Marine Biotechnology (2026) 28:87
https://doi.org/10.1007/s10126-026-10631-4

RESEARCH

Quantitative Proteomic Profiling of *Pinctada fucata* Shell Nacre Defines a Solubility-Based Type Classification of Shell Matrix Proteins

Keisuke Oshima[1] · Lumi Negishi[2] · Takashi Atsumi[3] · Hitoshi Kurumizaka[2] · Michio Suzuki[1]

Received: 23 March 2026 / Accepted: 5 May 2026

Abstract

Shell matrix proteins (SMPs) are key organic components of molluscan biominerals, yet previous nacre proteomics have remained largely qualitative, limiting evaluation of the abundance and fraction association of individual SMPs. Here, we established a quantitative proteomic approach for the nacreous layer of the pearl oyster *Pinctada fucata* by integrating optimized shell preservation, stepwise fractionation, and data-independent acquisition (DIA) proteomics. SMPs were separated into an ethylenediaminetetraacetic acid (EDTA)-soluble matrix (ESM), an EDTA-insoluble but sodium dodecyl sulfate/dithiothreitol (SDS/DTT)-soluble matrix (SSM), and an SDS/DTT-insoluble matrix (ISM). DIA outperformed data-dependent acquisition in proteome coverage and enabled quantification of 327 SMPs across a broad dynamic range. Fraction-resolved abundance profiling showed that each fraction was characterized by distinct SMP compositions. To summarize these distributions, we introduced a solubility-based type classification that grouped SMPs into four types according to their quantitative partitioning among fractions. Well-known SMPs, including nacrein, Pif 80, and MSI60, were assigned to intuitively consistent types, whereas proteases, protease inhibitors, and tyrosinases also showed biased type distributions. These results support a three-compartment model of the nacreous layer consisting of (i) an insoluble interlamellar membrane core, (ii) a relatively extractable interfacial layer, and (iii) a soluble matrix fraction enriched in proteins potentially involved in ionic regulation and protein maturation. This study provides a quantitative framework for understanding coordinated SMP functions during nacre formation and for comparative analyses of molluscan shell proteomes.

Keywords Nacreous layer · Biomineralization · Shell matrix proteins · Proteomics · Data-independent acquisition

Introduction

Mollusk shells are representative biominerals composed of $CaCO_3$ and organic macromolecules that protect soft tissues from environmental stress and predators (Cusack and Freer 2008). Among them, the nacreous layer of the pearl oyster *Pinctada fucata* has been studied as a representative model for understanding shell formation mechanisms in mollusks because of its highly ordered structure, consisting of regularly stacked aragonite crystals and organic matrix components (Saruwatari et al. 2009; Liu et al. 2015; Evans 2019). The nacreous layer is built from polygonal aragonite tablets arranged in a "brick-and-mortar" architecture (Fig. 1A–C), in which adjacent tablet layers are separated by thin interlamellar membranes composed primarily of β-chitin and proteins (Gerhard et al. 2017). Aragonite formation is thought to proceed not in an unconfined bulk solution, but within a restricted extracellular mineralization space compartmentalized by the interlamellar membranes (Saruwatari et al. 2009; Nudelman et al. 2006; Checa 2018). Thus, this membrane-bounded microenvironment serves not only as the site of crystal growth but also as an important setting that defines the local chemical environment and the properties of the mineral–organic interface. Within such a

✉ Michio Suzuki
amichio@g.ecc.u-tokyo.ac.jp

1 Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, the University of Tokyo, 1-1-1 Yayoi, Bunkyo- ku, Tokyo 113-8657, Japan

2 Institute for Quantitative Biosciences, the University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

3 Mie Prefecture Fisheries Research Institute, 3564-3 Hamajima, Hamajima- cho, Shima-city 517-0404, Mie, Japan

Published online: 20 May 2026

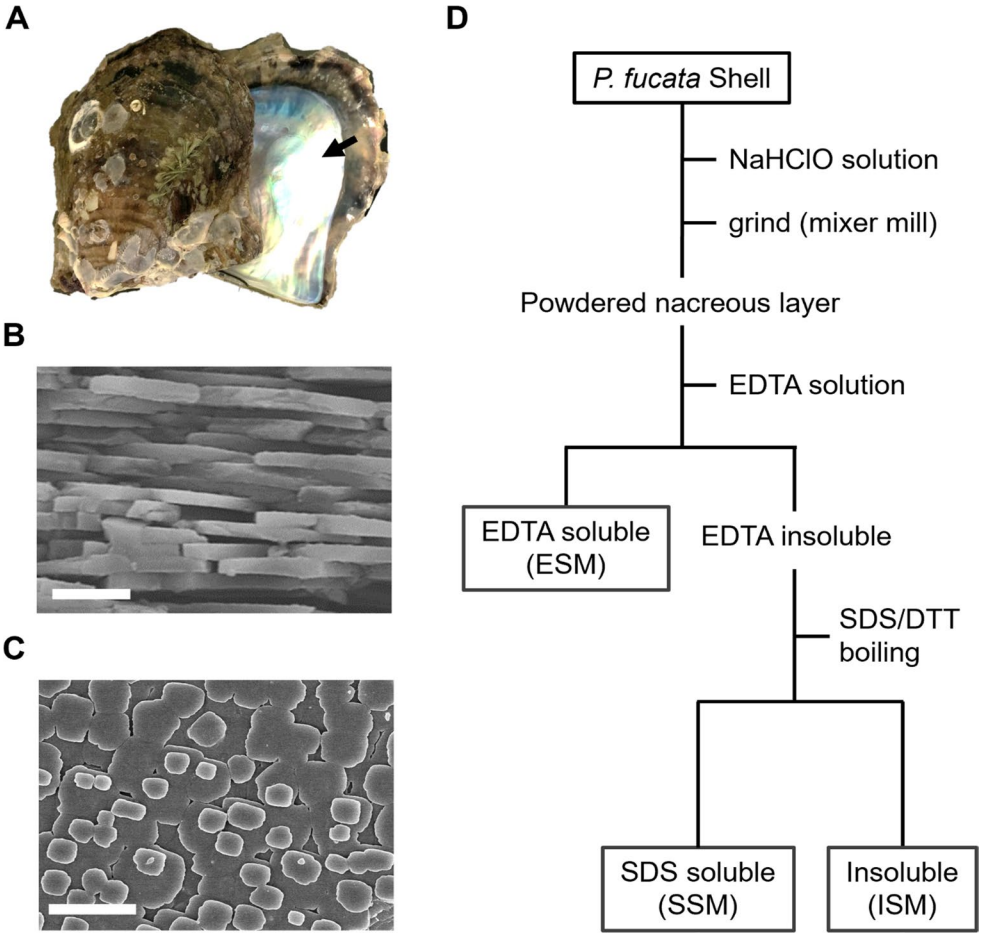

Fig. 1 Nacreous layer of *Pinctada fucata* shell and SMPs extraction scheme. **A**: Shell of *Pinctada fucata*. The black arrow indicates nacreous layer; **B**: Scanning electron microscopy (SEM) image of the nacreous layer cross-section; **C**: SEM image of the nacreous layer surface; **D**: A flow chart for the fractionation of SMPs

compartmentalized mineralization space, shell matrix proteins (SMPs) are thought to act cooperatively in regulating crystal polymorph selection, orientation, and the growth process (Xu et al. 2025; Bai et al. 2023; Otter et al. 2023).

