The degradation of intracrystalline mollusc shell proteins: a proteomics study of Spondylus gaederopus.

Jorune Sakalauskaite, Meaghan Mackie, Alberto J. Taurozzi, Matthew J. Collins, Frédéric Marin, Beatrice Demarchi

▶ **To cite this version:**

Jorune Sakalauskaite, Meaghan Mackie, Alberto J. Taurozzi, Matthew J. Collins, Frédéric Marin, et al.. The degradation of intracrystalline mollusc shell proteins: a proteomics study of Spondylus gaederopus.. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2021, 1869 (12), pp.140718. ⟨10.1016/j.bbapap.2021.140718⟩. ⟨hal-03379820⟩

HAL Id: hal-03379820

https://hal.science/hal-03379820v1

Submitted on 15 Nov 2022

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

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

HAL Authorization

The degradation of intracrystalline mollusc shell proteins: a proteomics study of *Spondylus gaederopus*

List of authors:

Jorune Sakalauskaite[1,2,3], Meaghan Mackie[3,4], Alberto J. Taurozzi[3], Matthew J. 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.

Corresponding authors:

Jorune Sakalauskaite, jorune@palaeome.org, Department of Life Sciences and Systems Biology, University of Turin, Via Accademia Albertina 13, 10123 Turin, Italy.

**Current address: Section for Evolutionary Genomics, GLOBE Institute, Faculty of Health and Medical Science, University of Copenhagen, Øster Farimagsgade 5, 1353 Copenhagen, Denmark*

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

Keywords

Protein degradation, peptide bond hydrolysis, TMT proteomics, liquid chromatography-tandem mass spectrometry

Abstract

Mollusc shells represent excellent systems for the preservation and retrieval of genuine biomolecules from archaeological or palaeontological samples. As a consequence, the *post-mortem* breakdown of intracrystalline mollusc shell proteins has been extensively investigated, particularly with regard to its potential use as a “molecular clock” for geochronological applications. But despite seventy years of ancient protein research, the fundamental aspects of diagenesis-induced changes to protein structures and sequences remain elusive. In this study we investigate the degradation of intracrystalline proteins by performing artificial degradation experiments on the shell of the thorny oyster, *Spondylus gaederopus*, which is particularly important for archaeological research. We used immunochemistry and tandem mass tag (TMT) quantitative proteomics to simultaneously track patterns of structural loss and of peptide bond hydrolysis.

Powdered and bleached shell samples were heated in water at four different temperatures (80, 95, 110, 140 °C) for different time durations. The structural loss of carbohydrate and protein groups was

investigated by immunochemical techniques (ELLA and ELISA) and peptide bond hydrolysis was studied by tracking the changes in protein/peptide relative abundances over time using TMT quantitative proteomics. We find that heating does not induce instant organic matrix decay, but first facilitates the uncoiling of cross-linked structures, thus improving matrix detection. We calculated apparent activation energies of structural loss: E_a (carbohydrate groups) = 104.7 kJ/mol, E_a (protein epitopes) = 104.4 kJ/mol, which suggests that secondary matrix structure degradation may proceed simultaneously with protein hydrolysis. While prolonged heating at 110 °C (10 days) results in complete loss of the structural signal, surviving peptide sequences were still observed. Eight hydrolysis-prone peptide bonds were identified in the top scoring shell sequence, the uncharacterised protein LOC117318053 from *Pecten maximus*. Interestingly, these were not the expected “weak” bonds based on published theoretical stabilities calculated for peptides in solution. This further confirms that intracrystalline protein degradation patterns are complex and that the overall microchemical environment plays an active role in protein stability. Our TMT approach represents a major stepping stone towards developing a model for studying protein diagenesis in biomineralised systems.

1. Introduction

Interest in ancient mollusc shell proteins and how they break down over time dates back to the middle of the 20th century, when chiral amino acids were detected in fossil shell samples [1]. Mollusc shells are organo-mineral nano-composite biomaterials that show excellent preservation potential in the palaeontological record due to their compact mineralised structures. The shell is composed of calcium carbonate (mainly calcite and/or aragonite) and a small fraction of organic matrix [2]: a suite of proteins, carbohydrates, lipids, free peptides, pigments and other metabolites [3]. The organics become occluded inside the mineral skeleton during shell formation (*i.e.* biomineralisation), either between mineral crystallites ('intercrystalline') or tightly bound to the mineral phase within the crystallites, ('intracrystalline') [4,5]. Intracrystalline proteins have attracted attention for their role in shell biomineralisation, for understanding mechanisms of protein-mineral interaction [3,6–8] and as a source of genuine geobiological information, since mineral-associated proteins can survive in deep time [2,4,9–14]. The analysis of the amino acid composition and extent of racemisation in shells has been extensively used as a chemical clock to determine the relative age of fossils [13,15–20].

Robust identification of ancient protein sequences became possible only in the last two decades with the advent of soft ionisation mass spectrometry techniques [21–24] (see also recent reviews in [25–27]). Among these, liquid chromatography coupled to electrospray ionisation (ESI) and tandem mass spectrometry (LC-ESI-MS/MS) has been especially fruitful for the characterisation of whole ancient proteomes (palaeoproteomics). Ancient protein studies have been carried out on a wide variety of bioarchaeological or palaeontological remains [27] enabling the identification of million-year-old proteins from dental enamel [28–30] as well as from ostrich eggshell (OES) dating to 3.8 million years ago [31]. We have also recently developed a palaeoproteomic approach for archaeological mollusc shells, which we dubbed "palaeoshellomics" [32]. However, the mechanisms of protein decay (i.e. diagenesis) in biomineral systems, particularly in mollusc shells, are still poorly understood [31,33–37].

Diagenesis is the sum of biological, chemical and physical *post mortem* processes taking place in the burial environment, which ultimately cause the transformation of a living organism into its constituent atoms. These processes are complex and governed by a number of different physico-chemical reactions that occur in the mineral and organic phases, interdependently [14,38]. Key factors affecting diagenesis in biominerals include: pH (carbonates dissolve in acidic soils) and temperature conditions of the burial environment; the age of the sample; its mineralogy, structure and compactness (calcite vs aragonite arranged into nacreous, prismatic, foliated, crossed-lamellar or other microstructures [39,40]); the composition of the shell proteome and the presence of other macromolecular compounds, such as saccharides [41] or lipids, as well as the presence of chemically reactive water.

Protein diagenesis occurs via peptide bond hydrolysis, amino acid racemisation and decomposition as well as via other diagenesis-induced modifications. Protein hydrolysis is the breakdown of peptide bonds into smaller peptides and ultimately free amino acids [18,42,43]. Amino acid racemisation is the *post-mortem* spontaneous interconversion between the D- and L- chiral forms of amino acids [15,16]. The extent of amino acid racemisation (AAR) determined on "fossil" intracrystalline mollusc shell proteins can be used as a relative geochronological tool (e.g. [19]). Decomposition refers to the decay of amino acids into their smaller molecular components, as for example the dehydration of Ser to Ala, which may also be relevant for geochronological applications [13,44–46]. There are a number of different diagenesis-induced modifications [14,47,48] and some of the best studied include the deamidation of asparagine (Asn) or glutamine (Gln) amide side chains [49], the oxidation of tryptophan (Trp) and methionine (Met) [37,50] and also Maillard-type condensation reactions [51–54]. The rate of Asn/Gln deamidation [49,55] can be used to validate sequences, where deamidation is assumed to be

a marker for preservation/authenticity [31,49,55–59]; however, the analysis is not straightforward since deamidation rates can be affected by a number of environmental/chemical factors [57,60,61]

Understanding the way in which early diagenesis affects protein fragmentation is particularly important in palaeoproteomic research in order to predict the survival of ancient sequences in archaeological and sub-fossil samples. However, in biomineral systems, modelling hydrolysis reactions is difficult, because the degradation of the same protein sequence carried out under laboratory conditions (e.g. in water solution, at a constant temperature) and in the environment (i.e. within the biomineral) differs significantly [18,53,62,63]. Degradation studies of synthetic peptides in aqueous solutions [34] fail to fully reflect protein behaviour in a biomineralised system due to the presence of structural factors, such as protein-mineral interactions [31]. For example, an acidic (Asp-rich) peptide from the main protein of ostrich eggshell (OES) was preserved in several 3.8 million-year-old OES fragments but was found to be unstable when heated in water in the absence of a mineral phase. Furthermore, the rate of Ser racemisation from a different (hydrophobic) OES peptide was found to be approximately six times faster in solution than in the biomineral [31,34]. Another crucial point is the limited knowledge on the role and nature of other (non-proteinaceous) organic components, e.g. glycosides, lipids, which can facilitate matrix cross-links and be transformed into highly insoluble advanced glycoxidation and lipoxidation end products (AGEs and ALEs - [41,47,51,64,65]). The key issue is that there are no quantitative real-time data on how mineral-bound protein sequences undergo breakdown in biominerals: all our knowledge is based on monitoring changes in amino acid concentrations and, at best, coupling these with qualitative information on sequence modifications.

In this study we investigate the diagenesis of intracrystalline shell matrix proteins at structural and proteomic level. First, we performed accelerated diagenesis experiments by subjecting powdered and bleached *Spondylus gaederopus* Linnaeus, 1758 (Pectinida, Spondylidae) mollusc shell samples to heating in the laboratory (Figure 1). We chose *S. gaederopus* as a model system due to its importance in archaeological research, as it was a widely used material in Europe from the Neolithic [66–69]. We are aware that the rapid degradation induced by laboratory conditions does not reflect the slow aging processes that occur naturally [63], but such experiments provide a solid foundation for elucidating some of the diagenetic pathways occurring in biominerals and as such, heating experiments have been exploited in a number of different kinetic studies on intracrystalline organics (e.g. [13,31,70–74]). Secondly, we extracted intracrystalline shell matrices from heated *Spondylus* samples and characterised them by immunochemistry and quantitative proteomics using tandem mass tag (TMT) labelling (Figure 1). We target the degradation patterns occurring within the intracrystalline fraction of proteins, as this typically displays a “closed-system” behaviour in many gastropods and bivalves [13,75,76]. In particular, it has been shown that “truly” intracrystalline proteins, which are isolated by an extensive bleaching step [11,13], do not ‘leach’ from the system in ‘leaching experiments’ carried out at high temperatures in aqueous environments [13]. This means that these proteins exhibit the highest potential for being retained (intact or degraded) in the system – in artificially aged shells or subfossil samples. For the TMT approach, we labelled the samples heated at multiple time-points with different tags, pooled the fractions and analysed them in a single run using high performance LC-MS/MS [77]. This approach has been successfully used to study other types of biological tissues [78,79], and in our study provides temporally-resolved measurements of peptide (protein) abundance in thermally-aged mollusc shell samples.

Overall, the main objectives of this work are:

i) to investigate how the intracrystalline *Spondylus* shell organic matrix is affected by accelerated aging and to calculate the rate of structural loss;

ii) to study patterns of intracrystalline protein degradation through quantitative proteomics and

iii) to identify some of the pathways of peptide bond hydrolysis in *Spondylus* shell proteins.

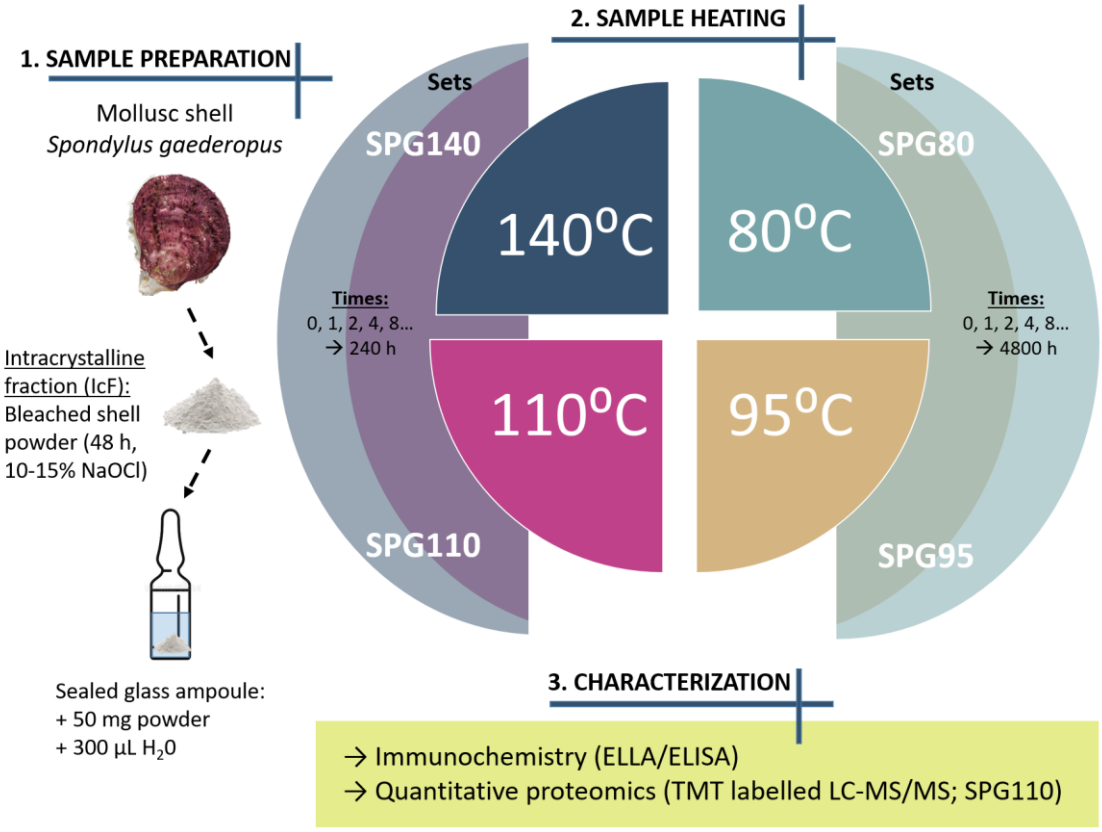


Figure 1. Experimental design of the methodology used in this study. The scheme displays the main steps of sample preparation, heating and matrix characterisation. Powdered shell samples were bleached with concentrated NaOCl (10-15%) for 48 h to isolate the intracrystalline fraction of organics (IcF). 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 (comprising four SPG sets - name stands for the species - ***Sp****ondylus* ***g****aederopus*; for more details, see materials and methods section 2.1). Each sample was prepared in triplicate. The samples were characterised by immunochemical techniques (ELLA/ELISA) and the set SPG110 was analysed by quantitative proteomics.

