparagine (N) deamidation.

Peptide 26.
Reaction: deamidation of N.
Linear fitting: $R^2 = 0.9365$
$k_{obs} = 1.04E-05$







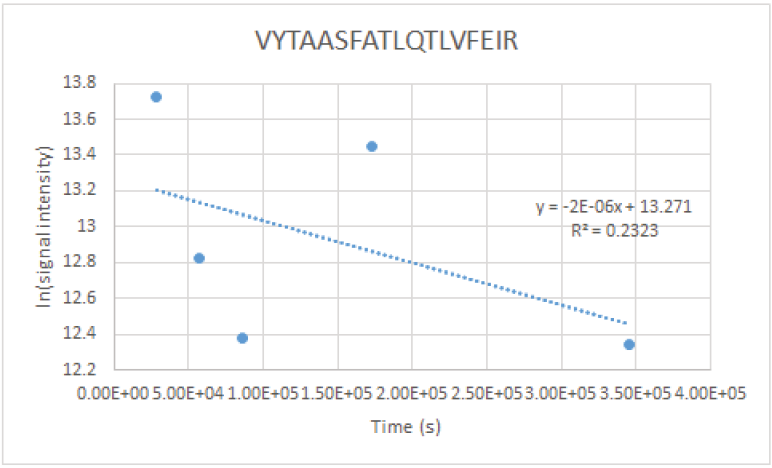


Kinetics of peptide bond hydrolysis

As detailed in the main manuscript, several peptides belonging to uncharacterized protein LOC117318053 [*Pecten maximus*] were identified with sequences of different lengths that indicate peptide bond hydrolysis due to diagenesis. Hence, we explored the possibility to apply a first order reaction rate kinetics model to find out if linear regression can be fitted to the observed hydrolysis reactions and obtain their kinetic rates. The fittings are displayed in the graphs below. Peptides and sequence segments are numbered as in Table 4 of the main manuscript. The reactions which showed linear regression fitting higher than $R^2 > 0.9$ were used to determine apparent reaction rate constants and are presented in the main manuscript. The majority of observed reactions did not show a sufficient linear regression fit.

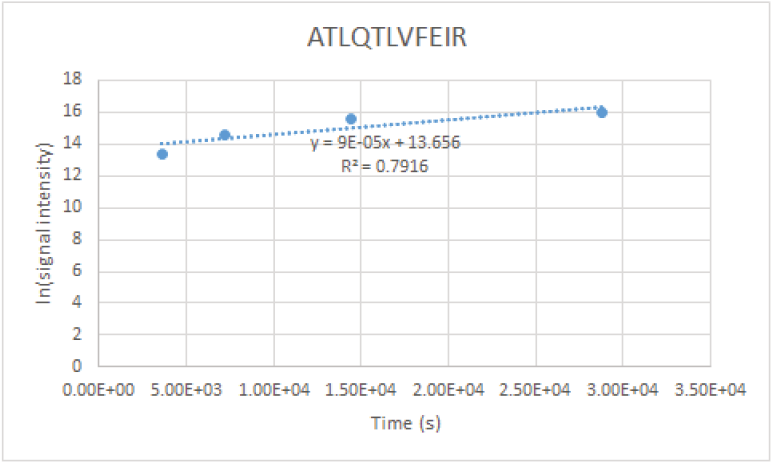
Sequence segment DC1.

Peptide no. 5. Hydrolysis between the F-A bond.

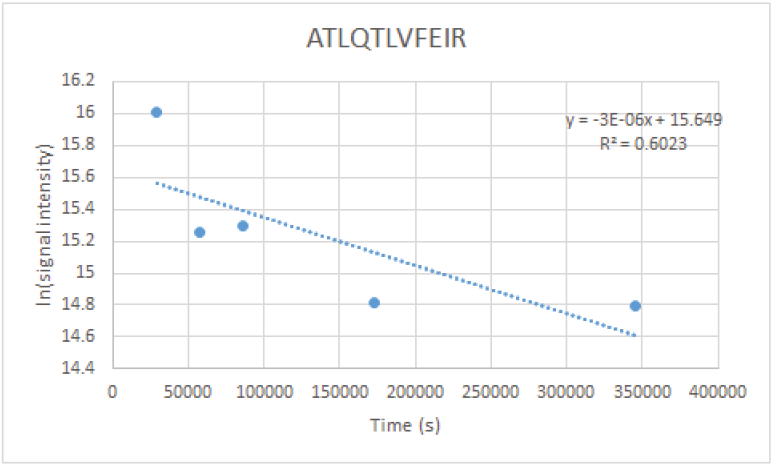


Peptide 6. Product formation due to the F-A bond hydrolysis*.

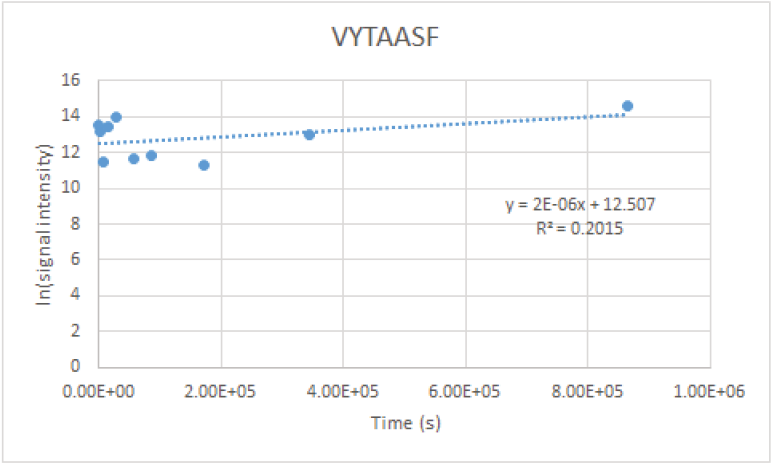
(* intermediate reactions may be involved)



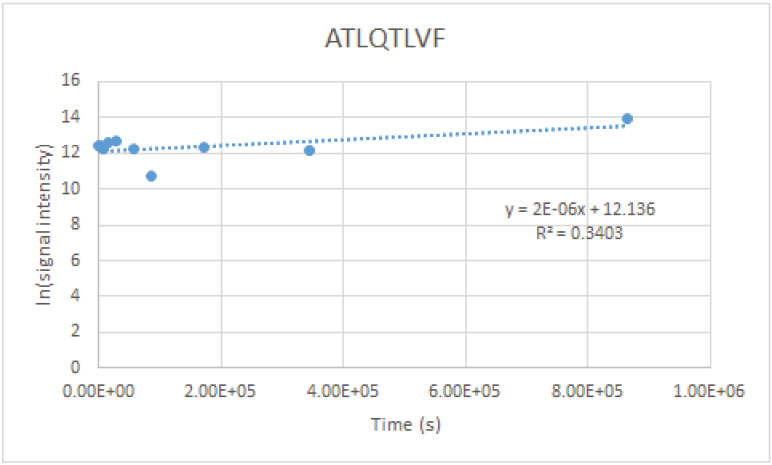
Peptide 6. Hydrolysis between the F-E bond*.
(* intermediate reactions may be involved)



Peptide 7. Product formation due to the F-A bond hydrolysis.



Peptide 8. Product formation due to the F-E bond hydrolysis.

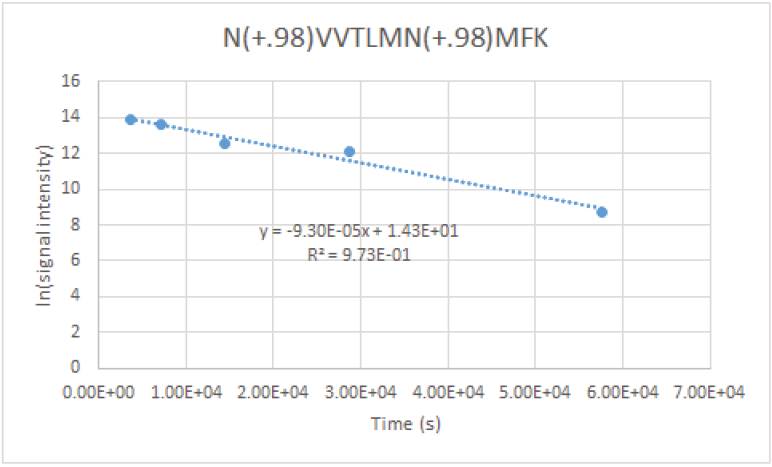


<u>Sequence segment DC2.</u>

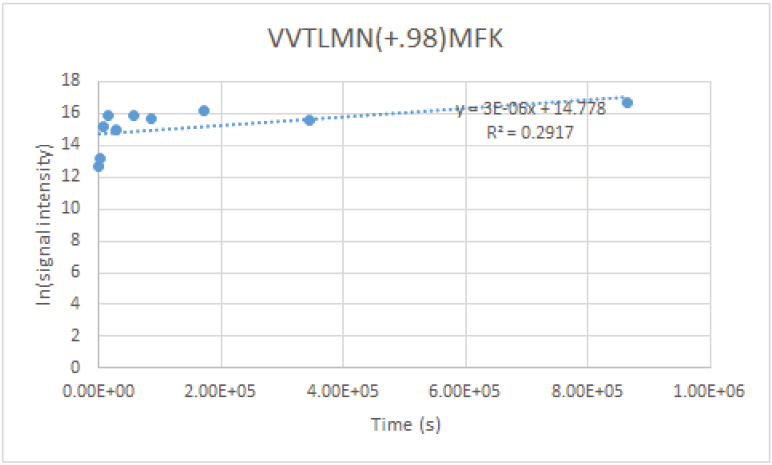
Peptide 18. Hydrolysis between the N(+.98)-V bond.

Linear fitting: $R^2 = 0.9734$

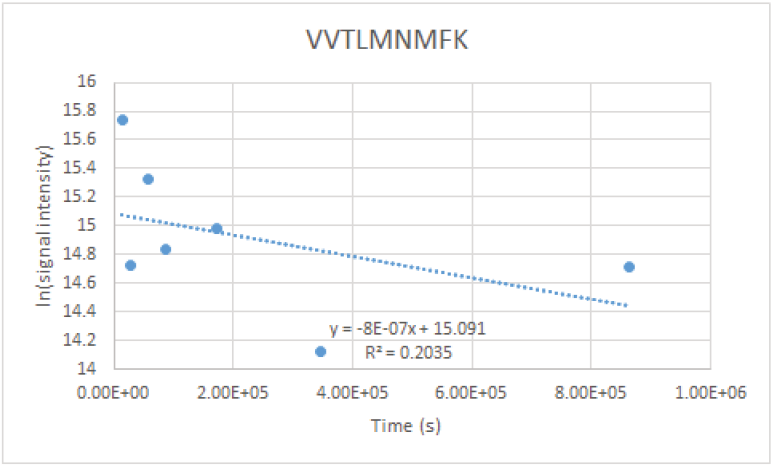
$k_{obs} = 9.30E-05$



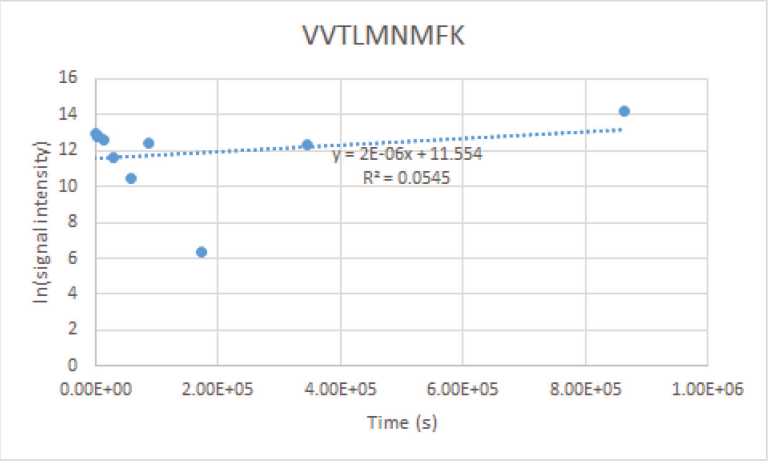
Peptide 20. Product formation due to the N(+.98)-V bond hydrolysis.



Peptide 19. Hydrolysis between the F-K bond*.
(* intermediate reactions may be involved)

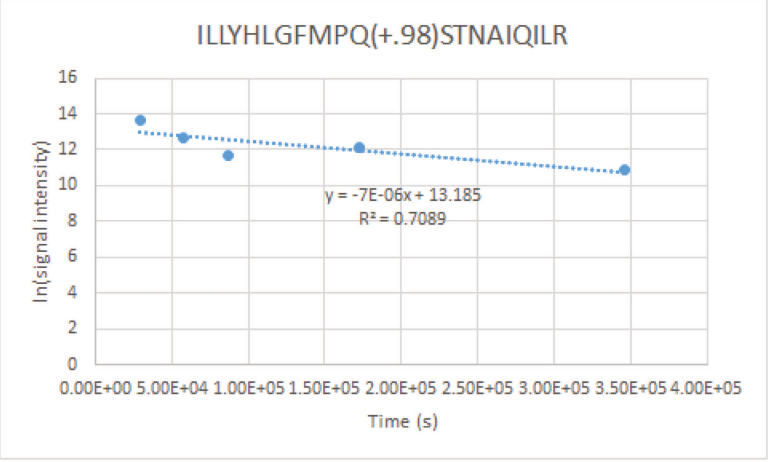


Peptide 21. Product formation due to the F-K bond hydrolysis.

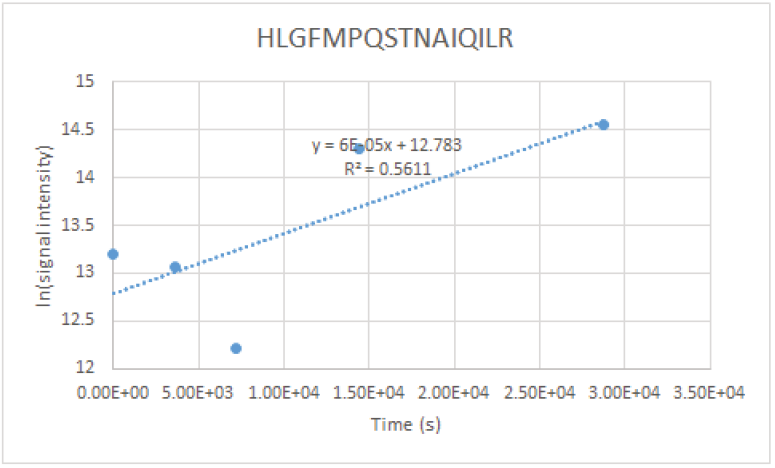


Sequence segment DC3a and DC3.

Peptide 22. Hydrolysis between the Y-H bond.

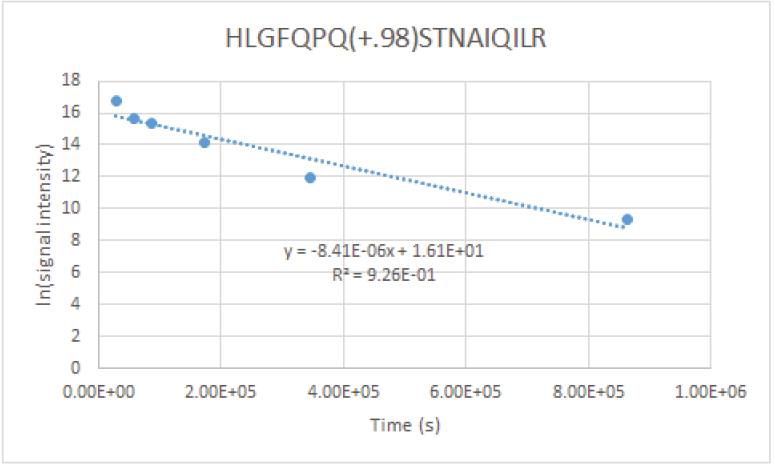


Peptide 23. Product formation due to the Y-H bond hydrolysis*.
(* intermediate reactions may be involved)

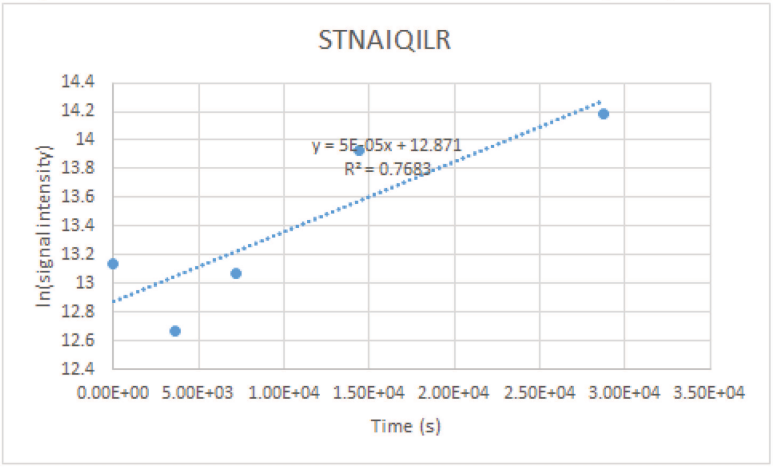


Peptide 24. Hydrolysis between the Q(+.98)-S bond.
(*alternative/additional deamidation reaction may be involved)

Hydrolysis Q(+.98)-S or deamidation of Asn
Linear fitting: $R^2 = 0.9257$
$k_{obs} = 8.41E-06$

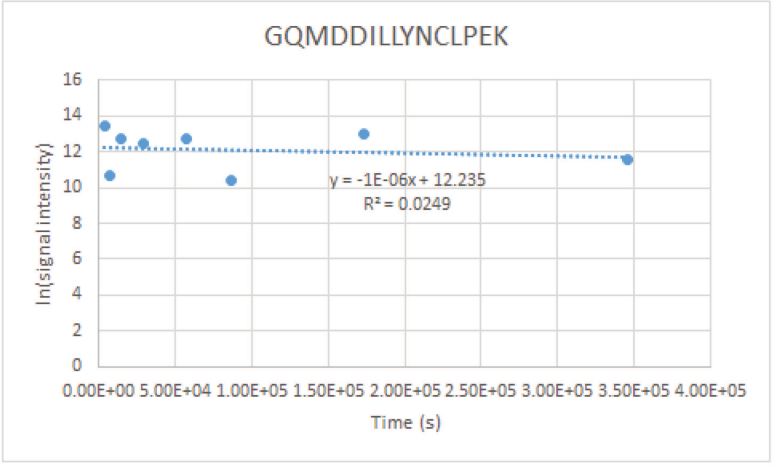


Peptide 26. Product formation due to the Q(+.98)S bond hydrolysis.

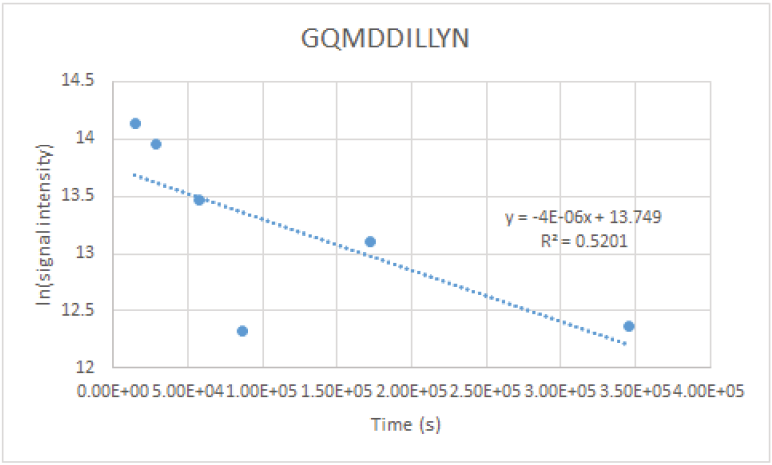


Sequence segment DC4

Peptide 30. Hydrolysis between the N-C bond.



Peptide 31. Hydrolysis between the Y-N bond.



Peptide 32. Product formation due to the Y-N bond hydrolysis.

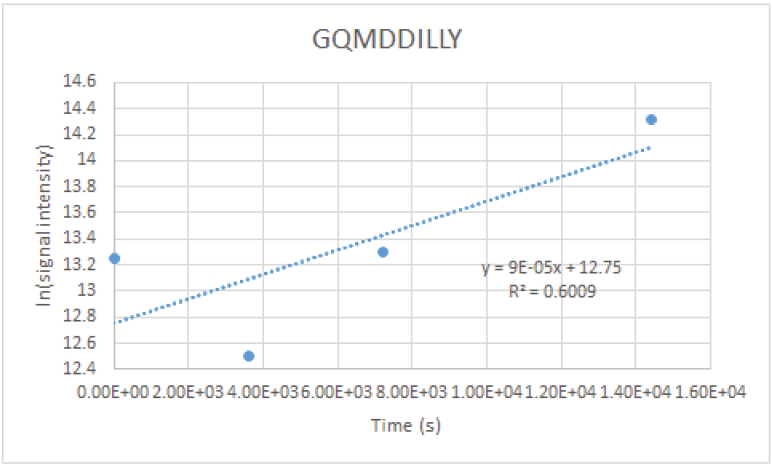


Table SI.1. Full sample list of artificially aged *Spondylus* shell.

Appendix 3. Table SI.1_Sample list.xlsx

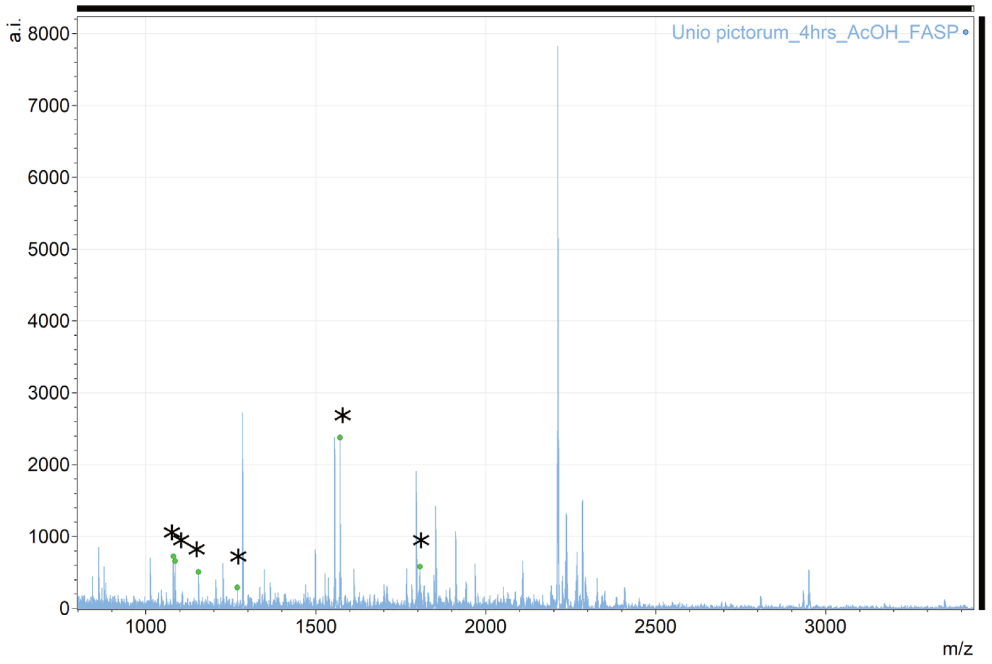
Appendix 4. Supplementary information.

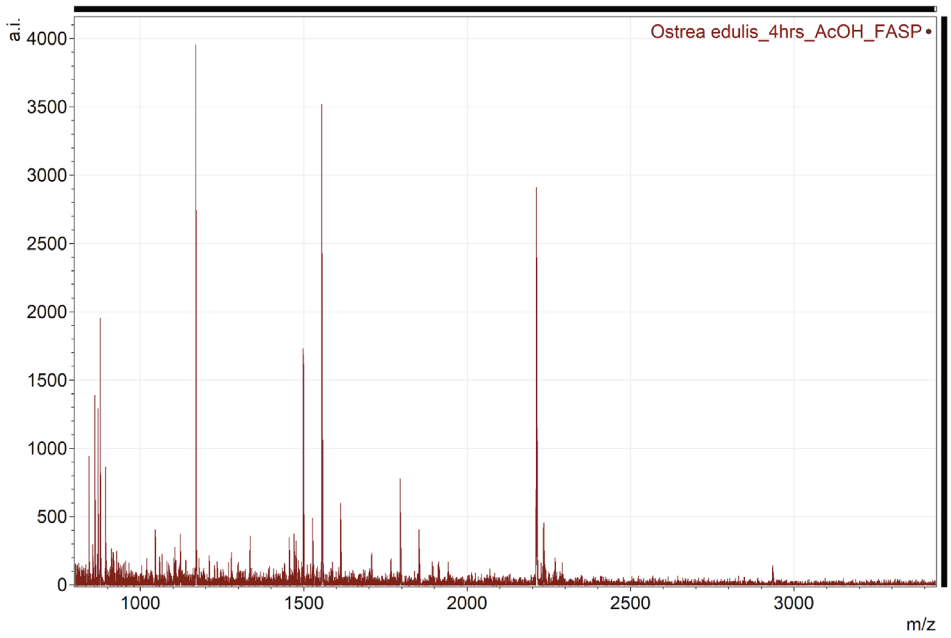
Shell palaeoproteomics: peptide mass fingerprinting of mollusc shells in archaeology

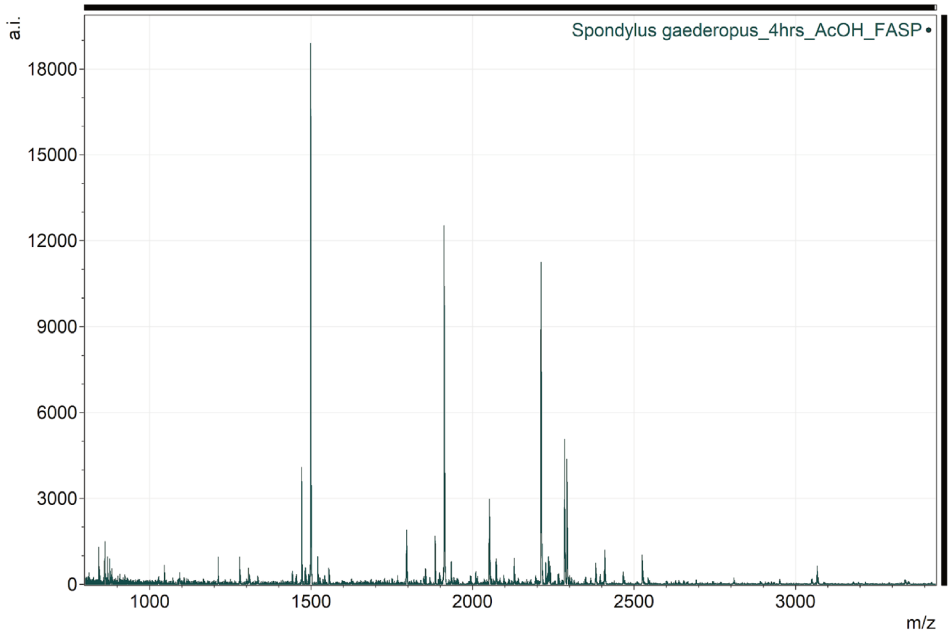
SI.1 Method development. Spectra obtained via different pre-treatment/extraction preparations

Asterisks indicate the marker peptides which were identified for these shells and are reported in the Table 2 (main manuscript). Spectra that do not show any presence of marker peaks represent those samples for which the extraction/ PMF characterization was not successful.

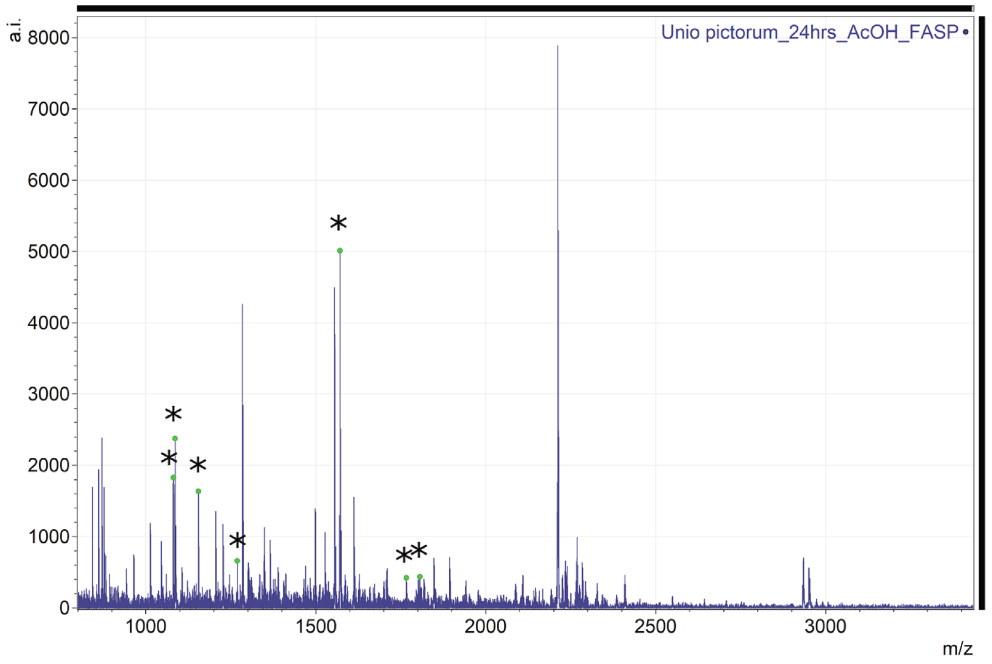
Method 1. 4h_AcOH_FASP

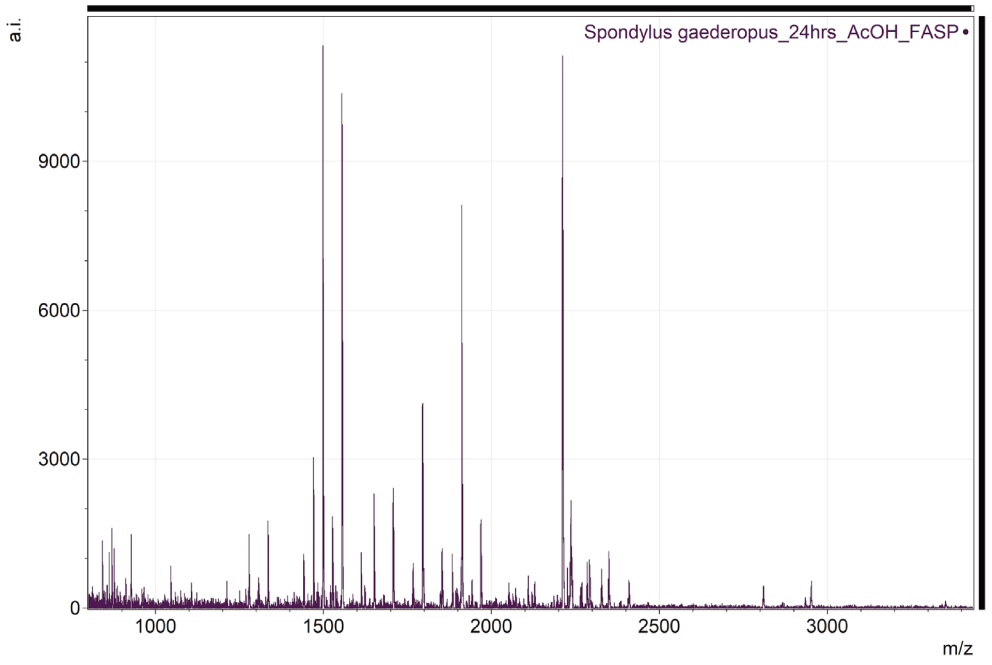


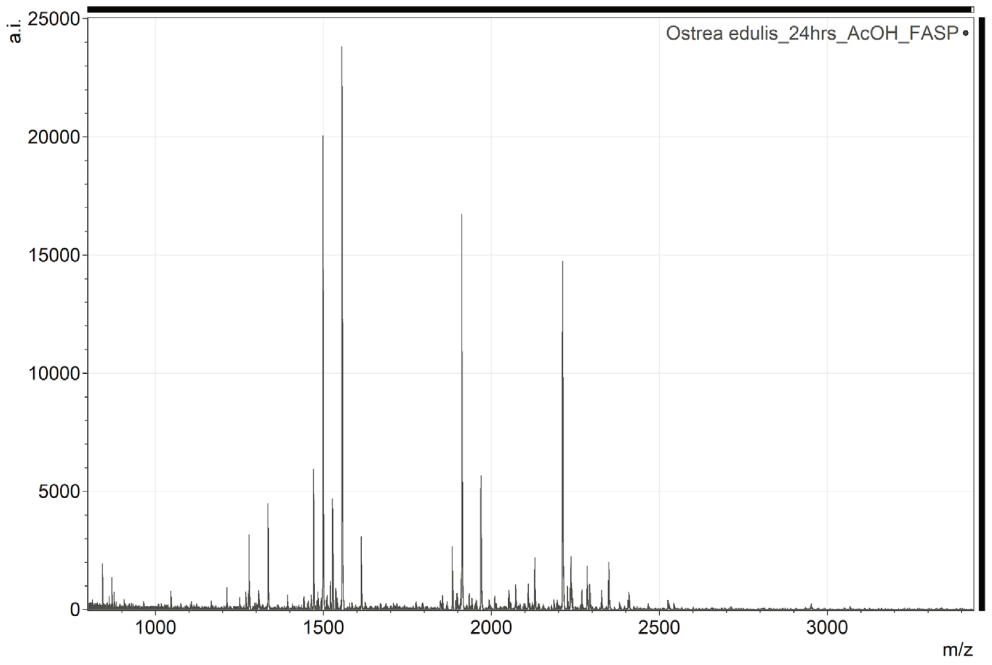

Ostrea edulis_4hrs_AcOH_FASP
a.i.
4000
3500
3000
2500
2000
1500
1000
500
0
1000
1500
2000
2500
3000
m/z


Spondylus gaederopus_4hrs_AcOH_FASP
a.i.
18000
15000
12000
9000
6000
3000
0
1000
1500
2000
2500
3000
m/z