In recent years, proteomics has become a major approach for investigating shell matrix proteins (SMPs) in diverse molluscan shells (Liu and Zhang 2021). For the nacreous layer, the liquid chromatography–tandem mass spectrometry (LC–MS/MS) study by Bédouet et al. (2007) marked a turning point, and subsequent proteomic surveys revealed that nacre contains far more diverse SMPs than previously appreciated (Mann et al. 2012). At the same time, shell proteomics has conventionally analyzed SMPs in two fractions: acid (water- or ethylenediaminetetraacetic acid (EDTA)-) soluble components and acid-insoluble but detergent-soluble components (Paleček et al. 2024; Liu et al. 2015). Based on the properties of the SMPs contained in these fractions, the former has been regarded as the intracrystalline organic matrix involved in controlling crystal growth and morphology, whereas the latter has been considered the components of the interlamellar membrane involved in regulating crystal nucleation, orientation, and polymorphism (Marin et al. 2012; Suzuki and Nagasawa 2013). The abundances of SMPs present in these fractions are expected to differ markedly. Therefore, understanding which SMPs most strongly define the characteristics of each fraction and which play important roles in shell formation requires not only identifying their presence or absence, but also quantitatively evaluating their abundances and preferential distributions among fractions. Nevertheless, most previous shell proteomic studies have remained largely qualitative inventories of SMPs in these fractions, often considering detected SMPs without sufficient regard to differences in abundance among them, and quantitative information for individual SMPs remains limited (Mann and Edsinger 2014). Likewise, comparative shell proteomic studies have often emphasized the presence or absence of conserved domains or protein families (Liu and Zhang 2021), even though large differences in their abundance within the proteome may substantially affect their relative importance. Quantitative shell proteomics is therefore important for characterizing individual SMPs and for advancing our understanding of shell formation mechanisms.

However, quantitative shell proteomics in mollusks has been limited by two major factors: uncertainty in sequence resources and the choice of data acquisition strategy.

Proteomic analyses require comprehensive and reliable protein sequence databases (Heck and Neely 2020; Griss et al. 2011; Guo et al. 2020), yet mantle transcriptomes have often been used as a practical alternative (Jin et al. 2023; Batzel et al. 2022). Transcriptome-derived sequence resources frequently contain incomplete amino acid sequences and uncertain assemblies, ultimately reducing the reliability of protein identification and quantification (Liu et al. 2024; Wong et al. 2025). In *P. fucata*, a complete genome has recently become available (Takeuchi et al. 2022), providing access to full-length protein sequences and markedly improving the reliability of LC–MS/MS-based shell proteomics.

Another major limitation is the LC–MS/MS data acquisition strategy (See Supplementary Information, Fig. S1), which strongly affects the accuracy and completeness of protein quantification (Barkovits et al. 2020). Data-dependent acquisition (DDA) has been widely used in conventional shell proteomics; however, the inherent bias of TopN sampling algorithms can reduce identification coverage and distort quantitative profiles (Mehta et al. 2022; Hu et al. 2016; Alves et al. 2008). In contrast, data-independent acquisition (DIA) fragments all precursor ions within predefined m/z windows, thereby enabling more comprehensive and less biased peptide detection (Venable et al. 2004; Bruderer et al. 2017; Li et al. 2021). Because this approach improves the consistency of peptide sampling, it is well suited for quantitative proteomics (Zou et al. 2025). Multiple studies have shown that DIA can outperform DDA in both proteome coverage and accuracy (Barkovits et al. 2020; Muntel et al. 2015; Tsou et al. 2016). However, despite these advantages, DIA has so far been applied only to a limited extent in shell proteomics. The frequent use of incomplete transcriptome-derived databases together with the selection bias inherent to DDA can substantially compromise both protein identification and quantification. Therefore, combining high-quality genome-derived databases with a DIA strategy represents an important step toward comprehensive and accurate proteomics in mollusks.

In addition to the analytical strategy itself, sample preservation conditions and extraction efficiency are also critical factors that can strongly influence the outcomes of shell proteomics. Although several protocols for extracting SMPs from the *P. fucata* nacreous layer have been reported, markedly different SDS–PAGE banding patterns have been observed even when the same protocol was used (Liu et al. 2015; Liang et al. 2015; Miyashita and Takagi 2011). This suggests that the effects of shell preservation and extraction conditions have not been systematically evaluated, and that differences in storage procedures and extraction methods may affect both integrity and extraction efficiency of SMPs. Therefore, achieving quantitative and comparable proteomic analyses of SMPs requires improvements in both the analytical platform itself and the pre-analytical workflow, from shell sample preservation to extraction.

The aim of this study is to establish a standardized and quantitative shell proteomics framework using the nacreous layer of the *P. fucata* shell that can be applied across studies. To this end, we first optimized shell preservation and subsequent SMP fractionation procedures—namely, EDTA decalcification of nacre followed by sodium dodecyl sulfate/dithiothreitol (SDS/DTT) treatment of the decalcification-insoluble residue—to clearly fractionate extracted SMPs while minimizing SMP degradation. We then integrated a high-confidence protein database constructed from full-length gene models of *P. fucata* (Takeuchi et al. 2022) with DIA mass spectrometry and intensity-based absolute quantification (iBAQ) to obtain fraction-resolved quantitative profiles for the EDTA-soluble matrix (ESM), EDTA-insoluble but SDS/DTT-soluble matrix (SSM), and insoluble matrix (ISM) extracted from the nacreous layer. This integrative approach enabled the identification and quantification of 327 SMPs, exceeding previously reported numbers (Liu et al. 2015; Du et al. 2017). Using these quantitative profiles, we developed a solubility-based type classification framework (types 1–4) linked to 14 inferred functional categories. In doing so, this study provides a quantitative basis for understanding the coordinated roles of individual SMPs in nacre formation by linking their biochemical extractability with functional predictions. Furthermore, the framework proposed here may serve as a basis for comparative studies of molluscan biomineralization and, in the longer term, may also provide fundamental information for biomimetic synthesis of organic–inorganic composite materials.

Materials and Methods

Ethics Statements

This study involved an invertebrate species that is not subject to animal ethics approval under the institutional guidelines applicable to this work; therefore, no specific ethical approval was required.

Shell Samples

The *P. fucata* shells used in this study were obtained from three-year-old oysters farmed in Ago Bay, Mie Prefecture. The soft tissues were removed, and the shells were washed sequentially with seawater and 70% ethanol. After washing, the shells were stored at room temperature for 1 month, 1 year, and at −30 °C for 4 years. Frozen storage at −30 °C was selected as a practical long-term preservation condition routinely used in our laboratory to minimize SMP

degradation and to compare its preservative effect with that of room-temperature storage. Freshly washed shells were prepared similarly and used for comparison.

SEM Imaging

Cross-sectional specimens of the nacreous layer were prepared using a low-speed cutter equipped with a diamond blade (Buehler Ltd., Tokyo, Japan), whereas small nacre fragments prepared with pliers were used for surface observations. Samples were mounted on SEM stubs using conductive carbon tape and sputter-coated with a thin Pt–Pd layer (E-1030, HITACHI, Tokyo, Japan). SEM imaging was performed using a field-emission SEM (S-4800, HITACHI) at an accelerating voltage of 10 kV.