2. Materials and Methods

2.1 Set up of the heating experiments and sample preparation

Modern *S. gaederopus* shells were purchased from Conchology, Inc (https://www.conchology.be/). According to the vendors the specimens were collected live at 15m depth in Saronikos (Greece) in 2010. One specimen had been prepared by cleaning the crushed shell with diluted bleach (1.0-1.5% NaOCl), 4 h of soaking [80], and was available as a finely ground powder (particle size < 200 µm). The intracrystalline fraction (IcF) was isolated by further bleaching for 48 h with concentrated NaOCl (10-15%; 1:50 w/v powder/bleach ratio). The bleach solution was centrifuged, the powder was rinsed with water (HPLC-grade, 5x times) and ethanol (HPLC-grade, 1x times) before being air-dried. The glass ampoules (Wheaton, 1 mL, clear) used for heating were also cleaned by bleaching (NaOCl 10-15%, 48 h), then rinsed with H_2O (HPLC-grade) and dried in the oven. 50 mg aliquots of *Spondylus* powder were placed into glass ampoules, 300 µL of H_2O (HPLC-grade) added, and the ampoules 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. We grouped the samples into sets SPG80, SPG95, SPG110, SPG140 according to these temperatures (the abbreviation SPG stands for the species – ***Sp****ondylus*

***g**aederopus*) and for simplicity, we use these set names to describe our samples in our study (see also Figure 1 and SI table 2). Heating times: samples of sets SPG110 and SPG140 were heated for 1, 2, 4, 8, 16, 24, 48, 96, 240 h; samples of set SPG80 were heated for 1, 2, 4, 8, 16, 24, 48, 96, 240, 480, 1200, 2400, 3600, 4800 h; samples of set SPG95 were heated for 1, 2, 4, 8, 12, 16, 24, 48, 96, 192, 240, 384, 480, 786, 1536, 4800 h. In addition, all sets contained non-heated samples ("0 hours"). Three replicates were prepared for each time point. After heating, the samples were cooled down at room temperature and stored at -20 °C. The samples were then defrosted at room temperature to separate the supernatant solutions. The powders were air-dried and weighed into separate plastic tubes (Eppendorf LoBind microcentrifuge tubes): 5 mg aliquots were set aside for immunochemical analysis (two replicates for each time point and temperature) and 35 mg were used for proteomics. The full list of samples is presented in SI table 2.

2.2 Immunochemistry

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 the (inter+intra) crystalline fraction (IIF, obtained by extracting the organics from the shell powder that was not subjected to the 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 had previously been shown to yield strong cross reactivity with *Spondylus* shell organic matrix [80]. Lectin jacalin is a carbohydrate-binding protein that recognises galactose/N-acetylgalactosamine residues and/or oligosaccharides terminating with D-galactose. Shell antibody K5090 was raised against the soluble matrix of the nacreous layer of *Pinna nobilis* [81], but the exact target epitopes of this antibody are undetermined. 5 mg of shell powders were demineralised with 200 µL 20% (w/v) EDTA solution (pH ~8.2) at room temperature, under constant agitation (1300 rpm), overnight. Upon demineralisation, 200 µL of TBS were added to reach a total volume of 400 µL and processed according to previously-described procedures [80,82]. In short, extracts were incubated in a 96-well microplates (MaxiSorp, Nunc/Thermo Scientific, Nunc A/S, Roskilde, Denmark) for 90 min at 37 °C, washed three times with TBS/Tween-20 and blocked for 30 min at 37 °C using Carbo-Free blocking solution (ref. SP 5040, Vector Laboratories) for ELLA and 1% gelatin solution in TBS for ELISA. 100 µL solutions of lectin jacalin (Vector Laboratories, Peterborough, UK, Ref. BK-3000; diluted 1:100 in TBS/Tween-20) or antibody K5090 (diluted 1:50 in 0.2% gelatin TBS/Tween-20) were added, incubated for 90 min at 37 °C and afterwards washed with TBS/Tween-20. 100 µL/well of alkaline phosphatase-conjugated avidin solution (A7294 Sigma, 1:70000) were added to ELLA microplates, and 100 µL of GAR/AP solution (A3687 Sigma, 1:30000) to ELISA plates. Microplates were incubated for another 90 min at 37 °C then thoroughly washed with TBS/Tween-20, incubated with aqueous diethanolamine solution (pH 9.8), containing phosphatase substrate (p-nitrophenylphosphate) at 37 °C. Non-specific binding of the secondary antibody was assessed by different negative controls performed in our previous study [80] and was not observed. Optical density (OD) was read with a BioRad Model 680 microplate reader at 405 nm. For calculations, background and blank absorptions (samples with no added shell matrix) were subtracted. The intensity values of heated samples were converted to the percentage of reactivity, with the non-heated sample (time point 0 h) representing 100% reactivity. The IcF matrix reactivity was calculated with respect to IIF (OD values of IIF given 100%).

To calculate apparent kinetic parameters, we used two different approaches: 1) we investigated if the degradation could be explained by a first order, non-reversible kinetic reaction; for this, we verified if a linear relationship is observed between the natural logarithm of cross-reactivity intensity (with jacalin and K5090) versus time; 2) we applied a 'model-free' approach to estimate kinetic parameters, which had been developed for other degradation reactions [63,70]. The 'model-free' approach is based on scaling the data and it uses numerical optimisation to estimate relative rate differences. Scaling was

achieved by fitting third order polynomial functions to the data at each temperature and using Generalized Reduced Gradient Algorithm (Microsoft Solver) to minimise the least squares difference. Scaling the data acquired at different temperatures in this manner yields relative rates of reaction; we chose to normalise to the mid-temperature point (110°C). In the case of the first approach (first order, non-reversible reaction), the observed reaction rate constants (k_{obs}) were calculated for the four different temperatures and for both cross-reactive substrates separately, using equation $\ln[A]=-kt+\ln[A]_o$, where [A] represents the relative abundance at time point t; $[A]_o$ is the relative abundance at time point 0; k is the reaction rate, t is time. These values were subsequently used to derive apparent activation energies (E_a) of these structural degradation reactions using the Arrhenius equation ($\ln k = -E_a/RT+\ln A$).

2.3 Proteomics

Quantitative proteomics with tandem mass tag (TMT) labelling was used to analyse sample set SPG110 (samples heated at 110 °C, see Figure 1). This set was chosen based on immunochemistry data: we selected a set where the degradation of the shell matrix was relatively rapid (in order to fit 10 samples), but also not too fast (as in the case of samples heated at 140 °C). Nine heated samples (for 1, 2, 4, 8, 16, 24, 48, 96, 240 h), one unheated sample (0 h) and one internal blank were selected for labelling with an 11-plex TMT kit (for more details, see SI table 2).

2.3.1 Extraction

Shell proteins were extracted using a FASP sample preparation method [83,84]. 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 at 4 °C and with frequent agitation (~100 rpm). Extracts were centrifuged at 13.4k rpm for 10 min to separate acid soluble (ASM) and acid insoluble (AIM) matrices; the ASMs 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 Ambic buffer (50 mM ammonium bicarbonate, pH 7.5-8). The AIMs were rinsed with water (HPLC grade, 1.5 mL aliquots, 5x times) and mixed with the ASM extracts. The proteins were reduced using 1 M stock solution of DL-Dithiothreitol (Sigma, Canada) for 1 hr at 65 °C, alkylated with 0.5 M stock solution of 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% trifluoroacetic acid (TFA, to a final TFA concentration of 0.1%), samples were purified using C18 solid-phase extraction tips (Pierce C18 tips; Thermo-Fisher) and evaporated to dryness.

2.3.2 TMT Labelling

Dried peptides were TMT labelled and analysed at the Novo Nordisk Foundation Center for Protein Research. The samples were resuspended in 50 µL 50% acetonitrile (ACN). Due to the difficulty of accurately determining the concentration of low amounts of protein by spectrophotometer analysis (at 280 nm and 205 nm), the samples were normalised by the starting weight of the shell used. 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 (for tag details, see SI table 2). 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 StageTips [85].

2.3.3 Mass Spectrometry (MS)

The StageTip was eluted using 20 µl each of 40% then 60% ACN and vacuum centrifuged at 40 °C until approximately 3 µl remained. It was then resuspended with 10 µL 0.1% TFA and 5% ACN solution. 5 µL were 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 [50], 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.

2.4 Bioinformatics

Bioinformatic analyses were performed using PEAKS Studio X software (Bioinformatic Solutions Inc [86,87], version released on January 31. 2019). 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 the 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 (any such matches were excluded from further analyses). The ProtParam tool on the Expasy portal was used to determine the Instability Index (an estimate of the stability of the protein in a test tube), the Aliphatic Index (the relative volume occupied by aliphatic side chains; may be regarded as a positive factor for the increase of thermostability of globular proteins) and the Grand Average of Hydropathicity (GRAVY, calculated as the sum of hydropathy values of all amino acids, divided by the number of residues in the sequence; increasing positive score indicates greater hydrophobicity, although no account is taken of the way the protein folds in three dimensions). Protein domains were identified using the InterPro tool [88]. Intrinsically disordered protein regions were determined by IUPred2A tool [89]. TMT quantification was 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 (%) 0.5, 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 Tables 3, 4 and in Figure 6).

3. Results

3.1 Characterisation of the intracrystalline shell matrix

The biomolecular and proteomic composition of intracrystalline (IcF) *Spondylus* shell matrix was characterised by immunochemistry assays (ELLA and ELISA) and liquid chromatography tandem mass spectrometry (LC-MS/MS). ELLA and ELISA were carried out using lectin jacalin and shell antibody K5090, respectively, and the main goal was to investigate if the IcF fraction showed different cross-reactivity compared to the bulk shell matrix fraction ((intra+inter)crystalline, IIF). In *Spondylus*, jacalin was previously found to bind a 30-34 kDa protein as well as to insoluble high molecular weight compounds [80]. Shell antibody K5090 cross-reacts with proteinaceous components of *Spondylus* organic matrix and in western blots it was shown to bind to ~50 kDa proteins as well as to some weakly soluble and high molecular weight components (for more details see SI material 1). Figure 2 shows that the cross-reactivity of the IcF matrix (i.e. that obtained after 48 h of bleaching) is moderately different compared to that of IIF, for both jacalin and antibody K5090: with jacalin the signal intensity of the IcF is lower by ~15±1%, whereas with K5090, it is higher by ~12±5%.