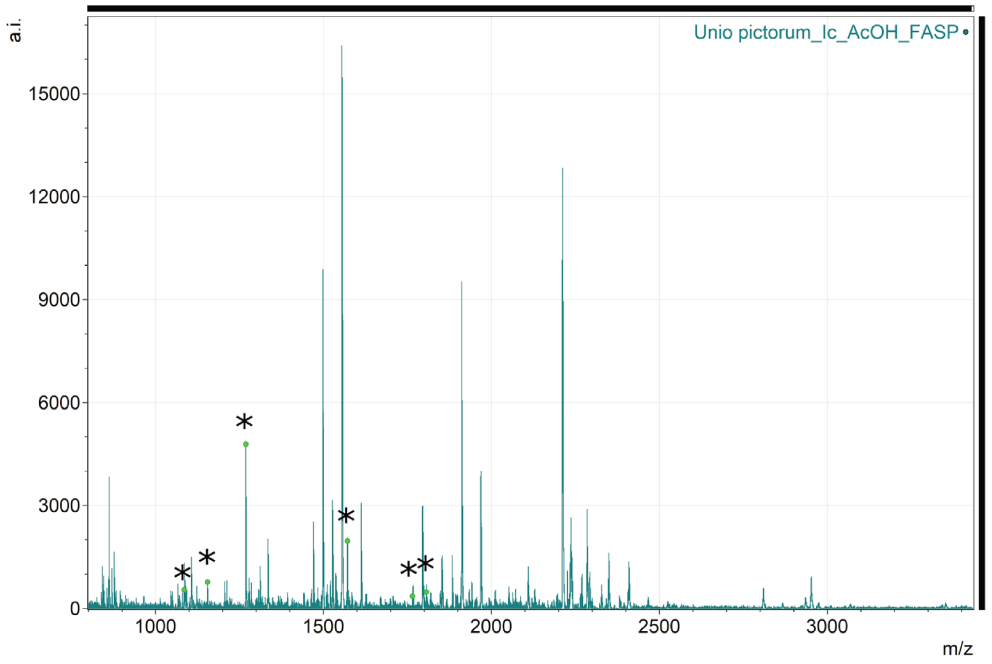
Method 2. 24h_AcOH_FASP

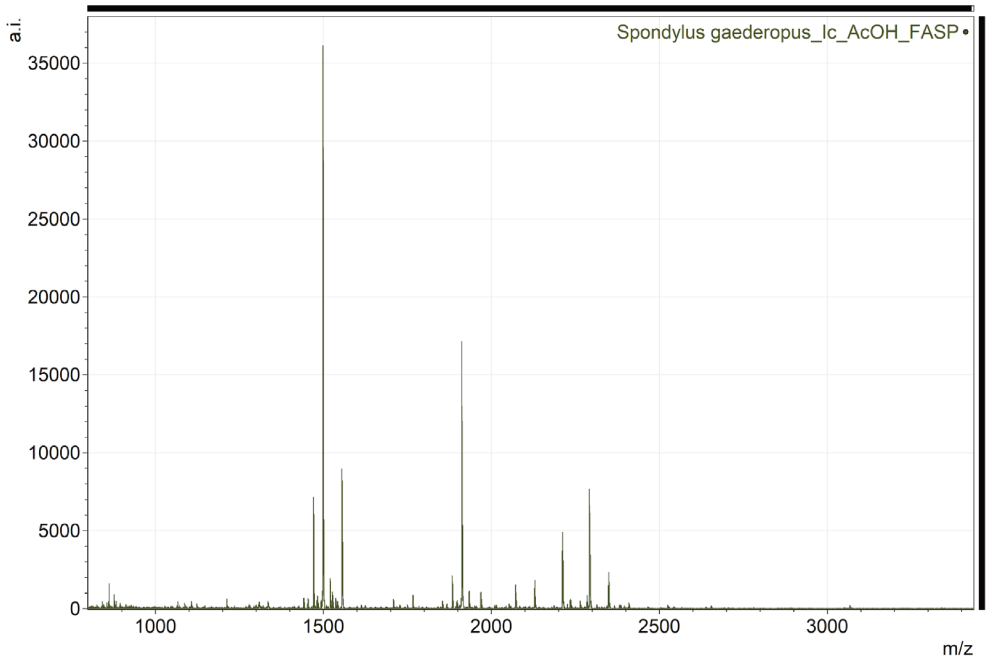


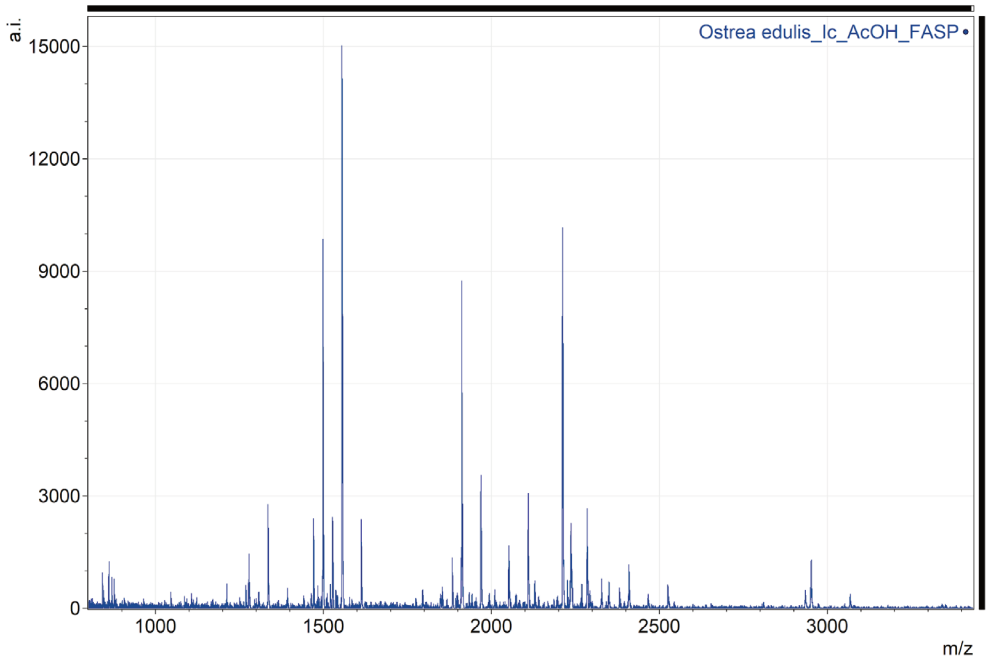




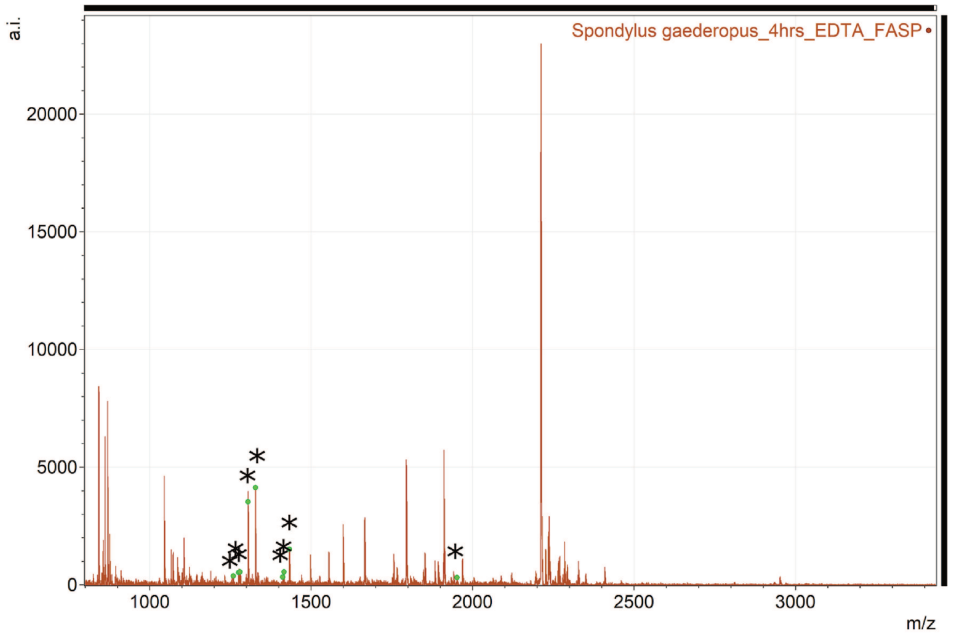
Method 3. lc_AcOH_FASP

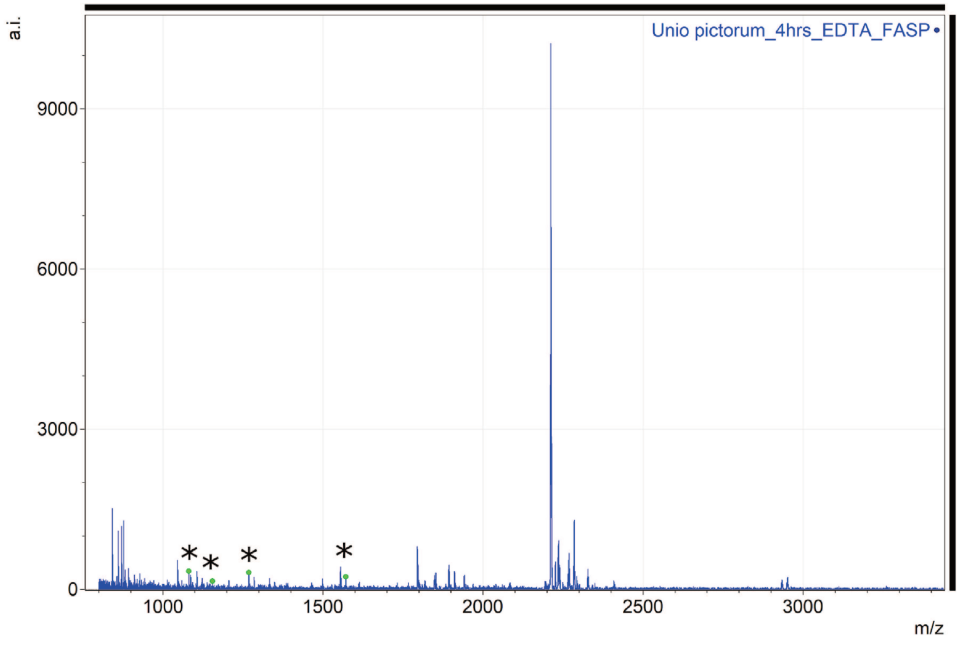


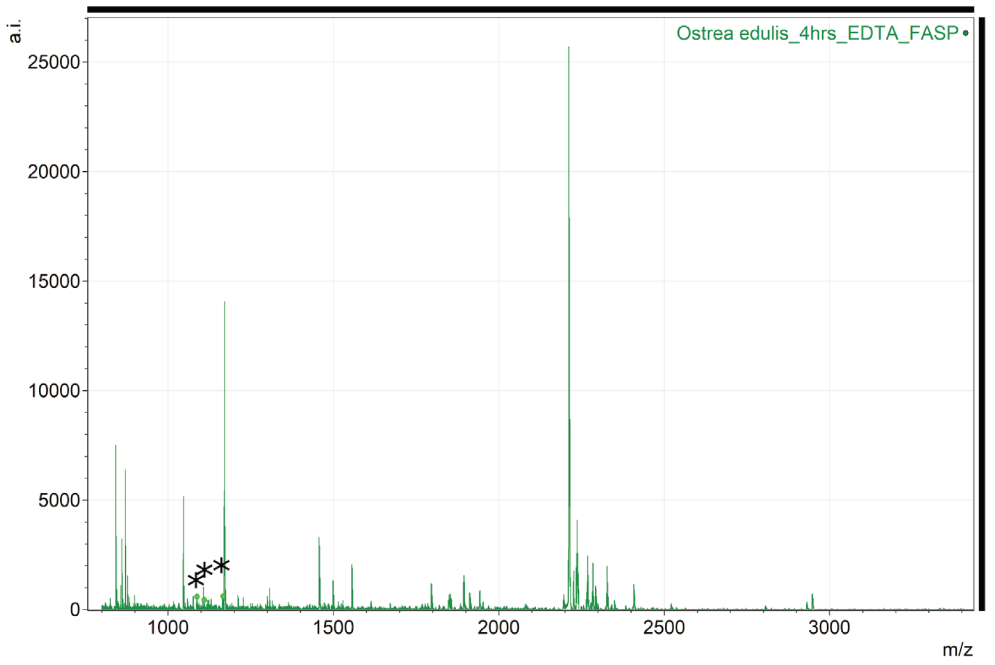

Spondylus gaederopus_lc_AcOH_FASP
a.i.
35000
30000
25000
20000
15000
10000
5000
0
1000
1500
2000
2500
3000
m/z


Ostrea edulis_lc_AcOH_FASP
a.i.
15000
12000
9000
6000
3000
0
1000
1500
2000
2500
3000
m/z

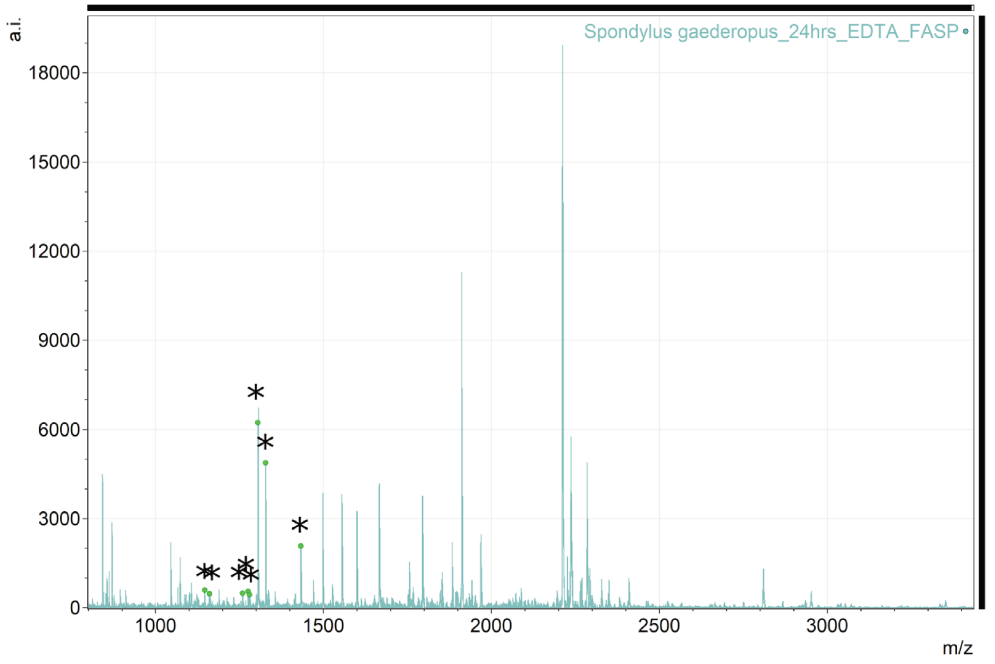
Method 4. 4h_EDTA_FASP

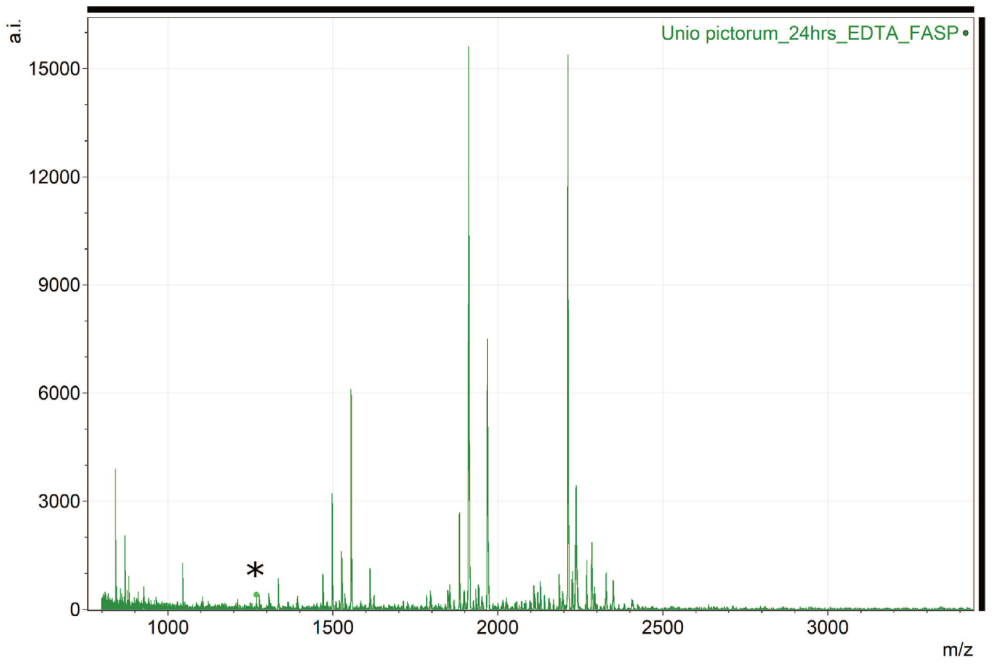


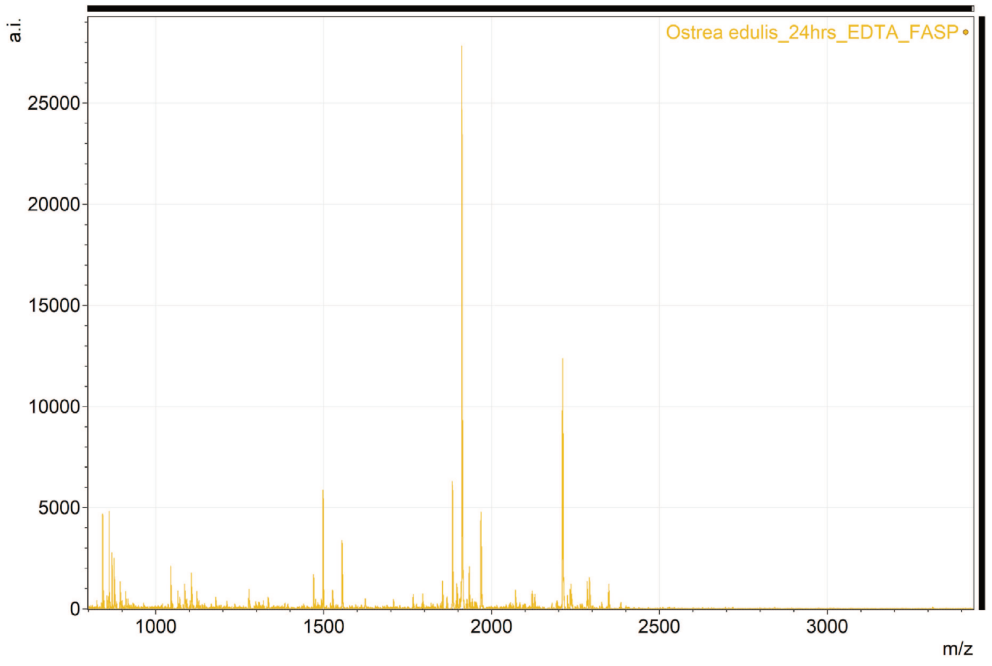




Method 5. 24h_EDTA_FASP



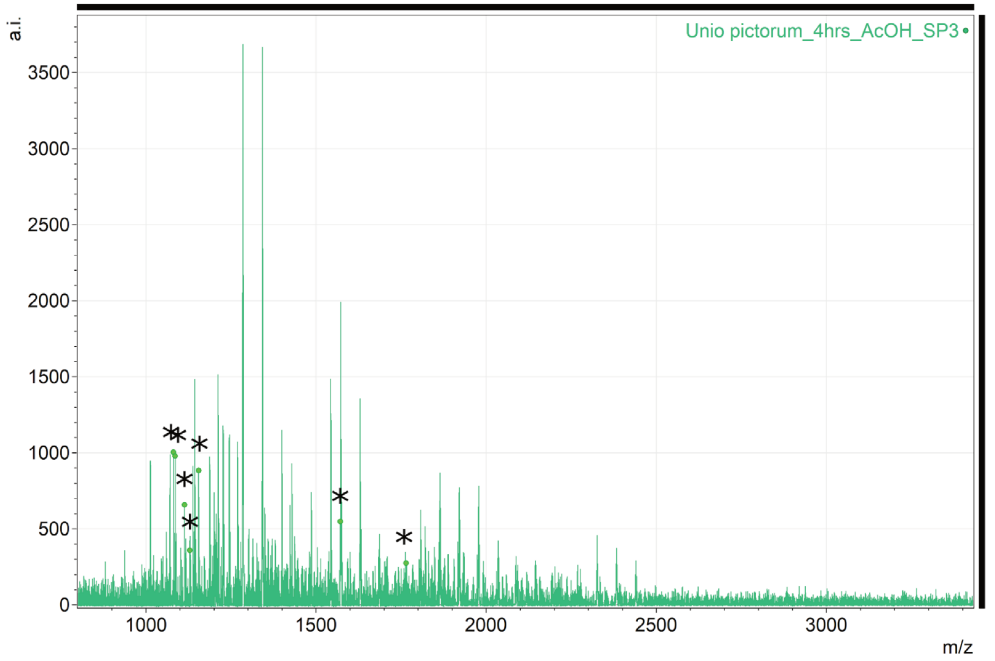

Unio pictorum_24hrs_EDTA_FASP
a.i.
15000
12000
9000
6000
3000
0
*
1000
1500
2000
2500
3000
m/z

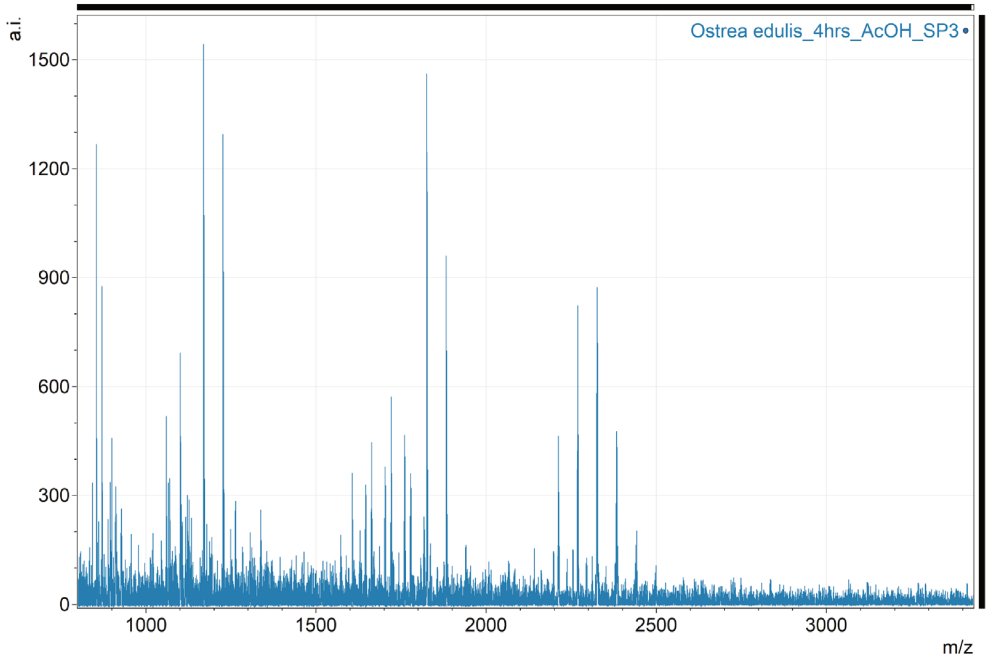

Ostrea edulis_24hrs_EDTA_FASP
a.i.
25000
20000
15000
10000
5000
0
1000
1500
2000
2500
3000
m/z

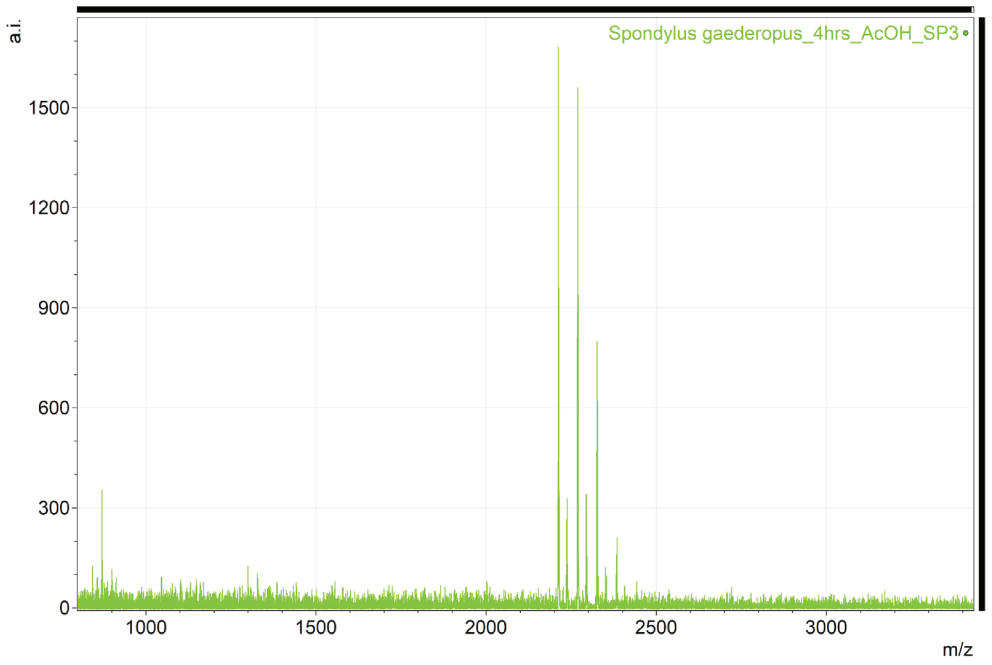
Method 6. lc_EDTA_FASP

Main manuscript

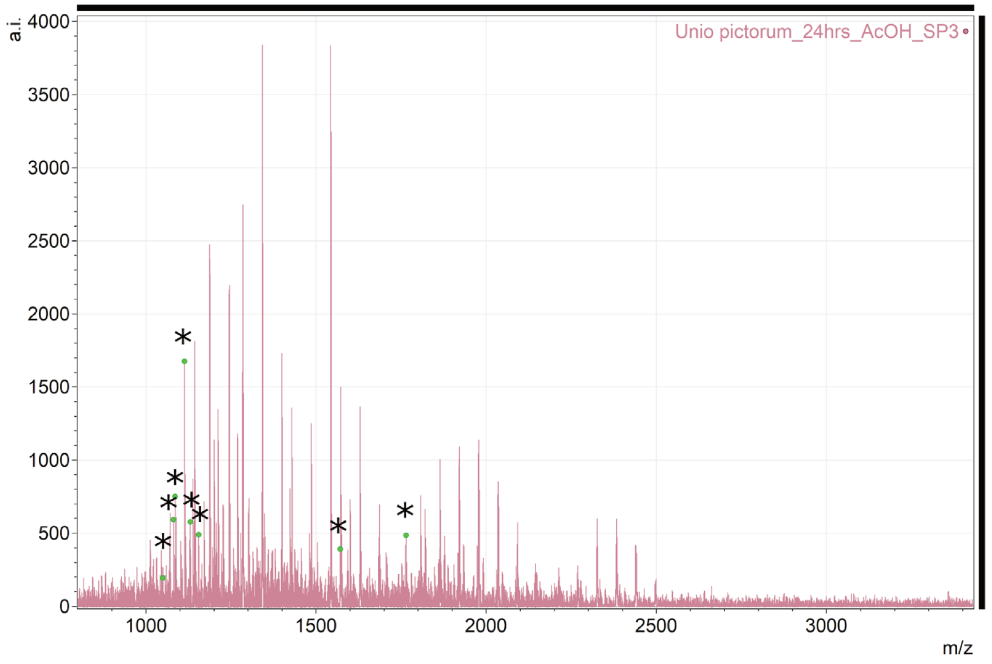
Method 7. 4h_AcOH_SP3

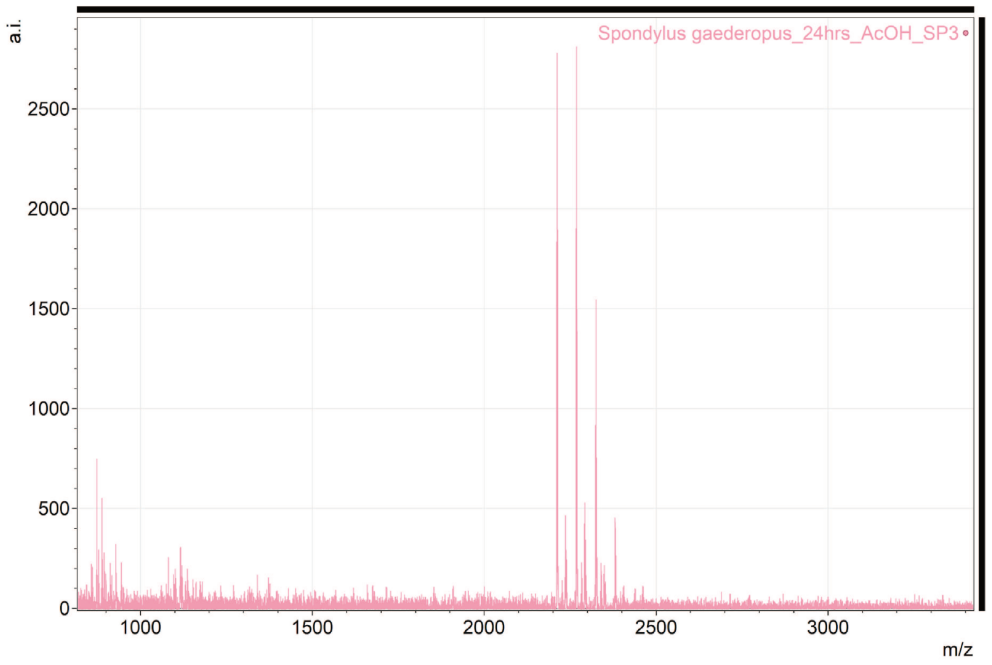


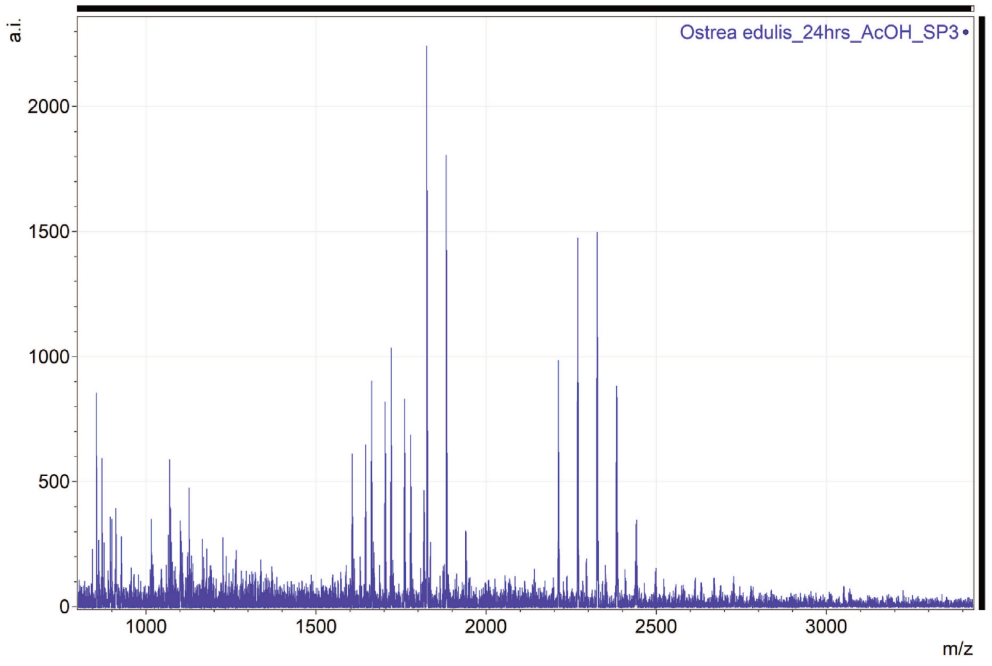




Method 8. 24h_AcOH_ SP3

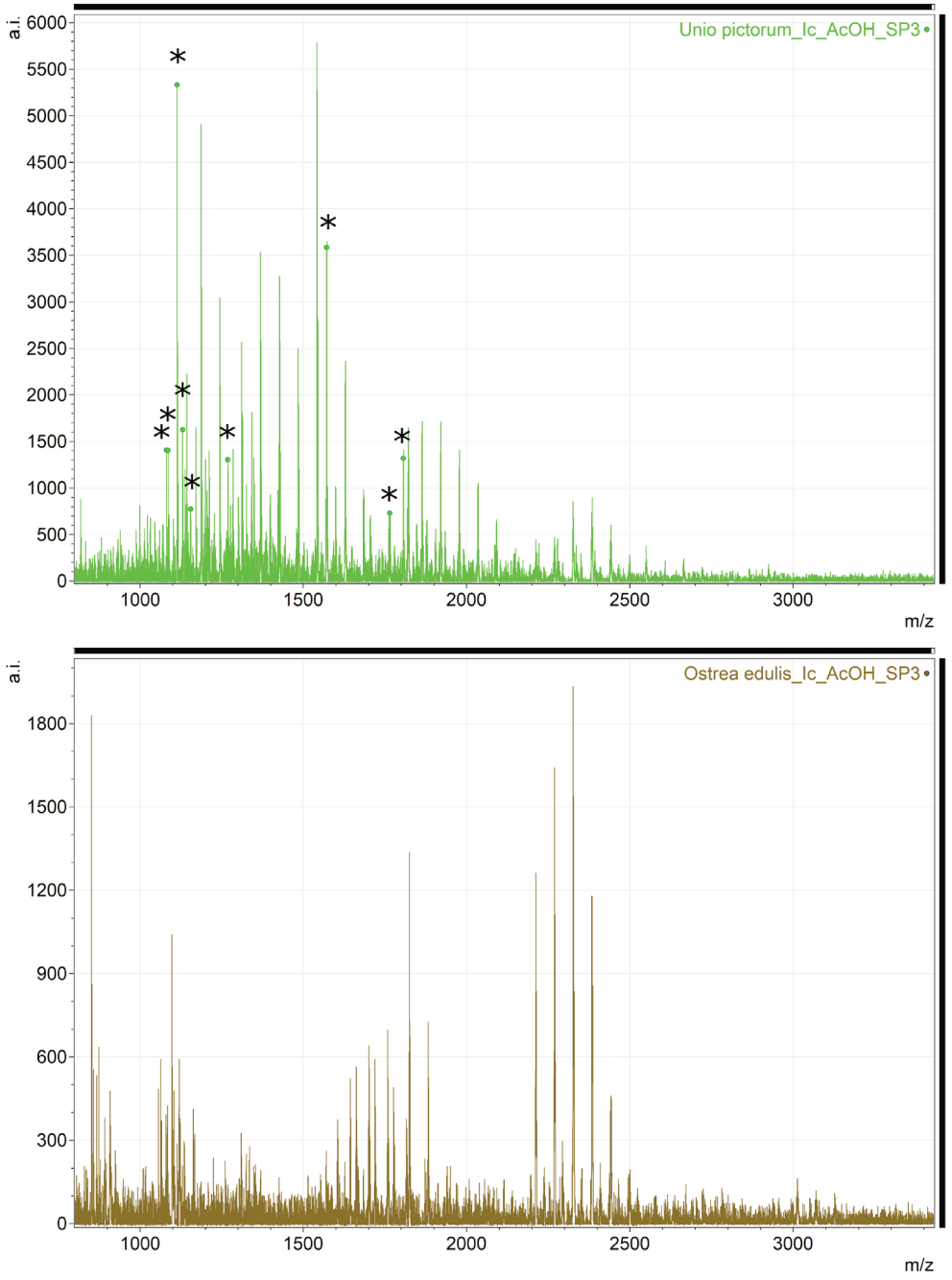



a.i.
Spondylus gaederopus_24hrs_AcOH_SP3
2500
2000
1500
1000
500
0
1000
1500
2000
2500
3000
m/z

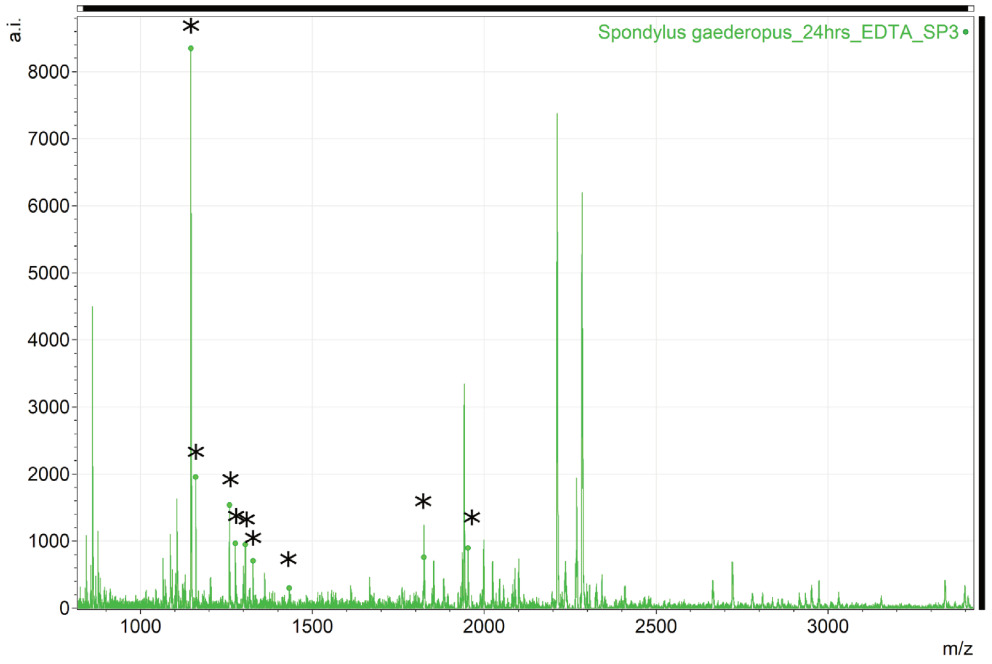

a.i.
Ostrea edulis_24hrs_AcOH_SP3
2000
1500
1000
500
0
1000
1500
2000
2500
3000
m/z

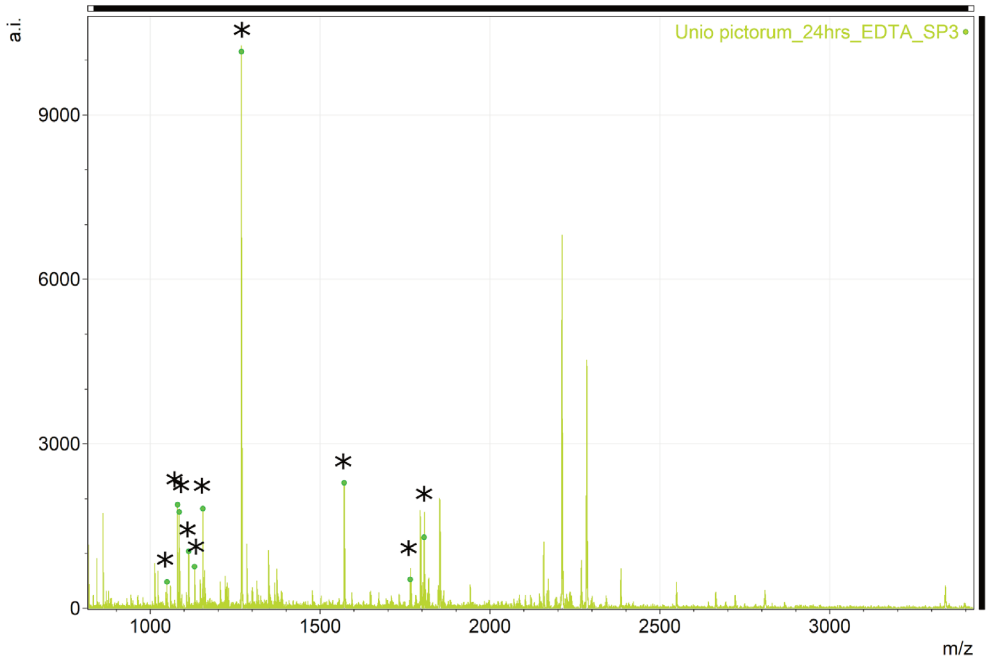
Method 9. Ic_AcOH_ SP3

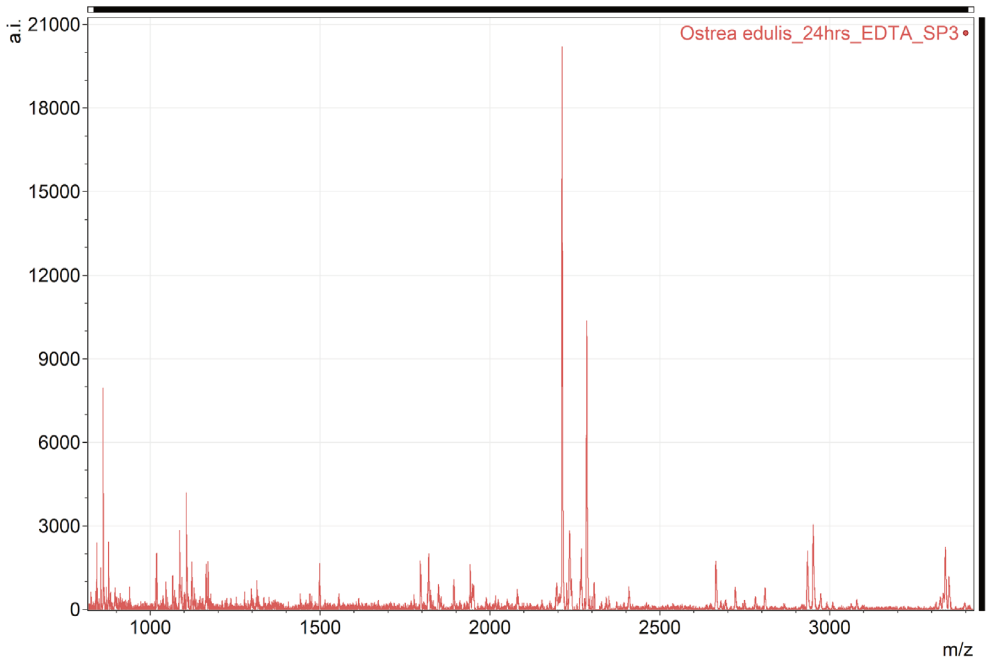
(*Spondylus* - main manuscript)



Method 11. 24h_EDTA_ SP3



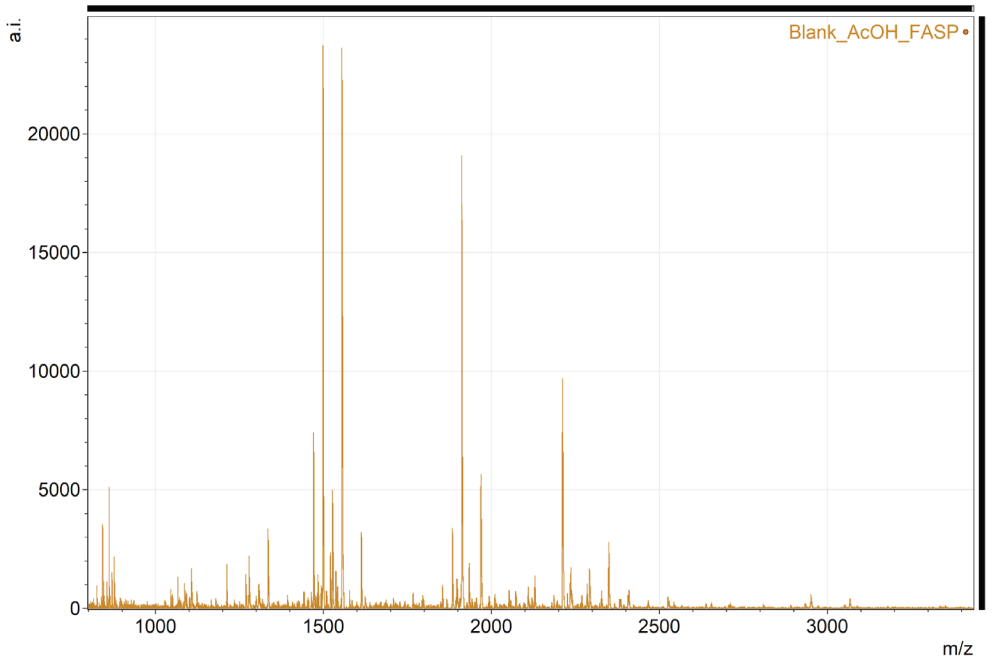




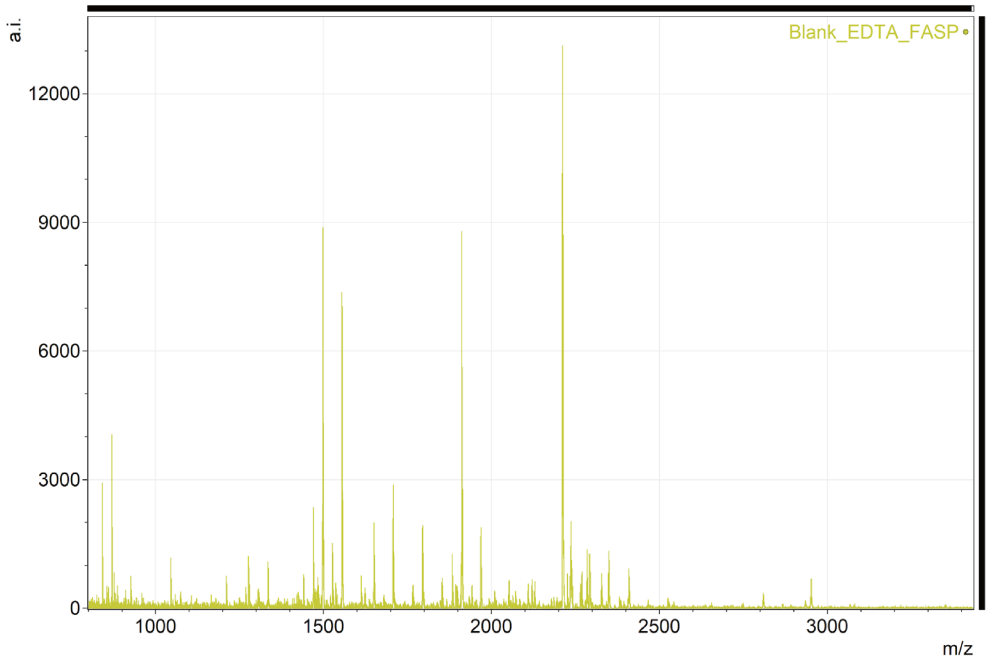
Method 12. lc_EDTA_ SP3

Main manuscript

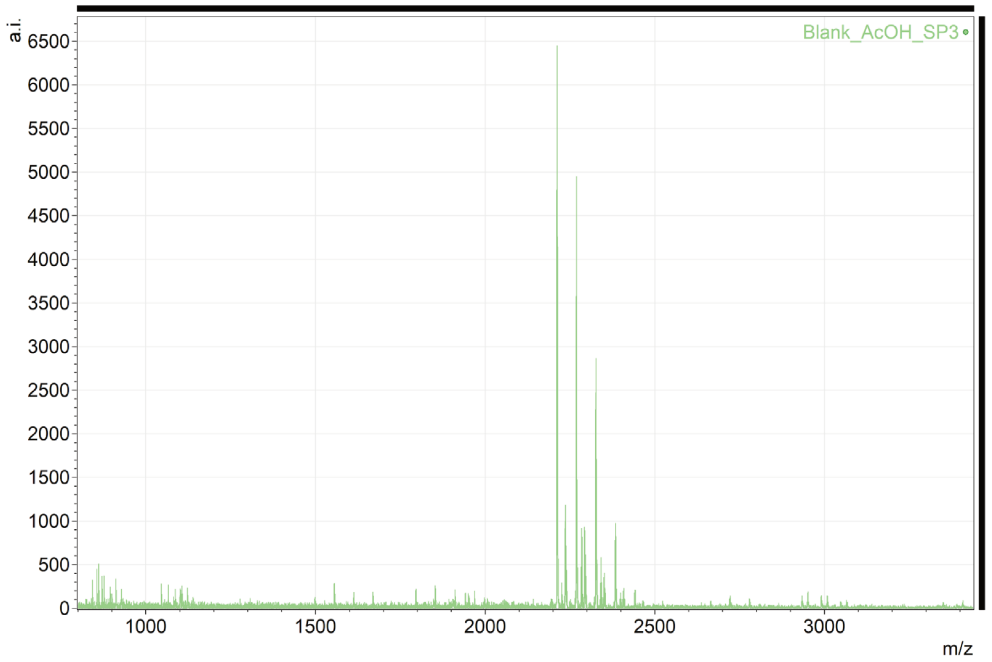
Blank_AcOH_ FASP



Blank_EDTA_ FASP



Blank_AcOH_ SP3



Blank_EDTA_ SP3

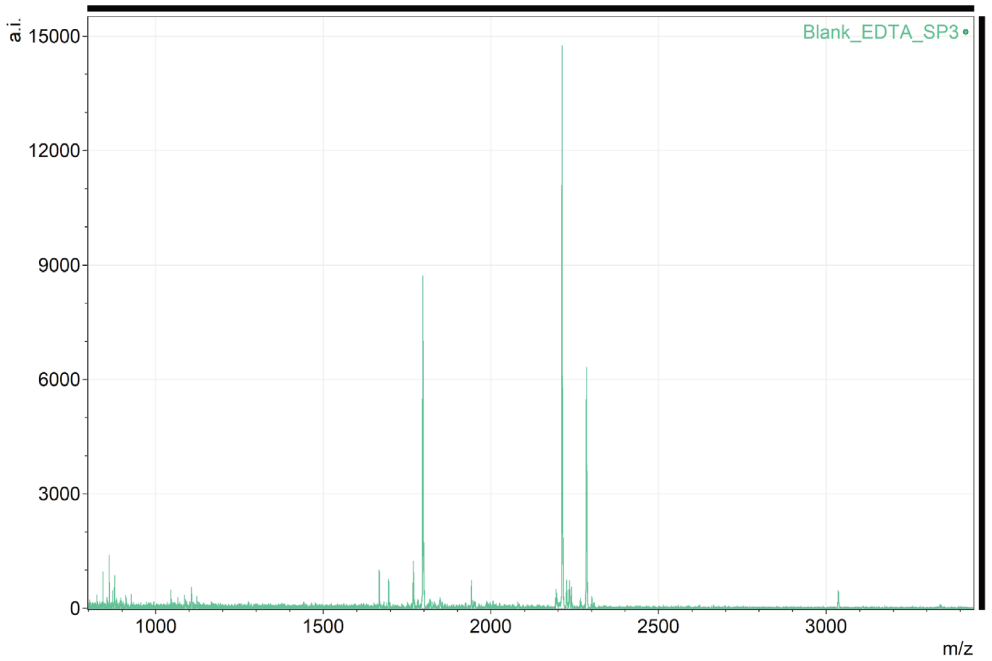


Table SI.1. List of reference m/z values corresponding to trypsin/keratin and MALDI matrix (HCCA) peaks that were used for internal calibration of the spectra.