Protein Extraction

Shell matrix proteins were extracted and fractionated following the workflow shown in Fig. 1D. *P. fucata* shells were immersed in a 30% NaClO to dissolve and completely remove the outer prismatic layer. The shells were then washed with distilled water to remove residual NaClO. The remaining nacreous layer was then dried overnight at room temperature. It was then crushed into approximately 3 cm × 3 cm pieces using a hammer, followed by grinding with a Wonder Crusher WC-3 (Osaka Chemical, Japan) at a speed of 5 for 1 min to produce a fine nacre powder.

Five grams of nacre powder was suspended in 100 mL of 0.5 M EDTA-Na (pH 8.0; adjusted with NaOH) and demineralized at 4 °C with magnetic stirring (500 rpm) for 96 h. This duration was selected based on a preliminary time-course optimization in which demineralization was monitored for up to 7 days. After demineralization, the mixture was centrifuged (12,000 × *g*, 10 min, 4 °C) to separate the supernatant and the demineralized residue. The supernatant was concentrated by ultrafiltration (Amicon Ultra-15, 10 kDa MWCO; Millipore) at 4,000 × *g* and 4 °C, diluted with 50 mM Tris-HCl (pH 8.0), and further concentrated by ultrafiltration to remove EDTA. The resulting fraction was designated as the ESM.

The demineralized residue was suspended in 10 mL of 50 mM Tris-HCl (pH 8.0) containing 1% SDS and 10 mM DTT and boiled for 10 min. After cooling the reaction mixture to room temperature, it was separated into the supernatant and insoluble matrix by centrifugation (12,000 × *g*, 4 °C, 10 min). The insoluble residue was treated similarly, and the supernatant was collected. This procedure was repeated five times, and the supernatants from each round were concentrated by ultrafiltration (MWCO: 10 kDa, 4,000 × *g*, room temperature). The supernatant was diluted with 50 mM Tris-HCl (pH 8.0) and concentrated by ultrafiltration, resulting in the SSM. The residue remaining after five SDS/DTT treatments was designated as the ISM. The protein content of the ISM was estimated to be 74% of the total amount (Kintsu et al. 2020).

The protein yield per 5 g of nacre powder was calculated based on the protein concentration of the extraction solution. Protein concentration was quantified on a Qubit 4 fluorometer using the Qubit Broad Range Protein Assay Kit (Invitrogen, USA).

LC-MS/MS Sample Preparation

ESM and SSM were diluted to 100 µL with 7 M guanidine hydrochloride, 0.01 M EDTA, and 0.05 M Tris-HCl buffer (pH 8.0), and 1 µL of 0.5 M DTT was added. The reduction reaction was carried out at 60 °C for 30 min. Afterward, 2 µL of 0.5 M iodoacetamide solution was added, and the mixture was stirred at room temperature in the dark for 1 h. Subsequently, 400 µL of methanol, 100 µL of chloroform, and 300 µL of pure water were added in this order and vortex-mixed. After centrifugation (14,000 × *g*, 5 min, 4 °C), the supernatant was discarded. On ice, 300 µL of methanol was added to the lower phase and mixed by inversion, followed by centrifugation (15,000 × *g*, 3 min, 4 °C). The supernatant was removed, and the protein pellet was then resuspended with 200 µL of 70% ethanol and further centrifuged (12,000 × *g*, 3 min, 4 °C). The supernatant was discarded, and the pellet was completely dried using a rotary evaporator. The sample was dissolved by adding 13 µL of pure water, followed by the addition of 26 µL of 50 mM NH_4HCO_3 (pH 8.0) and 5 µL of 100 ng/µL Trypsin Platinum (Promega, USA) solution. The protein was digested at 37 °C for 24 h, and 1 µL of 4% trifluoroacetic acid (TFA) was added. The digestion solution was concentrated via rotary evaporation and prepared as an LC-MS/MS sample.

The ISM was boiled in 10 mL of 10 mM DTT for 10 min, and iodoacetamide was added to a final concentration of 20 mM. The mixture was stirred at room temperature in the dark for 1 h. After centrifugation to remove unreacted iodoacetamide, the precipitate was suspended in 1 mL of 50 mM NH_4HCO_3 solution. Additionally, 6 µL of 1 µg/µL Trypsin Platinum solution was added, and the mixture was stirred at 37 °C for 48 h for protein digestion. The supernatant was applied to a Sep-Pak C18 column (Merck, Germany) to purify peptides. The purified peptides were dried and dissolved in 0.1% TFA.

LC-MS/MS DDA Measurement

Measurements were performed using 250 ng of the peptides. The separation was carried out with a Dionex Ultimate 3000 nano UPLC (Thermo Scientific, USA) equipped with a 3

μm C18 NANO HPLC CAPILLARY COLUMN NTCC-360/75-3-125 (75 μm × 12.5 cm, Nikkyo Technos, Japan). nano capillary microflow liquid chromatography (nano-LC) was performed at 40 °C with a flow rate of 300 nL/min. In nano-LC, peptides were separated and eluted over a 60-min gradient, from 98% solvent A (0.1% formic acid solution)/2% solvent B (0.1% formic acid acetonitrile solution) to 60% solvent A/40% solvent B. The peptides eluted from the column were ionized by nano-ESI and analyzed in the positive mode using a Thermo Orbitrap Fusion mass spectrometer (Thermo Scientific). The Top10 DDA method was set with a maximum cycle time of 1 s. The MS1 scan for peptide ions was performed in the range of m/z 375–1500, with a resolution of 60,000 and an automatic gain control set to "standard". The charge state was set from +2 to +7, and the mass tolerance was set to 10 ppm. The peptide ions were isolated using a quadrupole isolation window of 1.6 Da and subjected to CID fragmentation with 35% collision energy in the collision cell, followed by MS2 analysis. In MS2, the automatic gain control was set to standard with a maximum injection time of 35 ms.

LC-MS/MS DIA Measurement

The separation and mass analysis setup was the same as that used for the DDA measurement, with 250 ng of the measured sample. The solvents, flow rates, and gradients used for the nano-LC were also identical. Peptides eluted from the column were ionized by nano-ESI and analyzed in positive mode. MS1 peptide ions, with charge states from + 2 to + 3 and m/z values ranging from 390 to 1010, were isolated in 4 m/z increments using an isolation window in the collision cell. These ions were then fragmented by HCD at 30% energy. In MS2, the resolution was set to 30,000 with a maximum injection time of 60 ms.

Data Analysis and Pprotein Quantification

Both DDA and DIA analyses were performed using MaxQuant ver 2.6.0 (Cox and Mann 2008). In the DDA analysis, the type in the group-specific parameter tab was set to Standard, whereas in the DIA analysis, the type was set to MaxDIA. Additionally, the Match between runs and "iBAQ" options was selected in the Global parameters tab, and all other parameters were left as defaults. Database searches were conducted using the Andromeda search engine, with pfuV4.1HapA_ref_genemodels_aa (Takeuchi et al. 2022) and the common contaminant database implemented in MaxQuant as reference databases. The pfuV4.1HapA_ref_genemodels_aa database was manually modified so that the entry for Pif 177 (v4.1_scaf_03a.g7387.t1) was split into two sequences, Pif 80 and Pif 97, reflecting its known processing at a dibasic site in the nacreous layer. Trypsin/P was specified as the digestion enzyme, allowing up to two missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification, and oxidation of methionine and protein N-terminal acetylation were set as variable modifications. Peptide- and protein-level false discovery rates were controlled at 1% using a target–decoy strategy.