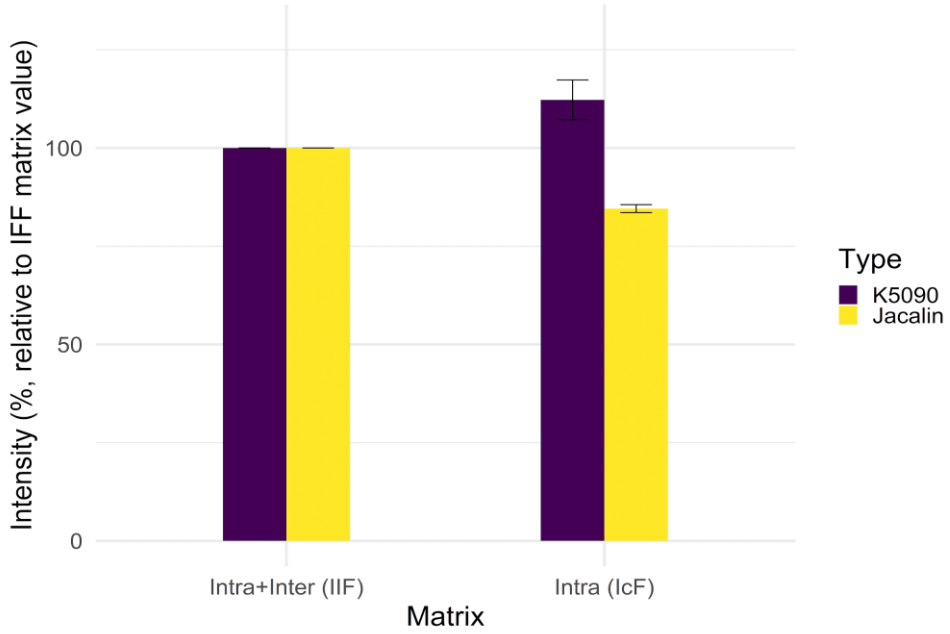


Figure 2. ELLA and ELISA assays, which compare the signal intensity between the two *Spondylus* shell matrices – intracrystalline (IcF) and (inter+intra)crystalline (IIF). Lectin jacalin was used for ELLA and antibody K5090 for ELISA. The IcF matrix is obtained via 48 h bleaching of fine shell powder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

The proteomic analysis of the IcF resulted in the identification of five protein sequences: uncharacterised protein LOC117318053, laccase-2-like, carbonic anhydrase 2-like, mucin-2-like isoform X1, aquaporin-10-like isoform X1 (Table 1). Three of the identified proteins are supported by more than two unique peptides and all sequences belong to the great scallop shell *Pecten maximus* (Pectinida, Pectinidae), which belongs to the same order as *Spondylus* sp. (Pectinida) We find that the protein sequence coverage ranges between 1-10% (maximum in uncharacterised 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* only' tags, i.e. peptide sequences that are

reconstructed by the PEAKS X algorithm but are not subsequently matched to any of the proteins provided in the database used for the search. For example, in uncharacterised protein LOC117318053, up to 155 *de novo* only tags are found.

Table 1. Intracrystalline shell proteins identified in *Spondylus gaederopus* and their characteristics. 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. *De novo* only tags refer to peptide sequences that were reconstructed by the PEAKS X software algorithm but could not be matched to any sequences provided in the molluscan protein database. BLASTp was used to verify that each supporting peptide does not match any exogenous sequences (contamination, bacterial proteins etc.).

Protein	**Accession number**	**Coverage (%)**	**Supporting peptides**	**Unique peptides**	***De novo* only tags**	**PSMs**	**pI**	**Acidic AAs (Asp/ Glu), (%)**	**Basic AAs (Arg/ Lys), (%)**	**Instability Index (II)**	**Aliphatic Index**	**GRAVY**
Uncharacterised protein LOC117318053 [*Pecten maximus*]	>gi|1835552843|ref|XP_033728946.1|	10	42	41	155	114	8.8	8.0	8.9	stable	65.84	-0.402
Laccase-2-like [*Pecten maximus*]	>gi|1835484459|ref|XP_033755522.1	3	2	2	37	4	6.4	9.9	7.8	unstable	65.55	-0.543
Carbonic anhydrase 2-like [*Pecten maximus*]	gi|1835482085|ref|XP_033754304.1|	3	2	2	19	11	6.6	14.5	13.9	stable	60.71	-0.852
Mucin-2-like isoform X1 *[Pecten maximus*]	gi|1835474913|ref|XP_033725463.1|	1	1	1	78	1	6.2	7.4	6.9	unstable	56.58	-0.52
Aquaporin-10-like isoform X1 [*Pecten maximus*]	gi|1835493482|ref|XP_033750578.1|	4	1	1	24	1	6.3	7.8	6.9	stable	93.95	0.259

The uncharacterised protein LOC117318053 from *Pecten maximus* was the best covered protein (10%) identified in the *Spondylus* IcF. Two enzymes were identified: laccase-2-like and carbonic anhydrase (CA) 2-like, as well as peptides that belong to mucin-2-like and aquaporin-10-like proteins (see also supplementary information).

Most of the identified intracrystalline *Spondylus* shell proteins have a slightly acidic pI, between ~6.2 and ~6.6 (Table 1), except for uncharacterised protein LOC117318053, which is basic (pI 8.8). The percentage of acidic/basic amino acids is <10% in all of the proteins, except for carbonic anhydrase 2-like (~15%). The majority of the sequences are classified as hydrophilic based on a negative GRAVY index, except aquaporin-10-like, which is hydrophobic. According to the Guruprasad protein instability index (II) [90], three proteins (uncharacterised 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.

3.2 Diagenesis of the intracrystalline shell matrix

We investigated the degradation patterns of the intracrystalline organic matrix in artificially-aged *Spondylus* shell samples. ELLA and ELISA were employed to assess structural degradation processes after exposure to heating. We used lectin jacalin and antibody K5090 to track the loss of reactive carbohydrate groups and protein epitope activity (Figure 3). The data show that the structural degradation of the intracrystalline shell matrix is temperature-dependent. Prolonged heating of the shell matrix decreases the affinity of both jacalin and K5090, but at different rates. Figure 3 shows that: i) at the highest temperature, 140 °C (SPG140), the binding affinity of jacalin and K5090 decreases exponentially; ii) at 110 °C (SPG110) it also decreases exponentially but during the first 48 hours (early diagenesis) the decrease is quasi-linear; and finally iii) at lower temperatures, 80 and 95 °C (SPG80 and SPG95) during the first 48 hours only a gradual linear decline of cross-reactivity is observed. Interestingly, during the first 48 hours the matrix cross-reactivity with K5090 is almost superimposable for samples heated at 80 and 95 °C, although variability is higher in the latter case.

It is also clear that the loss of signal with jacalin is much slower than with K5090:

- At 140 °C the jacalin signal disappears after 24 hours of heating. In the case of the K5090, it is almost 3 times faster.
- At 80 °C jacalin signal starts to decrease after 24 hours and it persists beyond 4800 h (200 days). In contrast, the K5090 signal starts to decrease just after a few hours and it is undetectable after 3600 h (150 days) of heating.

We also observe that in the first hours of heating, the intensity of the cross-reactivity with jacalin increases (Figure 3 a, c): after one hour of heating at 110 °C and 140 °C the signal is higher than in 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.

3.2.1 Immunochemistry

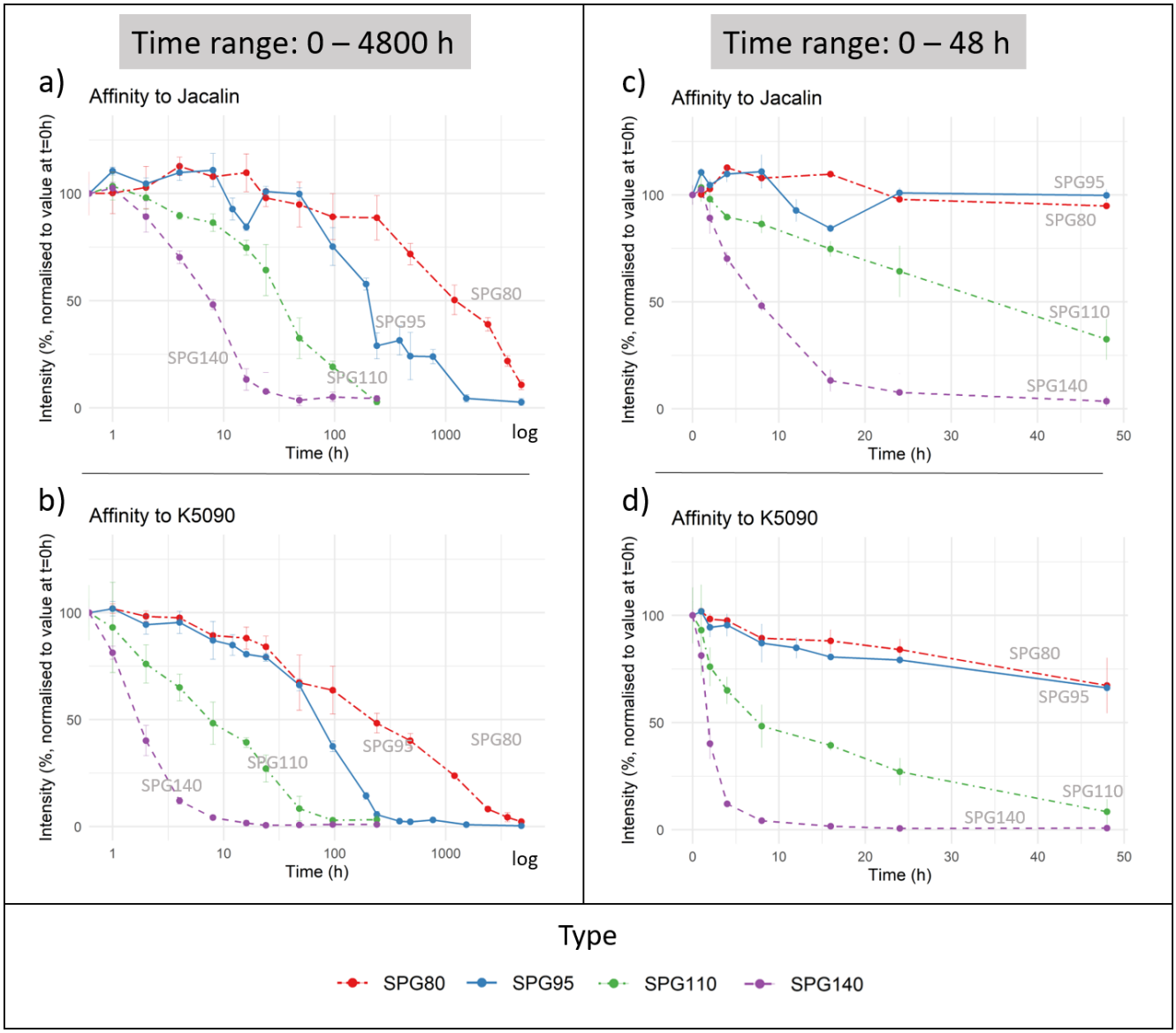


Figure 3. The line plots show the cross-reactivity changes (temperature-induced artificial diagenesis) of the intracrystalline (IcF) matrix, tracked by ELLA and ELISA using lectin jacalin and shell antibody K5090 respectively. Shell samples were heated at four different temperatures – 80, 95, 110, 140 °C. 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 degradation patterns over the full time range (note that time is plotted in logarithmic scale); graphs c), d) present results in real time for the duration of 1-48 h (early diagenesis).

The datasets were used to derive apparent kinetic parameters, i.e. observed reaction rate constants (k_{obs}) and activation energies (E_a) of the loss of cross-reactive groups/epitopes (Figure 4, Table 2). For this, we used two different approaches: 1) the first method was based on the assumption that the rate of degradation is a first order non-reversible reaction; 2) the second method was based on a 'model-free' approach, which uses numerical optimisation to estimate relative rate differences and makes no assumptions regarding the underlying kinetics of the system [63]. Using the first method we find that the matrix degradation is only partially explained by first order reaction kinetics – we observe that the linear regression shows an appropriate fit ($R^2 > 0.9$) only for the "early diagenesis" portion of the reaction (see supplementary material). We note that several time points from the "late diagenesis" portion of the

reaction had to be excluded in order to obtain linear fitting (excluded time points were the same for jacalin and K5090: 48 h (SPG140); 96, 240 h (SPG110, SPG140), 768, 1536, 4800 h (SPG95); cut-off of linear fitting $R^2 > 0.9$). Therefore, our rough estimates of apparent reaction rate constants (k_{obs}) and activation energies (E_a) only refer to early diagenesis. For this reason, we applied the alternative 'model-free' method, which is based on simple scaling of the data (Figure 4). The apparent activation energies derived using this approach are E_a (jacalin) = 104.7 kJ/mol; E_a (K5090) = 104.4 kJ/mol (Table 2). In Table 2 we also report the observed reaction rate constants (k_{obs}) and E_a for "early diagenesis' obtained using first-order reaction rate kinetics.

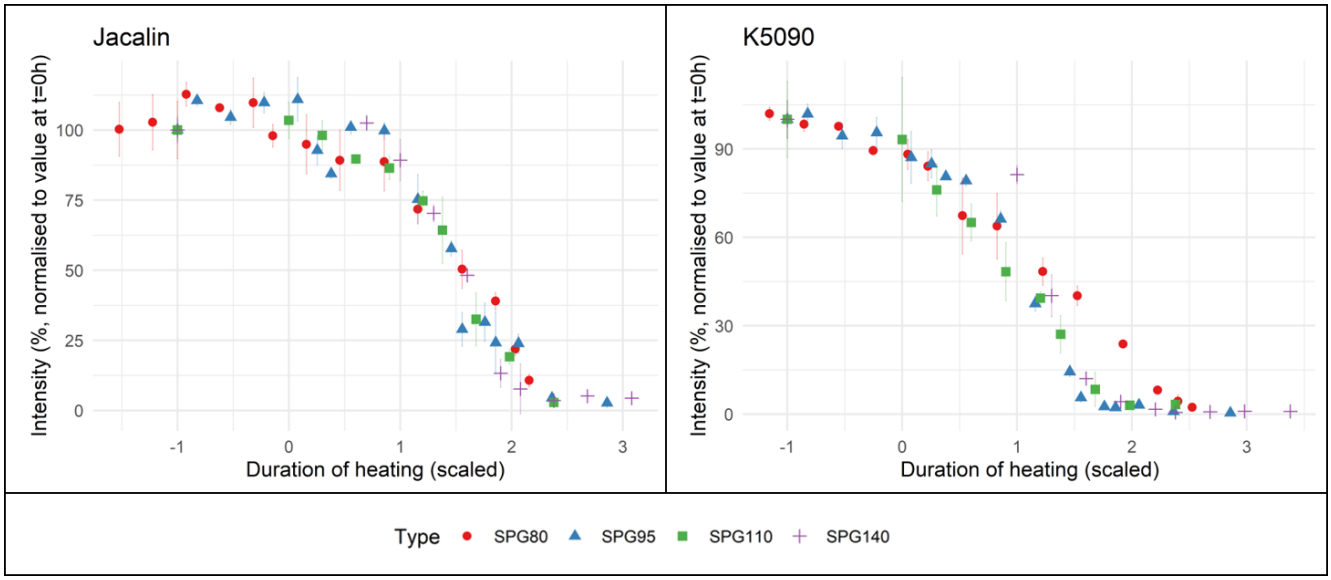


Figure 4. Overlay of the intensity of cross-reactivity values detected at different temperatures, plotted using the 'model-free' approach: with Jacalin (loss of reactive carbohydrate groups; analysed by ELLA) and antibody K5090 (loss of reactive proteinaceous epitopes; analysed by ELISA). The plots were used to derive apparent activation energies of structural degradation (presented in Table 2).