Description	*m/z*
HCCA cluster	861.0794
HCCA cluster	867.0874
HCCA cluster	1066.0559
Keratin (Human)	1475.749
Keratin	2705.1611
Keratin	1453.8373
Trypsin	1045.5637
Trypsin	2211.104
Trypsin	2283.1802

SI.2 PMF. Contamination

Table SI.2. Contaminant PMF markers peaks that were observed in sample PMFs, extracted by FASP and SP3 methods.

Appendix
4.Table.SI.2_Contamin

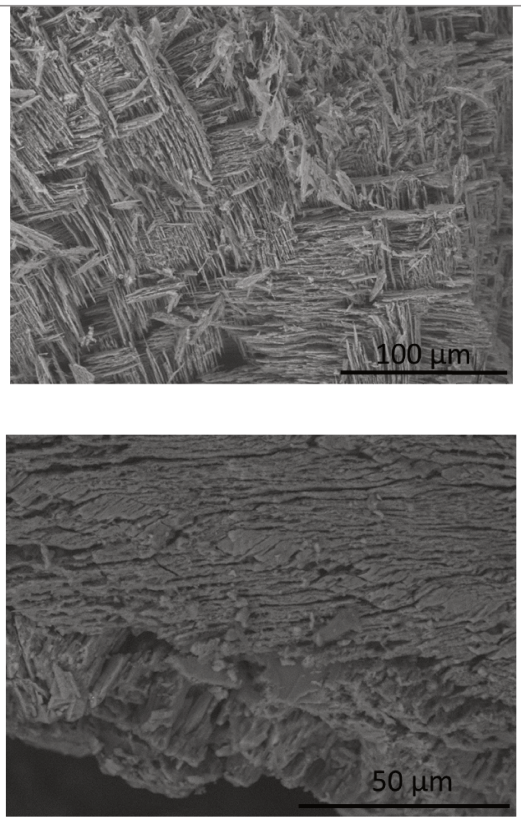
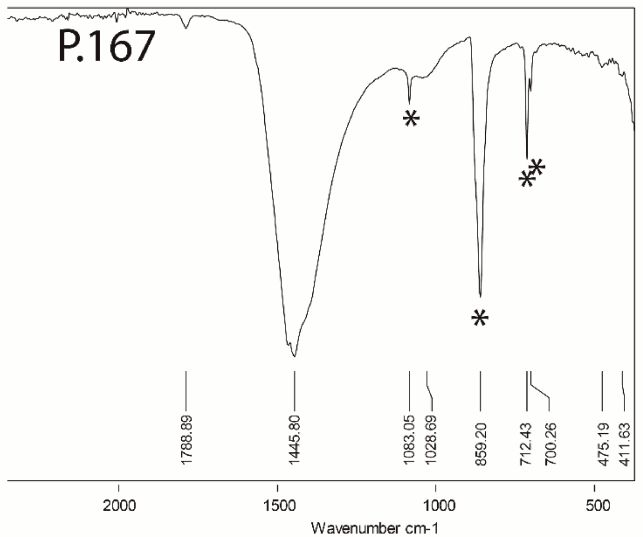
Appendix 5. Supplementary information.

Shell palaeoproteomics II: peptide mass fingerprinting of mollusc shells in archaeology

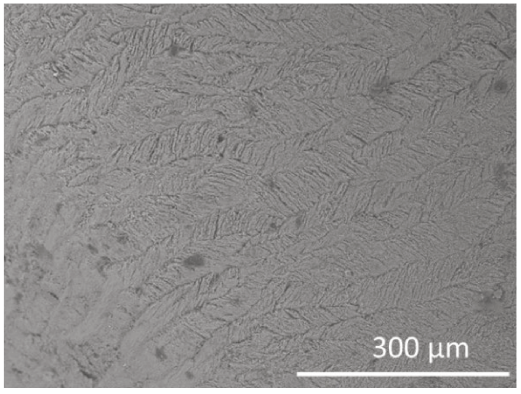
SI.1 Archaeological samples

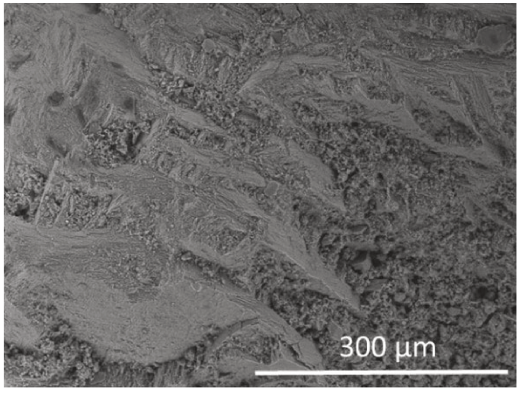
Data files SI.1. Micro-CT scans of archaeological samples from Le Taï and Yonne valley archaeological sites. The dataset includes raw mesh files which provide a 3D representation of the samples.

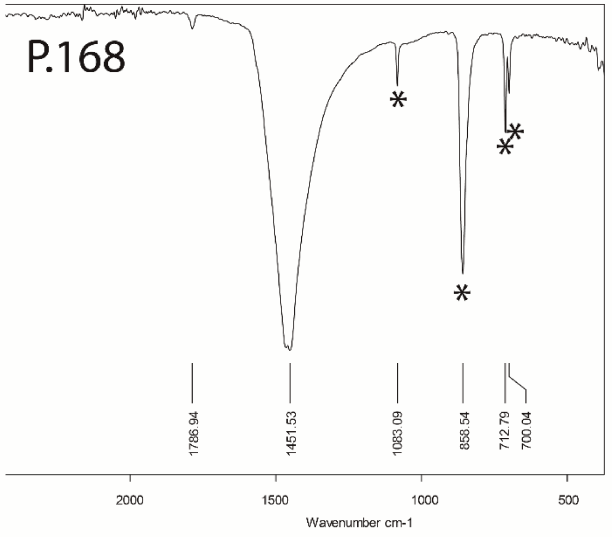
Table SI.1. SEM micrographs and corresponding FTIR-ATR spectra of archaeological samples from Le Taï and Yonne valley archaeological sites. SEM was used to identify sample microstructure. Asterisks mark distinctive FTIR bands that were identified for these samples which indicate presence of aragonite, calcite or calcium phosphate/collagen. The obtained results are presented in Table 3 of the main manuscript (Chapter 4).

Sample	*SEM*	*FTIR*
P.167		

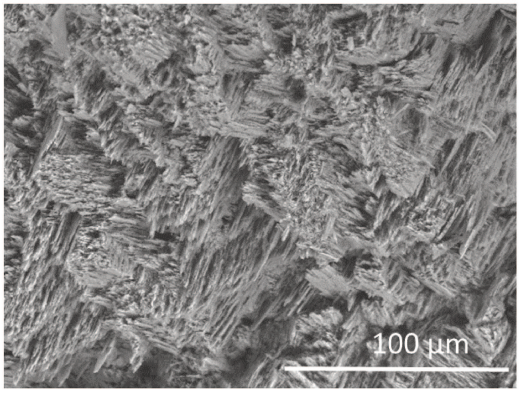
P.168

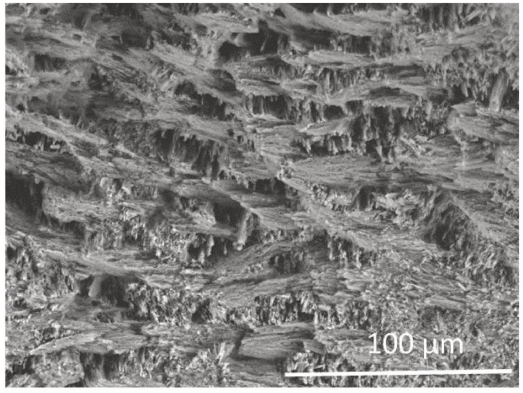

300 µm

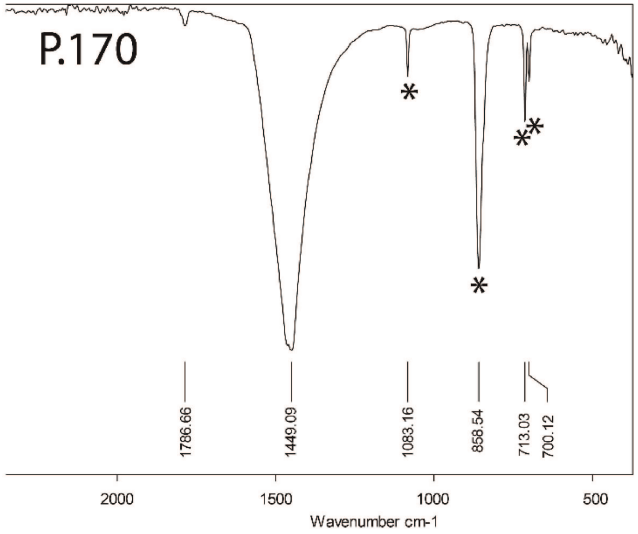

300 µm


P.168
1786.94
1451.53
1083.09
858.54
712.79
700.04
2000
1500
1000
500
Wavenumber cm-1

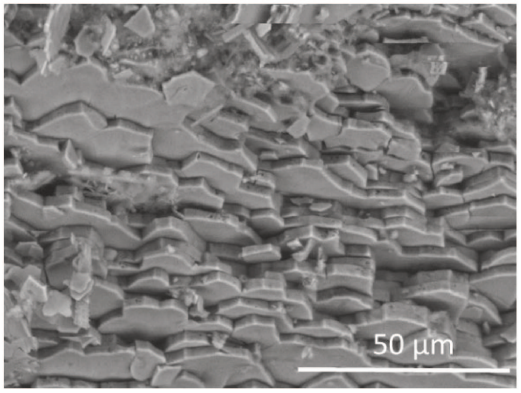
P.170

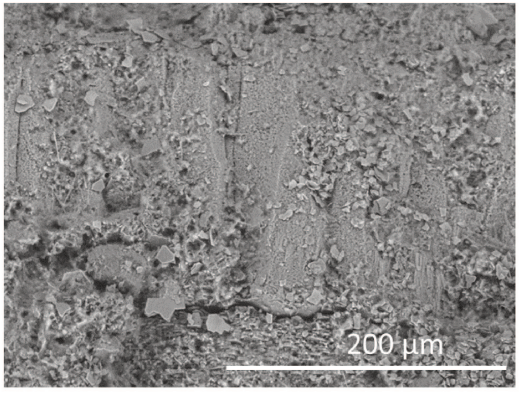

100 µm

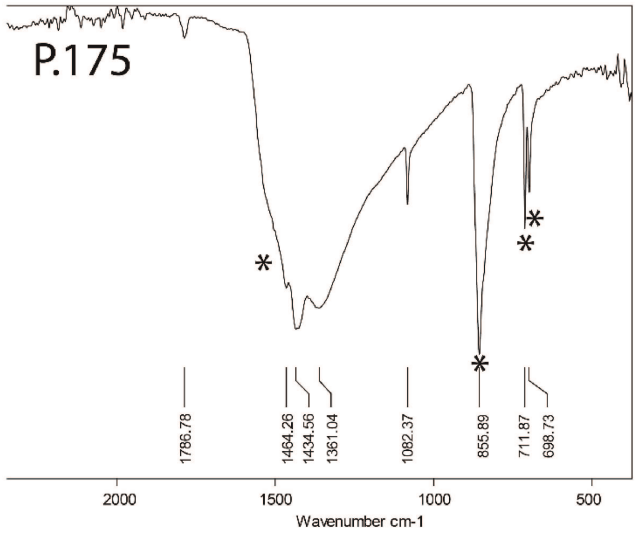

100 µm


P.170
1786.66
1449.09
1083.16
858.54
713.03
700.12
2000
1500
1000
500
Wavenumber cm-1

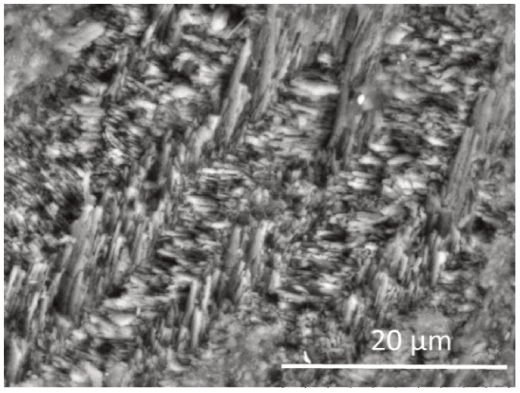
P.175

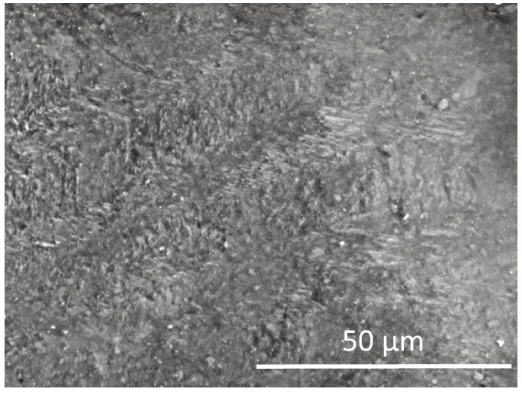

50 µm

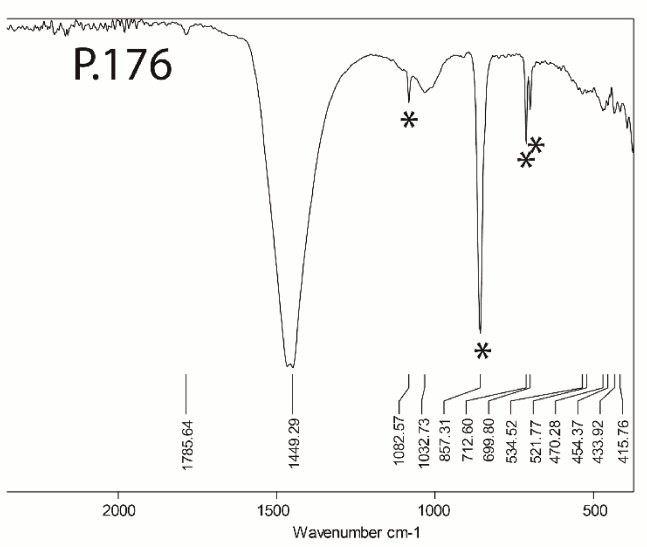

200 µm


P.175
1786.78
1464.26
1434.56
1361.04
1082.37
855.89
711.87
698.73
2000
1500
1000
500
Wavenumber cm-1

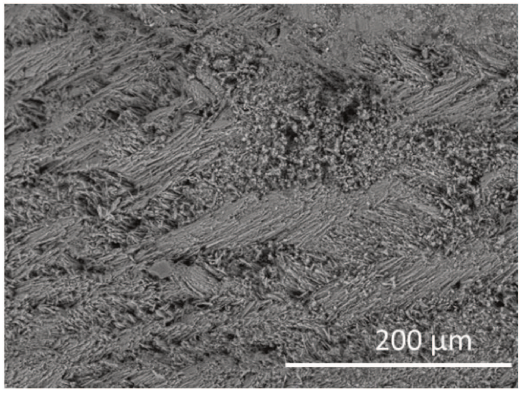
P.176

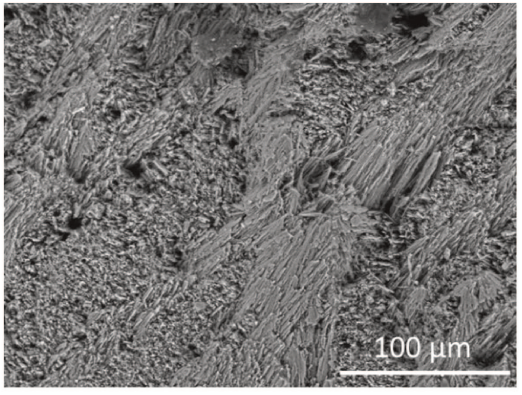

20 µm

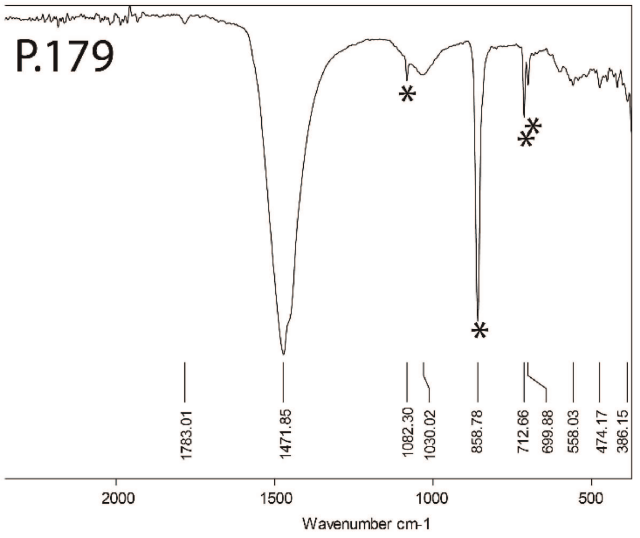

50 µm


P.176
1785.64
1449.29
1082.57
1032.73
857.31
712.60
699.80
534.52
521.77
470.28
454.37
433.92
415.76
2000
1500
1000
500
Wavenumber cm-1

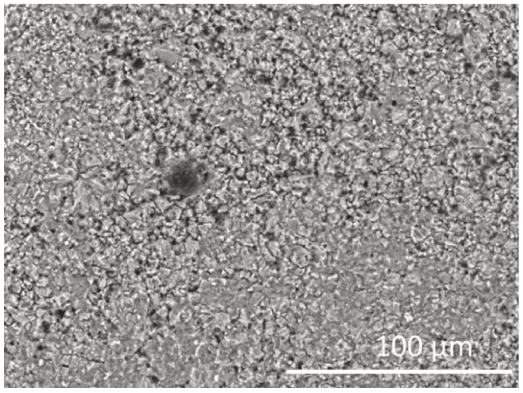
P.179

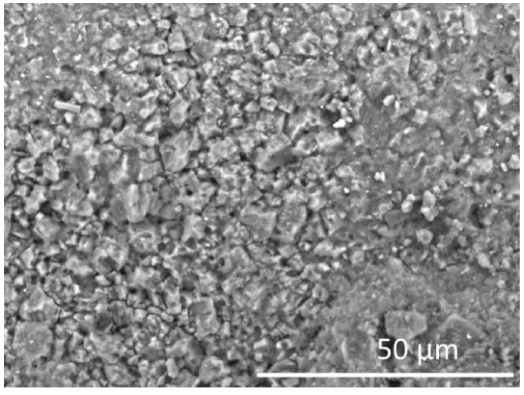

200 µm

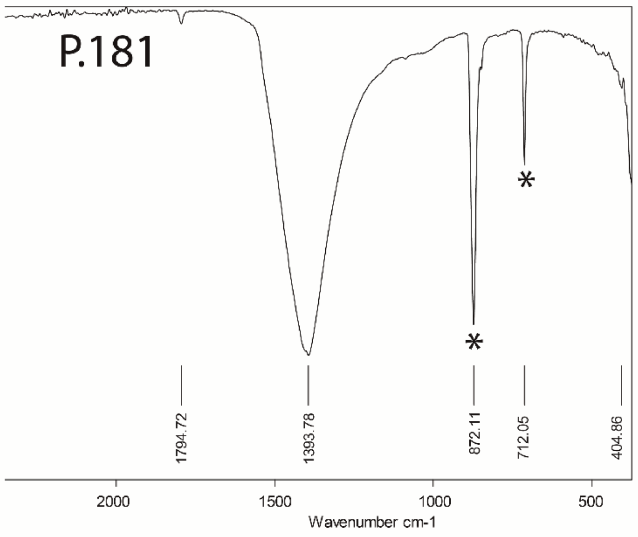

100 µm


P.179
1783.01
1471.85
1082.30
1030.02
858.78
712.66
699.88
558.03
474.17
386.15
2000
1500
1000
500
Wavenumber cm-1

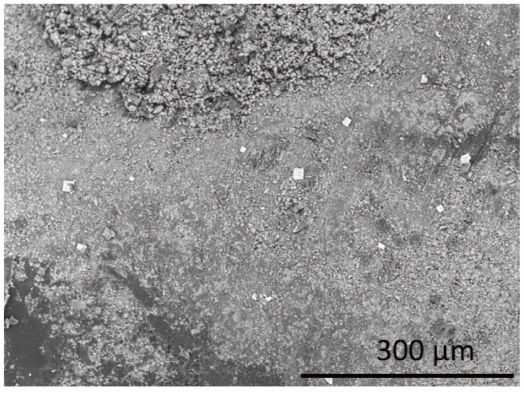
P.181

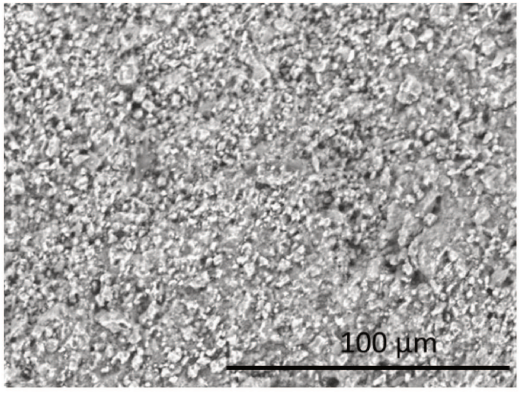

100 µm

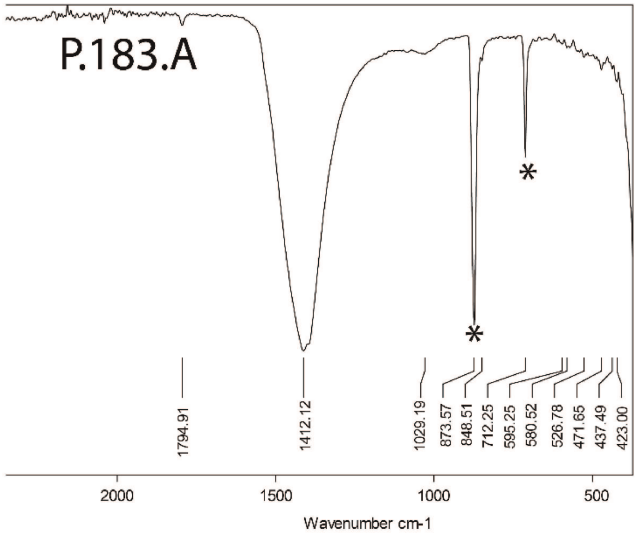

50 µm


P.181
1794.72
1393.78
872.11
712.05
404.86
2000
1500
1000
500
Wavenumber cm-1

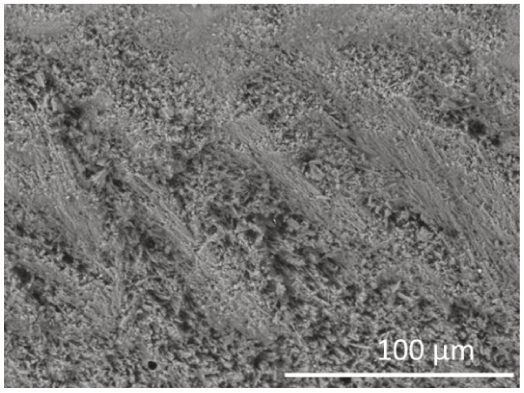
P.183.A

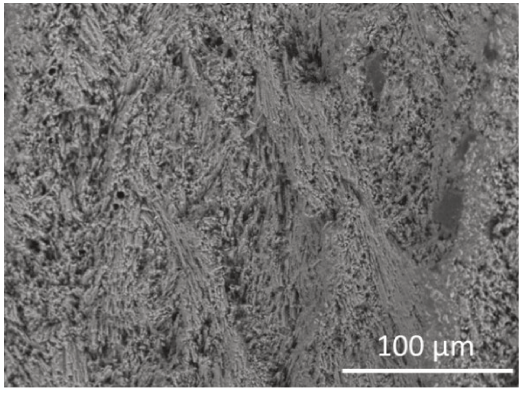

300 µm

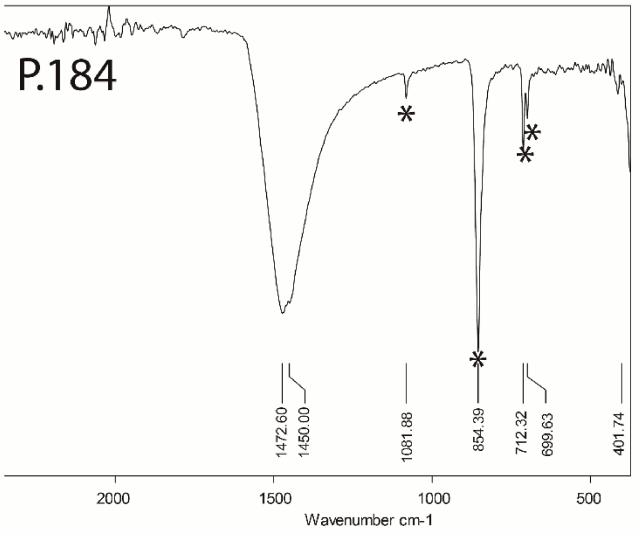

100 µm


P.183.A
1794.91
1412.12
1029.19
873.57
848.51
712.25
595.25
580.52
526.78
471.65
437.49
423.00
2000
1500
1000
500
Wavenumber cm-1

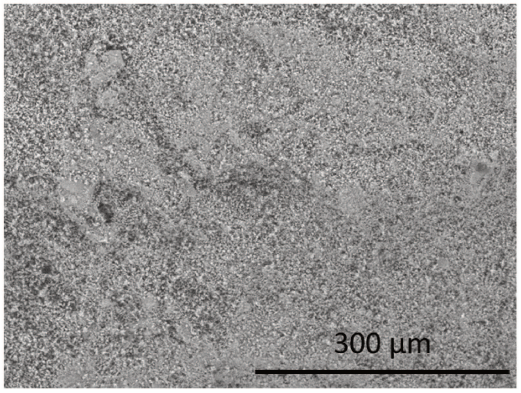
P.184

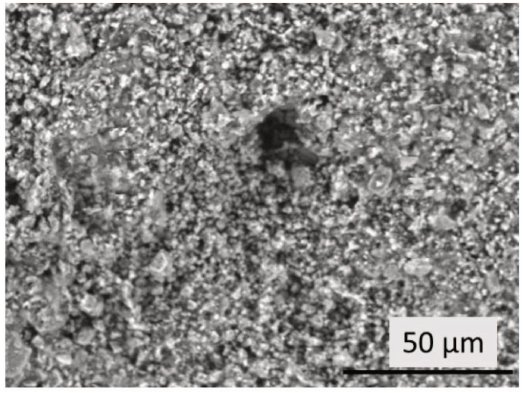

100 µm

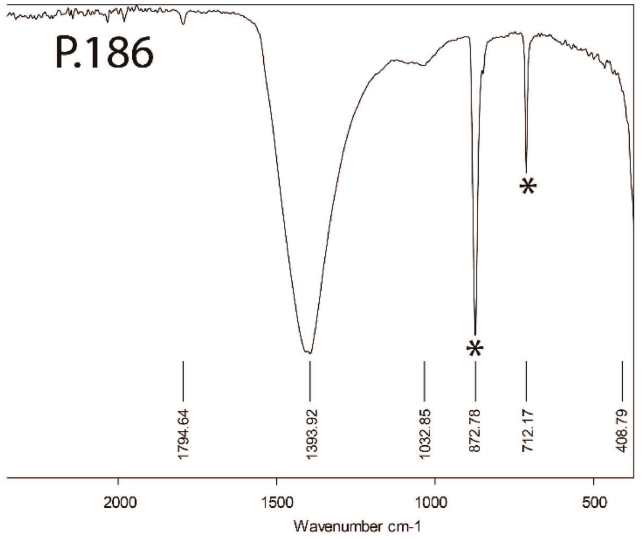

100 µm


P.184
1472.60
1450.00
1081.88
854.39
712.32
699.63
401.74
2000
1500
1000
500
Wavenumber cm-1

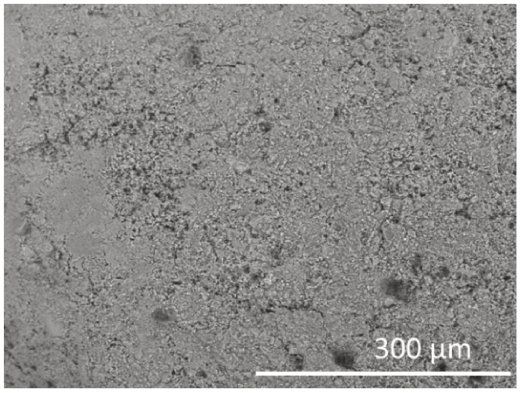
P.186

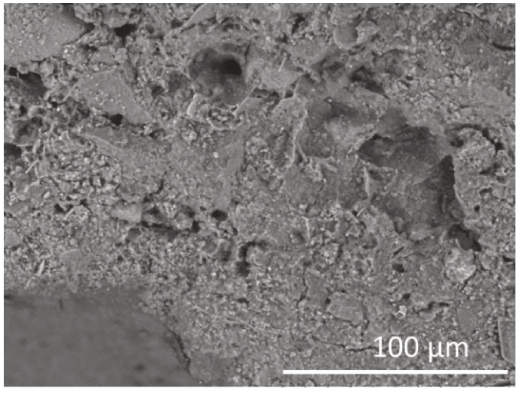

300 µm

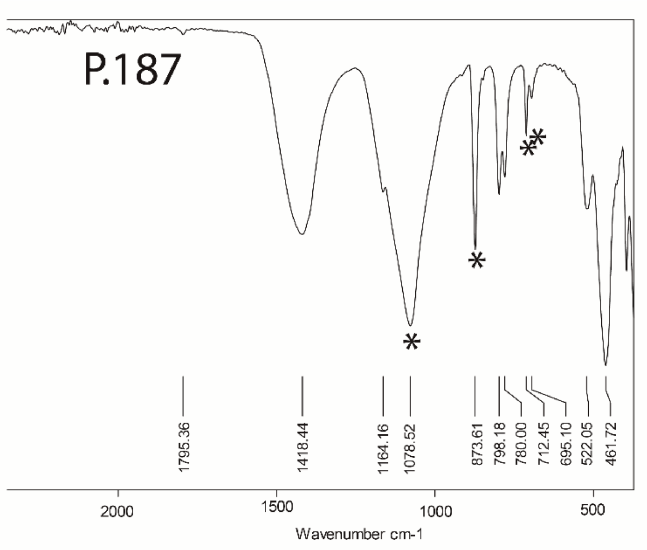

50 µm


P.186
1794.64
1393.92
1032.85
872.78
712.17
408.79
2000
1500
1000
500
Wavenumber cm-1

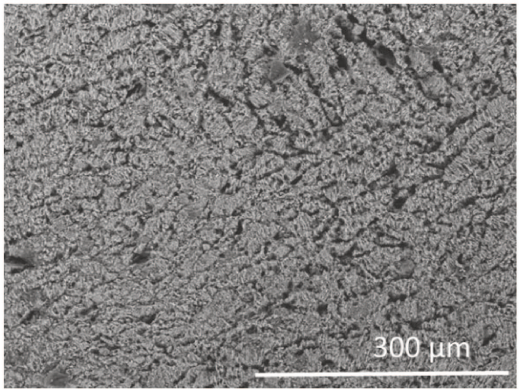
P.187

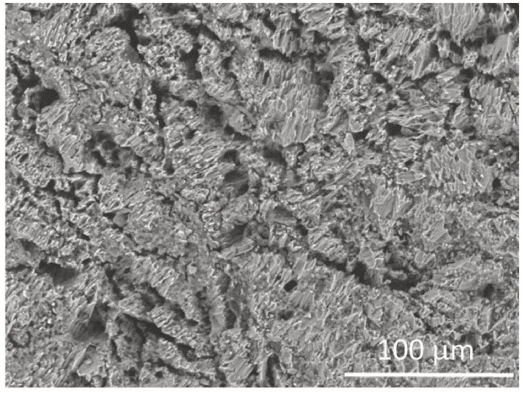

300 µm

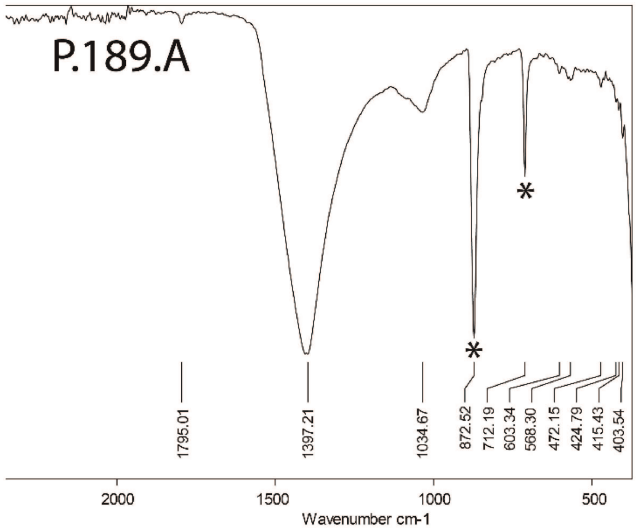

100 µm


P.187
1795.36
1418.44
1164.16
1078.52
873.61
798.18
780.00
712.45
695.10
522.05
461.72
2000
1500
1000
500
Wavenumber cm-1

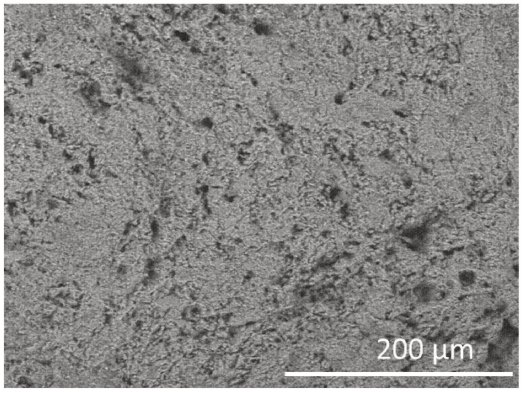
P.189.A

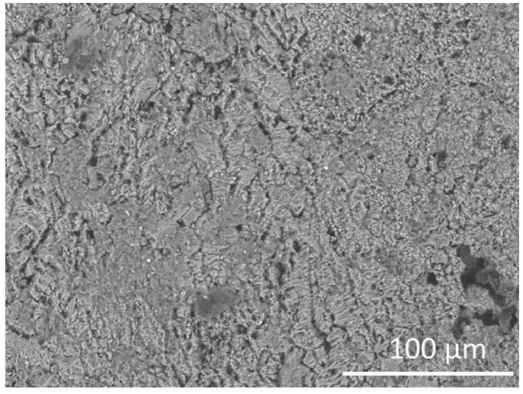

300 µm

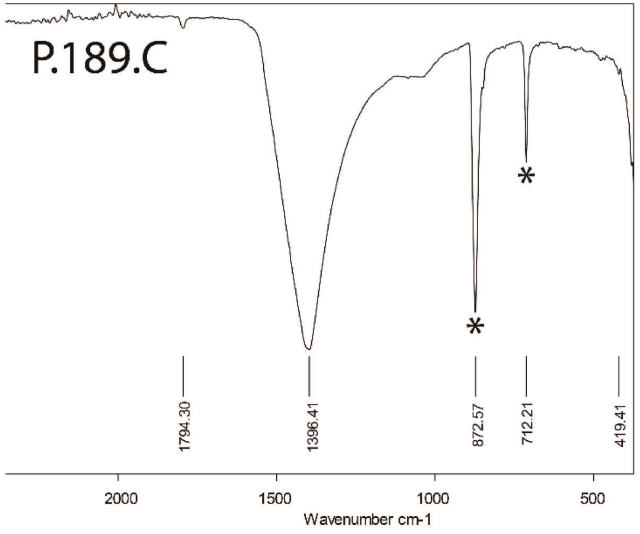

100 µm


P.189.A
1795.01
1397.21
1034.67
872.52
712.19
603.34
568.30
472.15
424.79
415.43
403.54
2000
1500
1000
500
Wavenumber cm-1