The ProteinGroups.txt file output from the analysis was used for downstream analysis. First, contaminants and reverse sequences were removed from the list of identified proteins, and proteins with fewer than two unique peptides were excluded. The iBAQ values were used for protein quantification. For missing values, the lowest observed value was substituted using a lowest-value replacement method. The iBAQ values were converted to protein mass-based relative abundances using the following equation (Heinemann et al. 2021; Li et al. 2020; Zhu et al. 2024).

$$mass\ abundance\ of\ protein_i\ (\%) = \frac{iBAQ_i \times\ molecular\ weight_i}{\sum_{i=1}^{n} iBAQ_i \times\ molecular\ weight_i} \times\ 100$$

$$protein\ mass_i\ (mg) = \frac{mass\ abundance\ of\ protein_i\ (\%)}{100} \times \text{total mass}\ (mg)$$

The relative abundance of each protein in the three fractions (ESM, SSM, and ISM) was calculated using the above method. The ratios of abundance in the ESM to SSM and in the SSM to ISM were then determined. Based on the patterns of these two ratios, proteins were classified into four types. For details on the correspondence between ratio patterns and solubility-based types, see Fig. S6.

The proteomaps were generated in R ver. 4.5.0 (R Development Core Team 2013) using custom scripts and the *WeightedTreemaps* package in RStudio, and intraclass correlation coefficients (ICC(3,1) were also calculated in R ver. 4.5.0 using custom scripts and the *psych* package, whereas all other statistical analyses and data visualizations were performed using Perseus ver. 2.0.7.0 (Tyanova et al. 2016).

Sequence-Based Functional Annotations and Structure Predictions

Protein functions were inferred from sequence similarities using NCBI BLAST searches (Altschul et al. 1997), conserved domain searches (Wang et al. 2023), and SignalP 6.0 (Nielsen et al. 2019). BLAST hits with E-values $< 1.0 \times 10^{-5}$ were considered significant and only the top hit was retained for annotation. Conserved domains with E-values below 1.0×10^{-5} were retained as high-confidence annotations.

The functional category of each protein was determined in a stepwise manner based on reference annotations of known SMPs, BLAST sequence similarity, conserved domain architecture, and the characteristics of low-complexity regions. First, when a protein showed high similarity to a known protein in BLASTp, with percent identity ≥ 75%, query coverage ≥ 80%, and subject coverage ≥ 80%, the functional category of the known protein was assigned. Next, when classification could not be made based on high sequence similarity, the combination of detected conserved domain names was defined as a domain signature, and the representative category corresponding to the same signature in known proteins was assigned. This representative category was defined as the category most frequently associated with known proteins sharing the same domain signature. When classification could not be achieved by these procedures, candidate categories were compiled based on information contained in the conserved domain names and the names of the representative BLAST hits, and the category receiving the greatest support was assigned. When multiple categories were equally supported, the final category was determined according to the following predefined priority order: nacrein = N16 and N19 = Pif-family = tyrosinase = oxidase = protein modifier > insoluble framework > carbohydrate active > protease > protease inhibitor > others > protein interaction > calcium binding > uncharacterized. Here, "insoluble framework" does not denote a conventional molecular function, but rather a structure-oriented category for proteins inferred to contribute to insoluble fibrous scaffold formation. Proteins were assigned to this category when they showed similarity to previously reported framework-forming shell proteins such as MSI60 (Sudo et al. 1997), shematrin, or N-U5, and/or when they contained extensive low-complexity regions over broad parts of the sequence enriched in amino acids associated with aggregation-prone fibrous proteins, such as glycine and alanine. In addition, similarity to fibrous proteins, including collagen, or the presence of conserved domains associated with filament-forming proteins, such as ASKHA_NBD_actin, FN3, or the Filament superfamily, was treated as supportive evidence for this category when no stronger alternative functional assignment was available.

Putative 3D structures of v4.1_scaf_11a.g26491.t1, v4.1_scaf_13a.g30686.t1, v4.1_scaf_13a.g31190.t1, and v4.1_scaf_13a.g31191.t1 were generated using AlphaFold3 (Abramson et al. 2024). In addition, for the N-terminal domain of v4.1_scaf_11a.g26491.t1 and the high-confidence N-terminal region of v4.1_scaf_13a.g30686.t1, structurally similar proteins were searched by performing 3D–3D structural alignment against the PDB database using the DALI server (Holm 2022).

Sequence Alignment

The putative chitin-binding domain of v4.1_scaf_13a.g30686.t1 was used as a query in the BLAST search implemented in UniProt to identify protein homologs within Mollusca. Multiple sequence alignments were performed using ClustalW (Thompson et al. 1994).

Results

Optimized Fractionation Preserves SMPs and Enables Reproducible Biochemical Partitioning

Because the nacreous layer of *P. fucata* consists predominantly of aragonite crystals with extremely low water content (Kong et al. 2019), conventional cell-lysis approaches are insufficient for recovering SMPs. We therefore decalcified the nacreous layer with EDTA solution to recover the co-eluted ESM, and subsequently extracted the decalcification precipitates with an SDS/DTT solution to obtain the SSM. The final residual fraction was defined as the ISM (Fig. 1D).

In the optimized workflow, EDTA-mediated decalcification was completed within 4 days. While the OD_{600} value stabilized within approximately 2 h, indicating rapid dissolution of aragonite, the ESM protein yield reached a plateau only after 96 h (Fig. 2A). The remaining matrix was then subjected to five cycles of SDS/DTT treatment, which recovered nearly all SMPs extractable under these conditions (Fig. 2B). Importantly, the stepwise extraction procedure used in this study was designed to recover, as completely as possible, the SMPs solubilized under each treatment condition. As a result, nacre SMPs could be separated relatively clearly into three fractions based on their solubility. The final yield of ESM was 2.2 ± 0.2 mg per 5 g of nacreous layer (Table 1). The cumulative yield of SSM was 9.2 ± 0.4 mg per 5 g of nacreous layer. ISM amounted to 11 ± 1 mg per 5 g of nacreous layer. Because the decalcified nacre matrix is known to consist of approximately 74% protein and 26% β-chitin (Kintsu et al. 2020), the protein