Table 2. Kinetic parameters which correspond to temperature-induced matrix degradation: the loss of reactive carbohydrate groups (affinity to lectin jacalin) and epitope cross-reactivity (affinity to shell antibody K5090). The parameters were obtained using a first order non-reversible reaction kinetics model and a 'model-free' approach. We note that the values obtained using the first approach correspond to the 'early diagenesis' processes. Cross-reactivity was assessed by ELLA/ELISA assays.

'Approach 1: 'First order' reaction				'Approach 2: 'Model-free' approach		
Temperature (ºC)	Lectin jacalin	AB: K5090			Lectin jacalin	AB: K5090
	k_{obs} (s^{-1})	k_{obs} (s^{-1})		E_a (kJ/mol); temperature range 80-140 ºC	104.7	104.4
80	1.25E-07	2.23E-07				
95	9.20E-07	2.46E-06				

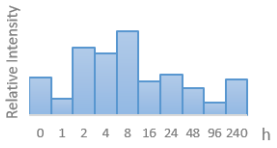
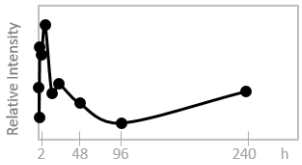
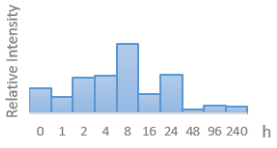
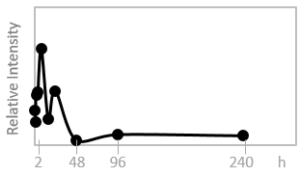
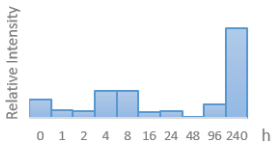
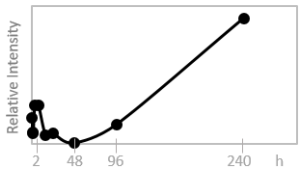
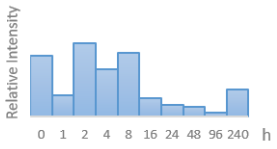
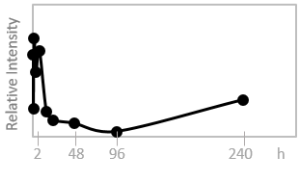
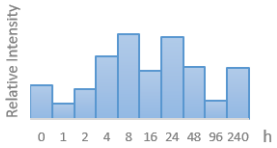
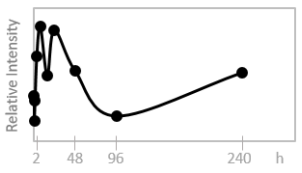
110	6.42E-06	1.37E-05				
140	3.22E-05	5.84E-05				
E_a (kJ/mol)	112.4	110.6				
R^2	0.96	0.94		R^2	0.96	0.99

3.2.2 Quantitative proteomics

3.2.2.1 Intracrystalline *Spondylus* shell matrix proteins

Proteomic analysis with TMT labelling was used to obtain qualitative and quantitative information on shell protein degradation from samples that were heated at 110 °C. We tracked the changes in protein/peptide abundance in samples that were heated for nine different time intervals (1, 2, 4, 8, 16, 24, 48, 96, 240 h) and compared to an unheated sample (the set also included an internal blank). The extracted intracrystalline shell proteins were labelled with isobaric TMT tags separately, pooled together and analysed simultaneously by LC-MS/MS. The abundance of different peptides and proteins was calculated relative to the internal blank sample (i.e. no shell proteins; any peptides/proteins identified in this sample were regarded as contaminants and were not considered further) and the obtained values were plotted against heating time (h). The obtained profiles are represented as bar plots and line graphs in tables 3 and 4. While line plots may not represent real trend lines, we find that they are useful for better visualisation and are interpreted cautiously. The tables also list all of the identified and quantified proteins, number of sequences that were used to create abundance profiles and the supporting peptides.

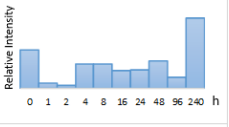
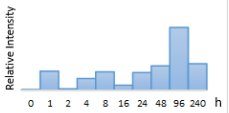
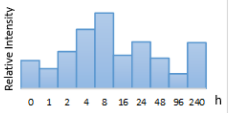
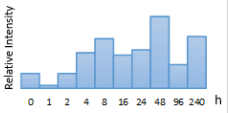
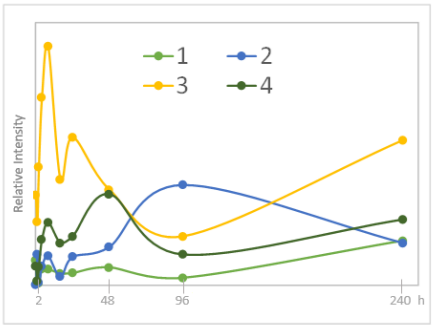
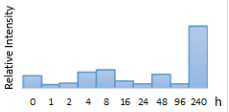
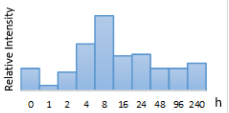
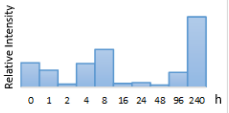
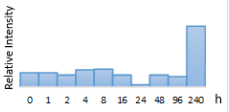
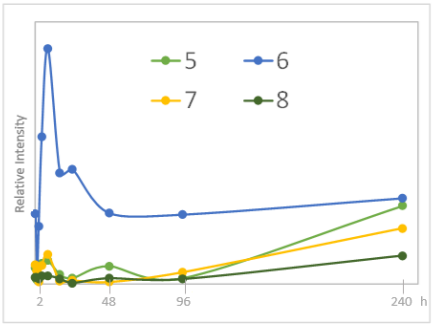
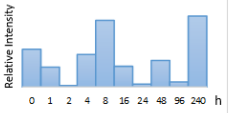
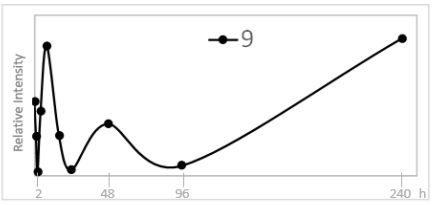
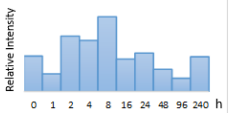
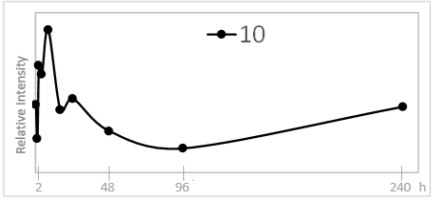
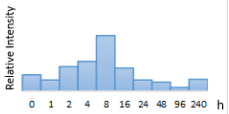
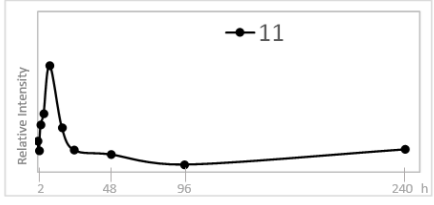
Table 3. Results from TMT proteomic analysis used to characterise and quantify intracrystalline shell proteins extracted from samples heated at 110 °C (set SPG110). Heating time points – 0 (unheated), 1, 2, 4, 8, 16, 24, 48, 96, 240 h. The 2nd and 3rd columns list the identified sequences used to create 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. The x-axis corresponds to heating time; the y-axis shows relative protein intensity at the specific time point.

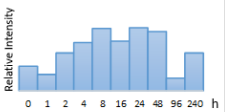
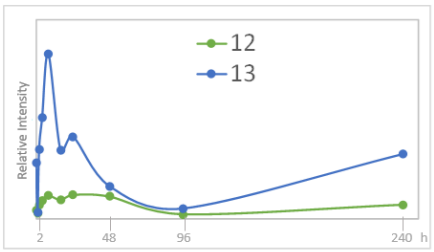
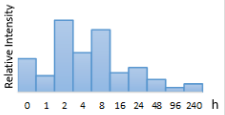
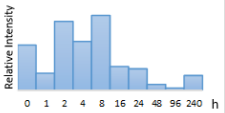
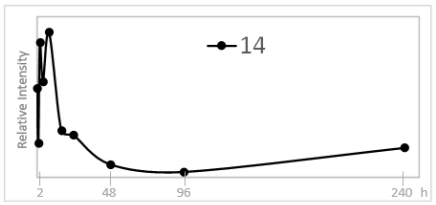
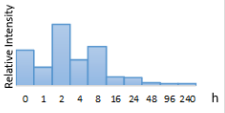
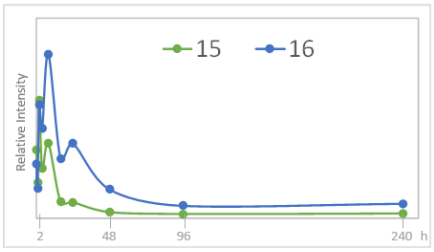
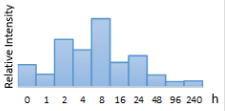
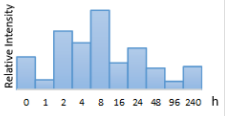
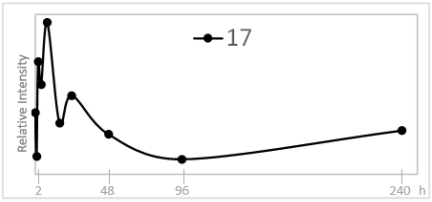
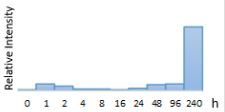
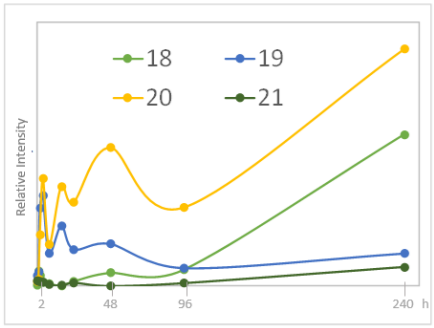
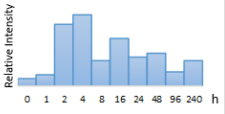
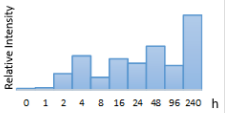
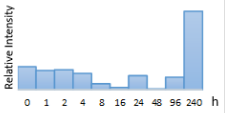
Protein	Quantified sequences	Peptides	Relative abundance profiles	
Uncharacterized protein LOC117318053 [*Pecten maximus*]	33 sequences: 25 unique peptides	Listed in Table 3		
Carbonic anhydrase 2-like [*Pecten maximus*]	2 sequences: 1 unique peptide	FGN(+.98)NRPIQR FGNNRPIQR		
Mucin-2-like isoform X1 [*Pecten maximus*]	1 sequence: 1 unique peptide	GSVNVALLNILPELR		
Aquaporin-10-like isoform X1 [*Pecten maximus*]	1 sequence: 1 unique peptide	SFVASILVFLVYF		
Laccase-2-like [*Pecten maximus*]	1 sequence: 1 unique peptide	ADGLFGALVIR		