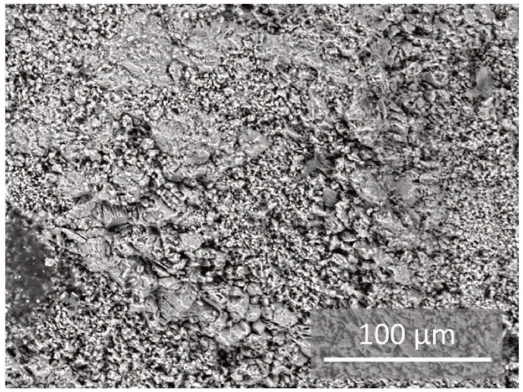
P.189.C

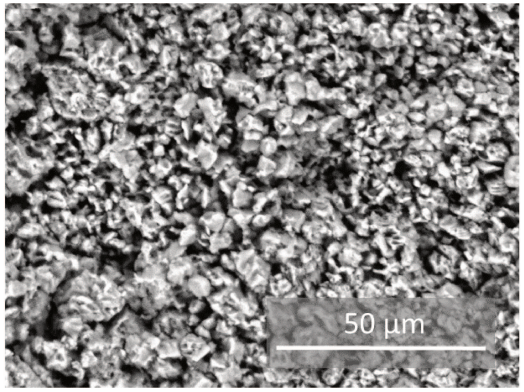

200 µm

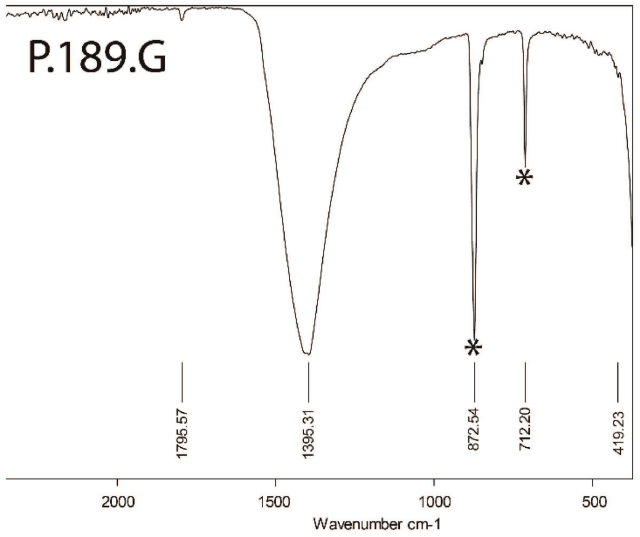

100 µm


P.189.C
1794.30
1396.41
872.57
712.21
419.41
2000
1500
1000
500
Wavenumber cm-1

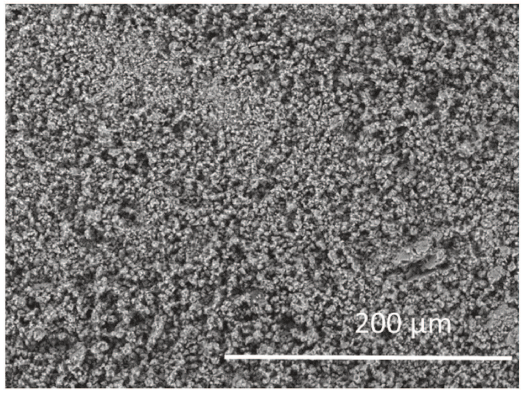
P.189.G

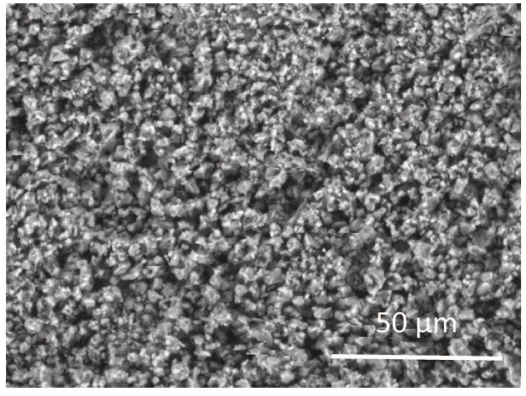

100 µm

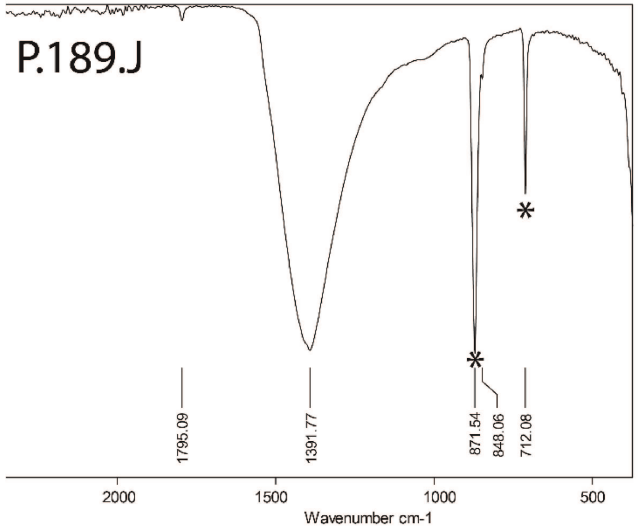

50 µm


P.189.G
1795.57
1395.31
872.54
712.20
419.23
2000
1500
1000
500
Wavenumber cm-1

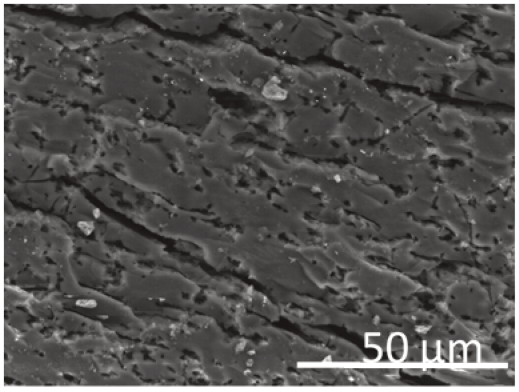
P.189.J

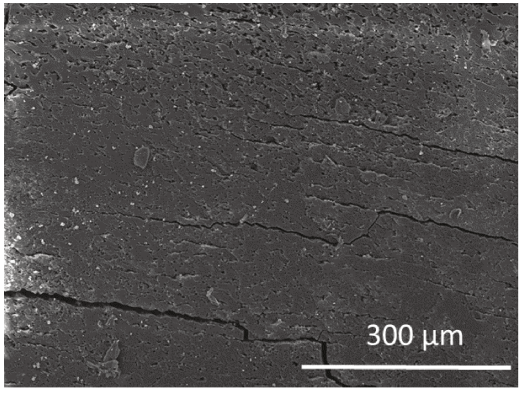

200 µm

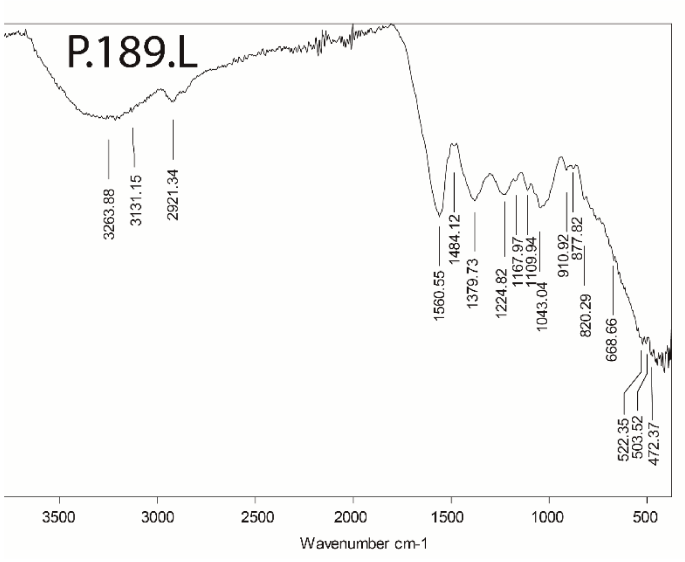

50 µm


P.189.J
1795.09
1391.77
871.54
848.06
712.08
2000
1500
1000
500
Wavenumber cm-1

P.189.L


50 µm


300 µm


P.189.L
3263.88
3131.15
2921.34
1560.55
1484.12
1379.73
1224.82
1167.97
1109.94
1043.04
910.92
877.82
820.29
668.66
522.35
503.52
472.37
3500
3000
2500
2000
1500
1000
500
Wavenumber cm-1

Appendix 6. Supplementary information.

'Palaeoshellomics' reveals the use of freshwater mother-of-pearl in prehistory

SI.1. Archaeological sites and double-buttons

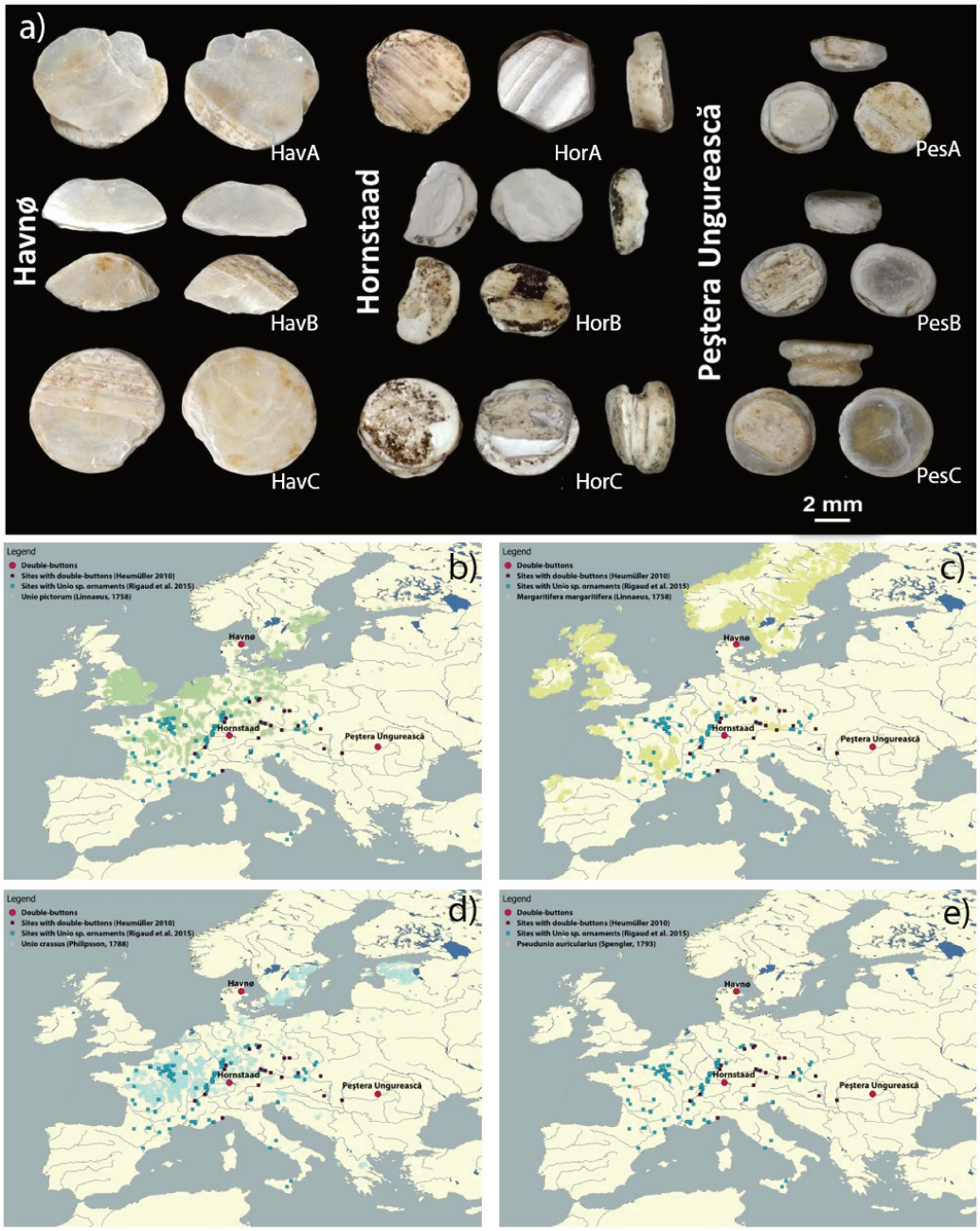


Figure S1.1: a) Double-button samples from the archaeological sites of Havnø (Denmark), Hornstaad-Hörnle IA (Germany) and Peştera Ungurească (Romania). Findings of double-buttons and of *Unio* ornaments as reported in the literature, compared to the present occurrence of *Unio pictorum* (b), *Margaritifera margaritifera* (c), *Unio crassus* (d), *Pseudunio auricularius* (e) (data obtained from GBIF, the Global Biodiversity Information Facility).

The three sets of double-buttons (*Doppelknöpfe*) analysed here come from the archaeological sites of Havnø (Denmark), Hornstaad-Hörnle IA (Germany) and Peştera Ungurească (Romania) and approximately span the period between 4200 and 3800 BCE.

1.1 Havnø

Havnø is a "stratified" shell midden, spanning the Late Mesolithic / Early Neolithic transition, i.e. Ertebølle and Funnel Beaker cultures, dated to 3950 cal BC. The site is on the East coast of Jutland c. 80 km north of Århus. The Gudenå river, which originated at the end of the Ice Age, a prime spot for fishing salmons and the major river in Denmark, has its estuary on the Randers Fjord, some 20 km to the South of Havnø.

Figure S1.2: The double-buttons from Havnø.

Now situated on the coast, Havnø was a small island during prehistoric times, covered in primeval forest of oak, elm, hazel and birch and an excellent base for fishing and marine hunting, as well as fowling and shell gathering (Andersen, 2008). The midden itself is 100 m long and 25-27 m wide and was used for a 1300-year period, demonstrating that this was a significant place during the Late Mesolithic and Early Neolithic. Indeed, among the abundant shells (oyster, cockle, mussel and periwinkle), one can find a variety of food remains as well as tools and ceramic vessels and traces of human settlements. Around 40 double-buttons were recovered from the Late Mesolithic horizon, dated to c. 4200-4000 cal BC

(Andersen, 2008, 2000). The buttons have a circular or oblong shape, one shiny surface and one matte, and a groove along the edge. Their diameters range between 0.69 - 0.79 cm. The working hypothesis has been that the raw material of the double-buttons was a marine shell: *Modiolus modiolus*, *Arctica islandica* or (according to Rowley-Conwy (2014)), *Ostrea edulis*, given that these are the most abundant species in the midden and that the site reflects the great importance of marine environments for the Ertebølle people. Interestingly, eight double-buttons, similar (but smaller, only around 5 mm in diameter) to the ones from Havnø and from the same period, have been found at the Nederst site, on the Djursland peninsula, 50 km south of Havnø (Kannegaard, 2013).

1.2 Hornstaad-Hörnle IA

Hornstaad-Hörnle IA is a pile-dwelling Neolithic site on Lake Konstanz, Baden-Württemberg, present-day Southern Germany. It is part of the UNESCO World Heritage “Prehistoric Pile Dwellings around the Alps”. Hornstaad is the typesite of the Hornstaad Group, an early cultural unit of the regional Late Neolithic (Jungneolithikum). The site contained several phases of wooden houses construction and large assemblages of artefacts and biofacts (Dieckmann et al., 2016, 2006; Schlichtherle, 1990). The economic basis of the village, which comprised 40-80 houses, was a combination of agriculture and animal husbandry, but hunting, fishing and the collection of wild plants also played an important role. The site has yielded extraordinarily rich assemblages of ornaments, including the (approximately) 564 double-buttons (diameter 0.46 - 0.58 cm) studied here, which come from an occupation level dated by dendrochronology to 3917- 3902 BC (Heumüller, 2012, 2009). Nearly all ornaments have been affected by the fire that destroyed the settlement in 3009 BC, and were found fragmented into pieces. The identification of the double-buttons from Hornstaad-Hörnle IA has been debated (Heumüller, 2012): macro- and micro-morphological observations resulted in the identification of the marine shell *Ostrea edulis* (edible oyster, also called European flat oyster) by one expert (H.-J. Niederhöfer) and of the freshwater pearl mussel *Margaritifera margaritifera* by another malacologist (A. Girod) (Borrello and Girod, 2008).

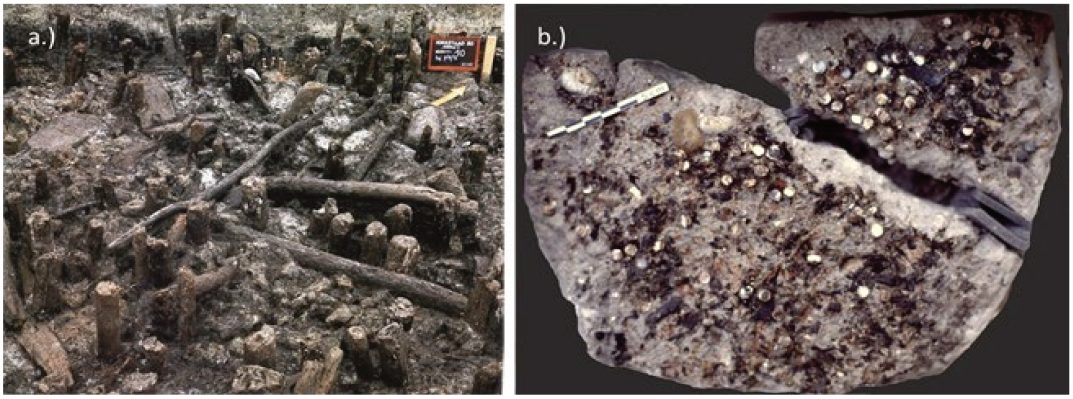


Figure S1.3: Excavations at Hornstaad D-Hörnle site (a) and the discovery of the double-buttons (b).

1.3. Peştera Ungurească

Peştera Ungurească is one of the cave sites in the karstic Cheile Turzii Gorge, approximately 9 kilometres west of Turda in central Transylvania, which was formed by the Hășdate river. The river currently flows at the bottom of the gorge, some 100 m below the cave (Biagi and Voytek, 2006). The archaeological sequence spans the transition between the Middle Neolithic and the Bronze Age (Cheile Turzii-/ Lumea Nouă-Iclod and Coţofeni).

The sediments from layers attributed to the Toarte Pastilate culture and the transition to Coţofeni were wet-sieved with a 1 mm mesh size and yielded a wealth of information on the environment and the subsistence strategies: hunting and rearing of large animals was supplemented by fishing and fowling, and both foraging and agriculture were practiced (Bartolomei, 2013; Biagi and Voytek, 2006; Boschian, 2010; Nisbet, 2010). Of interest, the finding of large fish vertebrae and freshwater turtles suggested the exploitation of a large river, probably not the Hășdate but the larger Ariesul, some 2.5 km east. The analysis of the knapped stone (chert and obsidian) assemblage revealed that the inhabitants of Peştera Ungurească sourced raw materials from western Hungary, north-western Ukraine and the Carpathians (Biagi and Voytek, 2006). Layer 2b (charcoal radiocarbon date GrN-29100: 5100±40 BP, 3980-3790 cal BC (2σ) (Biagi and Voytek, 2006)) yielded a kiln that was used for gold smelting, testified by the recovery of gold beads and gold platelets. This is interesting because, in other sites in the same area, a number of jewellery workshops have also been unearthed, which are attributed to Coţofeni culture and which yielded over 30 beads, made of bone, marble, shells (including *Spondylus*), as well as copper tools and small fragments of malachite (Lazarovici and Lazarovici, 2016).

In Peştera Ungurească, the small (diameter 0.41 - 0.52 cm) shell double-buttons were recovered from the Late Neolithic and transitional (Toarte Pastilate/Coţofeni) layers (2b and 1a.2) (Lazarovici and Lazarovici, 2013). The semi-worked double-button (PesB) selected for this study comes from layer 1a.2. The raw material had been identified as *Unio* sp., also because many broken *Unio crassus* valves (possibly associated with alimentation or tool-making) had been recovered throughout the sequence (Girod, 2010).

SI.2 Mollusc shells

Molluscs represent the second most diverse animal phylum in terms of numbers of described living species (Lydeard et al., 2004), however this global estimate may vary widely due to different synonym taxa names being used, the likelihood that there are many species not yet described from regions that are still unexplored, as well as issues in reconciling taxonomic classification based on morphology and molecular phylogenies. Molluscan “skeletons” have fascinated natural scientists for decades; in particular, bivalve shells have been fairly well investigated with regard to their microstructure, mineralogy and physicochemical characteristics (Boggild, 1930; Carter, 1990; Taylor et al., 1973, 1969). For example, mother-of-pearl (nacre) is one of the best-known and most widely studied shell microstructures. Nacre is present in four molluscan classes, including bivalves, gastropods, cephalopods and monoplacophorans. Among bivalves, nacre has a patchy distribution and is present in the following clades (orders): Nuculida, Mytilida (commonly known as true mussels, including *Modiolus*), Pterioida (which comprises, among others, the well-known and commercially-relevant pearl oyster *Pinctada* and the Mediterranean fan mussel *Pinna nobilis*), Trigoniida and Anomalodesmata. All these orders include exclusively marine animals. In addition, nacre is present in order Unionoida (freshwater bivalves). In bivalves, most of the nacreous microstructures belong to the “brickwall” type (sheet nacre (Taylor et al., 1969)).

Six different mollusc taxa have been selected for morphological and biomolecular comparative analysis for this study, choosing shells with structural features alike to archaeological ornaments and that are native to the local environment of the sites:

- two marine bivalves – *Ostrea edulis* and *Modiolus modiolus,* common in Northwestern European seas (also, oysters dominate the midden at Havnø).
- four freshwater bivalves belonging to order Unionida (Bolotov et al., 2016):
    - *Unio pictorum* (also known as painter's mussel) and *Unio crassus* (both belonging to genus *Unio*, family Unionidae).

- *Margaritifera margaritifera (*also known as freshwater pearl mussel) and *Pseudunio auricularius* (both belonging to family Margaritiferidae)*;*

Specimens of *U. crassus* and *P. auricularius* are archaeological (sub-fossil) shells from the sites of Peştera Ungurească and of Isorella (a Neolithic site in Italy (Girod, 2010)).

2.1. Ecology of the molluscan taxa investigated

Unionoids inhabit clean flowing waters (Bolotov et al., 2016) and freshwater molluscs account for a great portion of the total number of mollusc species facing extinction: a significantly higher proportion of freshwater molluscs is present in the IUCN Red List of Threatened Species than marine species. This is due to habitat degradation after the construction of dams, stream channelization, pollution and sediment toxicity, biological invasions and other human-driven impacts (Bolotov et al., 2016; Lopes-Lima et al., 2017). All the Unionoid shells studied in this research are present in the Red List of threatened species (http://www.iucnredlist.org) with *Unio crassu*s being in the endangered category and *Margaritifera* species at the point of critically endangered.

Ostrea edulis (the European flat oyster) has a natural geographical distribution ranging along the European Atlantic coast, from Norway to Morocco, and all along the Mediterranean as well as the Black Sea (Launey et al., 2002). This species is also becoming a concern to ecologists since oyster reefs and beds are, globally, one of the most endangered types of habitat. *Modiolus modiolus* (commonly named as horse mussel) is an Arctic-boreal species that is limited in distribution by warmer temperatures to the south, but occasionally specimens have been reported as far South as North West Africa.

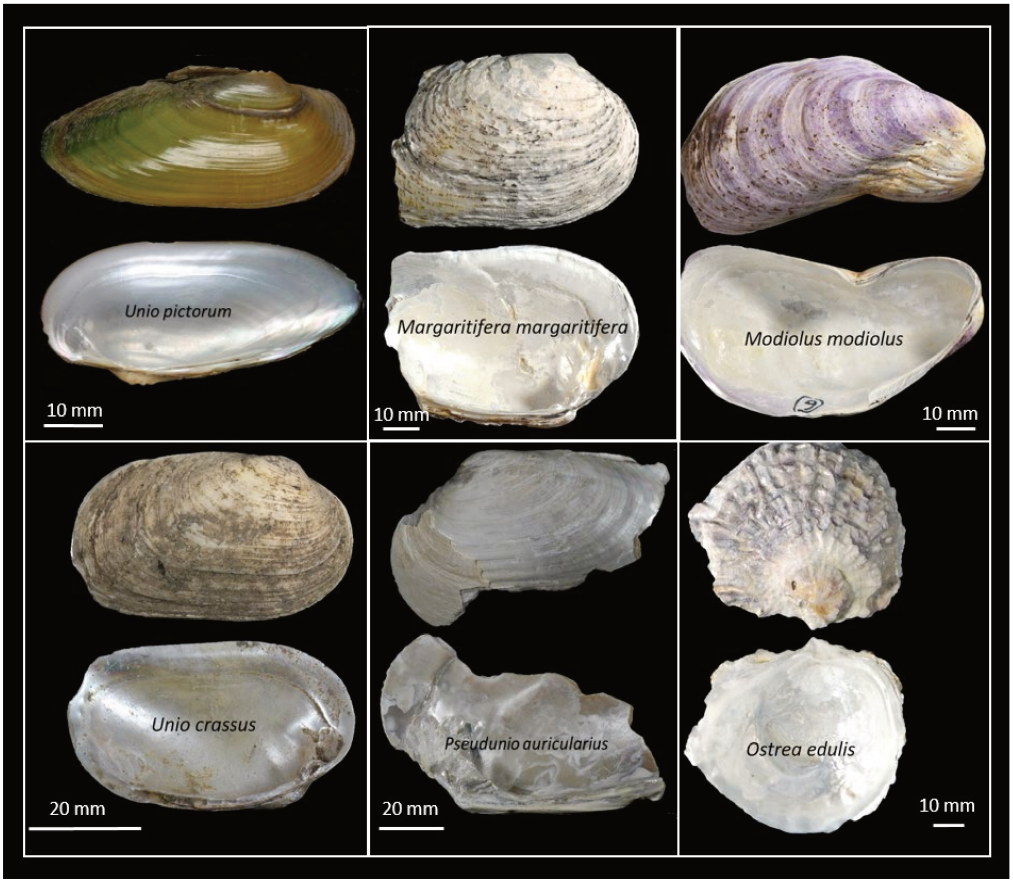


Figure S2.1: Marine and freshwater shells included in this study for comparative analysis: possible sources of raw material used for the manufacture of the double-buttons.

SI.3 Analytical methods and detailed results

The three sets of double-buttons were analysed using an integrated analytical approach across different facilities. Morphological analysis (SEM) was carried out at the Biogéosciences facility (mixed research unit 6282, CNRS/uB-FC, France); mineral phase characterization (FTIR in ATR mode) was obtained at the Department of Chemistry of the University Burgundy Franche-Comté (France). Isotope analysis of the Havnø and Hornstaad double-buttons were carried out the Light Stable Isotope Laboratory (School of Archaeological Sciences, University of Bradford, UK) and for Peştera Ungurească sample at the GISMO facility at the department of Biogéosciences (uB-FC, France). The AAR data were obtained at Department of Chemistry (University of York, UK). Sample preparation for proteomic analysis was carried out at CNRS/uB 6282 Biogéosciences facility (uB-FC) and Department of Life Sciences and Systems Biology (University of Turin,

Italy). LC-MS/MS analyses were performed at the MSAP CNRS laboratory, University of Lille and the Mass Spectrometry Biomolecules core facility, Molecular Biotechnology and Health Sciences Department, University of Turin.

3.1 Scanning Electron Microscopy

3.1.1 Shell morphology

The study of the morphology of shell microstructures is one of the most common approaches for taxonomic classification and can be applied also to the identification of archaeological samples. Furthermore, the mineralogical investigation of the shell layers, which can occur as either aragonite or calcite (polymorphs of calcium carbonate), can also help to distinguish different shell taxa.

- Unionoids are made of two fully mineralized layers, both aragonitic. The outer mineralized layer is composed of prisms that develop perpendicularly to the outer shell surface. The inner layer is nacreous, comprising extremely thin flat tablets that are superimposed in the typical “brickwall microstructure” arrangement, which is common in most nacreous bivalves.
- *Modiolus modiolus* shells have a finely prismatic calcitic outer layer and an aragonitic nacreous inner layer, the latter with alternating sheets of nacre and aragonitic myostracal prisms (Taylor et al., 1969).
- *Ostrea edulis* has a thin calcitic outer prismatic layer of regular and simple prisms and an inner layer of very fine folia (the "foliated calcite") parallel to the plane of the shell, except for the myostraca deposits, which are aragonitic (Carter, 1990). Discontinuous chalky layers are also observed. According to Bøggild, no trace of aragonite can be found in either recent or fossil Ostreidae (Boggild, 1930), but further research showed that oysters possess a small fraction of aragonite in their myostracum and inner ligament (Taylor et al., 1969).

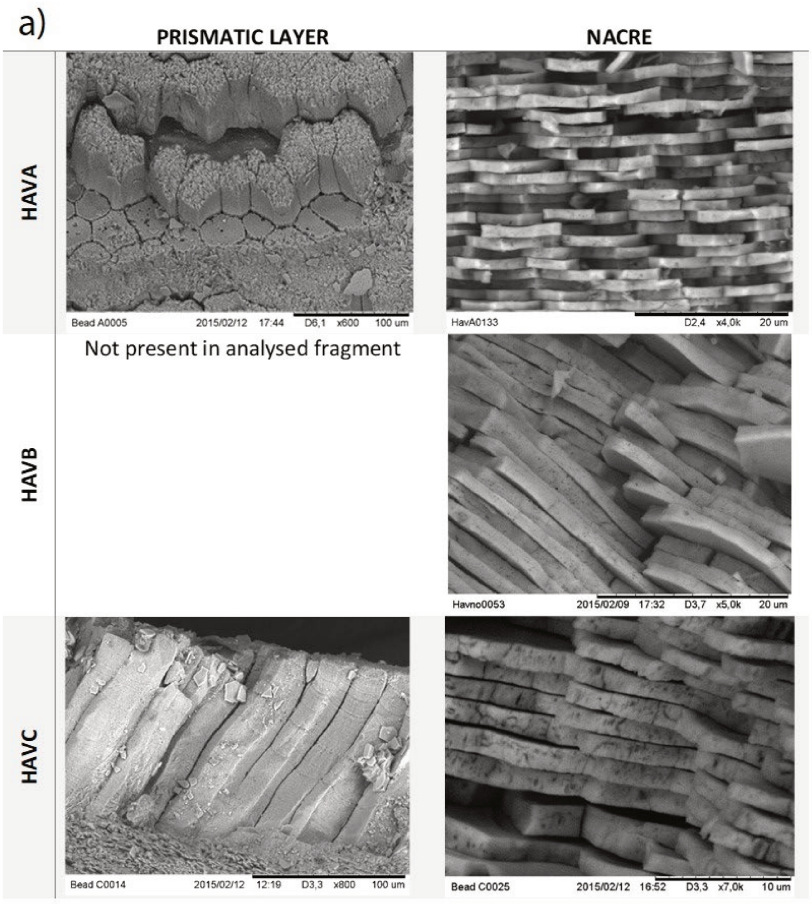
3.1.2 Analytical procedure

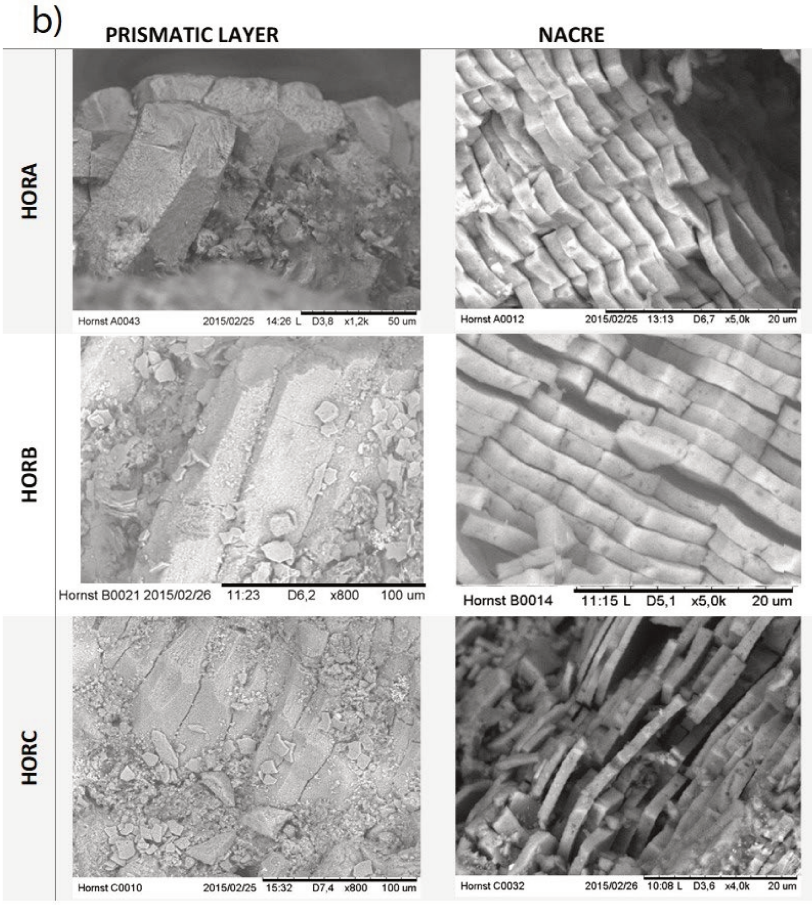
The microstructure of the archaeological double-buttons and the six reference shells was determined by means of SEM (Hitachi TM1000 Tabletop Microscope in low vacuum mode). Prior to the analysis, the samples were etched with 1 % (w/v) EDTA solution in a sonication bath, for two-three minutes for mollusc shells and up to one minute for archaeological ornaments, in order to provide better resolution of the topography of the surface. No carbon coating was applied.

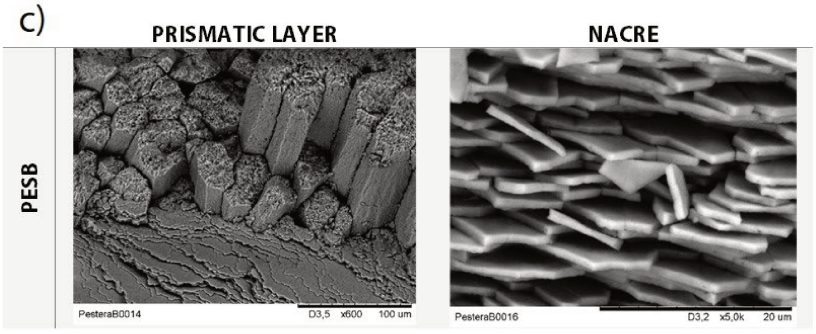
3.1.3 Results

All the double-buttons are clearly made of mollusc shell: nacre tablets and prisms are visible at both high and low magnification (Figure S3.1, a-c). All of them display a very similar nacro-prismatic microstructure, consisting of thick nacre platelets

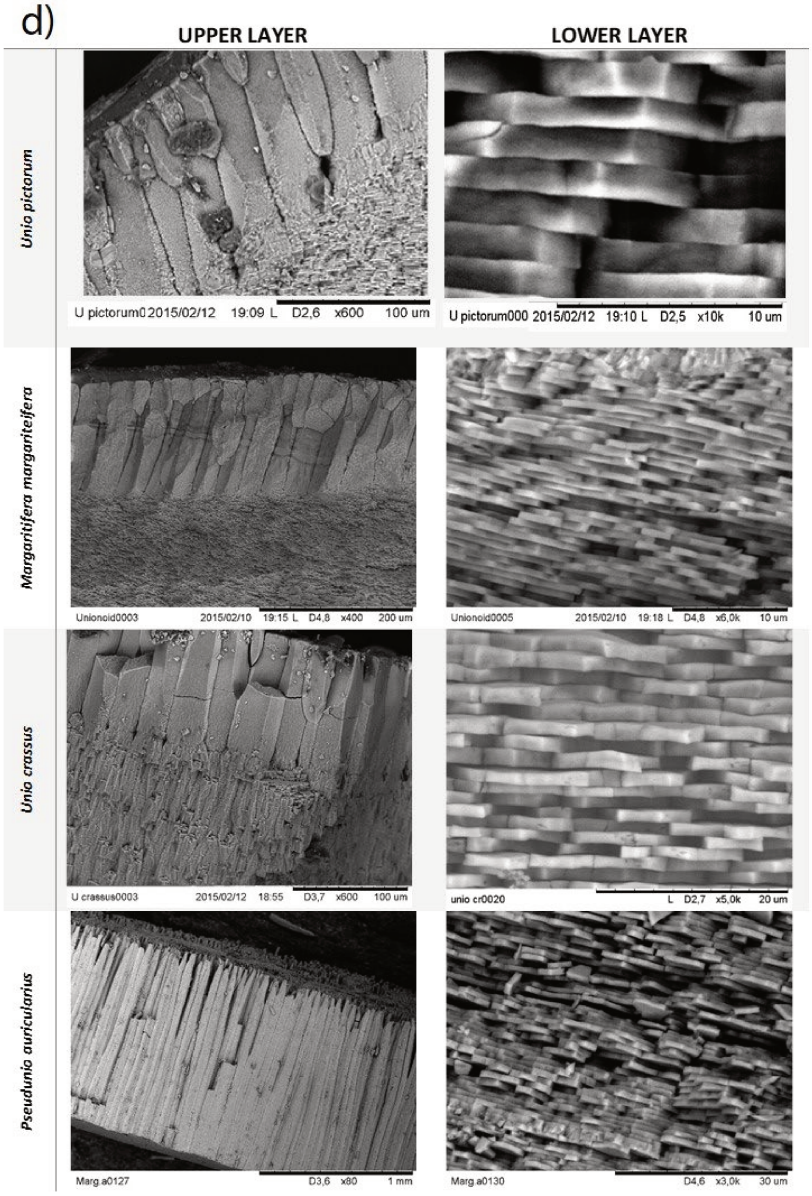
(around 1 µm) and long prisms (around 150 µm). The overall microstructure is similar that of Unionoid shells (Figure S3.1, d) and differs slightly from nacreous layer of *Modiolus modiolus* shell in terms of prism elongation. It does not resemble at all that of the marine oyster *Ostrea edulis* (Figure S3.1, e). All of the double-buttons, and especially the Hornstaad set (Figure S3.1, b), display clear signs of diagenetic alterations, resulting in the loss of connection between nacre tablets and the adjacent prisms.




b)
PRISMATIC LAYER
NACRE
HORA
HORB
HORC
Hornst A0043 2015/02/25 14:26 L D3,8 x1,2k 50 um
Hornst A0012 2015/02/25 13:13 D6,7 x5,0k 20 um
Hornst B0021 2015/02/26 11:23 D6,2 x800 100 um
Hornst B0014 11:15 L D5,1 x5,0k 20 um
Hornst C0010 2015/02/25 15:32 D7,4 x800 100 um
Hornst C0032 2015/02/26 10:08 L D3,6 x4,0k 20 um


c)
PRISMATIC LAYER
NACRE
PESB
PesteraB0014 D3,5 x600 100 um
PesteraB0016 D3,2 x5,0k 20 um

d)


UPPER LAYER
LOWER LAYER
Unio pictorum
U pictorum0 2015/02/12 19:09 L D2,6 x600 100 um
U pictorum000 2015/02/12 19:10 L D2,5 x10k 10 um
Margaritifera margaritefera
Unionoid0003 2015/02/10 19:15 L D4,8 x400 200 um
Unionoid0005 2015/02/10 19:18 L D4,8 x5,0k 10 um
Unio crassus
U crassus0003 2015/02/12 18:55 D3,7 x600 100 um
unio cr0020 L D2,7 x5,0k 20 um
Pseudunio auricularius
Marg.a0127 D3,6 x80 1 mm
Marg.a0130 D4,6 x3,0k 30 um

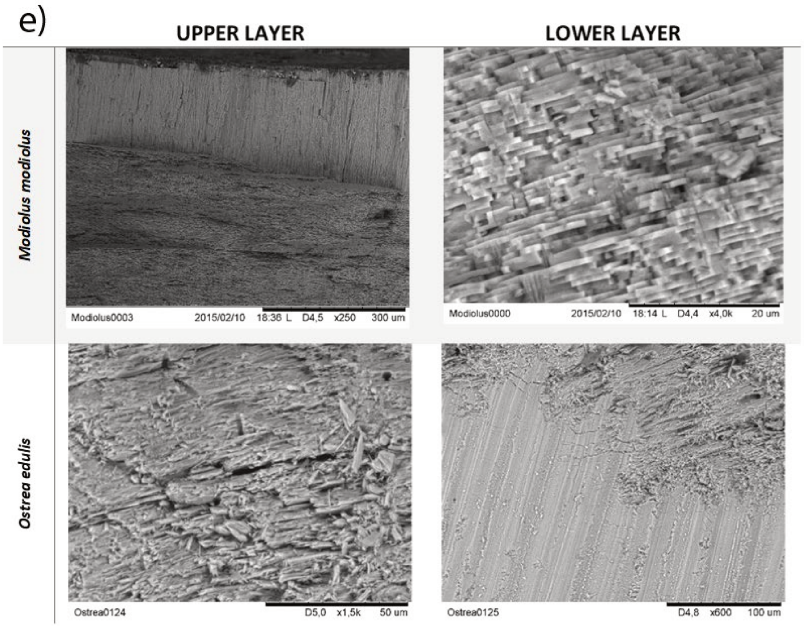


Figure S3.1: SEM microstructural analysis of archaeological double buttons (a-c) and mollusc shells (d, e). Double-buttons: a - Havnø (HavA, HavB, HavC), b - Hornstaad (HorA, HorB, HorC), c - Peştera Ungurească (PesB). Microstructure of shells: d - freshwater unionoid shells (modern *Unio pictorum*, *Margaritifera margaritifera* and sub-fossil *Unio crassus*, *Pseudunio auricularius*), e - marine taxa (*Modiolus modiolus* and *Ostrea edulis*).