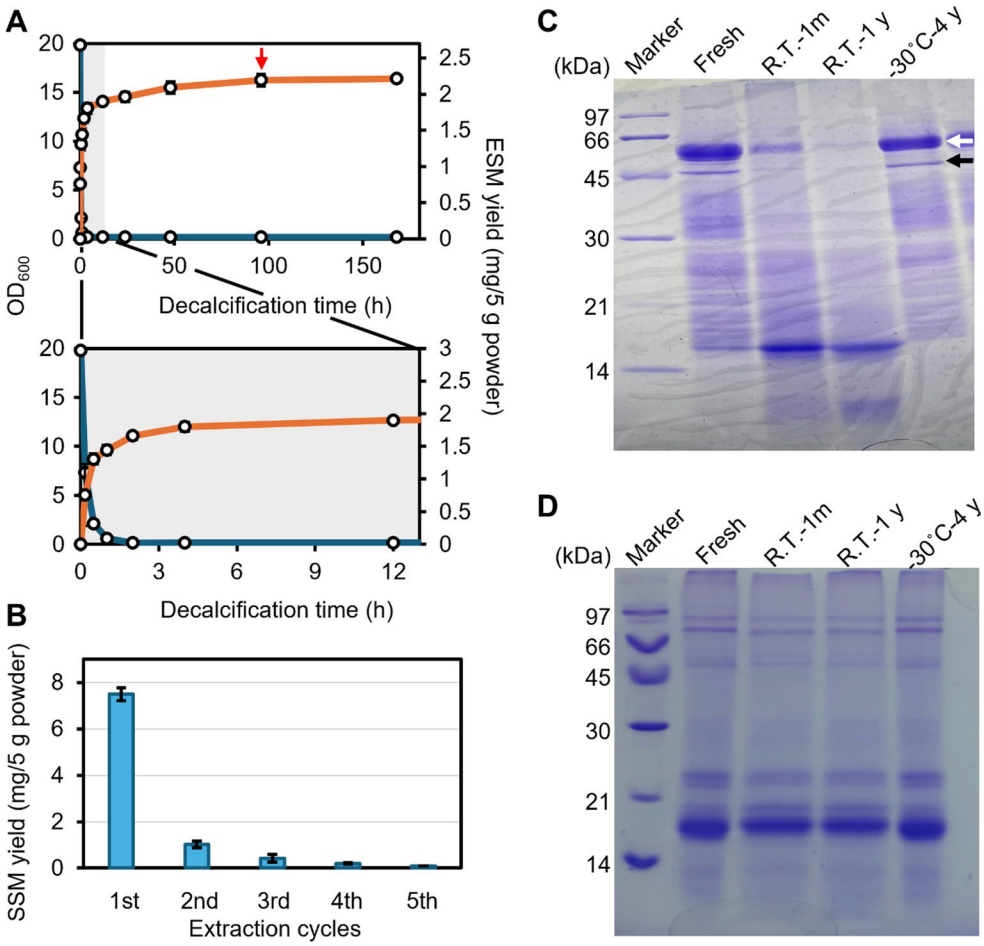

Fig. 2 Optimization of nacre SMP extraction and assessment of storage-dependent protein stability. **A**: Decalcification time course of 5 g nacre powder in 100 mL 0.5 M of EDTA, monitored by OD_{600} (blue line; aragonite dissolution) and ESM protein yield (orange line). The red arrow indicates the addition of a further 100 mL of EDTA solution, but this did not increase protein yield. All data are presented as mean ± S.D. (n = 3).; **B**: SSM yield across successive SDS/DTT extractions of the decalcified residue. All data are presented as mean ± S.D. (n = 3).; **C**: SDS–PAGE of ESM from nacre stored under the indicated conditions. The white and black arrows indicate the bands for nacrein and nacreous abundant protein 1 (NAP1), respectively.; **D**: SDS–PAGE of SSM from nacre stored under the indicated conditions. Gels were stained with CBB

Table 1 Protein yields (mg/5 g nacreous layer). Protein yields of each extraction fraction obtained by decalcifying 5 g of powdered nacreous layer followed by SDS/DTT treatment. Values represent mean ± S.D. (n = 3)

	ESM	SSM	Protein in ISM [a]
Fresh	2.2 ± 0.2	9.2 ± 0.4	8.5 ± 0.3
R.T. – 1 month	1.6 ± 0.1*	8.7 ± 0.3	8.3 ± 0.2
R.T. – 1year	1.5 ± 0.1*	8.8 ± 0.4	8.4 ± 0.5
−30°C – 4 years	2.0 ± 0.3	9.5 ± 0.5	8.7 ± 0.4

[a] ISM is a mixture of proteins and chitin fibers. The protein content in the nacreous ISM fraction was estimated as 74% of the total ISM weight, based on the report by Kintsu et al. 2020. The * indicates significant differences from the Fresh samples, assessed by Dunnett's test (p < 0.05)

content of the ISM was estimated to be 8.5 ± 0.3 mg per 5 g of nacreous layer.

Protein integrity depended strongly on shell storage temperature. Clear degradation of ESM proteins was observed in shells stored at room temperature, with a prominent ~ 15–16 kDa fragment detected on SDS–PAGE that was likely derived from nacrein (apparent molecular weight, ~ 60 kDa (Miyamoto et al. 1996) (Fig. 2C). In contrast, no comparable degradation was detected in samples stored at −30 °C. By contrast, the yield and overall banding pattern of the SSM were largely unaffected by storage conditions (Fig. 2D; Table 1).

These results indicate that both storage and fractionation conditions can affect proteomic analysis of nacre SMPs. All subsequent quantitative proteomic analyses were therefore performed using these optimized preservation and fractionation conditions.

DIA Improves SMP Coverage and Quantitative Depth Relative to DDA

To evaluate whether the data acquisition strategy limited SMP identification and quantification, we compared DDA and DIA analyses for the ESM fraction. DIA increased the number of identified SMPs from 94 in DDA to 221 (Fig. 3A). Furthermore, among the proteins identified by both methods, 68% showed an increase in the number of peptide fragments identified per protein in the DIA analysis. On average, DIA yielded 3.1 more peptides per protein than DDA (median increase: 3 peptides; Fig. 3B). As a result, sequence coverage of the identified proteins also improved, with average and median increases of 5.8% (DDA: 21.0%,

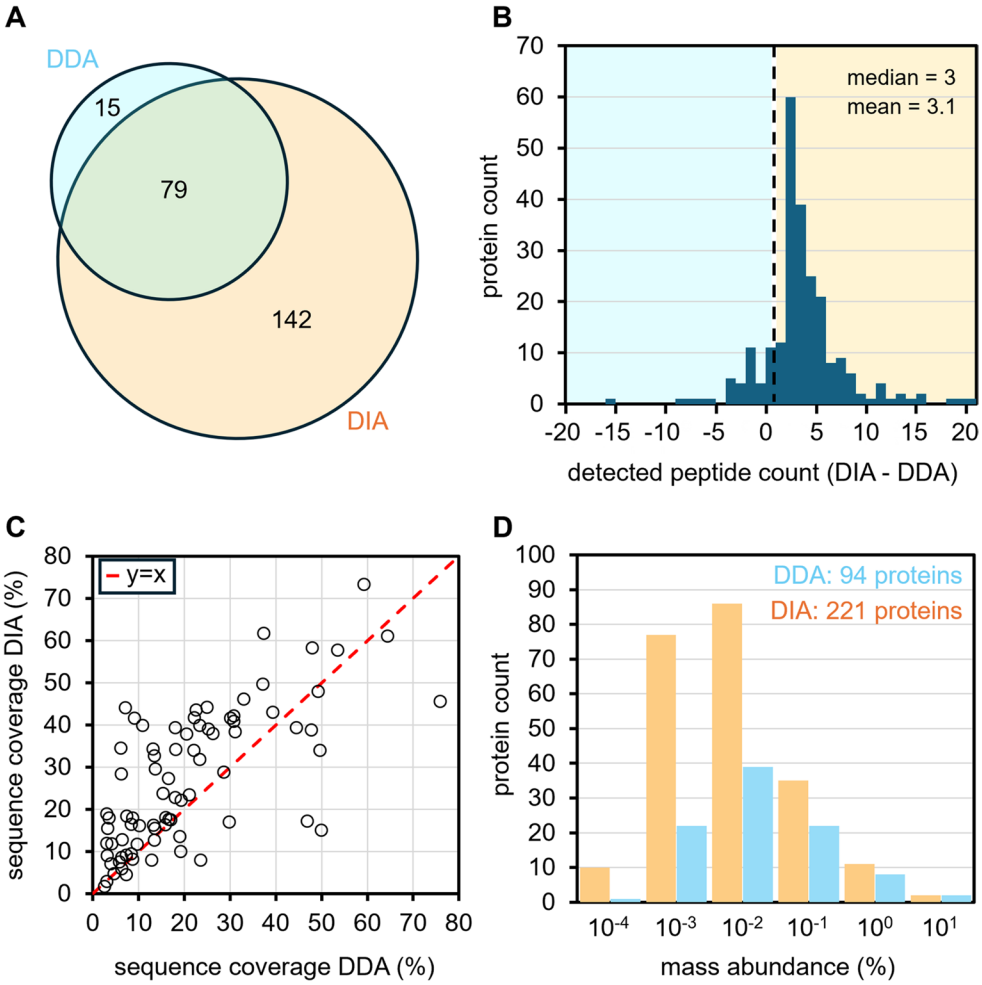

Fig. 3 Comparison of ESM profiling by DIA and DDA. **A**: Venn diagram of proteins identified from ESM by DDA and DIA; **B**: Distribution of unique peptides per protein; **C**: Distribution of sequence coverage; **D**: Distribution of protein mass abundance calculated from iBAQ value. The full list of identified proteins and the values used for comparison are provided in Table S1

DIA: 26.8%) and 6.6% (DDA: 16.7%, DIA: 23.3%), respectively (Fig. 3C).