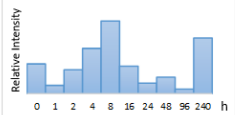
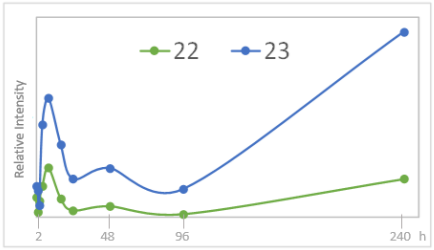
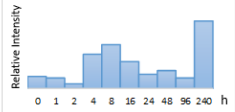
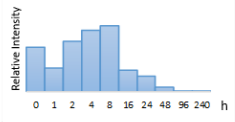
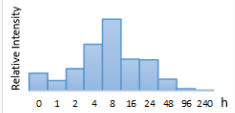
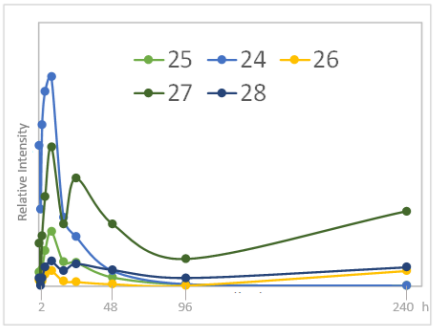
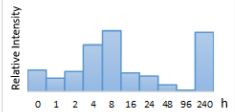
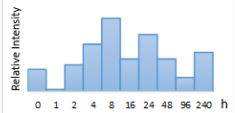
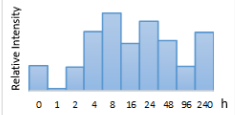
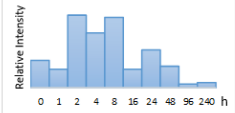
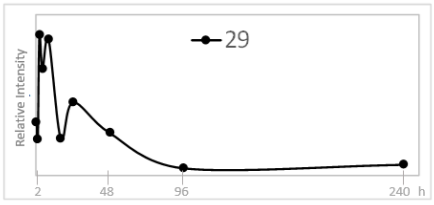
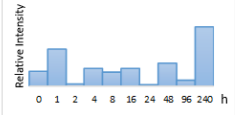
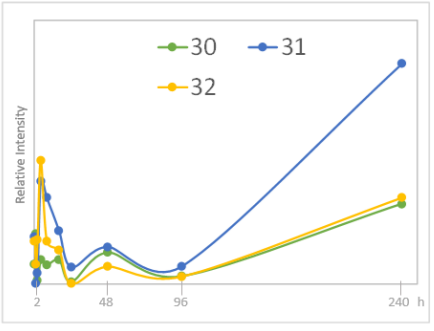
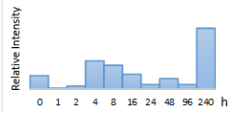
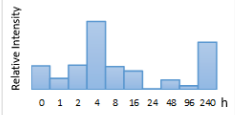
The quantification of uncharacterised protein LOC117318053 from *Pecten maximus*, which had the highest protein coverage (Table 1), was based on 33 peptide sequences, of which 25 were unique. The abundance profiles of the remaining four shell sequences were reconstructed based on only one unique peptide. Therefore, we stress that our interpretations regarding quantitative data for protein LOC117318053 (*Pecten maximus*) and for the other proteins, are valid only for these covered regions.

The data show that during the first hours of heating (at 110 °C) the relative abundance of uncharacterised protein LOC117318053 actually increases and only after several hours does the protein start to degrade. For example, the protein is most abundant after 8 h and its intensity starts to gradually decline after 16 h. After 240 h the protein is still present. Similar trends are also observed for peptide sequences of carbonic anhydrase 2-like (CA-like), laccase 2-like, aquaporin 10-like and mucin-2-like proteins. However, CA-like peptides appear to be fully degraded 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, excluding 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 interval used in this experiment and for this temperature). However, this 'secondary' effect could also be an analytical artefact due to limited data points.

Table 4. List of peptides that belong to uncharacterised protein LOC117318053 [*Pecten maximus*] and their relative abundance in shell samples that were heated at 110 °C 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 supplementary information) and their type: tryptic cleavage (T), peptides that have a deamidation modification (PTM.D; mass shift +0.98 Da) and diagenetic cleavages (DC 1-4). Detailed analysis of DC peptides is presented in Figure 6. 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 for 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		

3.2.2.2 The diagenesis of uncharacterised protein LOC117318053

A detailed investigation of protein degradation and peptide bonds hydrolysis patterns was carried out for the uncharacterised 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. the von Willebrand factor type A (vWA) and the chitin-binding domain (CBD) (Figure 5, see also supplementary information for full sequence); several peptides are also identified in the C-terminus portion. A large

part of the protein sequence is predicted to be intrinsically disordered and in *Spondylus,* these disordered regions (IDRs) are mainly uncovered.

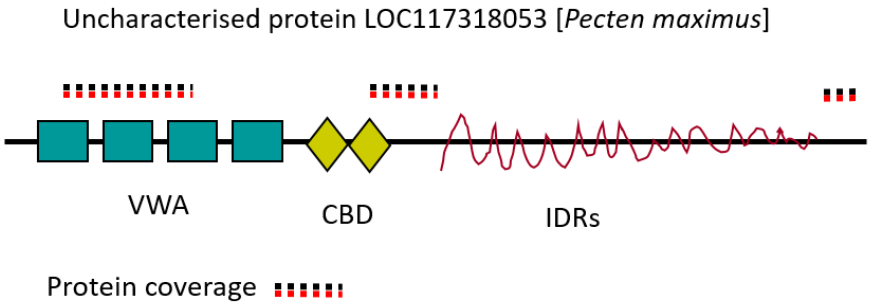


Figure 5. Schematic representation of uncharacterised protein LOC117318053 [*Pecten maximus*] and the obtained sequence coverage. Domains: VWA – von Willebrand factorA; CBD – chitin binding domain; IDR – intrinsically disordered region.

Table 4 lists all of the 32 supporting peptides that were labelled and quantified for this protein and the blue bar plots as well as line plots display their relative abundance after heating (line plots were used only for visualisation purposes). Here we highlight some of the most interesting observations and features:

a) At least 25 peptides 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 h before starting to decrease.

b) Fourteen peptides show a significant increase in relative abundance after 240 h hours of heating (the longest time point).

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 due to diagenetic cleavage (DC), DC1-4, in Figure 6 and also in Table 4). The obtained data allowed us to investigate peptide bond fragmentation (Figure 6) and is discussed in detail below.

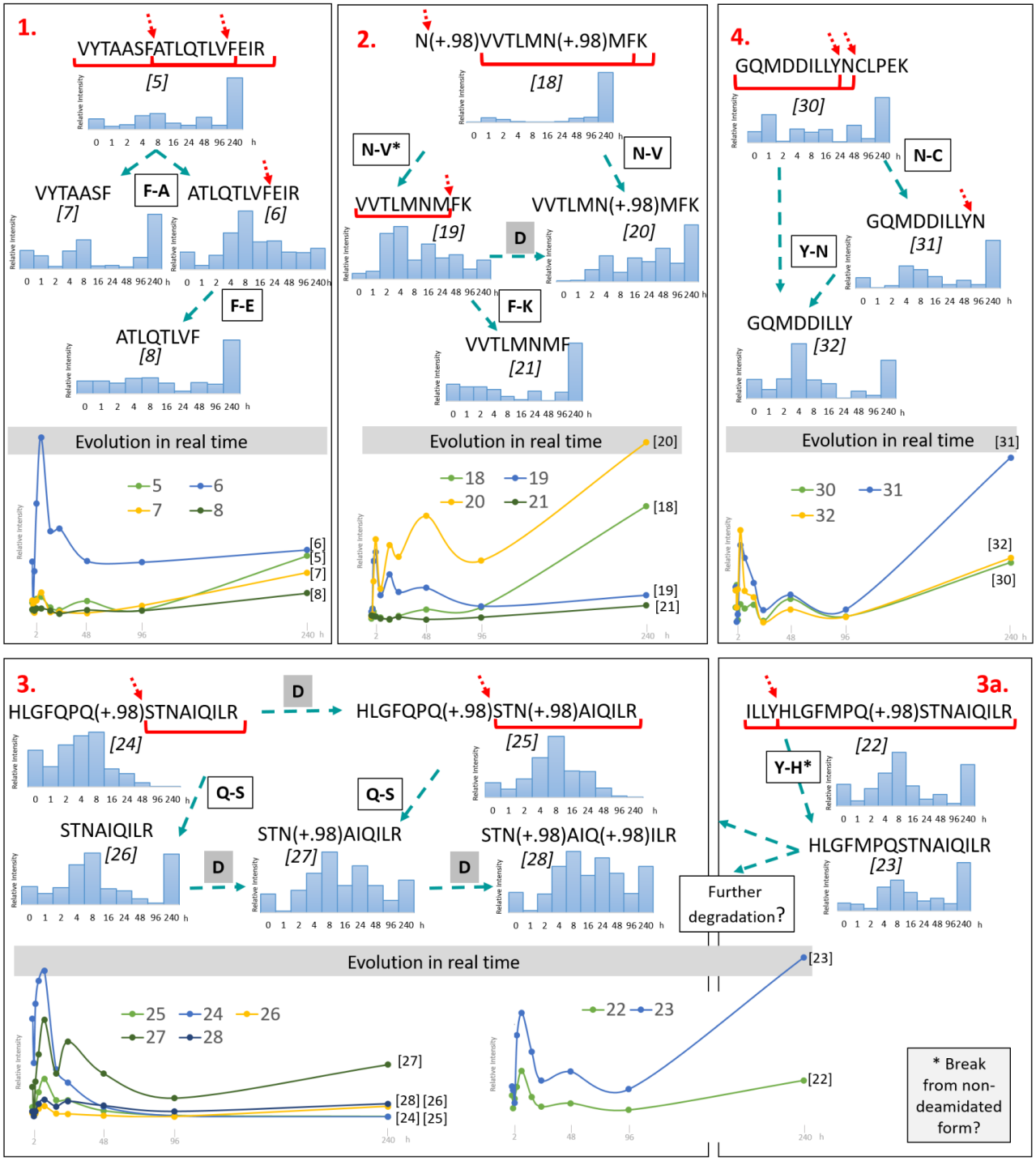


Figure 6. The scheme displays possible pathways of peptide bond fragmentation during shell protein diagenesis based on quantification data of multiple diagenetically fragmented sequences from uncharacterised protein LOC117318053 [*Pecten maximus*] (sequences DC 1-4, Table 4; peptides are numbered underneath the sequence). The quantification graphs display peptide abundance after specific heating time (0, 1, 2, 4, 8, 16, 24, 48, 96, 240 h) and are represented as bar and line plots. x-axis – sample heating time; y-axis – relative abundance, normalised against the internal blank sample.

The hydrolysis of the VYTAASFATLQTLVFEIR 'parent' peptide (peptide no. 5, DC 1) occurs between F-A residues resulting in two smaller fragments, VYTAASF and ATLQTLVFEIR (no. 6 & 7). The latter peptide fragment (ATLQTLVFEIR) accordingly shows a notable increase in abundance after the first hours of heating. Further hydrolytic damage occurs between F-E, resulting in a daughter peptide with

sequence ATLQTLVF (no. 8), which has the highest relative abundance after the longest heating time (240 h).
The hydrolysis of peptide N(+.98)VVTLMN(+.98)MFK (no. 18, DC 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 VVTLMNMFK (19) peptide implies that N-V hydrolysis also occurs from a non-deamidated parent peptide form (not observed in our samples) and suggests that the N-V bond hydrolysis occurs faster than the deamidation of all of the Asn residues, despite the fact that theoretically Asn deamidation is a very facile reaction. The fragment VVTLMNMFK (19) further loses C-terminal Lys (K) due to peptide bond breakage between F-K, and results in the shorter fragment VVTLMNMF (no. 21). After 240 h of heating (longest time interval) the relative abundance is highest for the smallest peptide fragment VVTLMNMF, the mono-deamidated fragment VVTLMN(+.98)MFK and, interestingly, the double-deamidated 'parent' sequence N(+.98)VVTLMN(+.98)MFK.
In the segment DC 3, the relative abundance profiles of peptide HLGFQPQ(+.98)STNAIQILR (no. 24) that has one deamidated Gln and HLGFQPQ(+.98)STN(+.98)AIQILR (no. 25) with an additional deamidated Asn, suggest that the second Gln residue is more easily deamidated than the other Gln and Asn residues present in this sequence. 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 [34] 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 (non-deamidated, 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, DC 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 non-deamidated, we hypothesise that the parent fragment was also non-deamidated.
The hydrolysis of GQMDDILLYNCLPEK (no. 30, DC 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 the 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 for this specific *Spondylus* shell protein:
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

4. Discussion

4.1 Structural degradation

The immunochemical analyses were used to characterise the structural degradation of the intracrystalline *Spondylus* shell matrix components by tracking the activity loss when samples were heated. First of all, ELLA and ELISA tests with jacalin and K5090 showed that the intracrystalline *Spondylus* matrix displayed cross-reactivity similar to that of its (inter+intra)crystalline fraction, indicating that bleaching does not considerably affect these cross-reactive groups and suggests that certain carbohydrate groups are also incorporated within the mineral framework [91]. In addition, we observed that K5090 showed slightly higher cross-reactivity (+12%) in the intracrystalline fraction, meaning that a harsh bleaching treatment may help to expose the reactive matrix epitopes, likely by breaking the insoluble organic matrix components (e.g. chitin or other crossed-linked structures [80,84]). Artificial diagenesis experiments showed that, as expected, structural matrix degradation is temperature dependent. At high temperatures (140 °C) degradation occurs very quickly and both carbohydrate and protein structures are lost after around 8-16 hours. In contrast, at lower temperatures (80 °C), the epitopes remain detectable for much longer, up to 150 days of heating, similar to that observed in other shell models [33,36]. We note that carbohydrate groups targeted in this study appear to be more stable, as the jacalin signal is still observed 200 days after heating. Interestingly, after the first hours of heating at temperatures below 100 °C, the matrix cross-reactivity with jacalin actually increased and was higher than that from the unheated sample, a feature that has also been observed in other studies by immunological techniques [41]. This suggests that heating first uncoils and denatures complex and cross-linked matrix structures, allowing a better exposure of target carbohydrate groups, whereas the actual degradation occurs only in later stages.