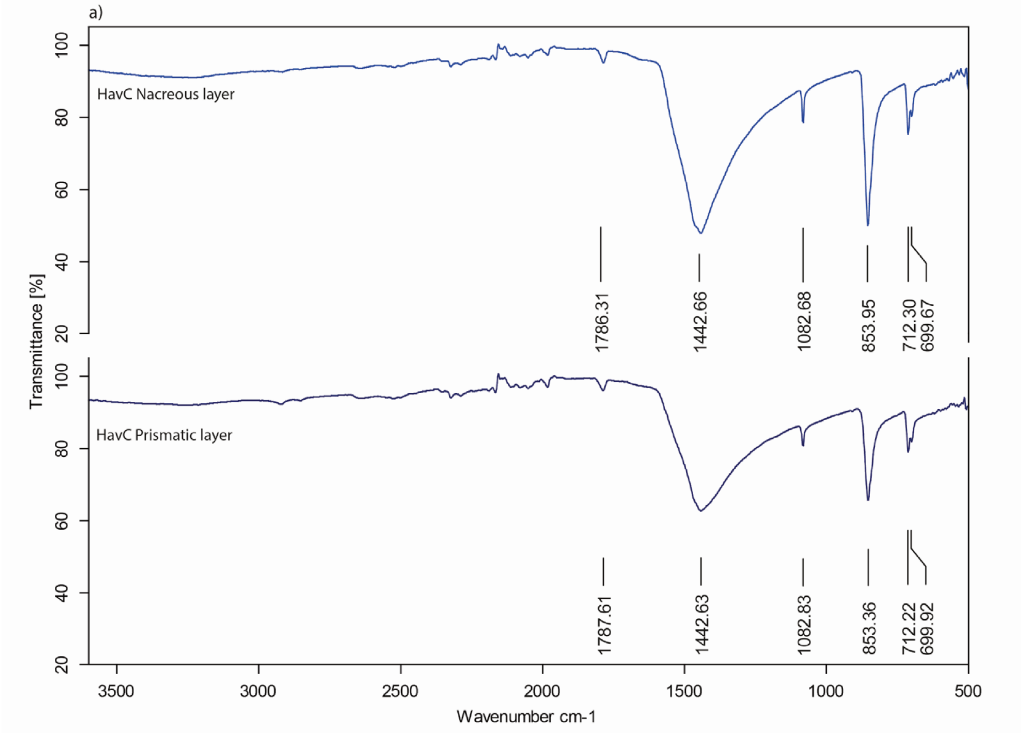
3.2 Infrared Spectroscopy

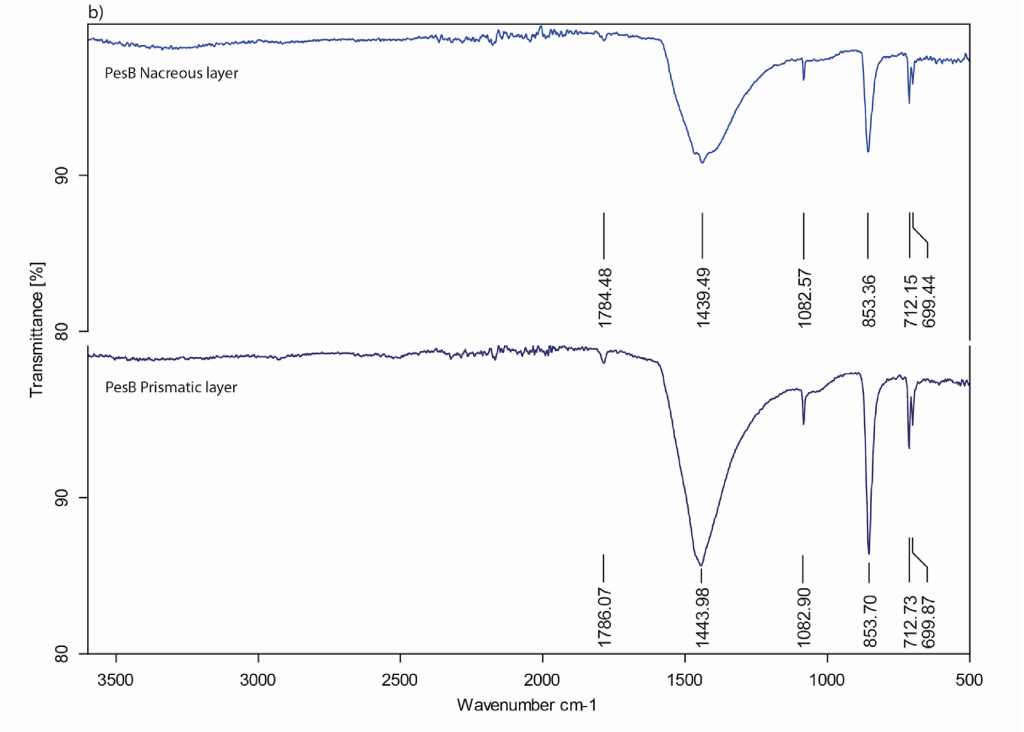
The identification of the mineral phases present in the archaeological double-buttons was obtained by Fourier Transform Infrared spectroscopy in attenuated total reflectance (ATR) mode. ATR spectra were recorded from small grains of material, delicately taken from the ornaments with a scalpel. The spectra were acquired with a Bruker Vector 22 instrument (BrukerOptics Sarl, France, Marne la Vallée) fitted with a GoldenGate attenuated total reflectance (ATR) device (SpecacLtd, Orpington, UK) in the 4000–500 cm^{-1} range (twelve scans at a spectral resolution of 4 cm^{-1}). Spectral analyses were performed with the OPUS software provided by the instrument manufacturer (BrukerOptics Sarl). The assignment of the different absorption bands was obtained by comparison with previous spectra descriptions available in the bibliography (Henry et al., 2017). Each double-button was sampled in “bulk” but the nacreous and prismatic layers of samples HavC and PesB were also analysed separately, in order to determine the mineralogical composition of each microstructural layer.

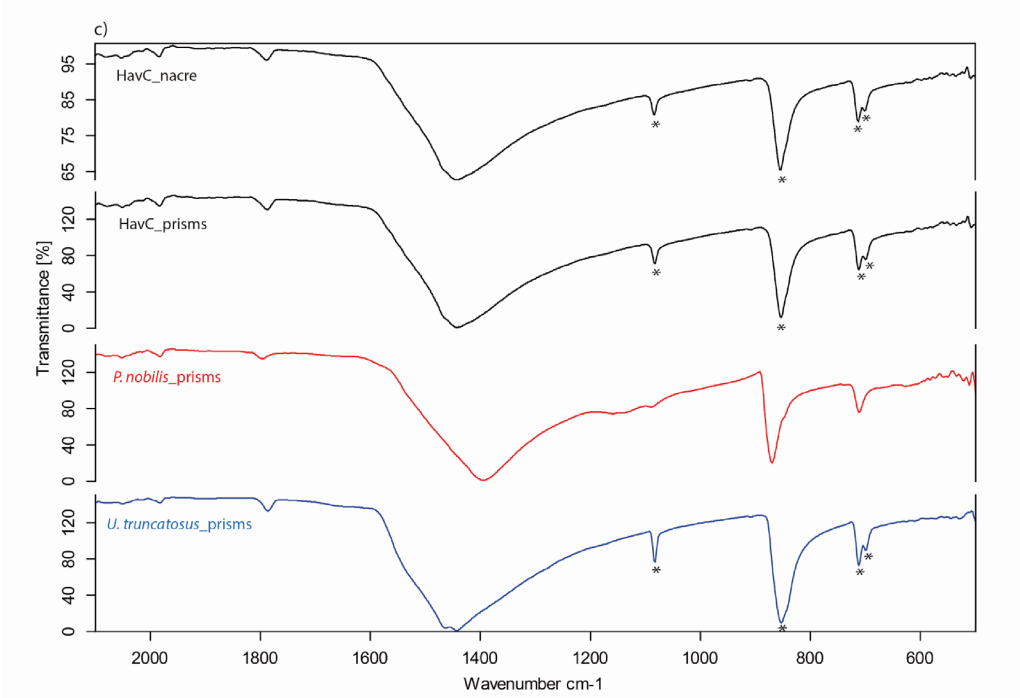
3.2.1. Results

The analysis of the nacre and prismatic layers from samples HavC and PesB showed the presence of distinctive aragonite absorption bands (Figure S3.2, a, b): a doublet at around 712 and 700 cm^{-1} (ν4, CO in-plane bending); a mid-size peak positioned at 853 cm^{-1}(ν2, out-of-plane bending) and a weak peak at 1082 cm^{-1} (ν1, symmetrical stretching). Calcite has only a single peak of ν4 at 711 cm^{-1}, its ν2 peak is shifted to 870 cm^{-1} and shows absence of the ν1 stretching at 1082 cm^{-1} (Loftus et al., 2015) - this is not observed in our samples, as can be seen comparing the double-button spectra to those of the purely-calcitic prisms of *Pinna nobilis* (Figure S3.2, c). The analysis confirms that the double-buttons are composed of aragonitic nacreous sheets with a thin aragonitic prismatic layer on top. The combination of these two aragonitic microstructures is common in unionoid bivalves.

Considering that aragonite is a less stable form than calcite and under natural diagenesis or strong heating biogenic aragonite re-crystallizes to calcite, infrared spectra can be used to assess the extent of degradation of the archaeological samples. It is particularly interesting for the Hornstaad double-buttons, as they have been exposed to heat (burning). FTIR-ATR spectra (Figure S3.2, d) from bulk samples obtained on each double-button show no presence of secondary calcite and thus no form of recrystallization - only the doublet (ν4, CO in-plane bending) characteristic of aragonite is present.


a)
HavC Nacreous layer
1786.31
1442.66
1082.68
853.95
712.30
699.67
HavC Prismatic layer
1787.61
1442.63
1082.83
853.36
712.22
699.92
Transmittance [%]
Wavenumber cm-1


b)
PesB Nacreous layer
1784.48
1439.49
1082.57
853.36
712.15
699.44
PesB Prismatic layer
1786.07
1443.98
1082.90
853.70
712.73
699.87
Transmittance [%]
Wavenumber cm-1


c)
HavC_nacre
HavC_prisms
P. nobilis_prisms
U. truncatosus_prisms
Transmittance [%]
Wavenumber cm-1

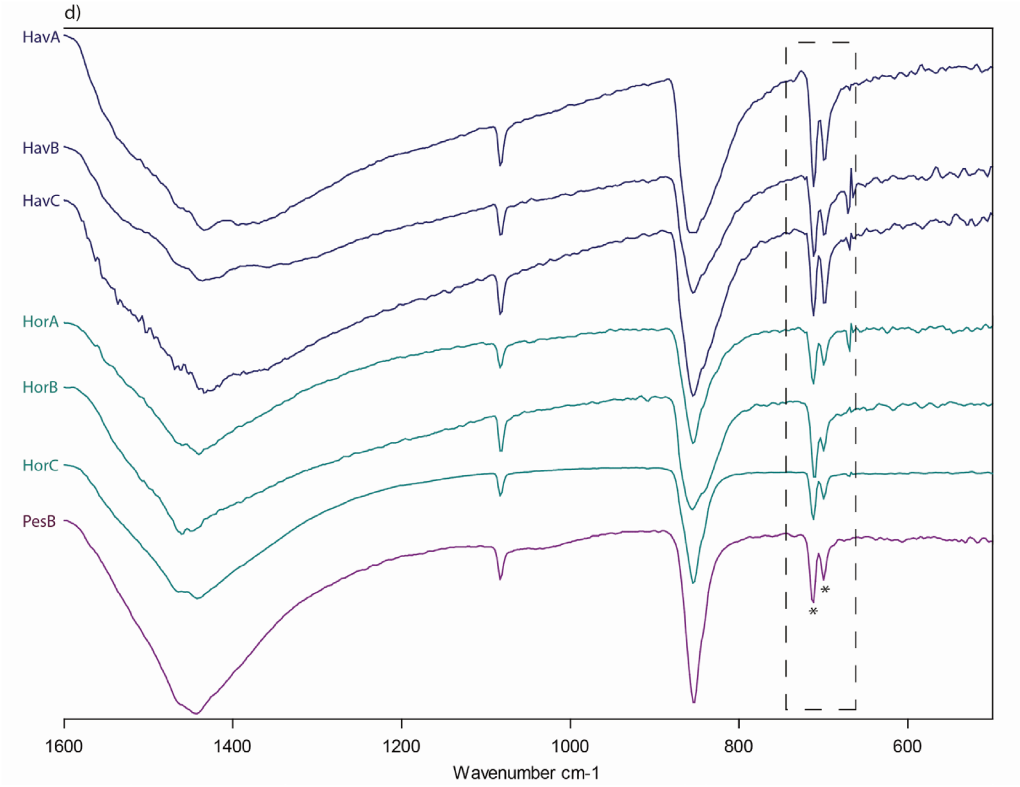

d)
HavA
HavB
HavC
HorA
HorB
HorC
PesB
Wavenumber cm-1

Figure S3.2: FTIR-ATR spectra of the double-buttons. Asterisks mark aragonite marker absorption bands: nacreous and prismatic layers of a) HavC, b) PesB; c) FTIR-ATR spectra comparison between HavC nacre and prismatic layers (black) with calcitic prismatic layer of *Pinna nobilis* (red) and aragonitic *Unio truncatosus* (blue), confirming the fully aragonitic mineralogy of both layers; d) FTIR-ATR spectra of all the double-buttons, sampled in "bulk": the presence of doublets (CO in-plane bending mode) at ~712 and ~700 cm^{-1} (dashed line) in all samples indicates absence of recrystallization of the biogenic carbonate.

3.3 Stable isotope analysis

Tracing the geographical origin of the raw material used to make archaeological ornaments is challenging; however, isotope geochemistry can provide palaeoenvironmental information on malacological fauna as well as shell ornaments (Bajnóczi et al., 2013; Bar-Yosef Mayer et al., 2012; Eerkens et al., 2005; Vanhaeren, 2004). This is because the stable isotopic composition of shells records the environmental conditions at the time when the organisms were active and deposited the shell mineral. In particular, the variation of $^{18}O/^{16}O$ ($\delta^{18}O$), which are relatively ^{18}O-enriched in oceans and ^{18}O-depleted in continental waters (Keith et al., 1964), can help to discriminate between freshwater *vs* marine carbonates. Similarly, the $^{13}C/^{12}C$ ($\delta^{13}C$) reflects the origin of the dissolved organic content in water and can be used to validate the inferences drawn from the analysis of $\delta^{18}O$.

3.3.1 Analytical procedure

Carbon and oxygen isotope analyses of biogenic carbonate (i.e. $\delta^{13}C_{carb}$ and $\delta^{18}O_{carb}$) for the Havnø and Hornstaad samples were measured at the Light Stable Isotope Laboratory (School of Archaeological Sciences, University of Bradford, UK) and for the Peştera Ungurească sample at the GISMO platform (UBFC, France). All samples were bleached for 48 hours using NaOCl (12% w/v) prior to analysis. The prismatic and nacreous layers of PesB were sampled separately.

Analytical procedure at Light Stable Isotope Laboratory (School of Archaeological Sciences, University of Bradford, UK): between 100 and 300 µg of the bleached calcium carbonate powders (Hornstaad and Havnø double-buttons) were loaded into 12 mL Exetainer ® tubes, and carbon and oxygen isotope values were determined by online phosphoric acid digestion at 70°C using a Thermo GasBench 2 preparation system coupled to a Thermo Delta V Advantage Isotope-Ratio mass spectrometer. Standardisation of $\delta^{18}O$ values against the V-PDB reference was undertaken using repeated measurements of international standards IAEA NBS-19, IAEA-CO-8 and IAEA-CO-1, as well as internal laboratory standards (Merck $CaCO_3$ and OES). The analytical precision of the instrument was better than ± 0.1 ‰.

Analytical procedure at GISMO platform (Biogéosciences, UBFC, France): small amounts of carbonate from each layer was removed from a bleached fragment of the Peştera Ungurească sample by scratching with a scalpel. The powders obtained (between 35 and 40 µg) were loaded into glass vials for isotopic analysis. Samples were reacted with 250 µl of 100 % phosphoric acid at 70 °C for 12 minutes. All isotopic values are reported in the standard δ-notation in permil (‰) vs VPDB. The reproducibility (2σ) of the IAEA NBS19 used as an external standard is better than 0.04 ‰ for the $\delta^{13}C$ and 0.08 ‰ for the $\delta^{18}O$.

3.3.2. Results

Stable isotope analyses yielded average $\delta^{18}O$ and $\delta^{13}C$ values ranging between -5.3±0.4 and -11.1±0.6 ‰ for Havnø (n = 3), -6.1±1.0 and -11.9±1.7 ‰ for Peştera Ungurească (n = 2) and -9.3±0.5 and -10.6±1.6 ‰ for Hornstaad (n = 3), respectively (Table S3.1).

The low average $\delta^{13}C$ and $\delta^{18}O$ values obtained for all of the samples suggest the inland provenance of the biogenic carbonates (Keith et al., 1964; Leng and Lewis, 2016). The average $\delta^{18}O$ values for Havnø and Hornstaad show a difference of ~4 ‰. The $\delta^{18}O$ of PesB is closer to that of Havnø, with an average of -6.1±1.0 ‰ between the two layers of the shell.

A sensible difference between $\delta^{18}O$ values probably implies different waters of provenance of the shells: $\delta^{18}O$ values obtained from molluscan carbonate reflect the average stable oxygen isotopic composition of the environmental waters and temperature (Verdegaal et al., 2005) and thus different palaeoenvironmental conditions in which the animals lived. Keith et al. (Keith et al., 1964) note that at progressively higher altitudes and higher latitudes there is a depletion of ^{18}O in continental waters. Indeed, Hornstaad is situated at the northern foot of the Alps and is influenced by waters running from the Swiss Alps, while Havnø lies at just 9 m above the sea level and Peştera Ungurească site can be characterized by Carpathian landscape but lies in a valley surrounded by low altitude (up to 700 m) mountains.

The $\delta^{13}C$ value of freshwater shells is mainly influenced by dissolved inorganic carbon, which typically display a wide range of values (e.g. -28‰ to -1‰) compared to seawater (0-1‰) (Aucour et al., 2003; Gillikin et al., 2009). However, the $\delta^{13}C$ values of freshwater bivalve shells can also be affected by some contribution of metabolic carbon (McConnaughey and Gillikin, 2008). The $\delta^{13}C$ in the three double-button sets show similar average values: -11.1±0.6 ‰ for Havnø, -10.6±1.6 ‰ for Hornstaad and -11.9±1.7 ‰ for Peştera Ungurească, and fall in the range of values reported for freshwater shells in European continental waters, including unionids (Aucour et al., 2003; Versteegh et al., 2010). The shell $\delta^{13}C$ values likely reflect similar environmental conditions, all indicating freshwater habitats. Taken together,

the $\delta^{18}O$ and $\delta^{13}C$ values suggest that the shells used to make the ornaments were probably sourced locally, from freshwater or freshwater-dominant environment (estuaries).

Table S3.1: Stable composition of the biogenic carbonate of the double-buttons.

Sample	$\delta^{13}C$ (‰)	$\delta^{18}O$ (‰)
HavA	-11.8 ± 0.19	-5.0 ± 0.09
HavB	-10.9 ± 0.09	-5.7 ± 0.17
HavC	-10.6 ± 0.07	-5.3 ± 0.08
HorA	-11.7 ± 0.07	-9.8 ± 0.11
HorB	-11.3 ± 0.07	-9.2 ± 0.07
HorC	-8.7 ± 0.07	-8.9 ± 0.06
PesB_n	-13.1 ± 0.01	-6.8 ± 0.03
PesB_p	-10.7 ± 0.01	-5.5 ± 0.03

3.4 Chiral Amino Acid Analysis

The chiral amino acids isolated and extracted from the intracrystalline fraction of proteins contained in shells can be used to obtain geochronological information, based on the extent of diagenesis or racemization (AAR) (Bosch et al., 2015; Demarchi et al., 2015, 2011; Ortiz et al., 2018; Penkman et al., 2011, 2007; Pierini et al., 2016). Additionally, the bulk amino acid composition can yield taxonomic information at a broad level (order), and this can be applied to shell ornaments (Demarchi et al., 2014). Here we used a well-established preparation procedure (see e.g. (Penkman et al., 2008)), which includes a 48-h bleaching step (NaOCl, 12% w/v) followed by acid hydrolysis (7M HCl, 24 h at 110°C) in order to obtain the composition and extent of racemisation of the total hydrolysable amino acids (THAA) of all shell ornaments. Hydrolysates were evaporated to dryness and rehydrated with a solution containing an internal standard (the non-protein amino acid L-homo-arginine) for analysis in duplicate by reverse-phase high-pressure liquid chromatography (RP-HPLC). One subsample was analysed for each of the Hornstaad and Peştera Ungurească, and two for each of the Havnø double-buttons (due to the larger size of the latter). Tables S3.2 and S3.3 report the THAA concentration and D/L values for each of the double-buttons.

The relative composition data (Figure S3.3) show that all double-buttons display similar bulk amino acid signatures, except for HorA, where the %Asx is lower than in the other examples. This suggests that the proteins in this sample have undergone rapid degradation, and this is likely due to their exposure to high

temperatures. PesB also displays low %Asx values, but this might be attributed to a species effect. The Principal Component Analysis (PCA) scores plot in Figure S3.4 shows that the bulk amino acid composition of the Havnø double-buttons is very similar to that of freshwater molluscs *Unio pictorum* and *Margaritifera margaritifera*. HorB and HorC are also falling in a similar area of the PCA plot (close to the tiny aragonitic gastropod *Rissoa*), while HorA and PesB both display different compositions. Diagenesis might be invoked to explain the results for HorA, but PesB cannot be easily classified using bulk amino acid composition data. The D/L values (Figure S3.5) are low for all amino acids. PesB is the sample with the lowest values, consistent with the slightly younger age of the Peştera Ungurească site compared to the other two. Notable exception are all the D/L values measured in HorA, as well as Asx D/Ls for HorB and HorC, which are higher than in all the other beads. This may indicate mild exposure to heat, as observed in other studies (Crisp, 2013; Demarchi et al., 2011). The fact that the double-buttons from Hornstaad display signs of possible burning is hardly surprising, given that the site was affected by a fire!, but this type of analysis is fundamental in order to assess whether the D/L values yield genuine information regarding the age of the specimen, and for any taxonomic inference.

Table S3.2: Total hydrolysable amino acid (THAA) concentrations measured in archaeological double-button samples (pmol/mg). Average and standard deviation were calculated on two analytical replicates. Values for Havnø include the average and standard deviation for the two subsamples taken from each double-button.

	[Asx]		**[Glx]**		**[Ser]**		**[Gly]**		**[Ala]**		**[Val]**		**[Phe]**		**[Ile]**	
	AV	SD	AV	SD	AV	SD	AV	SD	AV	SD	AV	SD	AV	SD	AV	SD
HOR-A	150	1	497	6	72	0	965	36	604	0	300	1	227	2	223	3
HOR-B	340	128	322	126	142	47	516	301	372	167	250	40	175	39	181	32
HOR-C	429	61	340	11	177	25	574	68	434	46	194	43	163	11	160	19
HAV-A	723	48	470	17	334	12	1274	82	644	57	262	25	212	20	163	11
HAV-B	711	87	470	45	346	25	1214	176	656	66	264	24	213	18	173	18
HAV-C	703	14	452	13	321	21	1132	228	602	19	241	5	204	6	138	8
PES-B	452	8	344	13	494	1	1605	66	699	1	368	1	282	1	264	13

Table S3.3: Total hydrolysable amino acid (THAA) D/L values measured in archaeological double-buttons. Average and standard deviation were calculated on two analytical replicates. Values for Havnø include the average and standard deviation for the two subsamples taken from each double-button.

	Asx D/L		**Glx D/L**		**Ser D/L**		**Ala D/L**	
	AV	SD	AV	SD	AV	SD	AV	SD
HOR-A	0.430	0.000	0.740	0.003	0.000	0.000	0.880	0.001
HOR-B	0.530	0.073	0.150	0.421	0.720	0.515	0.200	0.477
HOR-C	0.580	0.033	0.170	0.016	0.760	0.026	0.210	0.010
HAV-A	0.313	0.015	0.128	0.032	0.455	0.030	0.165	0.013
HAV-B	0.318	0.010	0.135	0.010	0.480	0.024	0.168	0.010
HAV-C	0.308	0.010	0.120	0.008	0.475	0.026	0.175	0.006
PES-B	0.310	0.000	0.080	0.000	0.280	0.000	0.100	0.000
	Val D/L		**Phe D/L**		**Ile D/L**			
	AV	SD	AV	SD	AV	SD		
HOR-A	0.740	0.001	0.730	0.004	0.850	0.029		

HOR-B	0.110	0.449	0.290	0.308	0.140	0.488		
HOR-C	0.080	0.019	0.320	0.026	0.100	0.039		
HAV-A	0.073	0.015	0.180	0.008	0.045	0.052		
HAV-B	0.075	0.013	0.193	0.013	0.065	0.044		
HAV-C	0.078	0.010	0.170	0.008	0.023	0.045		
PES-B	0.000	0.000	0.140	0.000	0.000	0.000		

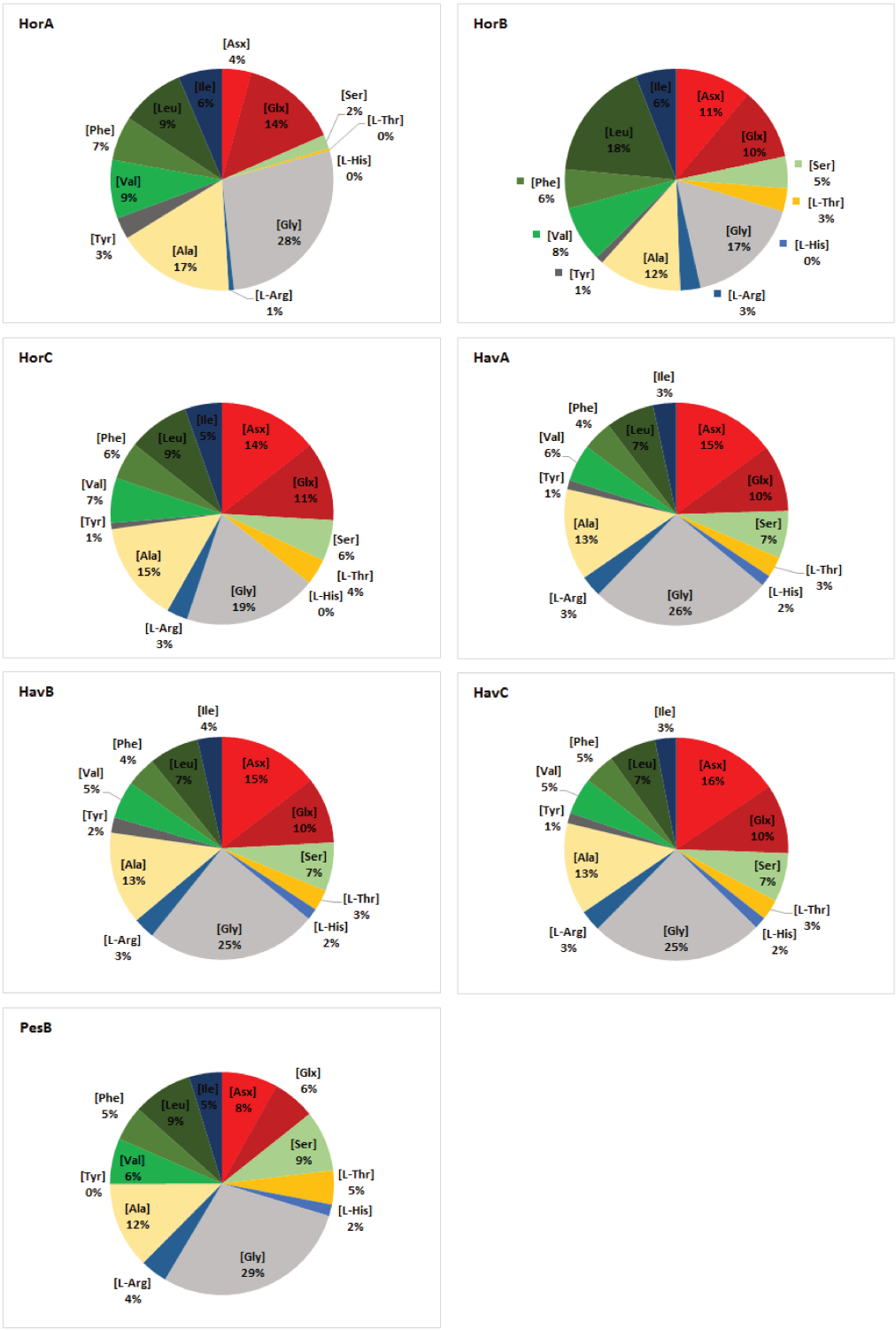


Figure S3.3: Relative THAA composition of double-button samples.

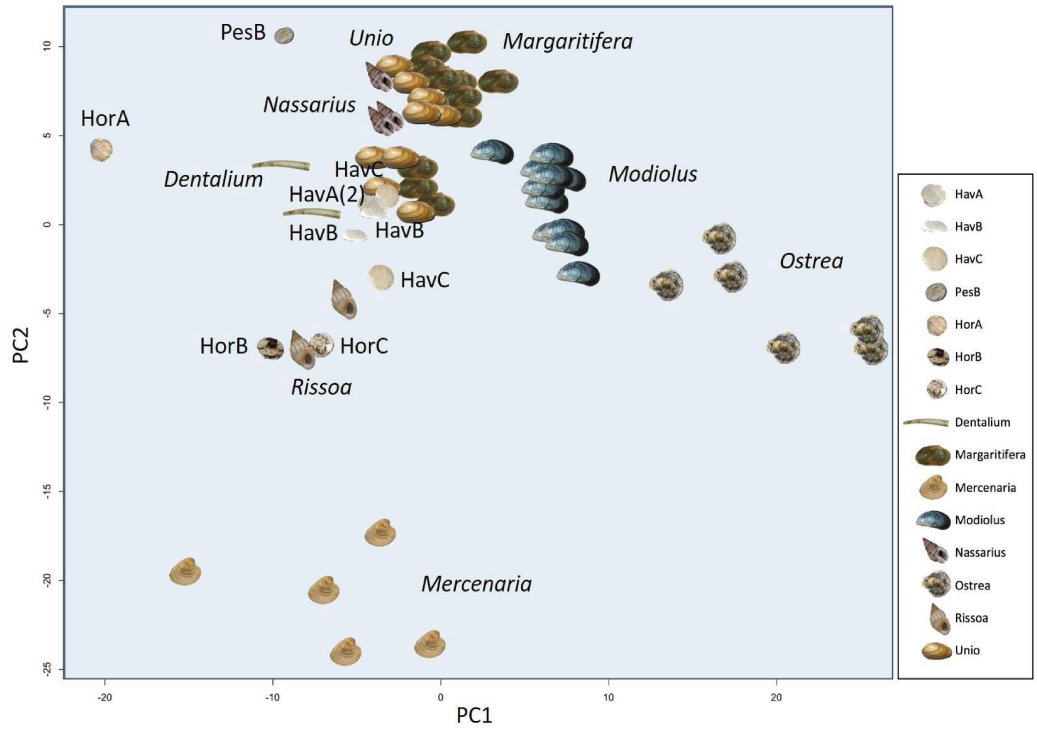


Figure S3.4: Principal Component Analysis (PCA) plot showing the similarity or differences between the amino acid composition of double-buttons and a range of shell taxa (reference taxa from (Demarchi et al., 2014)).

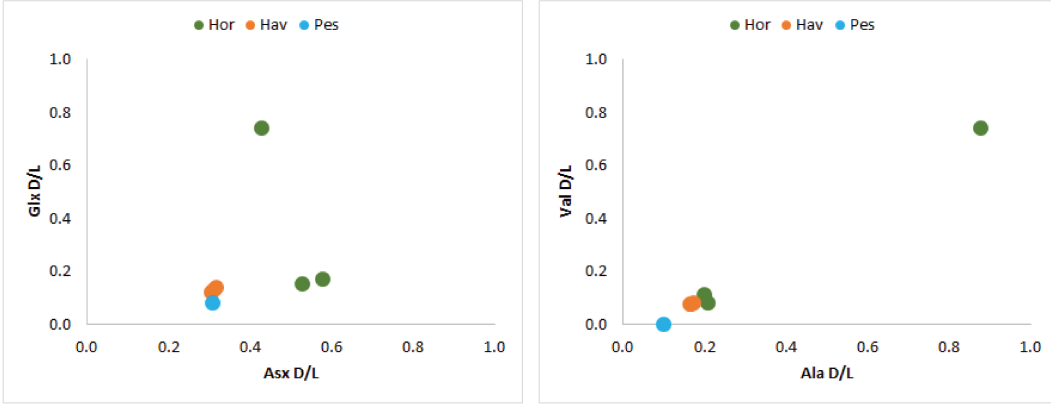


Figure S3.5: Total hydrolysable amino acid D/L values for all double-buttons (Glx vs Asx, left; Val vs Ala, right).

Overall, chiral amino acid data support the hypothesis that the shells used by the makers of the double-buttons were either fresh or recently dead and had not been collected from fossil deposits. This is important information in order to address the

debate over the added value of using fossil shells for making jewellery (e.g. Taborin, 1974).

3.5 “Shellomics”

3.5.1 Insights into the “Biomineralization toolkit”

‘Shellomics’ - a proteomics-based approach to study the processes of biomineralization, fills an important gap of knowledge in the evolution of shell biomineralization (Marin et al., 2016), and can also help to answer diverse environmental and archaeological questions. Despite the great attention that this field has gained in the past years and the numerous scientific articles published, considering there are at least 85,000 mollusc species documented, our knowledge of the proteins trapped inside the biomineral skeleton and their functions still remains in its infancy.

Just after the first studies of shell proteomes, which were driven by the idea of uncovering the ancestral shell “biomineralization toolkit” (Marin et al., 2012), there has been a steady increase in the numbers of mollusc shell proteins identified, owing to the development of ‘shellomic’ approaches, mainly by LC-MS/MS analysis (Marin et al., 2013). Shell matrix proteins, usually subdivided into acid-soluble (ASM) and acid-insoluble (AIM), have been extensively investigated for their role in biomineralization processes (Albeck et al., 1993; Arivalagan et al., 2017; Marin and Luquet, 2004) and with regard to their presence within the mineral skeleton (intracrystalline vs intercrystalline; related to specific structural layers) (Marin et al., 2007). Most of the research concerning biomineralization has been carried out on nacreous shells, especially marine pearl oysters (Marie et al., 2017; Marin et al., 2013, 2012), because of the importance of mother-of-pearl both commercially and as a source for biomimetic and bioinspired materials (Westbroek and Marin, 1998).

3.5.2. Protein extraction

Modern shell samples *Ostrea edulis, Modiolus modiolus, Unio pictorum, Margaritifera margaritifera* and sub-fossil shells *Unio crassus* and *Pseudunio auricularius* were scrupulously cleaned and abraded mechanically with a Dremel drill in order to remove the periostracal layer and surface contaminants. Shells were crushed and powdered to <200-micron grain size. Shell powders (Table S3.4) were bleached for 48 hours in NaOCl (2.8 % active chlorine) under constant agitation, then, bleach was removed by rinsing the powder in ultrapure water (5x) and ethanol (1x). Bleached dried shell powders were decalcified at 4°C using 10% acetic acid (ratio: 50 mL of acid for 1 g of shell powder) under constant stirring overnight. The solution was centrifuged for 30’ at 4500 RPM at 5°C in order to separate the acid-soluble matrix (ASM) from the acid-insoluble matrix (AIM). For modern shells, the supernatant (containing ASM proteins) was passed through a 5 µm filter, then

ultrafiltered through a 10kDa cut-off membrane (Amicon ultrafiltration cell) to reduce the volume to ~15 mL and finally dialysed (in a 1kDA MW cut-off membrane) for 2 days against 1 L water (5 water changes). The sub-fossil shell samples were ultrafiltered using 10kD ultrafiltration tube (centrifuged at 4500 RPM to reduce the volume to 0.75 mL) and washed with water (10 times). The AIM matrix was rinsed several times via cycles of resuspension in ultrapure water, centrifugation and removal of the supernatant. All matrices were lyophilized.

The three archaeological sample sets (HavABC, HorABC, PesB) were prepared as follows: fragments were crushed with a micro pestle directly in an eppendorf and submerged in 2.8 % NaOCl for 3 hours, then rinsed (5 times with water and once in EtOH). Bleached powders were decalcified in 10 % cold acetic acid under constant mixing (using the same proportion of 50 µL of acid for 1 mg of shell powder) and upon the completion of demineralization, the solution was centrifuged at 13 000 rpm for 30 min to separate AIM and ASM. AIM were rinsed several times in water and lyophilised. The ASM of HavABC & HorABC samples were passed through a 3 kDa MW cut-off VIVASPIN filter and rinsed with ultrapure water (using a total of 6 mL H_2O), to obtain a final volume of about 750 µL of ASM concentrate, which was then lyophilised. The ASM of PesB, due to the smaller amount of sample, was directly placed in a 3.5 kDa cut-off micro dialysis cassette (Slide-A-Lyzer™ G2 Dialysis Cassettes, 3 mL capacity, ThermoFisher) and dialysed in 1L of ultrapure water with gentle stirring (changing the water 5 times). The extract was then collected and lyophilised.

Table S3.4: Mass of biogenic carbonate analysed for proteomics

	Sample	**Powder mass (mg)**
Archaeological samples	HavA	92.9
	HavB	69.6
	HavC	172.5
	HorA	49.23
	HorB	47.34
	Horc	47.23
	PesB	32.48
	Sample	**Mass (g)**
Freshwater unionoid shells	*U. pictorum*	10
	U. crassus	3
	M. margaritifera	10
	P. auricularius	3
Marine shells	*O.edulis*	10
	M.modiolus	10

The purified protein extracts (of both reference shells and archaeological double-buttons) were resuspended in a buffer (50 mM ammonium bicarbonate, pH 7.5-8). The reduction and alkylation of disulphide bonds was achieved using 1M DL-Dithiothreitol (Sigma, Canada) for 1 hr at 60°C and 0.5M iodoacetamide (Sigma,

USA) for 45 min at room temperature in the dark. Each sample was split into two aliquots for the enzymatic digestion with trypsin ('T') and elastase ('E'). Digestion was carried out overnight at 37°C by adding: 4 µL trypsin (0.5 µg/µL; Promega, 2800 Woods Hollow Road Madison, WI 53,711 USA) for 'T' subsamples or 4 µL elastase (1 µg/µL; Worthington, Lakewood, NJ, USA) for 'E' subsamples. Digestion was stopped with 10 % TFA (to a final concentration of 0.1 % TFA), samples purified using C18 solid-phase extraction tips (Pierce zip-tip; Thermo-Fisher) and evaporated to dryness. Trypsin and elastase digests for each sample were resuspended in 0.1 % TFA and combined prior to LC-MS/MS analysis.

3.5.3. LC-MS/MS analysis

Reference shells *Ostrea edulis, Modiolus modiolus, Unio pictorum, Margaritifera margaritifera* and two sets of beads - Hav(ABC) and Hor(ABC) were analysed at MSAP CNRS laboratory, University of Lille; sample PesB and two sub-fossil reference shells, *Unio crassus* and *Pseudunio auricularius,* were analysed at the Mass Spectrometry Biomolecules core facility, Molecular Biotechnology and Health Sciences Department, University of Turin.