For protein quantification, we used iBAQ (Schwanhäusser et al. 2011; Shin et al. 2013) values to estimate the relative mass abundance of each SMP. The mass abundance of each SMP was calculated by multiplying its iBAQ value by its molecular weight and normalizing it to the total protein mass (Heinemann et al. 2021; Li et al. 2020; Zhu et al. 2024). DIA also expanded the detection of low-abundance SMPs with mass abundance below 0.1%, which are likely to be missed by TopN sampling in DDA (Fig. 3D). In this ESM-fraction benchmark, DIA outperformed DDA in proteome coverage and quantitative depth. These results indicate that DIA is well suited for analyzing protein populations with high complexity and a broad dynamic range, such as shell proteome. Based on these results, all subsequent analyses were performed using DIA. The MaxQuant output and calculated values are summarized in Table S1 (Supplementary Information).

Fraction-Resolved Quantitative Proteomics Reveals Distinct Abundance Biases Across ESM, SSM, and ISM

Using the optimized SMP fractionation procedure and DIA workflow, we identified 327 SMPs across the three fractions—ESM, SSM, and ISM—and quantified their relative mass abundances over a broad dynamic range (> 10^5; Fig. 4). The quantitative values, signal peptide predictions, BLAST annotations, conserved domains, and inferred functions of the identified SMPs are summarized in Table S2 (Supplementary Information). The mean coefficients of variation (CV) of the calculated protein mass abundances across replicates were 38% for ESM, 35% for SSM, and 36% for ISM, likely reflecting variability introduced by the multistep sample preparation process (Fig. S2). Nevertheless, Pearson correlation coefficients (R) between replicates exceeded 0.84 for all fractions, and intraclass correlation

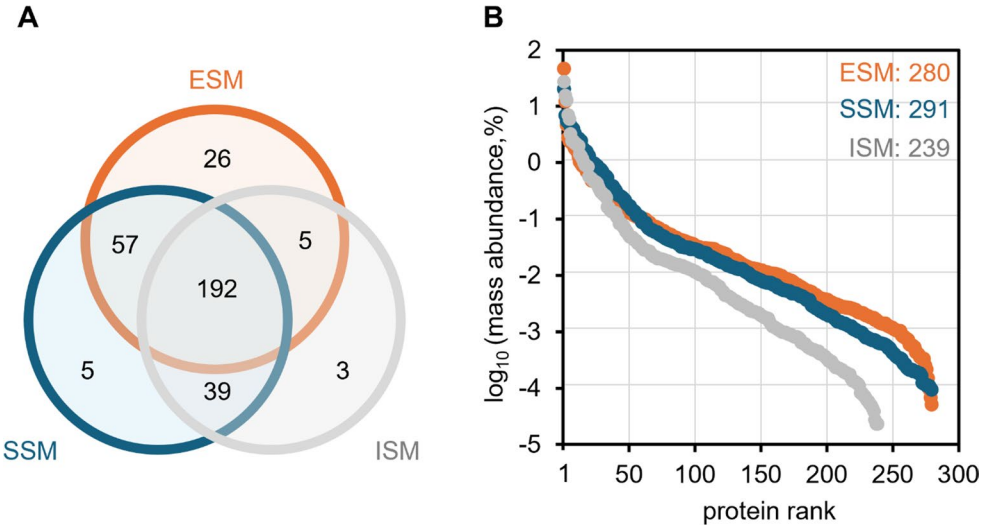

Fig. 4 DIA analysis of ESM, SSM, and ISM. **A**: Venn diagram of proteins identified from ESM, SSM, and ISM; **B**: Distribution of mass abundance of detected SMPs. All data are presented as mean ± S.D. (n = 3). The full list of identified proteins and quantitative values is provided in Table S2

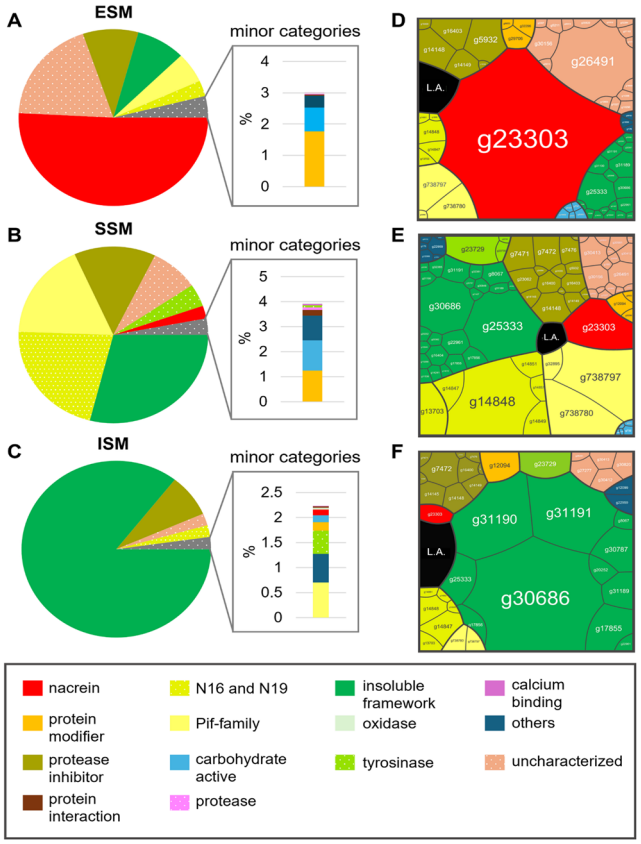

Fig. 5 Detailed protein profiles of the fractions. Inferred functions of each SMP were classified into 14 functional categories. **A**–**C**: Pie charts showing the total mass abundance (%) of proteins belonging to each functional category in ESM, SSM, and ISM, respectively; **D**–**F**: Proteomaps (Voronoi diagrams) of ESM, SSM, and ISM. In the proteomap, the mass abundance of each SMP is represented by the size of each cell. Only proteins with abundances > 0.1% are displayed, while low-abundance proteins are aggregated into a single black cell labeled "L.A.". Values represent the mean (n = 3)

coefficients (ICC (3,1)) were greater than 0.96, indicating good overall reproducibility of the abundance profiles.