The datasets obtained by immunochemistry approaches (ELLA and ELISA) were used to calculate the apparent activation energies for structure degradation using a 'model free' approach [63,70]. In simple terms, activation energy is the minimum amount of energy that must be provided for the reaction to occur and, in this case, the values represent the energy barrier after which the shell matrix (proteins, glycoproteins and/or carbohydrates) loses its original folding, structure and reactive groups/epitopes. We find that E_a (carbohydrate groups) = 104.7 kJ/mol and E_a (protein epitopes) = 104.4 kJ/mol. Both reactions show similar activation energies, although the loss of carbohydrate groups has slightly higher E_a value, in accordance with the degradation pattern observed for different temperatures. We also attempted to apply a first-order irreversible reaction kinetics model but we observed that this model showed an acceptable fit only for the 'early diagenesis' portion of the reaction and showed a poor fit for the experiments performed at the lowest (80 °C) and highest temperatures (140 °C). We interpret that as a result of different degradation rates of multiple cross-reactive epitopes/carbohydrate groups because the antibody/lectin we used were not monoclonal.

The apparent activation energies for matrix structure degradation in *Spondylus* shell were found to be comparable with the E_a of intracrystalline protein 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 and reported in the literature [70,71]. 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; for Asx (aspartic acid/asparagine) it was found to be around 108 kJ/mol in *Patella*. In the same shell, the racemisation 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 protein degradation follows similar models

in different shell systems, structural loss of intracrystalline shell matrix (i.e. proteins, glycoproteins, carbohydrates) in *Spondylus* shell should happen prior 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, we lack knowledge of the full biomolecular composition of the intracrystalline *Spondylus* shell matrix. For instance, our previous study on *Spondylus* shell showed that its matrix had a moderate cross-reactivity with another ten lectins in addition to jacalin [80] indicating the presence of different types of carbohydrate groups, such as those with N-linked mannose or N-acetylglucosamine sugars. Secondly, we do not precisely know the targeted epitopes of antibody K5090 and we do not know which *Spondylus* protein(s) (or glycoprotein(s)) it binds to. Lastly, as ELLA/ELISA allow us to characterise only the soluble fraction, we do not know how degradation affects the insoluble matrix, which represents an abundant part of shell organics [80] and may also show different diagenesis patterns [36]. Despite these nuances, the relative rates of cross-reactivity changes provide a useful proxy to monitor structural degradation.

4.2 Protein degradation

The analysis of *Spondylus* shell samples by LC-MS/MS allowed us to identify five proteins, all belonging to the pectinoid mollusc shell *Pecten maximus.* These include: 1) protein LOC117318053 which is characterised by the presence of four von Willebrand factor A domains (vWA), a chitin binding domain (CBD), intrinsically disordered region (IDR) and is homologous to other shell matrix proteins [92]; 2) carbonic anhydrase, which is an enzyme that is involved in catalysing the conversion of carbon dioxide and water into carbonic acid, protons and bicarbonate ions and which is ubiquitous in mollusc shell matrices [93]; 3) laccase-2-like enzyme, which displays copper ion binding and oxidoreductase activity and belongs to the cupredoxin superfamily, although its role in biomineralisation is poorly understood [94,95]; 4) aquaporin-like, a water channel protein that is involved in transmembrane transport processes facilitating water diffusion, albeit its role in shell biomineralisation is also not well known [96] and 5) mucin-like, a gel-forming protein; mucins are typically characterised by high molecular weight, are heavily glycosylated (up to 70–80% of their structure) with tandem repeat units [97] and they have also been identified in other matrices from molluscan shells [98,99]. However, among these sequences, only the uncharacterised protein LOC117318053 had reasonably extensive coverage (it was supported by 25 unique peptides) and this was further used to study peptide bond hydrolysis patterns. We note that protein identification, characterisation and the subsequent analysis of degradation patterns is limited due to the lack of 'omics' data (genomics, transcriptomics) for *Spondylus* sp., specifically because our previous studies showed that its proteome is represented by lineage-specific sequences [80]. However, in the future, as more molecular data are collected for different molluscan species, including spondylid bivalves, the same proteomic dataset will provide additional information by re-analysing the data *in silico*.

The degradation processes of intracrystalline *Spondylus* shell proteins (at 110 °C) were studied by quantitative proteomics using TMT labelling, the first application to study degradation patterns in a biomineralised system. The main advantage of this approach is that the qualitative information on protein/peptide fragmentation pathways [31,34] is combined with quantitative data obtained in "real time", thus yielding novel information on some pathways of peptide bond hydrolysis and deamidation. First of all, the proteomics data displayed complex shell protein and peptide degradation patterns. The majority of the identified peptides had higher relative abundance in samples that were heated for a short duration compared to an unheated shell sample; indeed, the highest relative abundance was observed after 8 hours of heating. An analogous effect was observed by immunochemistry, in which cross-reactivity increased after short term heating. Heating likely contributes to denaturation/uncoiling of the shell matrix by partially degrading cross-linked protein-carbohydrate structures, which thus leads to

better matrix solubility [12,100,101]. Denaturation may also increase the efficiency of the proteases (i.e. trypsin) used to cleave peptide bonds [6,102]. For example, in our study we find that a mucin-2-like peptide was most abundant in the sample heated for 240 hours and mucin proteins are typically heavily glycosylated (up to 80% of their sequence) and are characterised by insoluble polymerised structures [97,98,103]. A similar effect was also observed for naturally aged shells – for example, the protein coverage or the concentration of analysed amino acids was found to be higher in subfossil samples compared to modern specimens [32,70]. For some peptides we also observed a 'secondary' increase of their relative intensity (after around 240 h) which we interpret as a result of uncoiling/denaturation/solubilisation of initially insoluble, highly cross-linked shell matrix fraction. However, due to the lack of more data points we remain cautious and note that additional investigation is needed to confirm that it is not (or possibly is) an analytical artifact. We note that TMT proteomic experiment did not include a replicate and this will be reconsidered in our 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. Finally, we find that after 240 hours of heating, many peptides remained high in abundance, indicating that the proteins did not undergo complete hydrolysis at 110 °C. However, at the same time point (240 h heating), the immunochemistry data showed that both the reactive carbohydrate groups and the protein epitopes were already lost and inactive. This indicates that peptide bond hydrolysis reaction is slower than the loss of matrix structure. This is also interesting considering the analysis of ancient proteins from archaeological or sub-fossil shells – while our data obtained from thermally-aged samples cannot be directly translated to natural diagenesis, the results clearly demonstrate that the mineral-bound protein fraction is resistant even under harsh laboratory conditions. High-resolution LC-MS/MS techniques provide us with the ability to detect ancient peptides and reconstruct protein sequences from heavily degraded samples, compared to immunological approaches.

We studied multiple peptide sequence segments that were identified in uncharacterised protein LOC117318053 [*Pecten maximus*], and which revealed a complex and slightly unexpected pattern of peptide bond hydrolysis. First of all, we find that the 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 cross-reactivity of targeted carbohydrate groups decreases to around 80-75% of its initial intensity, whereas the proteinaceous signal falls to 50-40%. In the future, additional proteomic analyses and different temperature experiments could help us to link kinetic parameters and activation energies (E_a) of hydrolysis reaction with that obtained for structure degradation, and also to compare *Spondylus* with other model systems.

We have identified eight different peptide bond sites that were prone to hydrolysis in this system. These were next to the amino acids that were hydrophobic (Ala, Val, Phe) and polar with uncharged (Gln, Asn, Ser, Tyr) as well as charged (Glu, Lys, His) side chains. Typically, in aqueous solution, the hotspots for hydrolysis are next to hydrophilic amino acids with polar (Ser, Tyr, Asn, Gln) or electrically charged side chains (Arg, His, Lys, Asp, Glu). On the other hand, peptide bonds between aliphatic (Ala, Val, Ile, Leu) and 'bulky' amino acids (Phe, Tyr) are typically harder to break, hence are considered more stable [42]. Therefore, the hydrolysis of intracrystalline proteins in the *Spondylus* shell system does not show a simple correlation to the theoretical stability of peptide bonds in solution, which is reported in the literature. In our dataset we find that three out of eight identified peptide bond breakages actually occur between dipeptides with phenylalanine, a 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. Similar discrepancies between hydrolysis models in solution and actual protein degradation in biominerals have also been observed for other shell proteins when studying their thermal stability (e.g. [33]). In addition,

our dataset also does not correlate to the theoretical protein stability (*in vivo*) based on computational calculations [90].

It is well known that in biomineral systems, structural features, matrix cross-linkage, ligand binding and protein-mineral interaction play an important role in mediating protein diagenesis. The effect of surface stabilisation was found to be the main factor enabling the preservation of a 3.8-million-year-old protein sequence in the calcite biomineral system of ostrich eggshell [31]. Computational modelling suggested that protein-mineral binding involves acidic amino acids (Asp and Glu) that stabilise the adjacent water molecules and results in a higher energy barrier for hydrolysis, compared to the energy that would be required in solution. 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). While this may be a bias due to lack of proteomics data for *Spondylus* shell, we observe 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 [104,105]. 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 [106–108]. In addition, peptides with phosphorylated Ser residues have been identified in 1.77-million-year-old enamel proteomes, indicating that phosphoproteins are stable in time [30]. Therefore, in the *Spondylus* shell matrix, it is the overall microchemical and structural environment that mediates protein diagenesis.

The peptide quantification data also show a complex pattern of deamidation, which increases with heating time. The deamidation of Asn residues is known to be more than 60 times faster than that of Gln [49], but in *Spondylus*, some exceptions are also observed. For example, in peptide sequence HLGFQPQ(+.98)STNAIQILR, Asn undergoes deamidation after Gln. This could be due to the 3D structure of the protein that slows down Asn deamidation [55], or that the Gln deamidation is accelerated by the adjacent Ser which can undergo in-chain racemisation [34] or the combination of both effects. We also observe that 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. Neither of these exhibit any modifications, showing that, in that case, direct peptide bond hydrolysis is faster than deamidation. 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 [109] and the loss of amide side chain may contribute to making the structure more rigid, whereas the change in net charge may also impact structure stability [110,111].

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 at investigating diagenesis experimentally at the proteomics level and, since more data will be collected, shows excellent potential for uncovering the mechanisms that underlie the stability (or lack thereof) of biomineralised proteomes. Moreover, such data should help improve estimates of racemisation kinetics.

5. Conclusions

This study presents the first diagenesis investigation of intracrystalline shell proteins conducted by simultaneously tracking temperature-induced changes that affect both the structural properties of proteins and their sequences. Overall, we find that the intracrystalline *Spondylus* shell matrix, which comprises proteins and carbohydrates, is robust and persists for prolonged periods at elevated temperatures. TMT proteomics shows that intracrystalline proteins do not undergo complete hydrolysis even after 10 days of heating at 110 °C. At the same time, immunochemical proxies suggest that the matrix structure is largely lost, because the signal of cross-reactive carbohydrate groups and protein epitopes is no longer detected. We also observe that heating does not immediately destroy the total immunogenicity of the matrix but, as we hypothesise, mild or short temperature exposures first allow the uncoiling (denaturing) of cross-linked structures, facilitating the solubilisation and thus the detection of proteins and carbohydrate groups. A more in-depth sequence investigation was carried out for uncharacterised protein LOC117318053 [*Pecten maximus*], which enabled us to identify eight peptide bonds that were especially prone to hydrolysis. The observed peptide bond hydrolysis patterns do not fit with the theoretical stability of individual bonds and further expose the intrinsic peculiarities of protein aging in biomineralised tissues. Our data support the hypothesis that structural organisation between proteins, carbohydrates and the mineral surfaces, as well as the overall microchemical environment, all play key roles in stabilising the protein sequences and regulating degradation pathways. We present an analytical advancement which will be suited to track changes and pathways of degradation of individual shell proteins, peptides and their fragments in real time. Finally, this work suggests that *Spondylus* may serve as a good model system for protein preservation in sub-fossil substrates and therefore that finding ancient protein sequences in archaeological *Spondylus* artifacts is highly plausible.

Data deposition

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [112] partner repository with the dataset identifier PXD025068 and 10.6019/PXD025068.

Acknowledgements

JS, FM and BD would like to thank Jane Thomas-Oates (University of York) and Daniel Jackson (University of Göttingen) for valuable insights and discussions. JS, BD and FM also acknowledge the support from DBIOS (University of Turin) and Biogéosciences communities (University of Burgundy-Franche-Comté).
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 from program Eiffel.
BD was funded by the “Giovani Ricercatori - Rita Levi Montalcini” Programme (MIUR; Ministero dell'Istruzione dell'Università e della Ricerca).
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).