Proteomic analysis (Lille)

Analysis of peptides were performed using a nanoflow HPLC instrument (U3000 RSLC Thermo Fisher Scientific) coupled to a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) equipped with a nanoelectrospray ion source. A volume of 1 µL of peptide mixture was loaded onto the preconcentration trap (Thermo Scientific, Acclaim PepMap100 C18, 5 µm, 300 µm i.d x 5 mm) using a partial loop injection and a flow rate of 10 $\mu L.min^{-1}$ (duration 5 min) with buffer A (5% acetonitrile and 0.1% formic acid). The peptide mixture was then separated using a nanocolumn (Acclaim PepMap100 C18, 3 µm, 75 mm i.d. × 500 mm) and a linear gradient of 5–40% buffer B (75% acetonitrile and 0.1% formic acid) at a flow rate of 250 $nL.min^{-1}$ and a temperature of 45°C. The total duration of an LC MS/MS run was 120 min. MS data were acquired using a data-dependent top 20 method that selects the 20 most abundant precursor ions from the survey scan (400–1600 m/z range) for HCD fragmentation. The dynamic exclusion duration was 60 s. The isolation of precursors was performed with a 1.6 m/z window and MS/MS scans were acquired with a starting mass of 80 *m/z*. Survey scans were acquired at a resolution of 70,000 at *m/z* 400 (AGC set to 106 ions with a maximum fill time of 180 ms). Resolution for HCD spectra was set to 35,500 at *m/z* 200 (AGC set to 105 ions with a maximum fill time of 120 ms). Normalized collision energy was 28 eV. The underfill ratio, which specifies the minimum percentage of the target value likely to be reached at maximum fill time, was defined as 0.3%. The instrument was run with the peptide recognition mode (i.e. from 2 to 8 charges) and exclusions of singly

charged ions and unassigned precursor ions. Fifteen blanks of 1 h and one reference sample (80 fmol of Cytochrome C digest; Thermo Fisher Scientific) were injected and controlled (e.g. retention time, sensitivity, carry over, contaminations) between each sample injection.

Proteomic analysis (Turin)

An Ultimate 3000 Dionex nanoHPLC instrument coupled with an Orbitrap Fusion (Thermo Scientific, Milan, Italy) mass analyser were used. The separation was achieved using a PepMap RSLC C18, 2 µm, 100 Å, 75 µm × 50 cm column (Thermo Scientific) and a PepMap C18, 5 µm × 5 mm, 100 Å preconcentration column (Thermo Scientific). The eluent used for preconcentration step was 0.05 % trifluoroacetic acid in water/ acetonitrile 98/2 and the flowrate was 5 µL/min. The eluents used for chromatographic separation were 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile/water 8/2 (solvent B) in a program which was initially isocratic at 5:95 (A:B %) for 5 minutes, increased to 75:25 in 55 minutes, run up to 60:40 in 6 minutes, and to 10:90 in 5 minutes. Recondition time was 20 minutes. The injection volume was 1 µL and the flow rate 300 nL min^{-1}. The nanocolumn was provided with the ESI source. The mass spectrometry parameters were: positive spray voltage 2300 (V), sweep gas 1 (Arb) and ion transfer tube temperature 275 ºC. Full scan spectra were acquired in the range of *m/z* 375-1500 (resolution 120000 @ *m/z* 200). MS^n spectra in data dependent analysis were acquired in the range between the ion trap cut-off and precursor ion *m/z* values. HCD collision energy was fixed at 28%, orbitrap resolution 50000 and the isolation window was 1.6 *m/z* units.

3.5.4 Reference datasets used for proteomics data analysis

Publicly available reference sequences for molluscs are far from abundant, given the size of the phylum. Most of the proteomic studies on bivalve nacre proteins have been carried out on sub-class Pteriomorphia and very limited information is known of other nacre bearing bivalves such as freshwater pearl mussels (Marie et al., 2017).

We performed two separate searches for each sample, using:

- A “Protein” database, downloaded from NCBI (https://www.ncbi.nlm.nih.gov/) on 15/02/2018 and including 633,061 sequences (restricting the taxonomy to “Mollusca”)
- An “EST” database downloaded from NCBI (https://www.ncbi.nlm.nih.gov/) on 15/02/2018 and including 1,149,723 sequences (restricting the taxonomy to “Mollusca”)

The Protein database nominally includes a large number of species (e.g. more than 3000 bivalves), but only a few organisms are represented by a significant number of sequences (e.g. the marine bivalves *Crassostrea gigas*, *Mizuhopecten yessoensis*, *Crassostrea virginica*, *Mytilus galloprovincialis*, *Pinctada fucata*). The top-thirty freshwater bivalves represented in this database are (the number of sequences in brackets): *Cumberlandia monodonta* (561), *Popenaias popeii* (496), *Potomida littoralis* (357), *Elliptio hopetonensis* (269), *Dreissena presbensis* (200), *Cyprogenia aberti* (198), *Elliptio dariensis* (197), *Elliptio icterina* (182), *Obovaria jacksoniana* (158), *Unio tumidus* (129), *Anodonta anatina* (105), *Corbicula fluminea* (101), *Unio delphinus* (99), *Margaritifera falcata* (98), *Toxolasma parvus* (97), *Pyganodon grandis* (97), *Hyriopsis cumingii* (95), *Unio mancus* (89), *Utterbackia imbecillis* (89), *Utterbackia peninsularis* (86), *Lampsilis cardium* (84), *Strophitus radiatus* (84), *Unio elongatulus* (82), *Quadrula pustulosa* (76), *Venustaconcha ellipsiformis* (73), *Sinanodonta woodiana* (67), *Anodonta cygnea* (62), *Unio crassus* (61), *Elliptio complanata* (60). However, the sequences are usually not relevant to the shell or the shell mantle: for example, all 61 *Unio crassus* sequences refer to two proteins: cytochrome oxidase subunit I and NADH dehydrogenase subunit.

The EST sequence database contains 75 molluscan species (25 gastropods, 38 bivalves, 8 cephalopods, 2 polyplacophorans and 2 shell-less molluscs). For bivalves, the EST sequences include well-studied pearl-producing species, most of which belonging to the Pterioida order (*Pinctada* sp.), but also the edible oyster and the mussel (*Ostreoida* and *Mytilidae* sp.). Only 2 EST sets correspond to freshwater bivalves: *Lamellidens marginalis* (native to southeast Asia) and *Hyriopsis cumingii* (commonly known as Triangle Sail Mussel, native to China and Vietnam), both belonging to order Unionoida.

3.5.5 Peptide and protein identification

Product ion spectra of reference shells and archaeological double-buttons were analysed using PEAKS Studio (v. 8.5, Bioinformatics Solutions Inc. (BSI)) (Ma et al., 2003) and searched separately against the Mollusca Protein and Mollusca EST databases, including the search of common laboratory contaminants (cRAP; common Repository of Adventitious Proteins: http://www.thegpm.org/crap/). Search parameters were defined assuming no enzyme digestion, fragment ion mass tolerance of 0.05 Da and a parent ion tolerance of 10 ppm. Results obtained by SPIDER searches (i.e. including all possible modifications) were used for peptide identification and protein characterization, choosing the following threshold values for acceptance of high-quality peptides: false discovery rate (FDR) threshold 0.5%, protein scores $-10\lg P \geq 40$, unique peptides ≥ 2, de novo sequences scores (ALC %) ≥ 50. In some instances, (indicated in the main text), less stringent

parameters of −10lgP ≥ 20, unique peptides ≥ 1, de novo sequences scores (ALC %) ≥ 50 were also used to identify the presence or absence of specific proteins. The peptide sequences supporting shell protein identifications were individually checked using the BlastP tool (https://blast.ncbi.nlm.nih.gov/), and any sequences that were homologous to common laboratory contaminants were excluded from any further analysis.

3.5.6. Results

Supplementary File 4 “Proteomic dataset” summarizes the main proteins identified in the reference shell samples and the double-buttons, grouped by taxon. Here we include a brief overview of several freshwater shell ‘specific’ matrix proteins detected in our samples:

- Hic74 (GenBank: ARG42316.1) is an acidic (IP 4.68), alanine and glycine rich matrix protein, supposedly involved in nacreous layer formation (Liu et al., 2017a). Its structure consists of poly-alanine blocks (at least 17) with short acidic motifs, a “GS loop” type coil structure of lustrin A (Wustman et al., 2002) at the C-terminus, the C-terminal 30 residues of which consist of short acidic-basic motifs. It is probably a structural, silk-like protein, providing mechanical function (a system that dissipates the cracks) with short acidic sites to bind the mineral.
- Hic52 (GenBank: ARH52598.1) is a very basic (IP 10.82), glycine (almost 29 %) and alanine rich protein. It has poly-glutamine and poly-glycine blocks with several degenerate repeats of different lengths along the sequence. It was reported to have collagen-like structure (Liu et al., 2017b) and probably plays a structural role in nacre formation.
- Silkmapin (GenBank: AIZ03589.1) and its isoforms (nasilin 1 & 2 (GenBank: ASQ40996.1 & ASQ40997.1)) are glycine-rich (>34 %), non-acidic proteins with a structural function and probably play a role in the deposition of both nacre and prismatic layers (Liu et al., 2015).
- Chitin deacetylase enzyme (present in isoforms A & B (GenBank: AFO53262.1 & AFO53263.1)) is a slightly basic (IP 8.34) protein, with a chitin deacetylation function and possibly has a role in the 3D structuring of the mineral framework.
- Carbonic anhydrase 3 (*Hyriopsis cumingii*, GenBank: ARG42317.1), is an asparagine (10.2 %) and threonine (9.3 %) rich, acidic (IP 5.9) protein. The active enzymatic domain comprises most of the sequence, except for the C-terminal 40 residues, which have completely biased amino acid composition, enriched in asparagine and acidic residues. This protein is typical of a CA associated to calcium carbonate biominerals, as it possesses

a supernumerary low complexity domain (Le Roy et al., 2014), which would promote the binding of the enzyme onto the mineral surface.

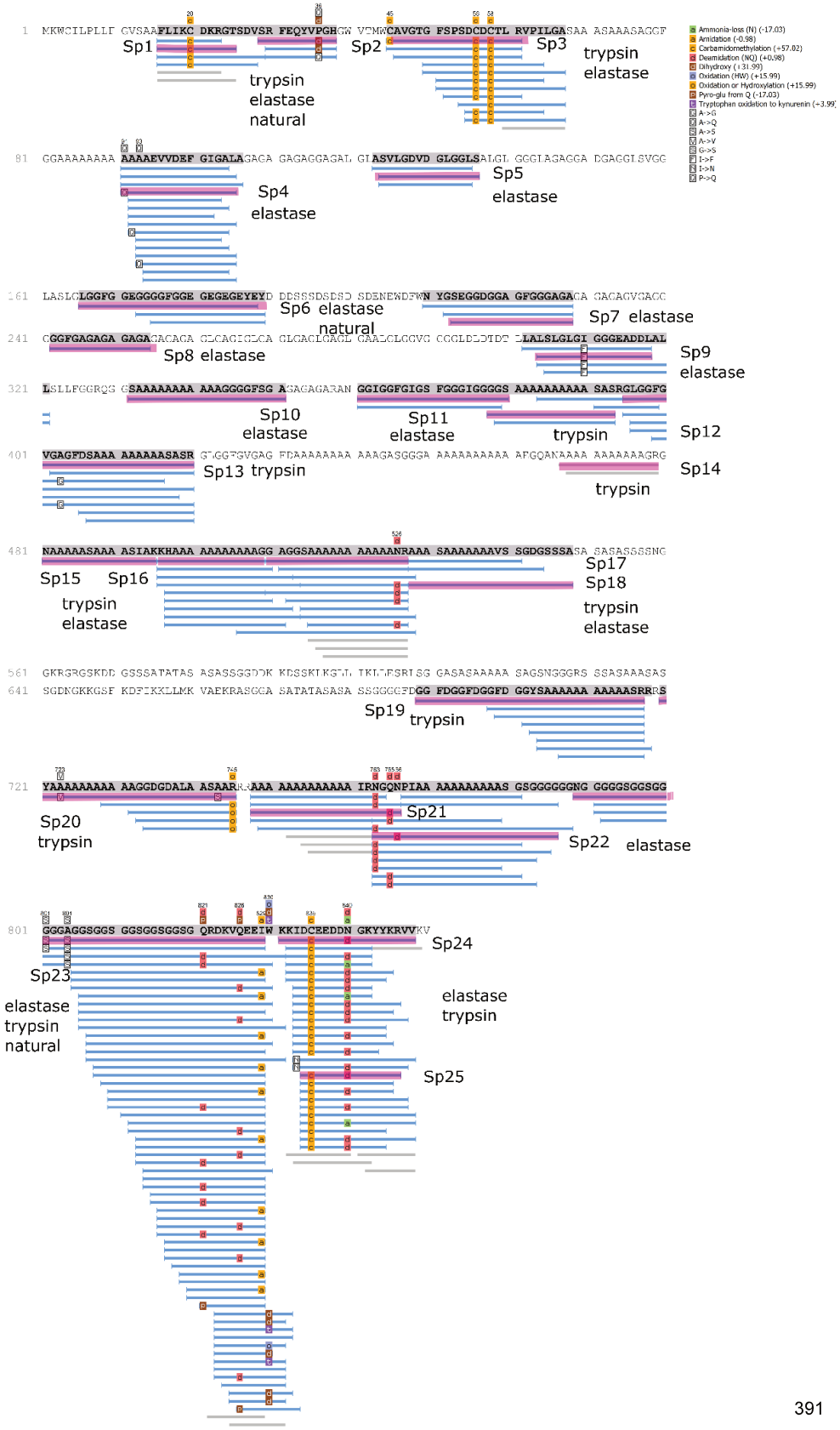
3.5.7 Protein identification in reference unionoid shells

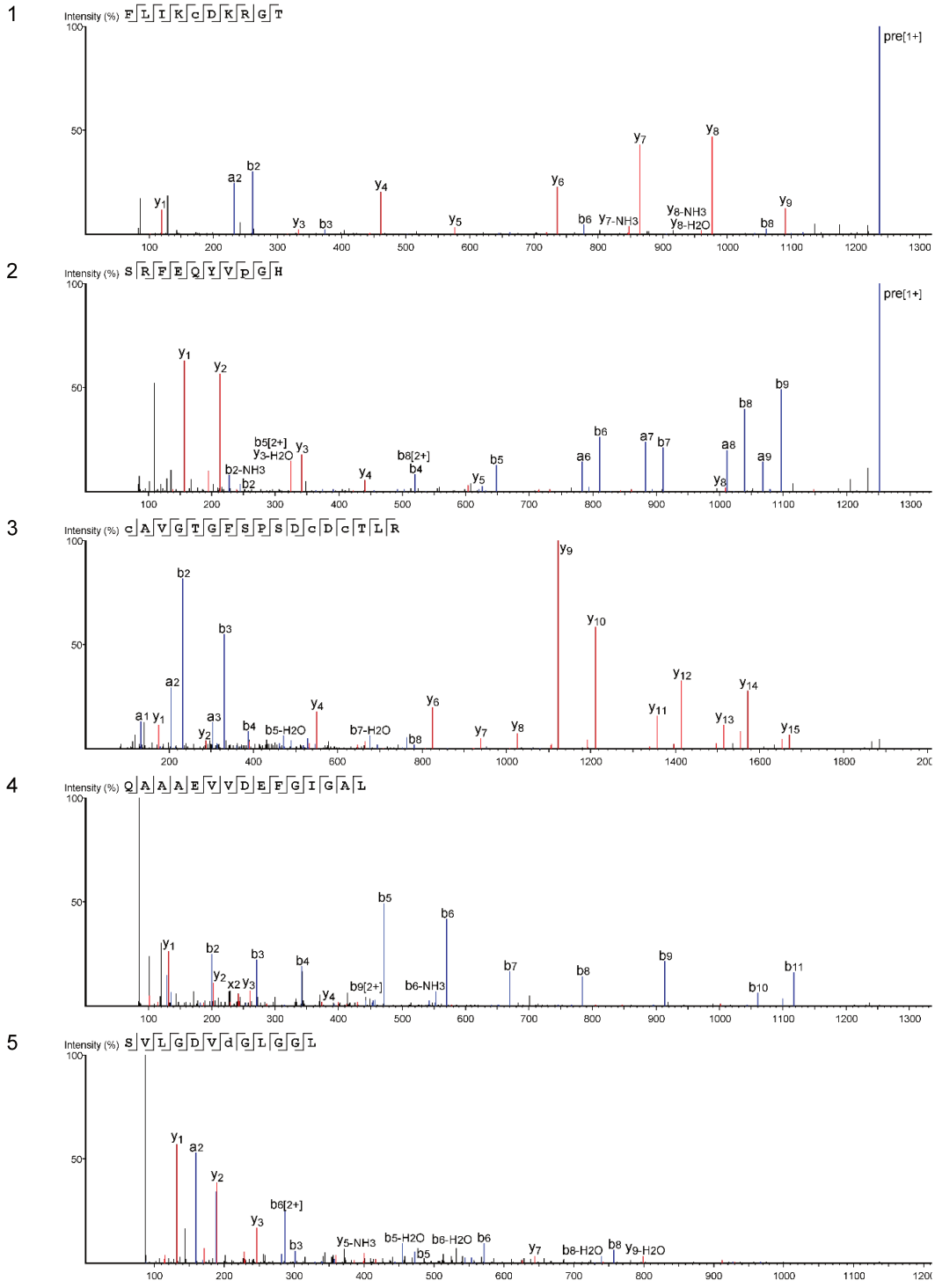
All of the reference unionoid shells gave hits to proteins of the freshwater bivalve *Hyriopsis cumingii.* Protein Hic74 was the top hit protein identified in all of the unionoids, with coverages varying from 34% in *Margaritifera margaritifera* to 54% in *Unio pictorum* (Figure S3.6) - including the sub-fossil shells of *Pseudunio auricularius* (49 %) and *Unio crassus* (50 %). Another significant *H. cummingii* protein in unionoids was Hic52, which was found in all the analysed freshwater shells with protein coverage up to 17 %. Silkmapin (and its isoforms nasilin 1, nasilin 2) were found in both *Unio* taxa (with coverage up to 17 %), while in *M. margaritifera* and *P. auricularius* they were only identified using less stringent protein identification parameters (see section SI-3.5.5). We found carbonic anhydrase of *H. cumingii* in *Unio* species and in *M. margaritifera*, while chitin deacetylase (isoform B) was only present in *U. pictorum*. Upsalin, a protein originally identified and described from the *U. pictorum* shell matrix (Ramos-Silva et al., 2012) was detected only in *Unio* species. In some of the unionoid samples we also detected some minor *Hyriopsis* proteins such as Krichin, Hic31, GTRPB5 and GTRPB7, which have not been described in the literature before and for which the function is currently unknown.

Proteins hits from shells other than *H. cummingii* were also found in the unionoid shell samples, such as MSI60-like and insoluble matrix proteins (from *Pinctada* species, the marine pearl oysters), poly-Ala protein Shelk2 and spidroin-1-like (from *Crassostrea* species, the marine oysters) and collagen-like protein precollagen D (*Mytilus* species). MSI60 is a structural protein of nacre and, together with Shelk2-like, spidroin-1-like and precollagen D, could be classified as a silk-like, repetitive low-complexity domain type (RLCD) protein exhibiting poly-Ala motifs and repetitive Gly domains alternating with Ala, Ser and Leu. RLCD type proteins are structure-specific, with a molecular function that is probably not exclusively restricted to nacre deposition (Marin and Luquet, 2004; Sudo et al., 1997) but to the overall biomineralization of nacro-prismatic structures (Gao et al., 2015; Marie et al., 2017). Supporting this, proteins such as MSI60-like was identified in this study in another nacreous shell, the marine mussel *Modiolus modiolus,* but were not present in the foliated calcitic shell of *Ostrea edulis.* This supported the lack of taxonomic specificity of RLC domains. Unfortunately, it still remains difficult to verify whether the RLCP-type shell matrix proteins are all true homologues, considering the limitations of the BlastP tools for analysing sequences with low complexity domains.

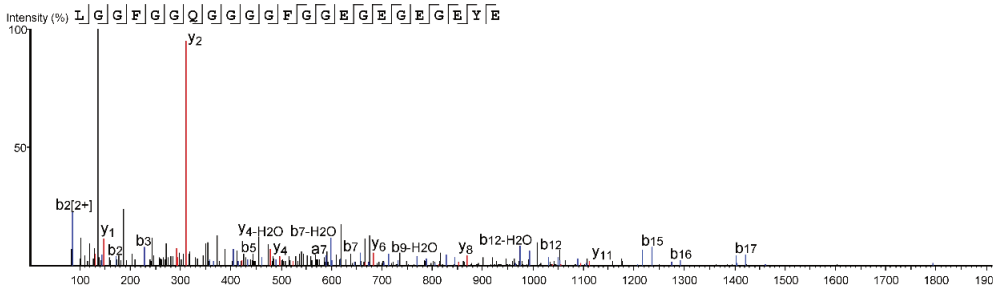
The analysis of the Hic 74 sequence from *U. pictorum* (Figure S3.6; all protein-peptide data from other shells are available in Supplementary File 4) shows several interesting points:

- cleavage sites for trypsin are relatively few, and therefore the use of elastase improved the coverage;
- “natural” cleavage sites (e.g. N in position 756) can be observed;
- among the most frequent modifications are the loss of ammonia, oxidation, dihydroxylation, deamidation;
- the evidence for amino acid substitutions is not usually strong (i.e. not all spectra show the occurrence of the substitution), except for Ile → Phe at residue 310. The same substitution is also seen in the Hic 74 sequence of *M. margaritifera* and *U. crassus* (but not *P. auricularius,* although Leu → Phe substitutions are observed in various positions of the same peptide LLALSLGLGI);
- the presence of RLCs (e.g. Ala-rich domains) means that spectra could cover any of the multiple domains, potentially biasing % coverage calculations.

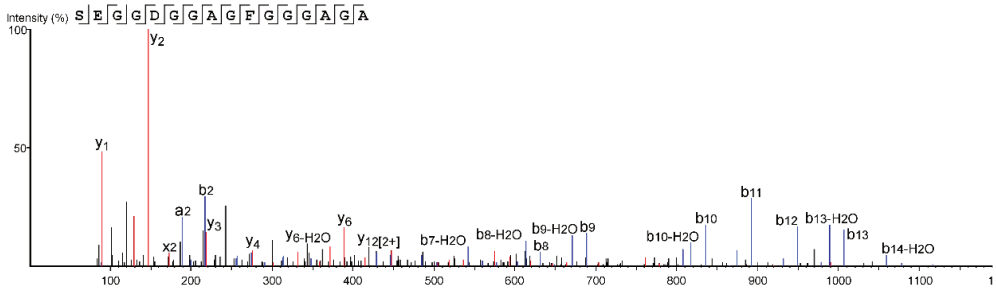

Sp1
Sp2
Sp3
trypsin
elastase
trypsin
elastase
natural
Sp4
elastase
Sp5
elastase
Sp6 elastase
natural
Sp7 elastase
Sp8 elastase
Sp9
elastase
Sp10
elastase
Sp11
elastase
trypsin
Sp12
Sp13 trypsin
Sp14
trypsin
Sp15
Sp16
trypsin
elastase
Sp17
Sp18
trypsin
elastase
Sp19
trypsin
Sp20
trypsin
Sp21
Sp22
elastase
Sp23
elastase
trypsin
natural
Sp24
elastase
trypsin
Sp25


1
Intensity (%)
F L I K c D K R G T
pre[1+]
y1
b2
a2
y3
b3
y4
y5
y6
y7
b6
y7-NH3
y8
y8-NH3
y8-H2O
y9
b8
2
Intensity (%)
S R F E Q Y V P G H
pre[1+]
y1
y2
b5[2+]
y3-H2O
y3
b2-NH3
b2
y4
b8[2+]
b4
y5
b5
b6
a6
a7
b7
a8
y8
b8
b9
a9
3
Intensity (%)
c A V G T G F S P S D c D c T L R
y9
b2
b3
y10
a2
y12
y14
a1
y1
a3
y2
b4
b5-H2O
y4
b7-H2O
b8
y6
y7
y8
y11
y13
y15
4
Intensity (%)
Q A A A E V V D E F G I G A L
b5
b6
y1
b2
b3
b4
y2
x2
y3
y4
b9[2+]
b6-NH3
b7
b8
b9
b10
b11
5
Intensity (%)
S V L G D V d G L G G L
y1
a2
y2
b6[2+]
y3
b3
y5-NH3
b5-H2O
b6-H2O
b6
b5
y7
b8-H2O
b8
y9-H2O

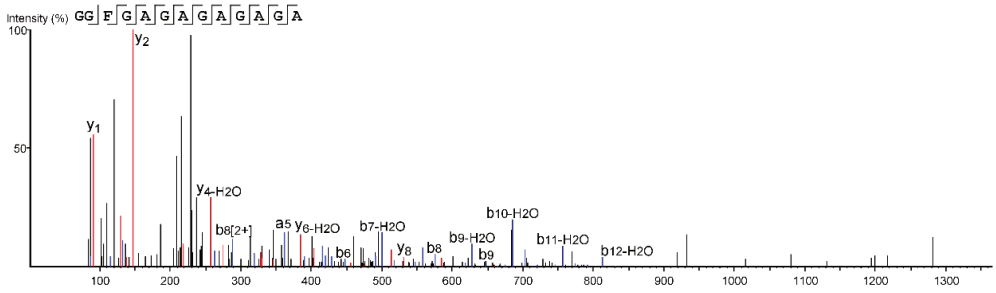
6


Intensity (%)
L G G F G G Q G G G G F G G E G E G E G E Y E
y2
b2[2+]
y1
b2
b3
y4-H2O
b7-H2O
b5
y4
a7
b7
y6
b9-H2O
y8
b12-H2O
b12
y11
b15
b16
b17

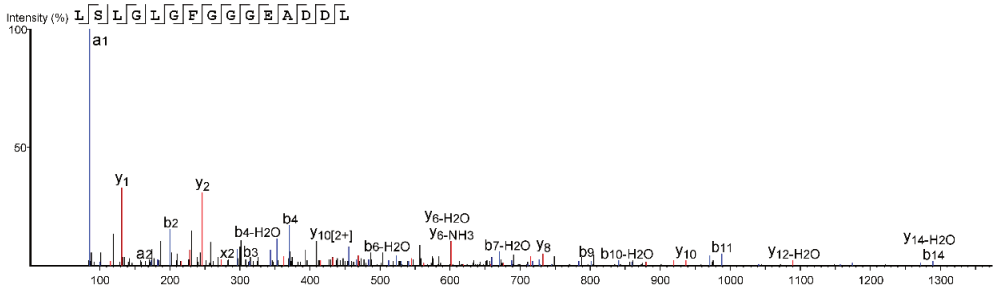
7


Intensity (%)
S E G G D G G A G F G G G A G A
y2
y1
b2
a2
y3
x2
y4
y6-H2O
y6
y12[2+]
b7-H2O
b8-H2O
b9-H2O
b9
b8
b10-H2O
b10
b11
b12
b13-H2O
b13
b14-H2O

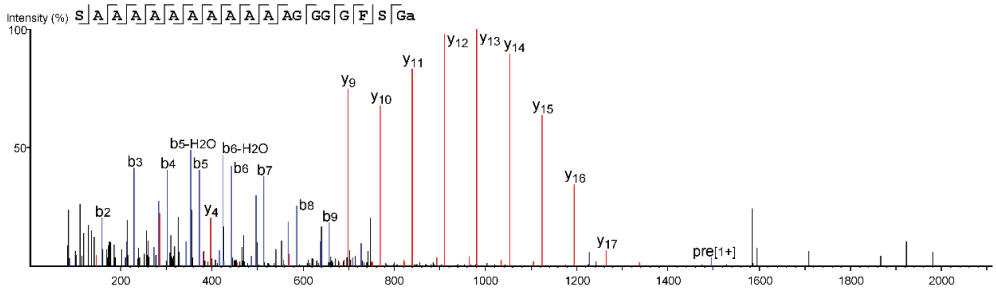
8


Intensity (%)
G G F G A G A G A G A G A
y2
y1
y4-H2O
b8[2+]
a5
y6-H2O
b7-H2O
b6
y8
b8
b9-H2O
b9
b10-H2O
b11-H2O
b12-H2O

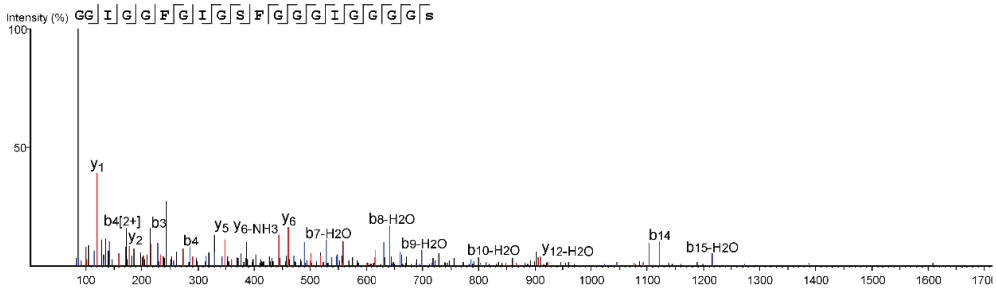
9


Intensity (%)
L S L G L G F G G G E A D D L
a1
y1
y2
b2
a2
x2
b3
b4
b4-H2O
y10[2+]
b6-H2O
y6-H2O
y6-NH3
b7-H2O
y8
b9
b10-H2O
y10
b11
y12-H2O
y14-H2O
b14

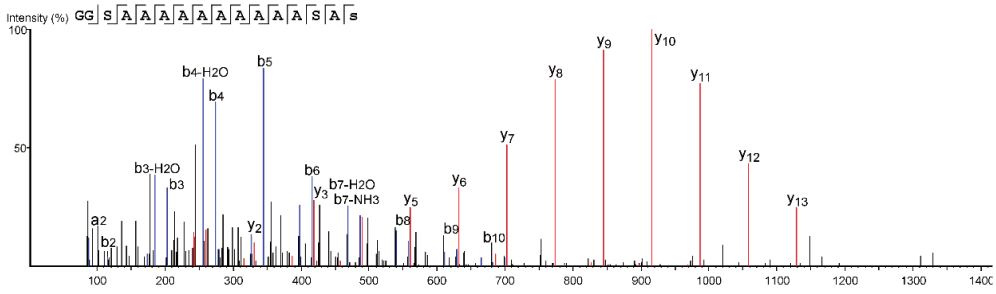
10


Intensity (%)
S A A A A A A A A A A G G G F S Ga
y12
y13
y14
y11
y9
y10
y15
y16
b5-H2O
b6-H2O
b3
b4
b5
b6
b7
b2
y4
b8
b9
y17
pre[1+]

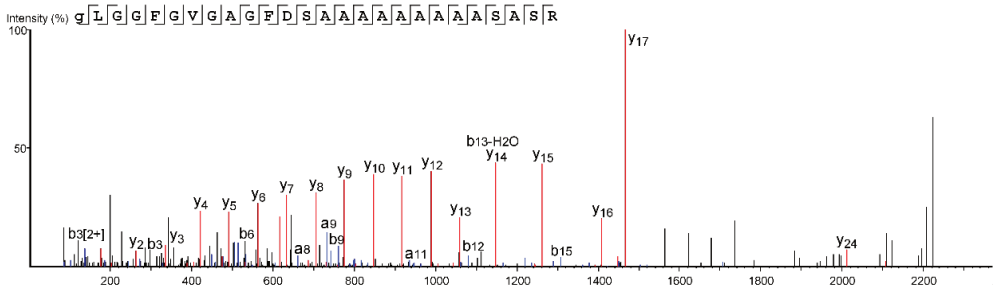
1
1


Intensity (%)
GG I G G F G I G S F G G G I G G G s

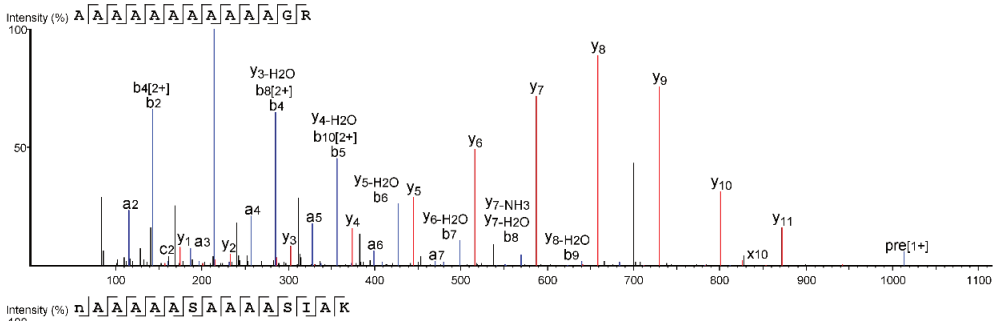
1
2


Intensity (%)
GG S A A A A A A A A A S A s

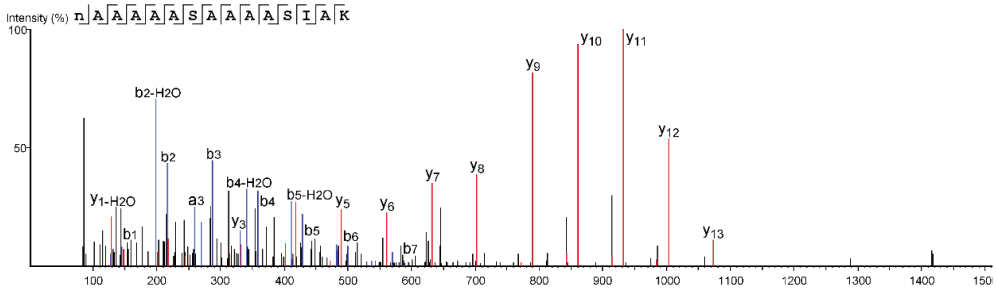
1
3


Intensity (%)
g L G G F G V G A G F D S A A A A A A A A S A S R

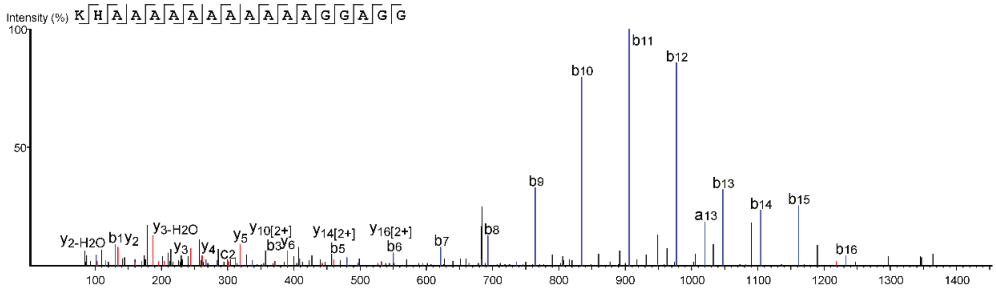
1
4


Intensity (%)
A A A A A A A A A A A G R

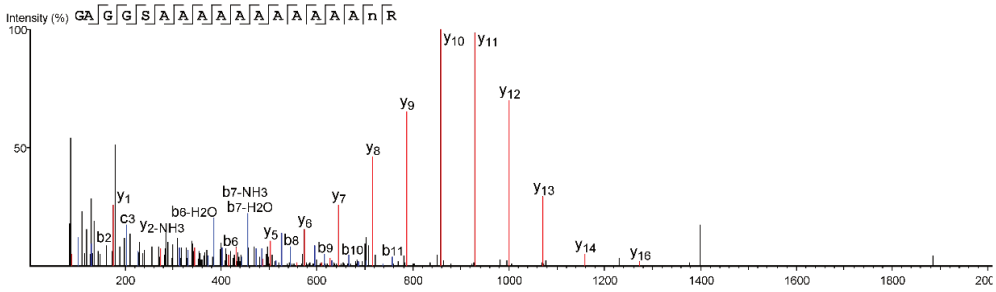
1
5


Intensity (%)
n A A A A A S A A A A S I A K

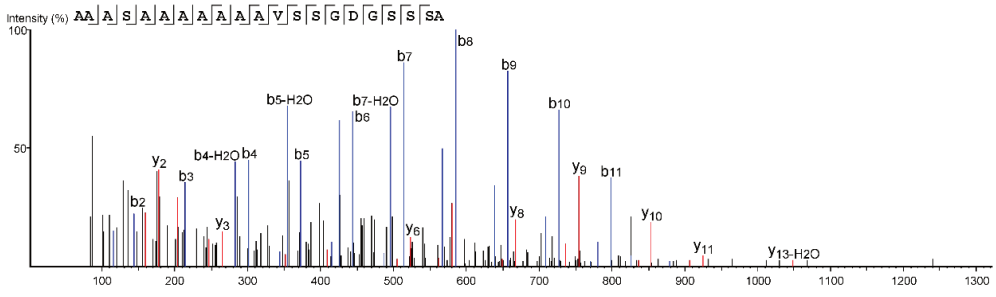
16


Intensity (%)
K H A A A A A A A A A A A G G A G G
b11
b10
b12
b9
b13
a13
b14
b15
b16
b8
b7
y2-H2O
b1y2
y3-H2O
y3
y4
y5
y10[2+]
b3y6
y14[2+]
b5
y16[2+]
b6

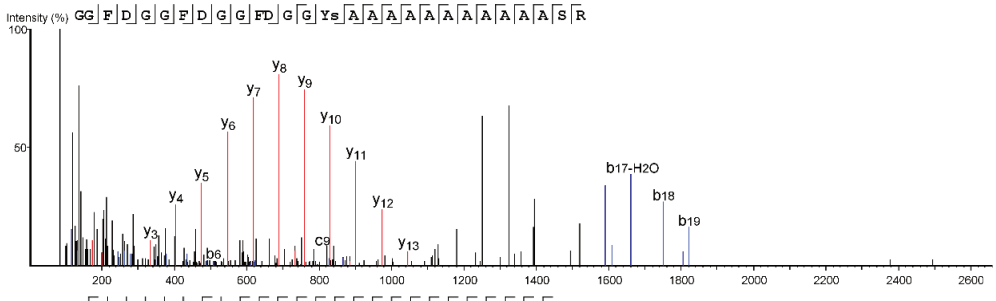
17


Intensity (%)
G A G G S A A A A A A A A A n R
y10
y11
y12
y9
y8
y13
y7
y1
b7-NH3
b7-H2O
b6-H2O
c3
y2-NH3
y5
y6
b2
b6
b8
b9
b10
b11
y14
y16

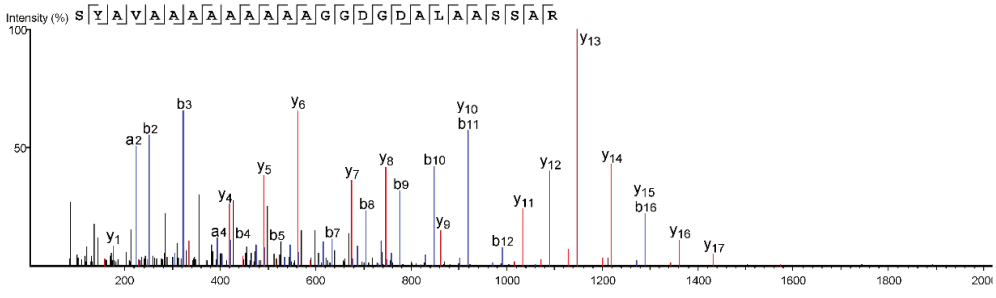
18


Intensity (%)
A A S A A A A A A V S S G D G S S S A
b8
b7
b9
b5-H2O
b7-H2O
b6
b10
b4-H2O
b4
b5
y2
b3
b2
y3
y9
b11
y8
y10
y6
y11
y13-H2O

19


Intensity (%)
G G F D G G F D G G F D G G Ys A A A A A A A A A A S R
y8
y9
y7
y6
y10
y11
y5
b17-H2O
y4
y12
b18
b19
y3
c9
b6
y13

20


Intensity (%)
S Y A V A A A A A A A A G G D G D A L A A S S A R
y13
b3
y6
y10
b11
b2
a2
y5
y8
b10
y12
y14
y7
b9
y4
b8
y11
y15
b16
y9
y1
a4
b4
b5
b7
b12
y16
y17

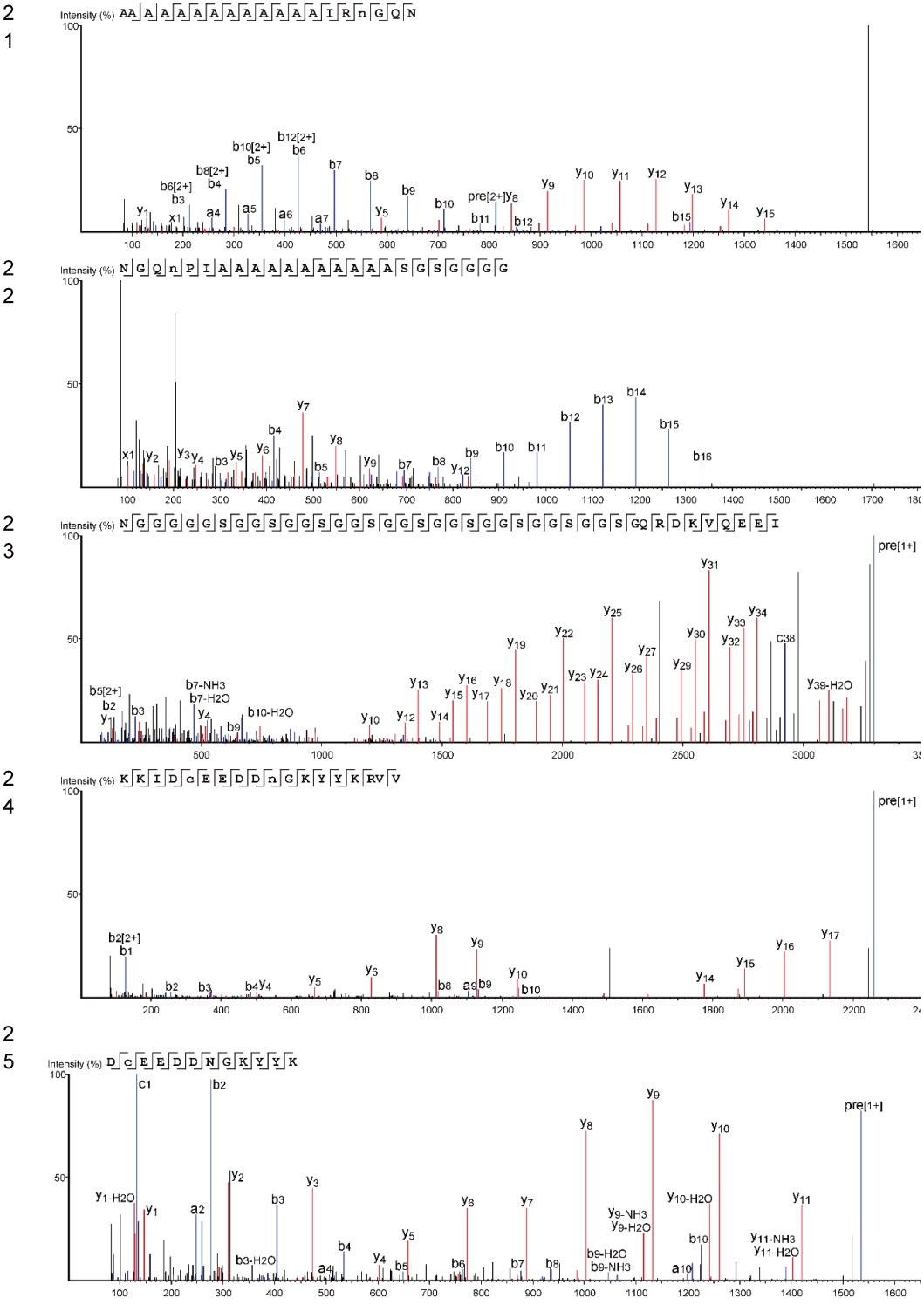


Figure S3.6: Protein Hic74 identified in Unio pictorum shell: sequence coverage, highlighting in pink the product ion spectra ("Sp") shown below. Sequences

reconstructed by assisted de novo on the basis of mono-charged ions mainly (spectra were acquired on the 400-1600 m/z range and multiply-charged ions were detected).