The 10 most abundant SMPs in each fraction and their domain architectures are shown in Fig. S3–S5. Among the SMPs containing conserved domains were several that have been implicated in crystal formation, including nacrein (v4.1_scaf_09a.g23303.t1: Miyamoto et al. 1996) and Pif (v4.1_scaf_03a.g7387.t1: Suzuki et al. 2009). In addition, several SMPs containing protease inhibitor domains (e.g., v4.1_scaf_05a.g14148.t1, v4.1_scaf_06a.g16403.t1, and v4.1_scaf_03a.g7472.t1) and an SMP containing a tyrosinase domain (v4.1_scaf_09a.g23729.t1) were also identified. High-abundance SMPs lacking conserved domains were likewise detected as major components, including N16 (v4.1_scaf_05a.g13703.t1: Samata et al. 1999), N19 (v4.1_scaf_05a.g14848.t1: Yano et al. 2007), the previously reported MSI60 (v4.1_scaf_13a.g31190.t1; Sudo et al. 1997), two additional MSI60-like proteins identified in this study (v4.1_scaf_13a.g30686.t1 and v4.1_scaf_13a.g31191.t1), and shematrin (v4.1_scaf_06a.g17855.t1: Yano et al. 2006). The detection of these SMPs as high-abundance components is consistent with the fact that many of them were originally identified in early SMP studies by isolating and partially sequencing individual proteins, rather than through comprehensive proteomic approaches.

However, the second most abundant SMP in the ESM (v4.1_scaf_11a.g26491.t1) was a previously unreported protein. This protein was not included in the sequence resources used in earlier *P. fucata* shell proteomics studies based on older genome assemblies (Takeuchi et al. 2012, 2016), and it was also readily degraded during room-temperature storage (Fig. 2C). The combination of a high-coverage genome-derived database with optimized preservation and extraction conditions enabled its robust detection and quantification. We therefore designated this newly identified

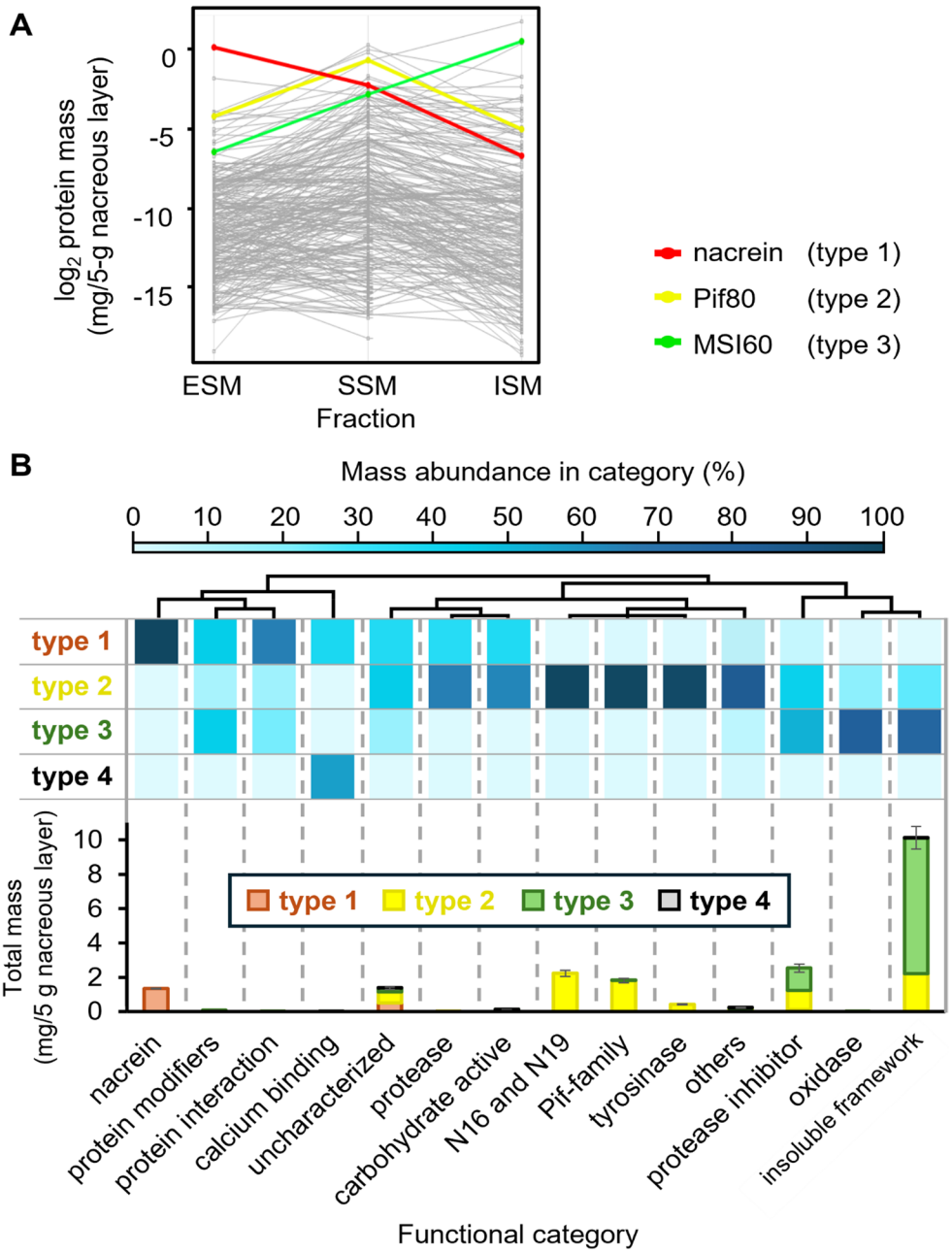

Fig. 6 Solubility-based type classification of nacreous layer SMPs. **A**: Fractional distribution of individual SMPs across ESM, SSM, and ISM based on iBAQ-derived protein mass. Representative proteins for types 1–3 are highlighted.; **B**: Association between solubility type (types 1–4) and functional categories shown as a heatmap (top) and stacked bar chart (bottom). The heatmap color indicates the proportion of the total mass of each functional category that is accounted for by SMPs of each type. The stacked bar chart is presented as mean ± S.D. ($n = 3$)

high-abundance component nacreous abundant protein 1 (NAP1).

Identified SMPs were assigned to 14 functional categories based on inferred functions derived from homology, conserved domain information, and sequence features. (see Methods). The relative proportions of these categories within each fraction, based on mass abundance, are shown as pie charts and proteomaps (Fig. 5). In the ESM, nacrein was the most abundant protein, accounting for nearly half of the total protein mass in this fraction. By contrast, the abundance of nacrein was markedly lower in the SSM and ISM, where Pif-family proteins together with N16 and N19 category were more prominent in the SSM, whereas SMPs assigned to the insoluble framework category predominated in the ISM. Thus, each fraction was characterized by a distinct set of SMPs, indicating that these fractions differ in their biochemical properties.

Solubility-Based Classification Groups SMPs Into Four Quantitative Partitioning Types

To examine the relationship between inferred function and fraction partitioning, we calculated the quantitative distribution of each SMP across the three fractions. The protein mass (mg) per 5 g of nacreous layer was estimated by multiplying the mass abundance of each protein by the total

protein yield of each fraction. Based on the resulting mass ratios among the ESM, SSM, and ISM, SMPs were classified into four solubility-based types (Fig. S6; Fig. 6A): type 1 (typically defined by an ESM/SSM ratio≥2 and an SSM/ISM ratio≥2), type 2 (ESM/SSM≤0.5 and SSM/ISM≥2), type 3 (typically defined by an ESM/SSM ratio≤0.5 and an SSM/ISM ratio≤0.5), and type 4 (proteins that did not clearly meet these criteria). Here, types 1, 2, and 3 were interpreted as SMPs with relatively high (ESM-enriched), intermediate (SSM-enriched), and low (ISM-enriched) solubility, respectively. Importantly, this solubility-based type classification not only enables the distinct properties of the fractions to be evaluated in terms of SMP partitioning, but also provides a comparable summary of the biochemical partitioning pattern of individual SMPs, that is, the fraction with which each SMP is predominantly associated.