Author contribution

Jorune Sakalauskaite: Conceptualization, Methodology, Investigation, Data curation, Writing - Original Draft. **Beatrice Demarchi**: Conceptualization, Methodology, Investigation, Data curation, Resources, Writing - original draft, Project administration, Funding acquisition. **Frédéric Marin**: Methodology,

Investigation, Data curation, Resources, Writing - original draft, Project administration, Funding acquisition. **Matthew Collins**: Investigation, Resources, Writing - Reviewing and Editing, Funding acquisition. **Meaghan Mackie**: Methodology, Investigation, Writing - Reviewing and Editing. **Alberto Taurozzi**: Methodology, Investigation, Writing - Reviewing and Editing

Competing interests statement

The authors declare that no competing interests exist.

References

[1] P.H. Abelson, Amino acids in fossils, Science. 119 (1954) 576–576.
[2] H.A. Lowenstam, S. Weiner, On Biomineralization, Oxford University Press, 1989.
[3] F. Marin, G. Luquet, Molluscan shell proteins, C. R. Palevol. 3 (2004) 469–492.
[4] S. Albeck, J. Aizenberg, L. Addadi, S. Weiner, Interactions of various skeletal intracrystalline components with calcite crystals, J. Am. Chem. Soc. 115 (1993) 11691–11697.
[5] S. Albeck, S. Weiner, L. Addadi, Polysaccharides of intracrystalline glycoproteins modulate calcite crystal growth in vitro, Chem. Eur. J. 2 (1996) 278–284.
[6] 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.
[7] Y.-Y. Kim, J.D. Carloni, B. Demarchi, D. Sparks, D.G. Reid, M.E. Kunitake, C.C. Tang, M.J. Duer, C.L. Freeman, B. Pokroy, K. Penkman, J.H. Harding, L.A. Estroff, S.P. Baker, F.C. Meldrum, Tuning hardness in calcite by incorporation of amino acids, Nat. Mater. 15 (2016) 903–910.
[8] F. Marin, Mollusc shellomes: Past, present and future, J. Struct. Biol. 212 (2020) 107583.
[9] M. Akiyama, R.W. Wyckoff, The total amino acid content of fossil pecten shells, Proc. Natl. Acad. Sci. U. S. A. 67 (1970) 1097–1100.
[10] S. Weiner, H.A. Lowenstam, L. Hood, Characterization of 80-million-year-old mollusk shell proteins, Proc. Natl. Acad. Sci. U. S. A. 73 (1976) 2541–2545.
[11] 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.
[12] M.J. Collins, D. Walton, G.B. Curry, M.S. Riley, T.N. Von Wallmenich, N.M. Savage, G. Muyzer, P. Westbroek, Long-term trends in the survival of immunological epitopes entombed in fossil brachiopod skeletons, Org. Geochem. 34 (2003) 89–96.
[13] 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.
[14] B. Demarchi, Amino acids and proteins in fossil biominerals: an Introduction for archaeologists and palaeontologists, John Wiley & Sons, 2020.
[15] E.P. Hare, Racemization of amino acids in fossil shells, Carnegie Institute of Washington Yearbook. 66 (1968) 526–528.
[16] R.A. Schroeder, J.L. Bada, A review of the geochemical applications of the amino acid racemization reaction, Earth-Sci. Rev. 12 (1976) 347–391.
[17] K. Endo, D. Walton, R.A. Reyment, G.B. Curry, Fossil intra-crystalline biomolecules of brachiopod shells: diagenesis and preserved geo-biological information, Org. Geochem. 23 (1995) 661–673.
[18] M.J. Collins, M.S. Riley, Amino acid racemization in biominerals: the impact of protein degradation and loss, in: G. Goodfriend, M. Collins, M. Fogel, S. Macko, J. Wehmiller (Eds.), Perspectives in Amino Acid and Protein Geochemistry, Oxford University Press, 2000: pp. 120–142.
[19] 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.
[20] B. Demarchi, M. Collins, Amino acid racemization dating, in: W.J. Rink, J. Thompson (Eds.), Encyclopedia of Scientific Dating Methods, Springer Netherlands, Dordrecht, 2014: pp. 1–22.
[21] P.H. Ostrom, M. Schall, H. Gandhi, T.-L. Shen, P.V. Hauschka, J.R. Strahler, D.A. Gage, New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry, Geochim. Cosmochim. Acta. 64 (2000) 1043–1050.
[22] M. Collins, E. Cappellini, M. Buckley, K. Penkman, R. Griffin, H. Koon, Analytical methods to detect ancient proteins, Bio-and Material Cultures at Qumran. (2006) 33.
[23] M. Buckley, M. Collins, J. Thomas-Oates, J.C. Wilson, Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry, Rapid Commun. Mass Spectrom. 23 (2009) 3843–3854.
[24] E. Cappellini, L.J. Jensen, D. Szklarczyk, A. Ginolhac, R.A.R. da Fonseca, T.W. Stafford, S.R. Holen, M.J. Collins, L. Orlando, E. Willerslev, M.T.P. Gilbert, J.V. Olsen, Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins, J. Proteome

Res. 11 (2012) 917–926.
[25] F. Welker, Palaeoproteomics for human evolution studies, Quat. Sci. Rev. 190 (2018) 137–147.
[26] T.P. Cleland, E.R. Schroeter, A comparison of common mass spectrometry approaches for paleoproteomics, J. Proteome Res. 17 (2018) 936–945.
[27] J. Hendy, Ancient protein analysis in archaeology, Sci Adv. 7 (2021). https://doi.org/10.1126/sciadv.abb9314.
[28] F. Welker, J. Ramos-Madrigal, M. Kuhlwilm, W. Liao, P. Gutenbrunner, M. de Manuel, D. Samodova, M. Mackie, M.E. Allentoft, A.-M. Bacon, M.J. Collins, J. Cox, C. Lalueza-Fox, J.V. Olsen, F. Demeter, W. Wang, T. Marques-Bonet, E. Cappellini, Enamel proteome shows that *Gigantopithecus* was an early diverging pongine, Nature. 576 (2019) 262–265.
[29] E. Cappellini, F. Welker, L. Pandolfi, J. Ramos-Madrigal, D. Samodova, P.L. Rüther, A.K. Fotakis, D. Lyon, J. Víctor Moreno-Mayar, M. Bukhsianidze, R.R. Jersie-Christensen, M. Mackie, A. Ginolhac, R. Ferring, M. Tappen, E. Palkopoulou, M.R. Dickinson, T.W. Stafford, Y.L. Chan, A. Götherström, Senthilvel K S, P.D. Heintzman, J.D. Kapp, I. Kirillova, Y. Moodley, J. Agusti, R.-D. Kahlke, G. Kiladze, B. Martínez-Navarro, S. Liu, M.S. Velasco, M.-H.S. Sinding, C.D. Kelstrup, M.E. Allentoft, L. Orlando, K. Penkman, B. Shapiro, L. Rook, L. Dalén, M.T.P. Gilbert, J.V. Olsen, D. Lordkipanidze, E. Willerslev, Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny, Nature. (2019) 1–5.
[30] F. Welker, J. Ramos-Madrigal, P. Gutenbrunner, M. Mackie, S. Tiwary, R. Rakownikow Jersie-Christensen, C. Chiva, M.R. Dickinson, M. Kuhlwilm, M. de Manuel, P. Gelabert, M. Martinón-Torres, A. Margvelashvili, J.L. Arsuaga, E. Carbonell, T. Marques-Bonet, K. Penkman, E. Sabidó, J. Cox, J.V. Olsen, D. Lordkipanidze, F. Racimo, C. Lalueza-Fox, J.M. Bermúdez de Castro, E. Willerslev, E. Cappellini, The dental proteome of Homo antecessor, Nature. 580 (2020) 235–238.
[31] 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.
[32] 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.
[33] F. Marin, V. Morin, F. Knap, N. Guichard, Caspartin: Thermal stability and occurrence in mollusk calcified tissues, in: J.L.A. Arias, M.S. Fernández (Eds.), Biomineralization: From Paleontology to Materials Science, researchgate.net, 2007: pp. 281–288.
[34] B. Demarchi, M. Collins, E. Bergström, A. Dowle, K. Penkman, J. Thomas-Oates, J. Wilson, New experimental evidence for in-chain amino acid racemization of serine in a model peptide, Anal. Chem. 85 (2013) 5835–5842.
[35] M. Buckley, C. Wadsworth, Proteome degradation in ancient bone: Diagenesis and phylogenetic potential, Palaeogeogr. Palaeoclimatol. Palaeoecol. 416 (2014) 69–79.
[36] A. Parker, F. Immel, N. Guichard, C. Broussard, F. Marin, Thermal stability of nacre proteins of the Polynesian pearl oyster: a proteomic study, in: F. Marin, F. Brümmer, A. Checa, G. Furtos, I.G. Lesci, L. Šiller (Eds.), Key Engineering Materials, Trans Tech Publ, 2015: pp. 222–231.
[37] T.P. Cleland, E.R. Schroeter, C. Colleary, Diagenetiforms: A new term to explain protein changes as a result of diagenesis in paleoproteomics, J. Proteomics. 230 (2021) 103992.
[38] P.E. Hare, T.C. Hoering, The organic constituents of fossil mollusc shells, Carnegie Institute of Washington Yearbook. 78 (1977) 625–631.
[39] R.J. Dodd, Processes of conversion of aragonite to calcite with examples from the Cretaceous of Texas, J. Sediment. Res. 36 (1966) 733–741.
[40] S. Milano, G. Nehrke, Microstructures in relation to temperature-induced aragonite-to-calcite transformation in the marine gastropod *Phorcus turbinatus*, PLoS One. 13 (2018) e0204577.
[41] M.J. Collins, P. Westbroek, G. Muyzer, J.W. de Leeuw, Experimental evidence for condensation reactions between sugars and proteins in carbonate skeletons, Geochim. Cosmochim. Acta. 56 (1992) 1539–1544.

[42] R.L. Hill, Hydrolysis of proteins, in: C.B. Anfinsen, M.L. Anson, J.T. Edsall, F.M. Richards (Eds.), Advances in Protein Chemistry, Academic Press, 1965: pp. 37–107.
[43] Y. Qian, M.H. Engel, S.A. Macko, S. Carpenter, J.W. Deming, Kinetics of peptide hydrolysis and amino acid decomposition at high temperature, Geochim. Cosmochim. Acta. 57 (1993) 3281–3293.
[44] J.L. Bada, M.-Y. Shou, E.H. Man, R.A. Schroeder, Decomposition of hydroxy amino acids in foraminiferal tests; kinetics, mechanism and geochronological implications, Earth Planet. Sci. Lett. 41 (1978) 67–76.
[45] J.L. Bada, E.H. Man, 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-Sci. Rev. 16 (1980) 21–55.
[46] B. Demarchi, M.G. Williams, N. Milner, N. Russell, G. Bailey, K. Penkman, Amino acid racemization dating of marine shells: A mound of possibilities, Quat. Int. 239 (2011) 114–124.
[47] T.P. Cleland, E.R. Schroeter, M.H. Schweitzer, Biologically and diagenetically derived peptide modifications in moa collagens, Proc. Biol. Sci. 282 (2015) 20150015.
[48] T.P. Cleland, E.R. Schroeter, R.S. Feranec, D. Vashishth, Peptide sequences from the first Castoroides ohioensis skull and the utility of old museum collections for palaeoproteomics, Proc. Biol. Sci. 283 (2016). https://doi.org/10.1098/rspb.2016.0593.
[49] N.E. Robinson, A. Robinson, Molecular clocks: deamidation of asparaginyl and glutaminyl residues in peptides and proteins, Althouse press, 2004.
[50] M. Mackie, P. Rüther, D. Samodova, F. Di Gianvincenzo, C. Granzotto, D. Lyon, D.A. Peggie, H. Howard, L. Harrison, L.J. Jensen, J.V. Olsen, E. Cappellini, Palaeoproteomic profiling of conservation layers on a 14th century Italian wall painting, Angew. Chem. Int. Ed Engl. 57 (2018) 7369–7374.
[51] L.C. Maillard, Action of amino acids on sugars. Formation of melanoidins in a methodical way, Compte-Rendu de l'Academie Des Sciences. 154 (1912) 66–68.
[52] T.C. Hoering, The organic constituents of fossil mollusk shells, in: Biogeochem. Amino Acids Conference Papers, bcin.ca, 1978: pp. 193–201.
[53] R.M. Mitterer, The diagenesis of proteins and amino acids in fossil shells, in: M.H. Engel, S.A. Macko (Eds.), Organic Geochemistry: Principles and Applications, Springer US, Boston, MA, 1993: pp. 739–753.
[54] C.M. Nielsen-Marsh, M.P. Richards, P.V. Hauschka, J.E. Thomas-Oates, E. Trinkaus, P.B. Pettitt, I. Karavanic, H. Poinar, M.J. Collins, Osteocalcin protein sequences of Neanderthals and modern primates, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 4409–4413.
[55] N.E. Robinson, A.B. Robinson, Prediction of protein deamidation rates from primary and three-dimensional structure, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 4367–4372.
[56] J. Wilson, N.L. van Doorn, M.J. Collins, Assessing the extent of bone degradation using glutamine deamidation in collagen, Anal. Chem. 84 (2012) 9041–9048.
[57] E.R. Schroeter, T.P. Cleland, Glutamine deamidation: an indicator of antiquity, or preservational quality?, Rapid Commun. Mass Spectrom. 30 (2016) 251–255.
[58] N. Procopio, M. Buckley, Minimizing laboratory-induced decay in bone proteomics, J. Proteome Res. 16 (2017) 447–458.
[59] A. Ramsøe, V. van Heekeren, P. Ponce, R. Fischer, I. Barnes, C. Speller, M.J. Collins, DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes, J. Archaeol. Sci. 115 (2020) 105080.
[60] J.P. Simpson, K.E.H. Penkman, B. Demarchi, H. Koon, M.J. Collins, J. Thomas-Oates, B. Shapiro, M. Stark, J. Wilson, The effects of demineralisation and sampling point variability on the measurement of glutamine deamidation in type I collagen extracted from bone, J. Archaeol. Sci. 69 (2016) 29–38.
[61] A. Ramsøe, M. Crispin, M. Mackie, K. McGrath, R. Fischer, B. Demarchi, M.J. Collins, J. Hendy, C. Speller, Assessing the degradation of ancient milk proteins through site-specific deamidation patterns, Sci. Rep. 11 (2021) 7795.
[62] P.E. Hare, Effect of hydrolysis on the racemization rate of amino acids, Carnegie Institute of Washington Yearbook. 70 (1971) 256–258.
[63] P.J. Tomiak, K.E.H. Penkman, E.J. Hendy, B. Demarchi, S. Murrells, S.A. Davis, P. McCullagh, M.J. Collins, Testing the limitations of artificial protein degradation kinetics using known-age massive Porites coral skeletons, Quat. Geochronol. 16 (2013) 87–109.