3.5.8. Protein identification in archaeological double-button samples

Protein Hic74 (*H. cummingii*) was the top hit for all the archaeological samples. Figure S3.7 shows the Hic74 sequence from sample HavC (for all other peptide-protein data see Supplementary File 4) and it highlights that:

- some regions of Hic74 that were not covered in the sequence retrieved from *Unio pictorum* were present in the archaeological sample, for example residues 107→135;
- the I→ F mutation in position 310 could also be seen in this sample (and were also present in HavA and HavB, but not in PesB - while all of the Hornstaad samples did not have spectra covering this region);
- the A→ S mutation at residue 804 is present in all spectra and also in *U. pictorum*
- the R→ G mutation in position 822 is present in most but not all spectra
- Diagenetically-induced modifications such as deamidation and oxidation were frequent.

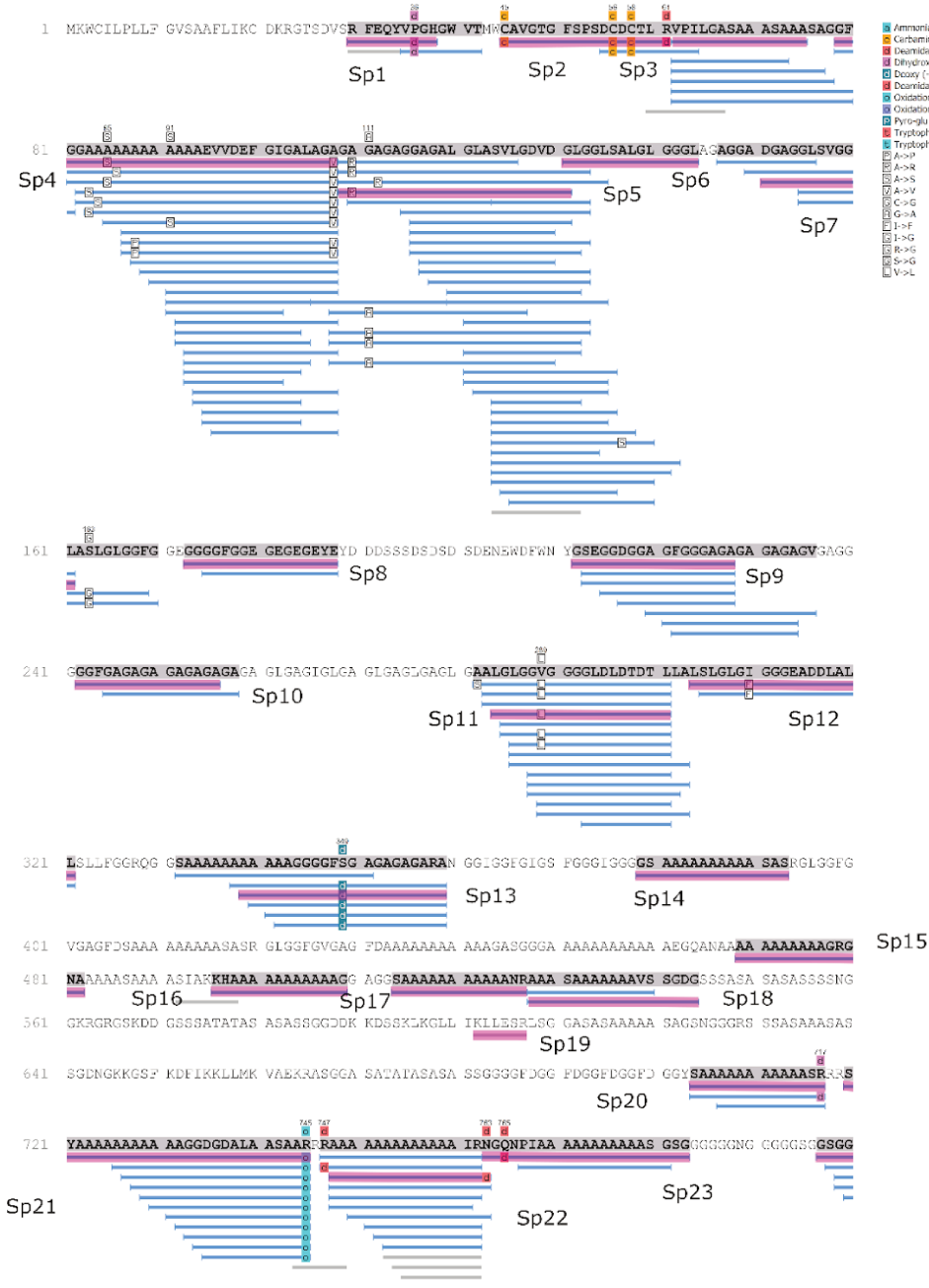
In particular, the extent of deamidation displayed by Gln (Q) and Asn (N) residues in Hic74, is broadly coherent with the relative rate of deamidation of N and Q, and with the age and the extent of degradation of the samples (Table S5):

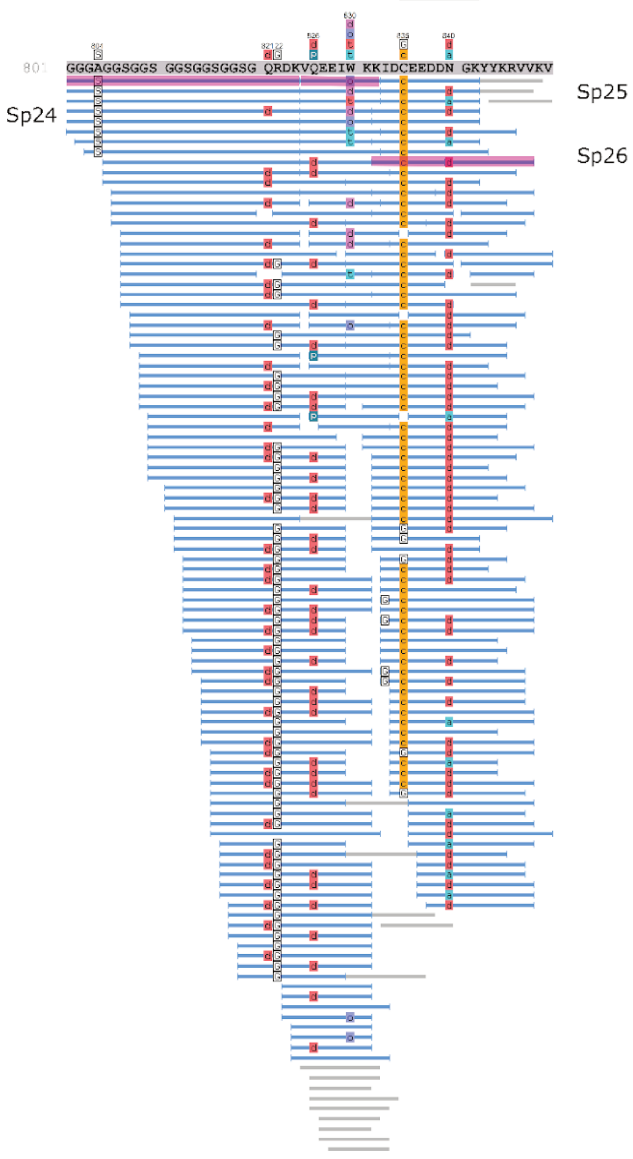
- Hornstaad double-buttons were too degraded to yield any surviving N or Q;
- All remaining archaeological samples (Havnø, PesB, *U. crassus* and *P. auricularius*) display similar extent of N deamidation (~50-70%), which is slightly higher than for the modern shells *M. margaritifera* and *U. pictorum;*
- The extent of Q deamidation is low in HavA, PesB as well as in the modern shells *M. margaritifera* and *U. pictorum* (~10%), and higher in all other samples (~30-40%).

Table S3.5: Extent of Asn (N) and Gln (Q) deamidation (N→D; Q→E) in the peptides identified in the Hic74 sequence.

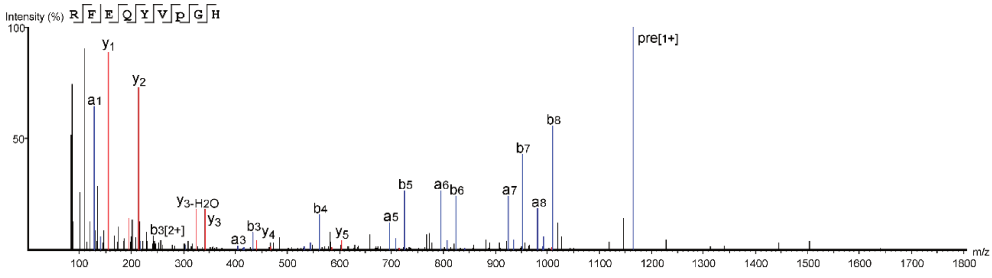
	Hav A	**Hav B**	**Hav C**	**Hor A**	**Hor B**	**Hor C**	**Pes B**	***U. pictorum***	***U. crassus***	***M. margaritifera***	***P. auricularius***
# Q	28	105	196	-	-	-	12	136	237	13	37
# N	38	31	104				4	71	96	8	88
# Q→E	3	33	71				1	18	67	1	12
# N→D	19	17	53				2	28	66	3	57
% Q→E	**11**	**36**	**36**				**8**	**13**	**28**	**8**	**32**
% N→D	**50**	**51**	**51**				**50**	**39**	**69**	**37**	**65**

Hic74 coverages varied from 35, 39 to 55 % respectively for HavA, HavB and HavC, where they were supported by a high number of peptides (132, 158, 260). In comparison, the coverage of this protein obtained from the shell matrix of reference unionoid shells (where the extraction was performed on samples with sizes at least 100 times higher) was almost the same: 34 and 54 % in modern *Margaritifera margaritifera* and *Unio pictorum* and 49 to 50 % in sub-fossil unionoids *Pseudunio auricularius* and *Unio crassus*. Such results indicate that this protein is stable during early diagenesis. This might be partly due to the chemical stability of the Ala-Ala bonds that dominate the low complexity poly-Ala domains, as well as the mineral stabilisation (68) of the acidic motifs at the C-terminus of the protein, which showed a very high number of identified peptides, most of which were deamidated (Figure S3.7). This result also suggests that over geological times (but probably during early diagenesis), slow denaturation (uncoiling) of protein structures or splitting of cross-linkage sites with other organic macromolecules (i.e glycosides) may aid protein release, extraction and subsequent identification. PesB showed relatively low % coverage (19 %) but this may be due to a species effect, as discussed elsewhere. Among the Hornstaad samples, the overall extent of preservation was low (for example Hic74 protein has a coverage of only 7 % in HorB and 12 % in HorC, supported by 6 and 11 peptides respectively), probably due to the effects of temperature. The exposure to heat of the archaeological biogenic carbonate did not induce recrystallization processes (as we did not observe any calcite in the ATR analyses), or extremely high D/L values in Hor B and HorC, however the temperature was high enough to induce protein modifications and degradation.

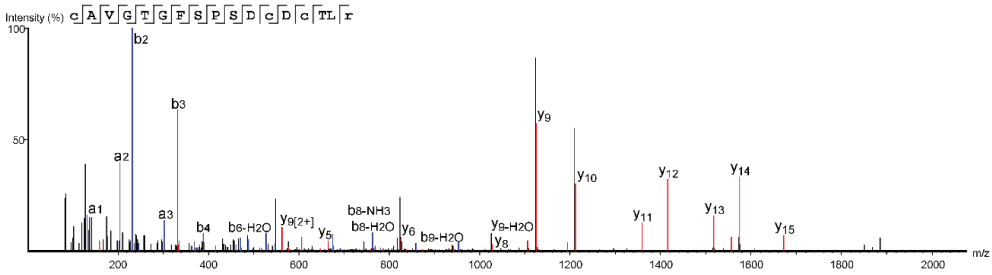

Ammonia-loss (N) (-17.03)
Carbamidomethylation (+57.02)
Deamidation (NQ) (+0.98)
Dihydroxy (+31.99)
Deoxy (-15.99)
Deamidation (R) (+0.98)
Oxidation or Hydroxylation (+15.99)
Oxidation (HW) (+15.99)
Pyro-glu from Q (-17.03)
Tryptophan oxidation to oxolactone (+13.98)
Tryptophan oxidation to kynurenin (+3.99)
A->P
A->R
A->S
A->V
C->G
G->A
I->F
I->G
R->G
S->G
V->L
Sp1
Sp2
Sp3
Sp4
Sp5
Sp6
Sp7
Sp8
Sp9
Sp10
Sp11
Sp12
Sp13
Sp14
Sp15
Sp16
Sp17
Sp18
Sp19
Sp20
Sp21
Sp22
Sp23


GGGAGGSGGS GGSGGGSGGSG QRDKVQEEIW KKIDCEEDDN GKYYKRVVKV
Sp24
Sp25
Sp26

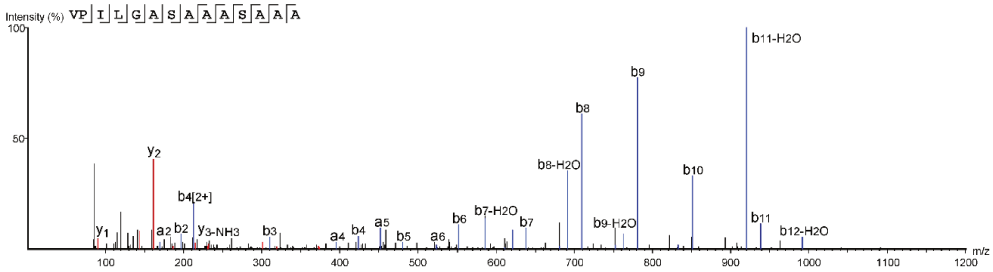
1


Intensity (%)
pre[1+]
y1
y2
a1
b8
b7
b5
a6
b6
a7
a8
y3-H2O
y3
b4
a5
b3[2+]
a3
b3
y4
y5
m/z

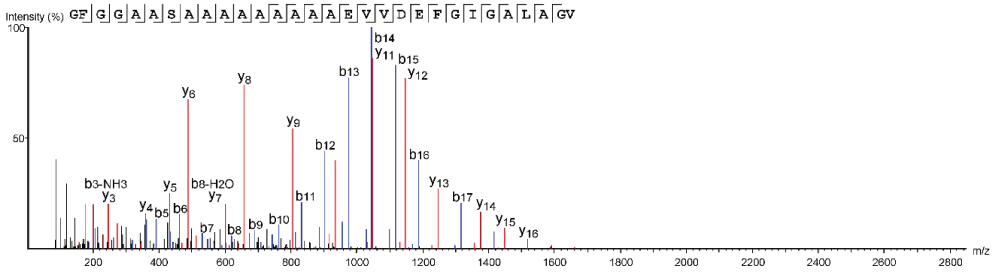
2


Intensity (%)
cAVGTGFSPSSDcDcTLr
b2
b3
y9
a2
y10
y12
y14
a1
a3
b4
b6-H2O
y9[2+]
y5
b8-NH3
b8-H2O
y6
b9-H2O
y9-H2O
y8
y11
y13
y15
m/z

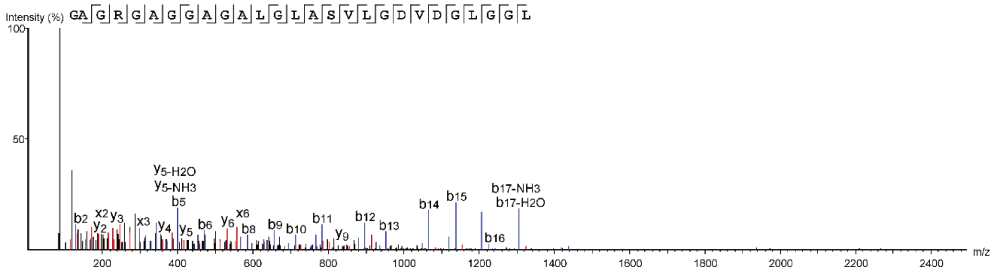
3


Intensity (%)
VPILGASAAASAAA
b11-H2O
b9
b8
y2
b8-H2O
b10
b4[2+]
b7-H2O
y1
a2
b2
y3-NH3
b3
a4
b4
a5
b5
b6
a6
b7
b9-H2O
b11
b12-H2O
m/z

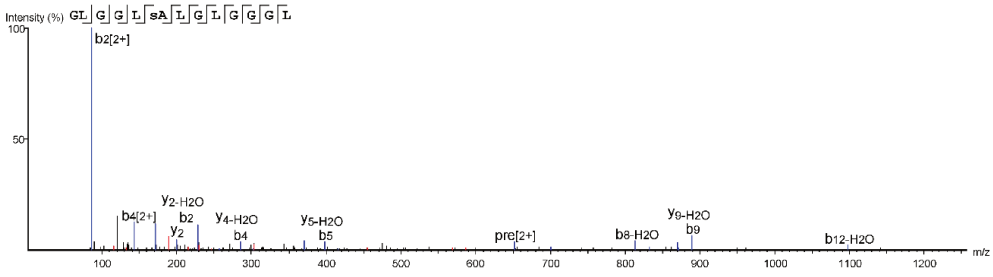
4


Intensity (%)
GFGGAASAAAAAAAAAEVVDEFGIGALAGV
b14
y11
b15
b13
y12
y8
y8
y9
b12
b16
b3-NH3
y5
b8-H2O
y3
y7
y13
b17
y4
b5
b6
b11
y14
y15
b7
b8
b9
b10
y16
m/z

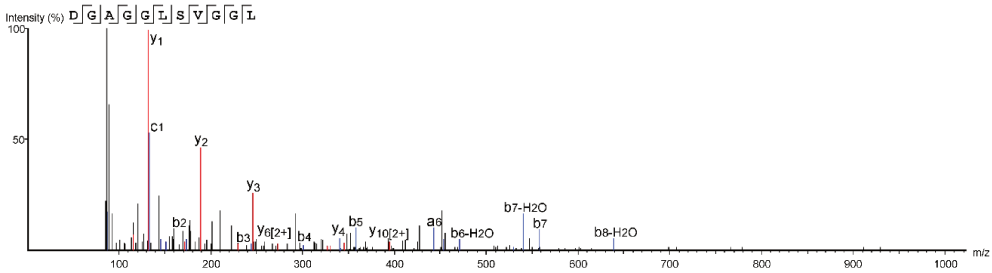
5


Intensity (%)
GAGRGAGGAGALGLASVLGDVDGLGGL
y5-H2O
y5-NH3
b5
b17-NH3
b17-H2O
b14
b15
b2
x2
y2
y3
x3
y4
y5
b6
y6
x6
b8
b9
b10
b11
b12
b13
y9
b16
m/z

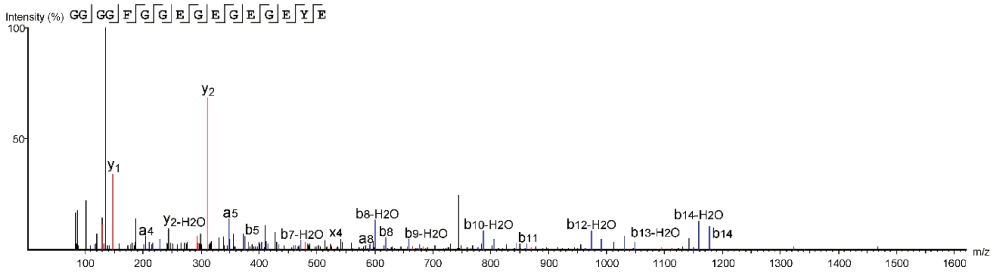
6


Intensity (%)
GLGGLsALGLGGGL
b2[2+]
y2-H2O
b4[2+]
b2
y4-H2O
y5-H2O
y2
b4
b5
pre[2+]
b8-H2O
y9-H2O
b9
b12-H2O
m/z

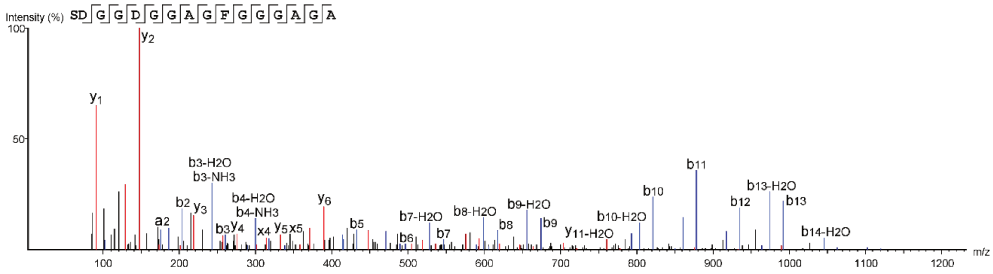
7


Intensity (%) D G A G G L S V G G L
y1
c1
y2
y3
b2
b3
y6[2+]
b4
y4
b5
y10[2+]
a6
b6-H2O
b7-H2O
b7
b8-H2O
m/z

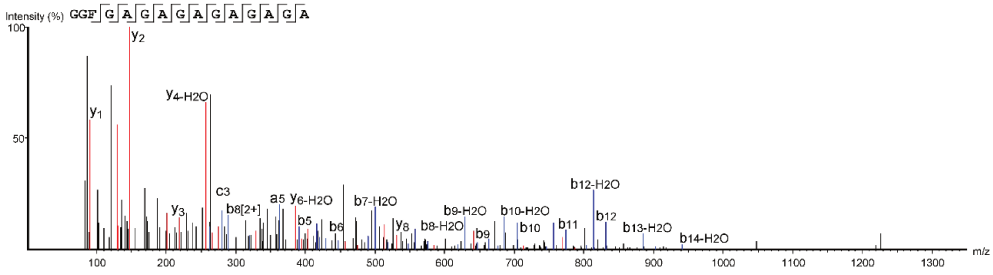
8


Intensity (%) GG GG F G G E G E G E G E Y E
y2
y1
y2-H2O
a4
a5
b5
b7-H2O x4
b8-H2O
a8
b8
b9-H2O
b10-H2O
b11
b12-H2O
b13-H2O
b14-H2O
b14
m/z

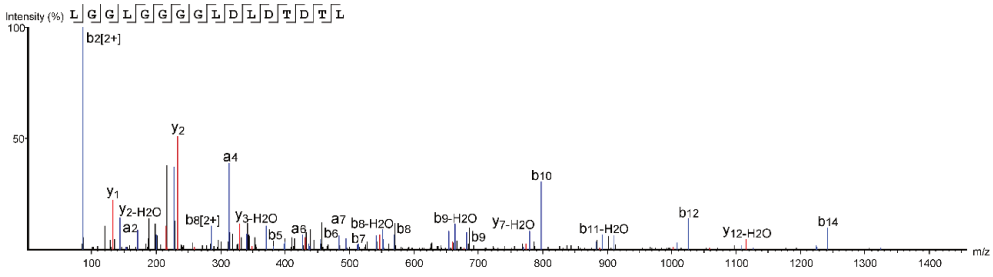
9


Intensity (%) SD G G D G G A G F G G G A G A
y2
y1
b3-H2O
b3-NH3
b2
y3
b4-H2O
b4-NH3
y6
a2
b3
y4
x4
y5
x5
b5
b7-H2O
b8-H2O
b9-H2O
b6
b7
b8
b9
b10-H2O
y11-H2O
b11
b10
b13-H2O
b12
b13
b14-H2O
m/z

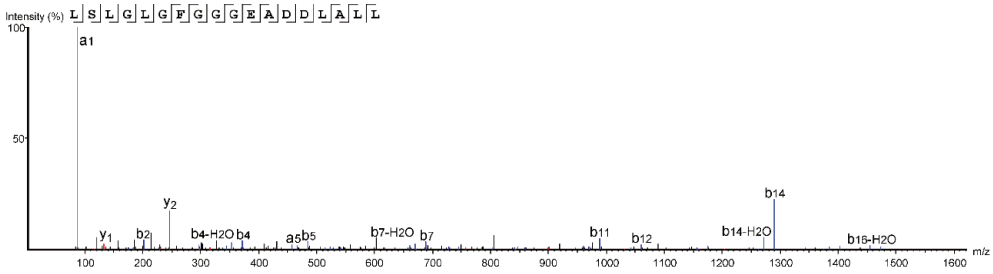
10


Intensity (%) GGF G A G A G A G A G A G A
y2
y4-H2O
y1
c3
b8[2+]
a5
y6-H2O
b7-H2O
y3
b5
b6
y8
b8-H2O
b9-H2O
b10-H2O
b9
b10
b11
b12-H2O
b12
b13-H2O
b14-H2O
m/z

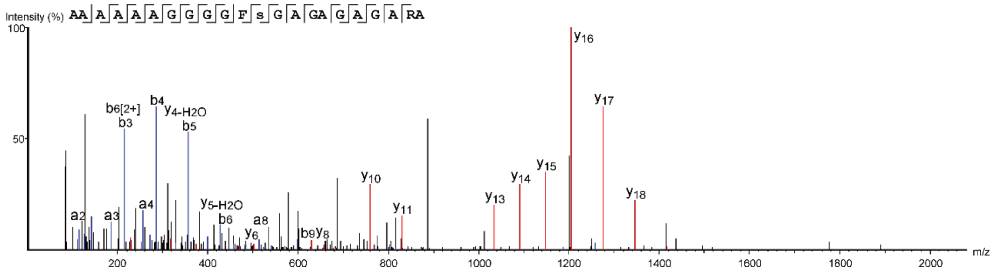
11


Intensity (%) L G G L G G G G L D L D T D T L
b2[2+]
y2
a4
y1
y2-H2O
a2
b8[2+]
y3-H2O
b5
a6
a7
b6
b8-H2O
b7
b8
b9-H2O
b9
y7-H2O
b10
b11-H2O
b12
y12-H2O
b14
m/z

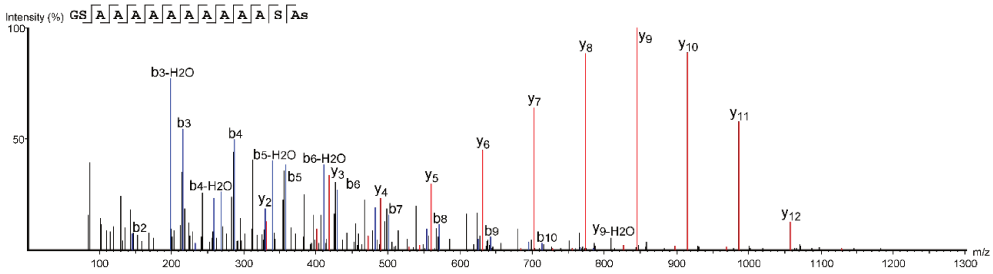
12


Intensity (%) L S L G L G F G G G E A D D L A L L
a1
y2
y1
b2
b4-H2O
b4
a5
b5
b7-H2O
b7
b11
b12
b14
b14-H2O
b16-H2O
m/z

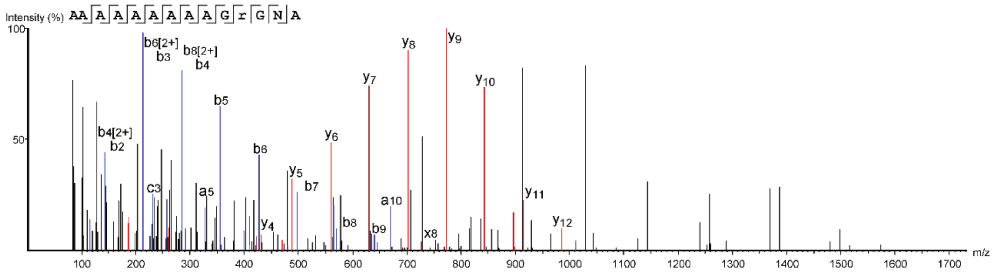
13


Intensity (%) AA A A A A G G G G F s G A GA G A G A RA
y16
y17
b6[2+]
b3
b4
y4-H2O
b5
y10
y15
y14
y13
y18
y11
a2
a3
a4
y5-H2O
b6
a8
y6
b9
y8
m/z

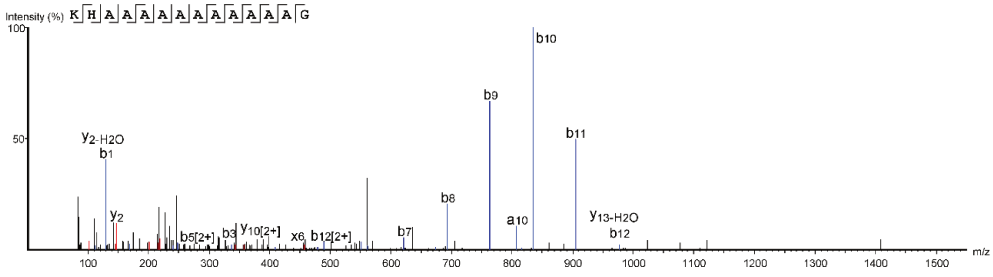
14


Intensity (%) GS A A A A A A A A A S As
y8
y9
y10
b3-H2O
y7
b3
y11
b4
b5-H2O
b6-H2O
y6
b5
y3
b4-H2O
b6
y5
y4
y2
b7
b8
y12
b2
b9
b10
y9-H2O
m/z

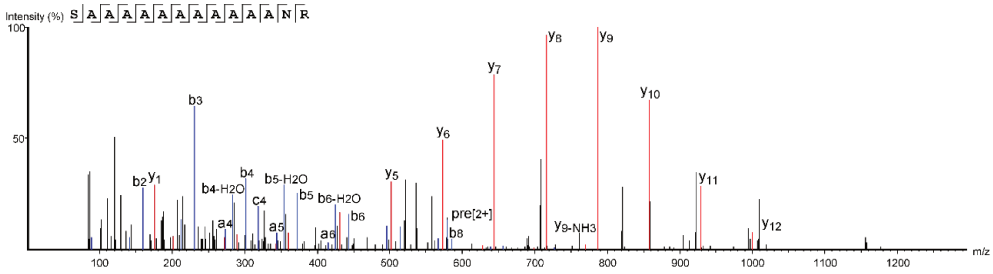
15


Intensity (%) AA A A A A A A A G r G N A
b6[2+]
b3
b8[2+]
b4
y8
y9
y7
y10
b5
b4[2+]
b2
y6
b8
y5
b7
c3
a5
a10
y11
y4
b8
b9
x8
y12
m/z

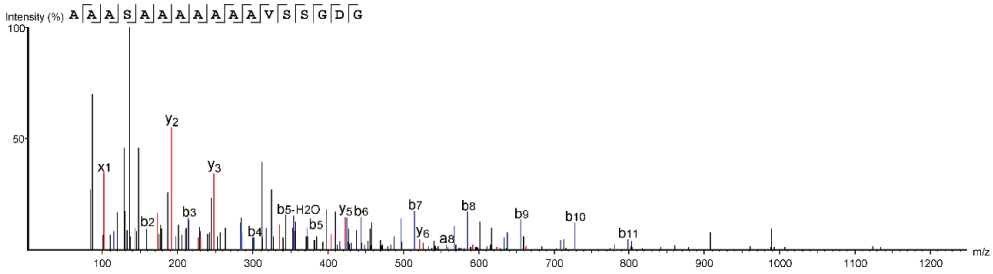
16


Intensity (%) K H A A A A A A A A A A G
b10
b9
y2-H2O
b1
b11
b8
y2
a10
y13-H2O
b12
b5[2+]
b3
y10[2+]
x6
b12[2+]
b7
m/z

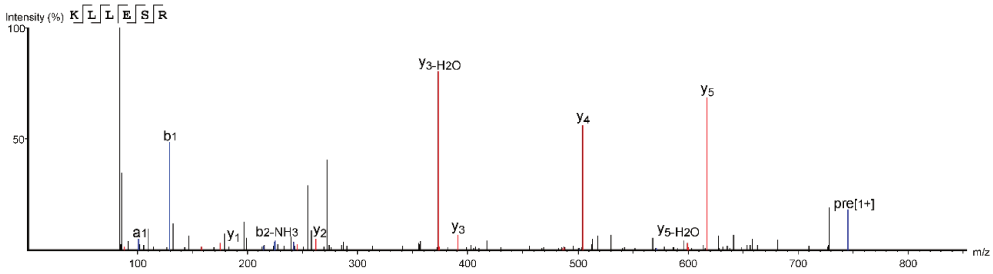
17


Intensity (%) SAAAAAAAAAAANR
100
50
b3
y8
y9
y7
y10
y6
b2
y1
b4
b4-H2O
b5-H2O
y5
b5
b6-H2O
c4
y11
pre[2+]
a4
a5
a6
b6
b8
y9-NH3
y12
100 200 300 400 500 600 700 800 900 1000 1100 1200
m/z

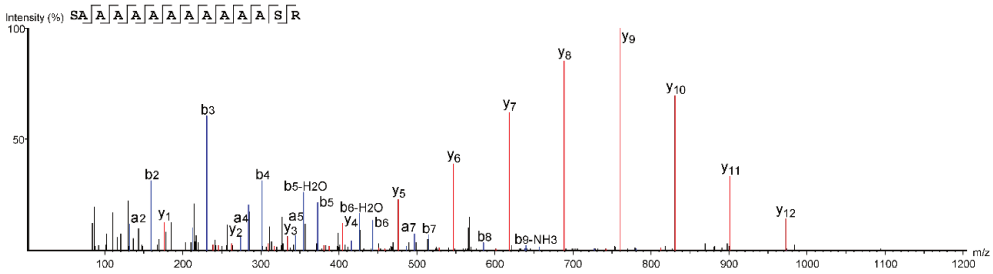
18


Intensity (%) AAASAAAAAAAVSSGDG
100
50
y2
x1
y3
b2
b3
b5-H2O
y5
b6
b7
b8
b9
b10
b4
b5
y6
a8
b11
100 200 300 400 500 600 700 800 900 1000 1100 1200
m/z

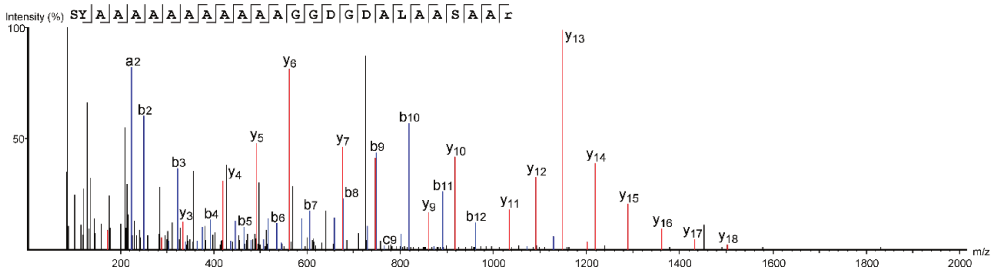
19


Intensity (%) KLLESR
100
50
y3-H2O
y5
y4
b1
pre[1+]
a1
y1
b2-NH3
y2
y3
y5-H2O
100 200 300 400 500 600 700 800
m/z

20


Intensity (%) SAAAAAAAAAAASR
100
50
y9
y8
y10
y7
b3
y6
y11
b2
b4
b5-H2O
y5
y12
b6-H2O
a2
y1
a4
a5
b5
y4
b6
a7
b7
y2
y3
b8
b9-NH3
100 200 300 400 500 600 700 800 900 1000 1100 1200
m/z

21


Intensity (%) SYAAAAAAAAAAAGGDGDALAASAAR
100
50
y13
a2
y6
b10
b2
y5
y7
b9
y10
b3
y12
y14
y4
b11
b8
y11
y15
y3
b4
b7
y9
b12
y16
b5
b6
y17
y18
c9
200 400 600 800 1000 1200 1400 1600 1800 2000
m/z

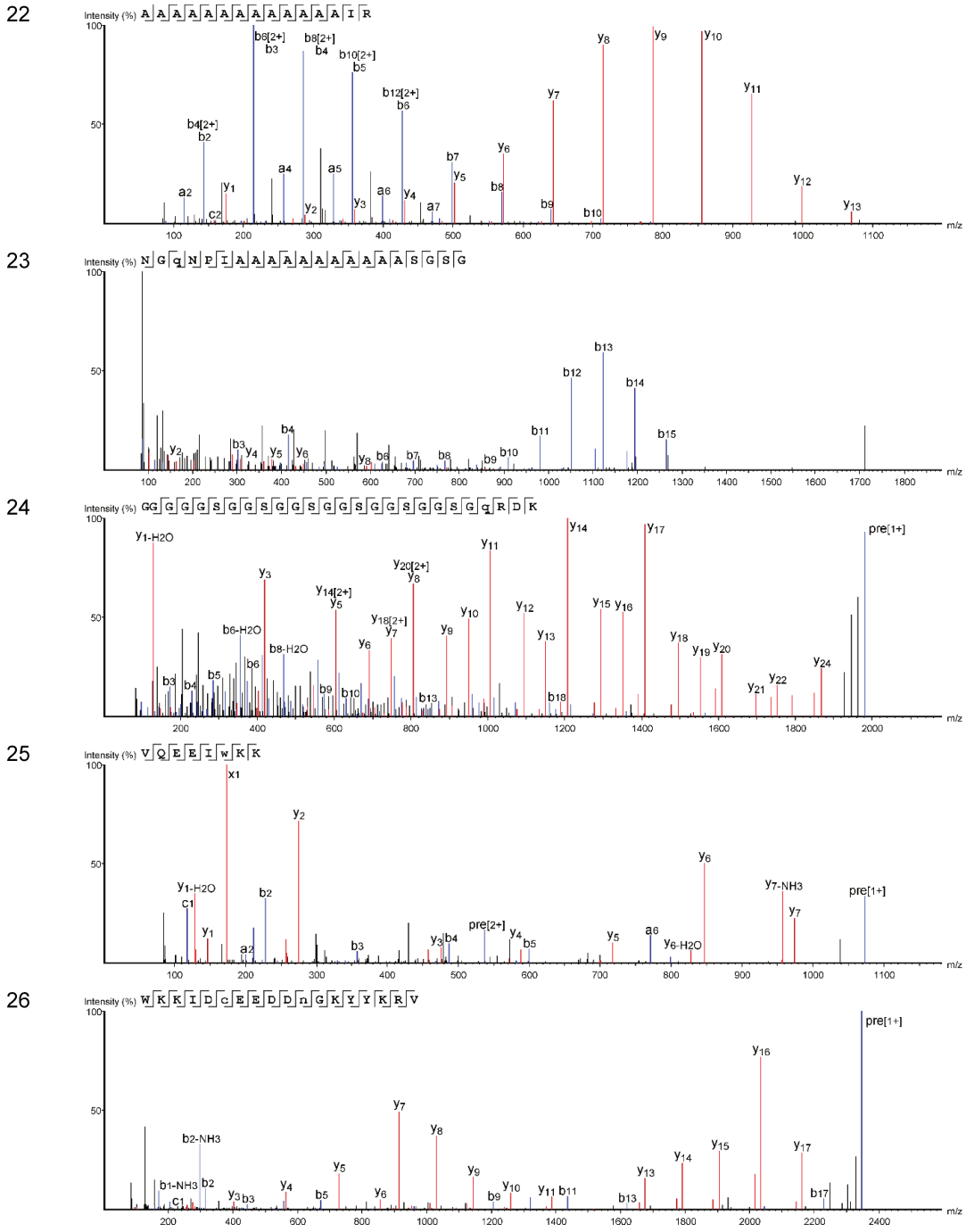


Figure S3.7: Protein Hic74 identified in the double-button HavC: sequence coverage, highlighting in pink the product ion spectra ("Sp") shown below. Sequences reconstructed by assisted *de novo* on the basis of mono-charged ions mainly (spectra were acquired on the 400-1600 *m/z* range and multiply-charged ions were detected).