[64] R.P. Evershed, H.A. Bland, P.F. van Bergen, J.F. Carter, M.C. Horton, P.A. Rowley-Conwy, Volatile compounds in archaeological plant remains and the Maillard reaction during decay of organic Matter, Science. 278 (1997) 432–433.
[65] G. Vistoli, D. De Maddis, A. Cipak, N. Zarkovic, M. Carini, G. Aldini, Advanced glycoxidation and lipoxidation end products (AGEs and ALEs): an overview of their mechanisms of formation, Free Radic. Res. 47 Suppl 1 (2013) 3–27.
[66] F. Ifantidis, M. Nikolaidou, eds., *Spondylus* in Prehistory: new data and approaches: contributions to the archaeology of shell technologies, Archaeopress, Oxford, 2011.
[67] M.A. Borrello, R. Micheli, *Spondylus gaederopus* in Prehistoric Italy: Jewels from Neolithic and Copper age sites, in: F. Ifantidis, M. Nikolaidou (Eds.), Spondylus in Prehistory. New Data and Approaches. Contributions to the Archaeology of Shell Technologies, Archaeopress, 2011: pp. 25–37.
[68] 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.
[69] A. Windler, From the Aegean Sea to the Parisian Basin: *Spondylus* shell exchange in Europe during the process of Neolithisation, in: P. Eisenach, T. Stöllner, A. Windler (Eds.), The RITaK Conferences 2013–2014, Bochum, 2017: pp. 95–110.
[70] B. Demarchi, M.J. Collins, P.J. Tomiak, B.J. Davies, K.E.H. Penkman, Intra-crystalline protein diagenesis (IcPD) in *Patella vulgata*. Part II: Breakdown and temperature sensitivity, Quat. Geochronol. 16 (2013) 158–172.
[71] M. Crisp, B. Demarchi, M. Collins, M. Morgan-Williams, E. Pilgrim, K. Penkman, Isolation of the intra-crystalline proteins and kinetic studies in *Struthio camelus* (ostrich) eggshell for amino acid geochronology, Quat. Geochronol. 16 (2013) 110–128.
[72] J.E. Ortiz, T. Torres, Y. Sánchez-Palencia, M. Ferrer, Inter- and intra-crystalline protein diagenesis in Glycymeris shells: Implications for amino acid geochronology, Quat. Geochronol. 41 (2017) 37–50.
[73] 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.
[74] M.R. Dickinson, A.M. Lister, K.E.H. Penkman, A new method for enamel amino acid racemization dating: A closed system approach, Quat. Geochronol. 50 (2019) 29–46.
[75] B. Demarchi, K. Rogers, D.A. Fa, C.J. Finlayson, N. Milner, K.E.H. Penkman, Intra-crystalline protein diagenesis (IcPD) in *Patella vulgata*. Part I: Isolation and testing of the closed system, Quat. Geochronol. 16 (2013) 144–157.
[76] 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.
[77] H. Erdjument-Bromage, F.-K. Huang, T.A. Neubert, Sample preparation for relative quantitation of proteins using tandem mass tags (TMT) and mass spectrometry (MS), Methods Mol. Biol. 1741 (2018) 135–149.
[78] K.A. Welle, T. Zhang, J.R. Hryhorenko, S. Shen, J. Qu, S. Ghaemmaghami, Time-resolved analysis of proteome dynamics by tandem mass tags and stable isotope labeling in cell culture (TMT-SILAC) hyperplexing, Mol. Cell. Proteomics. 15 (2016) 3551–3563.
[79] R.R. Jersie-Christensen, L.T. Lanigan, D. Lyon, M. Mackie, D. Belstrøm, C.D. Kelstrup, A.K. Fotakis, E. Willerslev, N. Lynnerup, L.J. Jensen, E. Cappellini, J.V. Olsen, Quantitative metaproteomics of medieval dental calculus reveals individual oral health status, Nat. Commun. 9 (2018) 4744.
[80] 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 (2020) 107497.
[81] F. Marin, G. Muyzer, Y. Dauphin, Caractérisations électrophorétique et immunologique des matrices organiques solubles des tests de deux Bivalves Ptériomorphes actuels, *Pinna nobilis* L. et *Pinctada margaritifera* (L.), C. R. Acad. Sci. 318 (1994) 1653–1659.
[82] T. Takeuchi, L. Plasseraud, I. Ziegler-Devin, N. Brosse, C. Shinzato, N. Satoh, F. Marin, Biochemical characterization of the skeletal matrix of the massive coral, *Porites australiensis* - The saccharide moieties and their localization, J. Struct. Biol. 203 (2018) 219–229.
[83] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for

proteome analysis, Nat. Methods. 6 (2009) 359–362.
[84] J. Sakalauskaite, F. Marin, B. Pergolizzi, B. Demarchi, Shell palaeoproteomics: First application of peptide mass fingerprinting for the rapid identification of mollusc shells in archaeology, J. Proteomics. 227 (2020) 103920.
[85] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, Nat. Protoc. 2 (2007) 1896–1906.
[86] B. Ma, K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby, G. Lajoie, PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry, Rapid Commun. Mass Spectrom. 17 (2003) 2337–2342.
[87] N.H. Tran, X. Zhang, L. Xin, B. Shan, M. Li, De novo peptide sequencing by deep learning, Proc. Natl. Acad. Sci. U. S. A. (2017). https://doi.org/10.1073/pnas.1705691114.
[88] M. Blum, H.-Y. Chang, S. Chuguransky, T. Grego, S. Kandasaamy, A. Mitchell, G. Nuka, T. Paysan-Lafosse, M. Qureshi, S. Raj, L. Richardson, G.A. Salazar, L. Williams, P. Bork, A. Bridge, J. Gough, D.H. Haft, I. Letunic, A. Marchler-Bauer, H. Mi, D.A. Natale, M. Necci, C.A. Orengo, A.P. Pandurangan, C. Rivoire, C.J.A. Sigrist, I. Sillitoe, N. Thanki, P.D. Thomas, S.C.E. Tosatto, C.H. Wu, A. Bateman, R.D. Finn, The InterPro protein families and domains database: 20 years on, Nucleic Acids Res. 49 (2021) D344–D354.
[89] G. Erdős, Z. Dosztányi, Analyzing Protein Disorder with IUPred2A, Curr. Protoc. Bioinformatics. 70 (2020) e99.
[90] K. Guruprasad, B.V. Reddy, M.W. Pandit, 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 Eng. 4 (1990) 155–161.
[91] A. Lang, S. Mijowska, I. Polishchuk, S. Fermani, G. Falini, A. Katsman, F. Marin, B. Pokroy, Acidic monosaccharides become incorporated into calcite single crystals, Chem. Eur. J. (2020) 2020.08.03.234310.
[92] B. Marie, J. Arivalagan, L. Mathéron, G. Bolbach, S. Berland, A. Marie, F. Marin, Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the *Unionoida Elliptio complanata* and *Villosa lienosa*, J. R. Soc. Interface. 14 (2017). https://doi.org/10.1098/rsif.2016.0846.
[93] N. Le Roy, D.J. Jackson, B. Marie, P. Ramos-Silva, F. Marin, The evolution of metazoan α-carbonic anhydrases and their roles in calcium carbonate biomineralization, Front. Zool. 11 (2014) 75.
[94] A. Luna-Acosta, E. Rosenfeld, M. Amari, I. Fruitier-Arnaudin, P. Bustamante, H. Thomas-Guyon, First evidence of laccase activity in the Pacific oyster Crassostrea gigas, Fish Shellfish Immunol. 28 (2010) 719–726.
[95] X. Yue, S. Zhang, J. Yu, B. Liu, Identification of a laccase gene involved in shell periostracal tanning of the clam *Meretrix petechialis*, Aquat. Biol. 28 (2019) 55–65.
[96] R.N. Finn, J. Cerdà, Evolution and functional diversity of aquaporins, Biol. Bull. 229 (2015) 6–23.
[97] M.A. McGuckin, D.J. Thornton, J.A. Whitsett, Chapter 14 - Mucins and Mucus, in: J. Mestecky, W. Strober, M.W. Russell, B.L. Kelsall, H. Cheroutre, B.N. Lambrecht (Eds.), Mucosal Immunology (Fourth Edition), Academic Press, Boston, 2015: pp. 231–250.
[98] F. Marin, P. Corstjens, B. de Gaulejac, E. de Vrind-De Jong, P. Westbroek, 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), J. Biol. Chem. 275 (2000) 20667–20675.
[99] V.A. Sleight, B. Marie, D.J. Jackson, E.A. Dyrynda, A. Marie, M.S. Clark, An Antarctic molluscan biomineralisation tool-kit, Sci. Rep. 6 (2016) 36978.
[100] C. Ma, C. Zhang, Y. Nie, L. Xie, R. Zhang, Extraction and purification of matrix protein from the nacre of pearl oyster *Pinctada fucata*, Tsinghua Science & Technology. 10 (2005) 499–503.
[101] H. Kintsu, R. Nishimura, L. Negishi, I. Kuriyama, Y. Tsuchihashi, L. Zhu, K. Nagata, M. Suzuki, Identification of methionine -rich insoluble proteins in the shell of the pearl oyster, *Pinctada fucata*, Sci. Rep. 10 (2020) 18335.
[102] 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.
[103] 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.), International Symposium on Pearl Research, Terrapub Tokyo, 2013: pp. 149–166.
[104] C. Li, G.D. Botsaris, D.L. Kaplan, Selective in vitro effect of peptides on calcium carbonate crystallization, Cryst. Growth Des. 2 (2002) 387–393.
[105] D. Gebauer, A. Verch, H.G. Börner, H. Cölfen, Influence of selected artificial peptides on calcium carbonate precipitation - a quantitative study, Cryst. Growth Des. 9 (2009) 2398–2403.
[106] J.E. Borbas, A.P. Wheeler, C.S. Sikes, Molluscan shell matrix phosphoproteins: Correlation of degree of phosphorylation to shell mineral microstructure and to in vitro regulation of mineralization, J. Exp. Zool. 258 (1991) 1–13.
[107] T. Samata, D. Ikeda, A. Kajikawa, H. Sato, C. Nogawa, D. Yamada, R. Yamazaki, T. Akiyama, A novel phosphorylated glycoprotein in the shell matrix of the oyster *Crassostrea nippona*: A novel acidic glycoprotein from the oyster shells, FEBS J. 275 (2008) 2977–2989.
[108] J. Du, G. Xu, C. Liu, R. Zhang, The role of phosphorylation and dephosphorylation of shell matrix proteins in shell formation: an in vivo and in vitro study, CrystEngComm. 20 (2018) 3905–3916.
[109] A.A. Kossiakoff, Tertiary structure is a principal determinant to protein deamidation, Science. 240 (1988) 191–194.
[110] T. Silva, A. Kirkpatrick, B. Brodsky, J.A.M. Ramshaw, Effect of deamidation on stability for the collagen to gelatin transition, J. Agric. Food Chem. 53 (2005) 7802–7806.
[111] K.L. Zapadka, F.J. Becher, A.L. Gomes Dos Santos, S.E. Jackson, Factors affecting the physical stability (aggregation) of peptide therapeutics, Interface Focus. 7 (2017) 20170030.
[112] J.A. Vizcaíno, R.G. Côté, A. Csordas, J.A. Dianes, A. Fabregat, J.M. Foster, J. Griss, E. Alpi, M. Birim, J. Contell, G. O'Kelly, A. Schoenegger, D. Ovelleiro, Y. Pérez-Riverol, F. Reisinger, D. Ríos, R. Wang, H. Hermjakob, The Proteomics Identifications (PRIDE) database and associated tools: status in 2013, Nucleic Acids Res. 41 (2013) D1063–D1069.