Some proteins from other species, such as MSI60-related protein (*Pinctada*), glycine rich structural-like protein (*Crassostrea*) and precollagen D (*Mytilus*) were identified in PesB and all of the Havnø samples (Figure S3.8). However, the identification of these silk-like and collagen-like proteins in the archaeological samples were supported only by low complexity, Ala and Gly rich domains, exactly as for the freshwater unionoid shells, and could thus be attributed generally to the nacro-prismatic structure of the shell material.

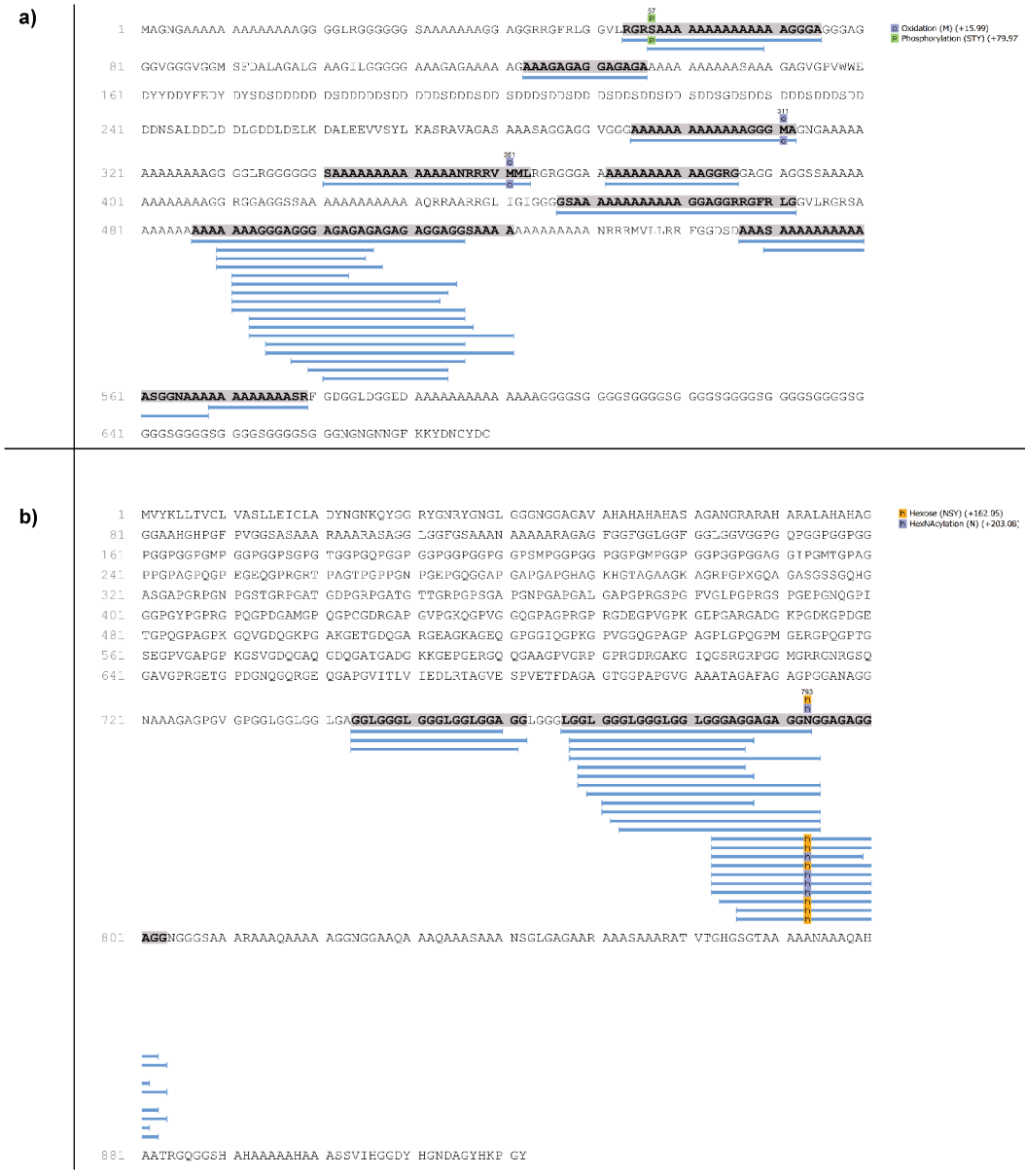


Figure S3.8: Marine shell proteins identified in double-button HavB: a) MSI60-related protein (*Pinctada fucata*); b) Precollagen D (*Mytilus edulis*). Note that both are supported only by repetitive low complexity (RLC) domains. Sequences reconstructed by assisted *de novo* on the basis of mono-charged ions mainly (spectra were acquired on the 400-1600 *m/z* range and multiply-charged ions were detected).

3.5.9 Proteome similarities

Searching the *de novo*-reconstructed peptides against the expressed sequence tag (EST) database yielded additional peptide-protein matches, particularly in freshwater unionoid shells and archaeological samples. These included many hits from *Hyriopsis cumnigii* (e.g more than 20 EST hits in *U. pictorum*), mostly unannotated sequences (a search of *Hyriopsis cumingii* on NCBI will retrieve 246 protein sequences, but 10156 EST sequences).

The EST-derived dataset, consisting of lists of the unique identifiers of the ESTs retrieved from each sample, was used to explore the similarities between the proteomes of the reference shells and of the double-buttons. First, we attempted to understand whether it is possible to discriminate the six reference shells taxa. The circular graph (Figure S3.9) represents the number of identified EST sequences in each of the shells and shows the proportion of sequences shared between species.

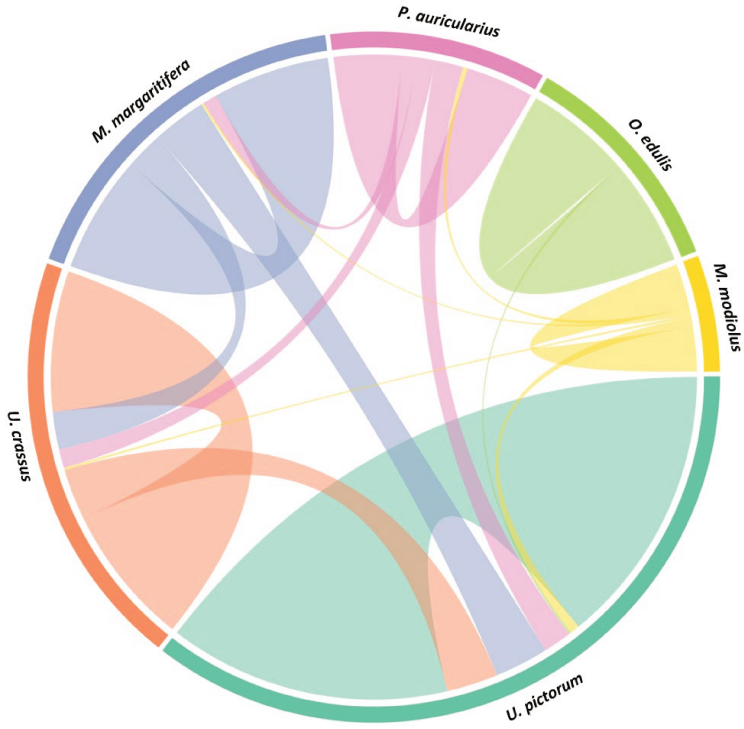


Figure S3.9: Circular diagram representing the extent of similarity between the proteomes of six reference mollusc shells based on the identified EST sequences.

Unio pictorum samples yielded the highest number of EST hits (as seen from the proportion of the occupied circular area), while the marine shell *Modiolus modiolus* yielded the least. The four freshwater shells - *U. pictorum, U. crassus, M. margaritifera, P. auricularius* shared the highest proportion of sequences, while the

two marine shells - *M. modiolus* and *O. edulis*, seemed to have very distinctive proteomes and only few sequences in common with unionoid shells. These results indicate that there is a definite similarity between the proteomes of freshwater nacre shells here studied: *U. pictorum* is mostly similar to *U. crassus* and *M. margaritifera* and slightly less to *P. auricularius.* We cannot exclude, however, that this may be due to the fact that *U. pictorum* yielded the highest number of identified EST sequences out of all the unionoid shells. Few sequences were shared between freshwater *U. pictorum* and marine *M. modiolus,* owing to the nacro-prismatic shell structure, present in both. Overall, the data show that marine and freshwater reference shells can be easily discriminated on the basis of their proteome profiles (even if they share the same nacro-prismatic structure). However, we stress that it is difficult to retrieve a simple phylogenetic signal within the group of unionoids.
The same analysis was repeated for the seven archaeological double-buttons and the results are presented in Figure S3.10. Searching against the EST database, the Havnø samples yielded the highest number of identified sequences (as can be seen in the figure, more than two thirds of the circular area is occupied by Havnø). While the highest degree of closeness was observed within the same set of samples (Havnø vs Hornstaad), a degree of similarity was also observed between the three different sets. This was particularly evident for HavA, which shared sequences not only with HavB and HavC, but also with all the three Hornstaad samples and with PesB. In the Hornstaad double-buttons we observed a lower amount of identified EST sequences and a more limited extent of similarity between sample sets. This is in accordance with our previous analysis, which showed that the preservation of the Hornstaad double-buttons was the least optimal. Nevertheless, the criss-crossing displayed by the circular diagram implies that the same raw material was likely used to manufacture the ornaments in all of the archaeological sites.

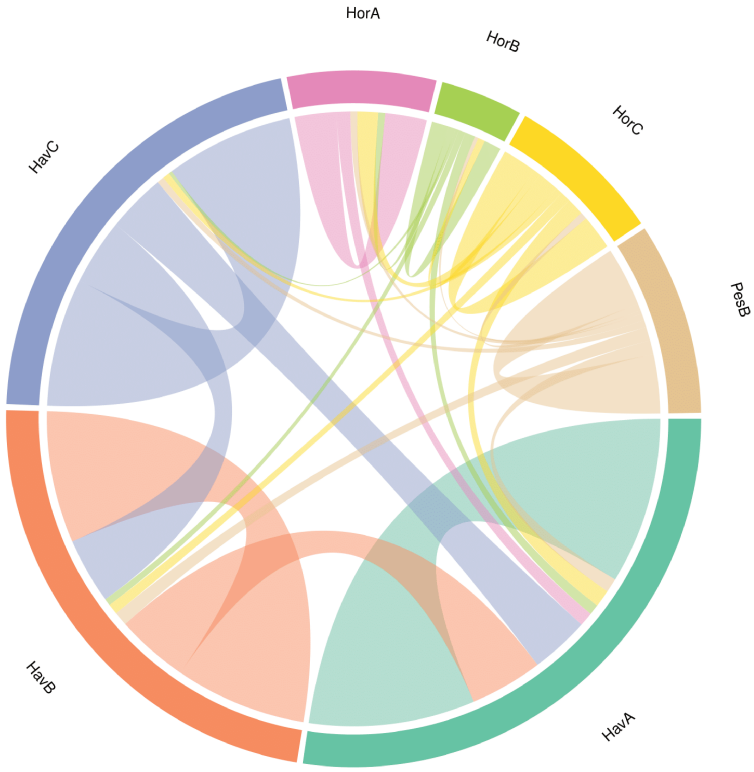


Figure S3.10: Circular diagram representing the similarity between the proteomes of the seven double-buttons based on the identified EST sequences.

3.6 Database independent biomolecular comparisons

The classification of samples on a molecular level has been widely applied for phylogenetics to study the evolutionary relationships between species. However, database-dependant search algorithms of DNA and/or protein sequences cannot take into account peptides and proteins that do not yet *exist* in databases. In recent years, as shotgun proteomics has become a routine tool for biochemistry, and huge amounts of data have been generated on novel systems and organisms (including extinct), there has been an increasing interest in developing methods enabling to compare proteome-wide measurements that would be flexible and allow database-free workflows (Palmblad and Deelder, 2012; Rieder et al., 2017; Yılmaz et al., 2016).

3.6.1 Peptide sequence similarity

The peptide sequences generated by *de novo* algorithm of the software PEAKS Studio 8.5 (Bioinformatics Solutions Inc.) were used in a database-independent

biomolecular comparison. Any peptides belonging to contaminants were removed prior to the analysis, by searching each sample against a common laboratory contaminant database (cRAP; https://www.thegpm.org/crap/). The lists generated by this analysis contained peptide sequences reconstructed by the PEAKS software on the basis of the raw product ion spectra. Only common post translational (phosphorylation) and diagenesis-induced (deamidation, oxidation and dioxidation, pyro-Glu formation) modifications were taken into account by the PEAKS algorithm. Each sequence is present only once (unique), thus the number of peptide sequences does not affect the overall analysis.

The algorithm, developed in-house in C language, provides a score for the sequence similarity between two lists of peptides. This is achieved by finding the best-matching peptide in the second file for each peptide in the first, taking into account the number of consecutive matches and mismatches within the overlapping region of the two peptides. The scores for the best matches (which may be zero) are then combined to give a score for the similarity between the two files. However, the similarity score from file 1 to file 2 is not necessarily the same as the score from file 2 to file 1 as the two files may contain different numbers of peptides. Therefore, a score is also calculated from the best match in the first file to each peptide in the second and the two scores averaged to provide the similarity metric. A similarity matrix is obtained from all pairwise comparisons and converted to a distance matrix.

Multidimensional scaling (MDS) is a technique that allows the similarity of individual observations (here lists of peptides) to be visualized (Gower, 1966). The method provides new coordinates in just two dimensions that represent the information in a distance matrix. Figure 4 (main text) shows the map obtained from the distance matrix calculated by scoring peptide matches between the buttons and several mollusk species. The plot illustrates the great level of similarity of all the Havnø archaeological samples with the freshwater unionoid shells, especially *Unio pictorum*, *Unio crassus* and *Pseudunio auricularius*. The Hornstaad samples cluster slightly apart, likely because of altered proteome profiles due to burning, and the PesB sample falls a bit further from the biggest unionoid cluster, but far from the marine shells. Therefore, the MDS plot obtained on the basis of the reconstructed *de novo* sequences was coherent with the pattern highlighted by the bulk amino acid compositions (Figure S3.5).

3.6.2 Proteome wide distance calculations based on product ion spectra

We attempted to perform comparative shell proteome analysis based solely on the proteome-wide distance calculation of different product ion spectra using a DISMS2 algorithm implemented in R language (Rieder et al., 2017). The algorithm is a four-step procedure, consisting of spectra filtering, checking constraints for matching,

matching of product ion spectra and calculation of the distance matrix with pairwise distances between runs. Raw data of each sample were processed using PEAKS Studio 8.5 (Bioinformatics Solutions Inc.) - searched against a common laboratory contaminant database cRAP (https://www.thegpm.org/crap/) to eliminate the possibility of matching contaminant spectra among the different samples. The exported *de novo only* peptide spectral data were converted to mzXML format using ProteoWizard and used directly as a primary input in R. Since the parameters are flexible, during the first pre-processing step, the topn function was disabled - considering that some sub-fossil samples and archaeological beads may be altered, all of the peaks (and not only those with the highest intensities) were included in the analysis. Binning was done with a fixed bin size where bin = 0.01 (a trial run without binning did not show any changes for the distance matrix calculated, only the time of analysis was prolonged). Cosine distance (dcos) was used to calculate the distance between the mass spectra and the cut-off value (cdis) was set at 0.3. The obtained dissimilarity matrix was visualised by hierarchical clustering using the package *factoextra*, and clustering performed with the *ward.D2* method.

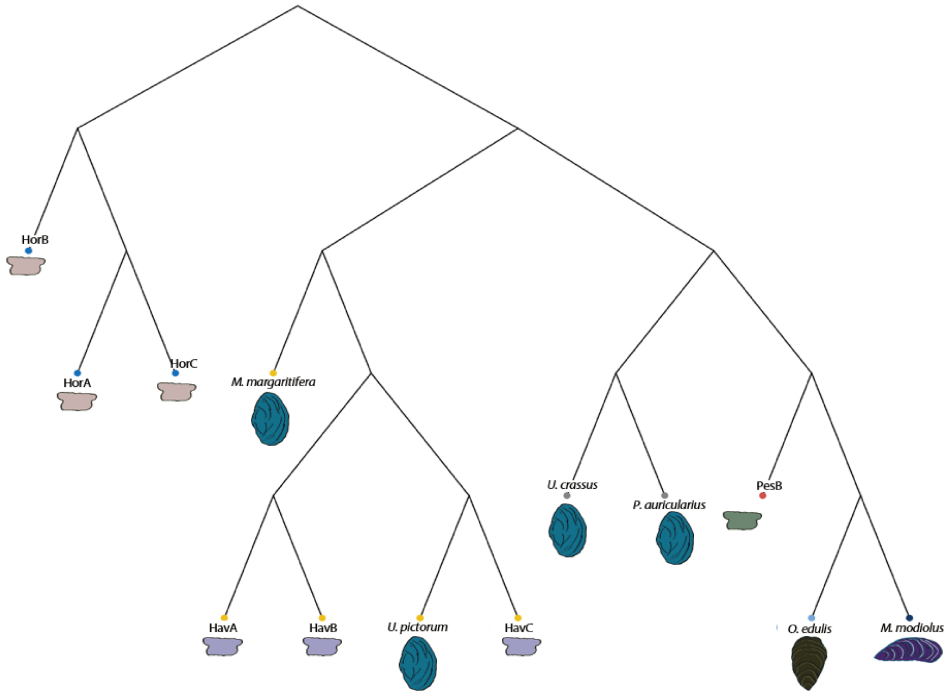


Figure S3.11: Pairwise MS/MS comparison of the seven archaeological double-buttons and six reference shells (freshwater and marine): the cluster dendrogram is obtained from a distance matrix from proteome-wide distance calculations of product ion spectra implemented in R using the DISMS2 code.

A dendrogram representation was used to explore the shell proteome differences between samples (Figure S3.11). The position of *M. modiolus*, which also has a nacreous inner structure with an upper prismatic layer, clustering apart from all of the freshwater unionoids, is interesting, as it may be related to the evolution and adaptation of different calcification processes (Arivalagan et al., 2017; Marie et al., 2009). In fact, *M. modiolus* clustering together with *O. edulis* supports our previous results, i.e. that marine shells have a distinct proteome profile, diverging from freshwater molluscs. The archaeological double-buttons from Hornstaad and Havnø cluster close to *Unio pictorum,* except for HavA, which falls together with the freshwater pearl mussel *Margaritifera margaritifera*. Sub-fossil samples of *Unio crassus* and *Pseudunio auricularius* cluster together independently and this could be explained by the vicinity of shell proteomes and possibly the alteration of the profiles due to diagenesis, as well as a possible effect of the analytical platform used (both samples were analysed in Turin). PesB sample, also analysed in Turin, falls in the middle, not showing a clear and evident proximity to one particular sample, suggesting phylogenetic distance or alteration of the sample, coherently with chiral amino acid data (Figure S3.5) and all other database-dependant analyses. Overall, the consistent pattern revealed by database-dependent and database-independent protein data analysis as well as chiral amino acid analysis (these were all performed in the same laboratory in York) strongly support a biological effect (phylogenetic or diagenetic) and not an analytical effect for the clustering.

References

Albeck, S., Aizenberg, J., Addadi, L., Weiner, S., 1993. Interactions of various skeletal intracrystalline components with calcite crystals. Journal of the American Chemical Society 115, 11691–11697.

Andersen, S.H., 2000. “Køkkenmøddinger” (shell middens) in Denmark: a survey. In: Proceedings of the Prehistoric Society. Cambridge University Press, pp. 361–384.

Arivalagan, J., Yarra, T., Marie, B., Sleight, V.A., Duvernois-Berthet, E., Clark, M.S., Marie, A., Berland, S., 2017. Insights from the Shell Proteome: Biomineralization to Adaptation. Molecular biology and evolution 34, 66–77.

Aucour, A.-M., Sheppard, S.M.F., Savoye, R., 2003. δ 13 C of fluvial mollusk shells (Rhône River): A proxy for dissolved inorganic carbon? Limnology and oceanography 48, 2186–2193.

Bajnóczi, B., Schöll-Barna, G., Kalicz, N., Siklósi, Z., Hourmouziadis, G.H., Ifantidis, F., Kyparissi-Apostolika, A., Pappa, M., Veropoulidou, R., Ziota, C., 2013. Tracing the source of Late Neolithic Spondylus shell ornaments by stable

isotope geochemistry and cathodoluminescence microscopy. Journal of archaeological science 40, 874–882.
Bartolomei, G., 2013. I micromammiferi della grotta Peştera Ungurească nella gola di Cheile Turzii. Atti della Società per la Preistoria e Protostoria della Regione Friuli-Venezia Giulia 2010-2011, 107–117.
Bar-Yosef Mayer, D.E., Leng, M.J., Aldridge, D.C., Arrowsmith, C., Gümüş, B.A., Sloane, H.J., 2012. Modern and early-middle Holocene shells of the freshwater mollusc Unio, from Çatalhöyük in the Konya Basin, Turkey: preliminary palaeoclimatic implications from molluscan isotope data. Journal of archaeological science 39, 76–83.
Biagi, P., Voytek, B., 2006. Excavations at Pestera Ungurească (Caprelor) (Cheile Turzii, Petresti de Jos, Transylvania) 2003-2004: A preliminary report on the chipped stone assemblages from the Chalcolithic Toarte Pastilate (Bodrogkeresztur) layers. Analele Banatului 14, 177–202.
Boggild, O.B., 1930. The shell structure of the mollusks. A.F. Host, Kjobenhavn.
Bolotov, I.N., Vikhrev, I.V., Bespalaya, Y.V., Gofarov, M.Y., Kondakov, A.V., Konopleva, E.S., Bolotov, N.N., Lyubas, A.A., 2016. Multi-locus fossil-calibrated phylogeny, biogeography and a subgeneric revision of the Margaritiferidae (Mollusca: Bivalvia: Unionoida). Molecular phylogenetics and evolution 103, 104–121.
Borrello, M., Girod, A., 2008. Bivalves d'eau douce, matière première pour la fabrication de parures en Suisse, du Néolithique à l'âge du Bronze. Annual review of Swiss Archaeology 108–115.
Boschian, G., 2010. Soil micromorphological observations on the anthropogenous deposits of Peştera Ungurească. Atti della Società per la Preistoria e Protostoria della Regione Friuli-Venezia Giulia 2008-2009, 157–169.
Bosch, M.D., Mannino, M.A., Prendergast, A.L., O'Connell, T.C., Demarchi, B., Taylor, S.M., Niven, L., Plicht, J. van der, Hublin, J.-J., 2015. New chronology for Ksâr 'Akil (Lebanon) supports Levantine route of modern human dispersal into Europe. Proceedings of the National Academy of Sciences of the United States of America 112, 7683–7688.
Carter, J.G. (Ed.), 1990. Skeletal Biomineralization: Pattern, Processes, and Evolutionary Trends. Amer Geophysical Union.
Crisp, M., Demarchi, B., Collins, M., Morgan-Williams, M., Pilgrim, E., Penkman, K., 2013. Isolation of the intra-crystalline proteins and kinetic studies in Struthio camelus (ostrich) eggshell for amino acid geochronology. Quaternary geochronology 16, 110–128.
Demarchi, B., Williams, M.G., Milner, N., Russell, N., Bailey, G., Penkman, K., 2011. Amino acid racemization dating of marine shells: A mound of possibilities. Quaternary international: the journal of the International Union for Quaternary Research 239, 114–124.

Demarchi, B., O'Connor, S., Lima Ponzoni, A. de, Almeida Rocha Ponzoni, R. de, Sheridan, A., Penkman, K., Hancock, Y., Wilson, J., 2014. An integrated approach to the taxonomic identification of prehistoric shell ornaments. PloS one 9, e99839.

Demarchi, B., Clements, E., Coltorti, M., Locht, R. van de, Kröger, R., Penkman, K., Rose, J., 2015. Testing the effect of bleaching on the bivalve Glycymeris: A case study of amino acid geochronology on key Mediterranean raised beach deposits. Quaternary geochronology 25, 49–65.

Dieckmann, B., Harwath, A., Heumüller, M., Hoffstadt, J., Maier, U., Matuschik, I., Schwoerbel, A., Stephan, E., Schweizer-Strobel, P., Styring, A., Schlichtherle, H., Theune-Großkopf, B., 2016. Eine kurze Dorfgeschichte. Hornstaad-Hörnle I am Bodensee. In: 4000 Jahre Pfahlbauten. J. Thorbecke.

Eerkens, J.W., Herbert, G.S., Rosenthal, J.S., Spero, H.J., 2005. Provenance analysis of Olivella biplicata shell beads from the California and Oregon Coast by stable isotope fingerprinting. Journal of archaeological science 32, 1501–1514.

Gao, P., Liao, Z., Wang, X.-X., Bao, L.-F., Fan, M.-H., Li, X.-M., Wu, C.-W., Xia, S.-W., 2015. Layer-by-Layer Proteomic Analysis of Mytilus galloprovincialis Shell. PloS one 10, e0133913.

Gillikin, D.P., Hutchinson, K.A., Kumai, Y., 2009. Ontogenic increase of metabolic carbon in freshwater mussel shells (Pyganodon cataracta). Journal of geophysical research 114, 2186.

Girod, A., 2010a. Land and freshwater snails from Peştera Ungurească (Cheile Turzii Gorge, Transylvania, Romania). Atti della Società per la Preistoria e Protostoria della Regione Friuli-Venezia Giulia 2008-2009, 187–199.

Girod, A., 2010b. Lavorazioni su conchiglia di Pseudunio auricularius (Mollusca, Bivalvia) nel Neolitico antico a Isorella (BS), Italia settentrionale. Atti della Società italiana di scienze naturali e del museo civico di storia naturale di Milano 151, 89–98.

Gower, J.C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53, 325–338.

Henry, D.G., Watson, J.S., John, C.M., 2017. Assessing and calibrating the ATR-FTIR approach as a carbonate rock characterization tool. Sedimentary geology 347, 36–52.

Heumüller, M., 2009. Der Schmuck der jungneolithischen Seeufersiedlung Hornstaad-Hörnle IA im Rahmen des mitteleuropäischen Mittel- und Jungneolithikums. K. Theiss.

Heumüller, M., 2012. Schmuck als Zeichen weit gespannter und lang andauernder Kommunikationsräume im 5. und frühen 4. Jahrtausend. In: Ralf, G., Valeska, B. (Eds.), Mitteleuropaim 5. Jahrtausend Vor Christus. Presented at the Internationalen Konferenz, Universität Münster, pp. 359–387.

Kannegaard, E., 2013. Småting fra Nederst. Årbog fra museum Østjylland 207–208.

Keith, M.L., Anderson, G.M., Eichler, R., 1964. Carbon and oxygen isotopic composition of mollusk shells from marine and fresh-water environments. Geochimica et cosmochimica acta 28, 1757–1786.

Launey, S., Ledu, C., Boudry, P., Bonhomme, F., Naciri-Graven, Y., 2002. Geographic structure in the European flat oyster (Ostrea edulis L.) as revealed by Microsatellite polymorphism. The Journal of heredity 93, 331–351.

Lazarovici, C.M., Lazarovici, G., 2016. The jewelry workshop at Cheile Turzii: Cave Binder, Copper Age. Annales Universitatis Apulensis. Series Historica 20, 335–357.

Lazarovici, G., Lazarovici, C.M., 2013. Etapa veche din atelierul de bijuterii de la Cheile Turzii - Peştera Ungurească. Archeovest I.

Leng, M.J., Lewis, J.P., 2016. Oxygen isotopes in Molluscan shell: Applications in environmental archaeology. Environmental Archaeology 21, 295–306.

Le Roy, N., Jackson, D.J., Marie, B., Ramos-Silva, P., Marin, F., 2014. The evolution of metazoan α-carbonic anhydrases and their roles in calcium carbonate biomineralization. Frontiers in zoology 11, 75.

Liu, X., Dong, S., Jin, C., Bai, Z., Wang, G., Li, J., 2015. Silkmapin of Hyriopsis cumingii, a novel silk-like shell matrix protein involved in nacre formation. Gene 555, 217–222.

Liu, X., Jin, C., Wu, L., Dong, S., Zeng, S., Li, J., 2017a. Hic74, a novel alanine and glycine rich matrix protein related to nacreous layer formation in the mollusc Hyriopsis cumingii. Aquaculture and Fisheries 2, 119–123.

Liu, X., Pu, J., Zeng, S., Jin, C., Dong, S., Li, J., 2017b. Hyriopsis cumingii Hic52-A novel nacreous layer matrix protein with a collagen-like structure. International journal of biological macromolecules 102, 667–673.

Loftus, E., Rogers, K., Lee-Thorp, J., 2015. A simple method to establish calcite:aragonite ratios in archaeological mollusc shells. Journal of Quaternary Science 30, 731–735.

Lopes-Lima, M., Sousa, R., Geist, J., Aldridge, D.C., Araujo, R., Bergengren, J., Bespalaya, Y., Bódis, E., Burlakova, L., Van Damme, D., Douda, K., Froufe, E., Georgiev, D., Gumpinger, C., Karatayev, A., Kebapçi, Ü., Killeen, I., Lajtner, J., Larsen, B.M., Lauceri, R., Legakis, A., Lois, S., Lundberg, S., Moorkens, E., Motte, G., Nagel, K.-O., Ondina, P., Outeiro, A., Paunovic, M., Prié, V., Proschwitz, T. von, Riccardi, N., Rudzīte, M., Rudzītis, M., Scheder, C., Seddon, M., Şereflişan, H., Simić, V., Sokolova, S., Stoeckl, K., Taskinen, J., Teixeira, A., Thielen, F., Trichkova, T., Varandas, S., Vicentini, H., Zajac, K., Zajac, T., Zogaris, S., 2017. Conservation status of freshwater mussels in Europe: state of the art and future challenges: Conservation of European freshwater mussels. Biological Reviews, Invertebrates (Part 1) 92, 572–607.

Lydeard, C., Cowie, R.H., Ponder, W.F., Bogan, A.E., Bouchet, P., Clark, S.A., Cummings, K.S., Frest, T.J., Gargominy, O., Herbert, D.G., Hershler, R., Perez, K.E., Roth, B., Seddon, M., Strong, E.E., Thompson, F.G., 2004. The Global Decline of Nonmarine Mollusks. Bioscience 54, 321–330.

Marie, B., Marin, F., Marie, A., Bédouet, L., Dubost, L., Alcaraz, G., Milet, C., Luquet, G., 2009. Evolution of nacre: biochemistry and proteomics of the shell organic matrix of the cephalopod Nautilus macromphalus. Chembiochem: a European journal of chemical biology 10, 1495–1506.

Marie, B., Arivalagan, J., Mathéron, L., Bolbach, G., Berland, S., Marie, A., Marin, F., 2017. Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the Unionoida Elliptio complanata and Villosa lienosa. Journal of the Royal Society, Interface / the Royal Society 14. doi:10.1098/rsif.2016.0846

Marin, F., Luquet, G., 2004. Molluscan shell proteins. Comptes rendus. Palevol 3, 469–492.

Marin, F., Pokroy, B., Luquet, G., Layrolle, P., De Groot, K., 2007. Protein mapping of calcium carbonate biominerals by immunogold. Biomaterials 28, 2368–2377.

Marin, F., Le Roy, N., Marie, B., 2012. The formation and mineralization of mollusk shell. Frontiers in bioscience 4, 1099–1125.

Marin, F., Marie, B., Hamada, S.B., Ramos-Silva, P., Le Roy, N., Guichard, N., Wolf, S.E., Montagnani, C., Joubert, C., Piquemal, D., Saulnier, D., Gueguen, Y., 2013. “Shellome”: Proteins involved in mollusk shell biomineralization-diversity, functions. In: Watabe, S., Maeyama, K., Nagasawa, H. (Eds.), International Symposium on Pearl Research. Terrapub Tokyo, pp. 149–166.

Marin, F., Bundeleva, I., Takeuchi, T., Immel, F., Medakovic, D., 2016. Organic matrices in metazoan calcium carbonate skeletons: Composition, functions, evolution. Journal of structural biology 196, 98–106.

McConnaughey, T.A., Gillikin, D.P., 2008. Carbon isotopes in mollusk shell carbonates. Geo-Marine Letters 28, 287–299.

Nisbet, R., 2010. Copper age wild and cultivated plants from Peştera Ungurească cave (Transylvania, Romania). Atti della Società per la Preistoria e Protostoria della Regione Friuli-Venezia Giulia 2008-2009, 171–185.

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

Palmblad, M., Deelder, A.M., 2012. Molecular phylogenetics by direct comparison of tandem mass spectra. Rapid communications in mass spectrometry: RCM 26, 728–732.

Penkman, K.E.H., Preece, R.C., Keen, D.H., Maddy, D., Schreve, D.C., Collins,

M.J., 2007. Testing the aminostratigraphy of fluvial archives: the evidence from intra-crystalline proteins within freshwater shells. Quaternary science reviews 26, 2958–2969.

Penkman, K.E.H., Kaufman, D.S., Maddy, D., Collins, M.J., 2008. Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells. Quaternary geochronology 3, 2–25.

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

Pierini, F., Demarchi, B., Turner, J., Penkman, K., 2016. Pecten as a new substrate for IcPD dating: The quaternary raised beaches in the Gulf of Corinth, Greece. Quaternary geochronology 31, 40–52.

Ramos-Silva, P., Benhamada, S., Le Roy, N., Marie, B., Guichard, N., Zanella-Cléon, I., Plasseraud, L., Corneillat, M., Alcaraz, G., Kaandorp, J., Marin, F., 2012. Novel molluskan biomineralization proteins retrieved from proteomics: a case study with Upsalin. Chembiochem: a European journal of chemical biology 13, 1067–1078.

Rieder, V., Blank-Landeshammer, B., Stuhr, M., Schell, T., Biß, K., Kollipara, L., Meyer, A., Pfenninger, M., Westphal, H., Sickmann, A., Rahnenführer, J., 2017. DISMS2: A flexible algorithm for direct proteome- wide distance calculation of LC-MS/MS runs. BMC bioinformatics 18, 148.

Rowley-Conwy, P., 2014. Foragers and farmers in Mesolithic/Neolithic Europe, 5500-3900 cal BC: beyond the anthropological comfort zone. In: Foulds, F., Drinkall, H., Perri, A., Clinnick, D. & Walker, J. (Ed.), Wild Things: Recent Advances in Palaeolithic and Mesolithic Research. Oxford: Oxbow Books, pp. 185–201.

Schlichtherle, H., 1990. Die Sondagen 1973–1978 in den Ufersiedlungen Hornstaad-Hörnle I. In: Siedlungsarchäologie Im Alpenvorland I. Forsch. u. Ber. zur Vor- u. Frühgesch. Baden-Württemberg 36. K. Theiss.

Sudo, S., Fujikawa, T., Nagakura, T., Ohkubo, T., Sakaguchi, K., Tanaka, M., Nakashima, K., Takahashi, T., 1997. Structures of mollusc shell framework proteins. Nature 387, 563–564.

Taborin, Y., 1974. La parure en coquillage de l'Épipaléolilhique au Bronze ancien en France. Gallia Préhistoire 101–179.

Taylor, J.D., Kennedy, W.J., Hall, A., 1969. The shell structure and mineralogy of Bivalvia: introduction, Nuculacea-Trigonacea, Bulletin of the British Museum (Natural History), Zoology, Supplement. The British Museum (Natural History), London.

Taylor, J.D., Kennedy, W.J., Hall, A., 1973. The shell structure and mineralogy of the Bivalvia. II. Lucinacea-Clavagellacea, Conclusions, 9. The British Museum (Natural History), London.

Vanhaeren, M., 2004. Tracing the source of Upper Palaeolithic shell beads by strontium isotope dating. Journal of archaeological science 31, 1481–1488.

Verdegaal, S., Troelstra, S.R., Beets, C.J., Vonhof, H.B., 2005. Stable isotopic records in unionid shells as a paleoenvironmental tool. Netherlands Journal of Geosciences 84, 403–408.

Versteegh, E.A.A., Vonhof, H.B., Troelstra, S.R., Kaandorp, R.J.G., Kroon, D., 2010. Seasonally resolved growth of freshwater bivalves determined by oxygen and carbon isotope shell chemistry: freshwater bivalve intra-annual growth. Geochemistry, Geophysics, Geosystems 11. doi:10.1029/2009GC002961

Wang, S., Zhang, J., Jiao, W., Li, J., Xun, X., Sun, Y., Guo, X., Huan, P., Dong, B., Zhang, L., Hu, X., Sun, X., Wang, J., Zhao, C., Wang, Y., Wang, D., Huang, X., Wang, R., Lv, J., Li, Y., Zhang, Z., Liu, B., Lu, W., Hui, Y., Liang, J., Zhou, Z., Hou, R., Li, X., Liu, Y., Li, H., Ning, X., Lin, Y., Zhao, L., Xing, Q., Dou, J., Li, Y., Mao, J., Guo, H., Dou, H., Li, T., Mu, C., Jiang, W., Fu, Q., Fu, X., Miao, Y., Liu, J., Yu, Q., Li, R., Liao, H., Li, X., Kong, Y., Jiang, Z., Chourrout, D., Li, R., Bao, Z., 2017. Scallop genome provides insights into evolution of bilaterian karyotype and development. Nature ecology & evolution 1, 120.

Westbroek, P., Marin, F., 1998. A marriage of bone and nacre. Nature 392, 861–862.

Wustman, B.A., Morse, D.E., Evans, J.S., 2002. Structural Analyses of Polyelectrolyte Sequence Domains within the Adhesive Elastomeric Biomineralization Protein Lustrin A. Langmuir: the ACS journal of surfaces and colloids 18, 9901–9906.

Yılmaz, Ş., Victor, B., Hulstaert, N., Vandermarliere, E., Barsnes, H., Degroeve, S., Gupta, S., Sticker, A., Gabriël, S., Dorny, P., Palmblad, M., Martens, L., 2016. A Pipeline for Differential Proteomics in Unsequenced Species. Journal of proteome research 15, 1963–1970.