Several well-studied SMPs were classified into types that were intuitively consistent with their known properties. For example, nacrein, originally identified as a major component of the ESM (Miyamoto et al. 1996), was classified as type 1; Pif 80, which is abundant in the SSM (Suzuki et al. 2009), was classified as type 2; and MSI60, which is abundant in the ISM and has been proposed to form insoluble fibers (Sudo et al. 1997), was classified as type 3 (Fig. 6A). When the functional categories were mapped onto the solubility-based types (Fig. 6B), type 1 was enriched in SMPs including nacrein and Ca^{2+}-binding categories, type 2 in crystal-modulation-related proteins such as N16 and Pif-family, and type 3 in insoluble framework category including MSI60 and shematrin. By contrast, proteases were distributed mainly in types 1–2, whereas protease inhibitors tended to be distributed mainly in types 2–3. Because the partitioning patterns of these groups were less intuitively predictable from previous studies than those of nacrein, Pif 80, and MSI60, this classification was also useful for showing that fraction-to-fraction distribution biases exist among inferred functional groups.

Discussion

A Standardized Workflow Improves the Depth of Nacre Proteomics

In this study, we developed a quantitative analytical framework for nacre SMPs by integrating optimization of shell preservation and SMP fractionation, a high-quality genome-derived database, and DIA-based quantitative proteomics. In particular, the marked degradation observed in the ESM after room-temperature storage, in contrast to the absence of comparable degradation in samples stored at −30 °C, indicates that soluble SMPs are highly sensitive to preservation conditions. The low-molecular-weight fragment observed in Fig. 2C, likely derived from nacrein, provides a representative example of the susceptibility of the ESM to degradation during storage. By contrast, the SSM was relatively insensitive to storage conditions, suggesting that protein stability differs among fractions.

In the comparison using the ESM fraction, DIA outperformed DDA in terms of the number of identified proteins, the number of peptides per protein, and sequence coverage, and it was also more effective for detecting low-abundance SMPs. Previous proteomic studies of the *P. fucata* nacreous layer have relied largely on DDA-based analysis (Liu et al. 2015; Du et al. 2017), which may have contributed to limitations in proteome coverage and quantitative resolution. The identification of 327 SMPs in the present study indicates that the combination of optimized preservation and extraction conditions with DIA analysis based on a genome-derived database substantially improves the depth of nacreous layer proteome.

Solubility-Based Type Classification Provides a Biochemical Framework for Interpreting SMP Functional Differentiation

In this study, SMPs were classified into four solubility-based types according to their quantitative distribution across the ESM, SSM, and ISM fractions. Proteins such as nacrein and Ca^{2+}-binding proteins were more frequently assigned to type 1. Nacrein is an SMP containing a carbonic anhydrase domain and is thought to be involved in supplying bicarbonate ions for $CaCO_3$ formation (Miyamoto et al. 2005). In addition, the glycine-x-asparagine (GxN) repeat region inserted within its catalytic domain has been proposed to regulate crystallization through interactions with Ca^{2+} and the mineral phase (Miyamoto et al. 2005; Gong et al. 2008). SMPs containing Ca^{2+}-binding motifs, such as EF-hand, have likewise been suggested to participate in biomineralization by buffering, transporting, or locally regulating ionic species in the mineralization environment (Hanif et al. 2023; Namikawa et al. 2025). Taken together, these observations suggest that type 1 is enriched in proteins potentially involved in regulating the ionic environment of calcification.

Type 1 also contained many protein modifier category SMPs, including peptidyl–prolyl cis–trans isomerases and protein disulfide isomerases, and protease category SMPs. Given that SMPs are produced in mantle epithelial cells and secreted into the mineralization space, these modifying factors may support SMP folding and proteolytic processing before and/or during secretion (Yarra et al. 2021). Indeed, several nacre SMPs have been reported to undergo protease-dependent processing (Lao et al. 2007; Suzuki et al. 2009). Interestingly, proteases were more common in types 1–2, whereas protease inhibitors were enriched in types 2–3.

This distribution may be related to the greater susceptibility of ESM proteins to degradation during room-temperature storage, although direct evidence for protease activity in stored shells will be required to test this possibility.

NAP1 did not show detectable similarity to any functionally characterized sequence in BLAST or conserved domain searches. However, structural prediction suggested that NAP1 adopts a well-defined fold composed of two structural units: domain A, consisting mainly of α-helices, and domain B, containing both α-helices and β-sheets (Fig. S7). In 3D–3D structural alignment analysis, domain A showed structural similarity to a domain involved in oligomerization of the Vegetative Insecticidal Protein 3B (Vip3B, Zheng et al. 2020). The structural similarity of domain A to an oligomerization-related domain raises the possibility that this region contributes to protein–protein interactions. By contrast, no structurally similar protein was detected for domain B, raising the possibility that it represents a structurally unique domain. Given that NAP1 is the second most abundant component in the nacreous ESM fraction, it may play an important but as yet uncharacterized role in nacre formation, which warrants further investigation.

By contrast, SMPs involved in crystal modulation, such as N16 and Pif, were classified predominantly as type 2. These proteins are associated with the interlamellar membrane and are thought to regulate the oriented crystallization of $CaCO_3$ by acting at the mineral–organic interface (Samata et al. 1999; Suzuki et al. 2009). These proteins, as well as their functional fragments, have also been extensively studied for their ability to induce aragonite formation on substrates (Du et al. 2016; Bahn et al. 2017; Keene et al. 2010). This distribution suggests that type 2 is enriched in proteins previously implicated in crystal nucleation, polymorph selection, and mineral–organic interfacial regulation.

Structural Organization of the Interlamellar Membrane Revealed by SMP Classification

Previous studies have generally classified the decalcification-soluble fraction of shells as the intracrystalline component, whereas the decalcification-insoluble residue has been regarded as the component of the interlamellar membrane together with β-chitin fibers (Suzuki and Nagasawa 2013; Paleček et al. 2024). In the present study, further subdivision of this interlamellar membrane fraction into the SSM and ISM enabled a more detailed view of interlamellar membrane organization. Given that β-chitin fibers are known to be insoluble under SDS/DTT treatment (Kintsu et al. 2020), the interlamellar membrane can be considered as consisting of two distinguishable components: SDS/DTT-insoluble type 3 SMPs together with β-chitin fibers, and SDS/DTT-soluble type 2 SMPs.

Although Pif-family proteins contain chitin-binding domains, most of them were classified as type 2, suggesting that these proteins are more readily extractable from the interlamellar membrane than type 3 SMPs. This behavior is consistent with a surface-associated or interfacial mode of interaction rather than deep incorporation into the insoluble core. This interpretation is further supported by the observation that many other chitin-binding proteins were likewise rich in types 1 and 2 (Table S2). In contrast, type 3 was strongly enriched in SMPs assigned to the insoluble framework category, suggesting that type 3 proteins, together with β-chitin fibers, are likely candidates for the structurally stable core of the interlamellar membrane.

Insoluble framework category SMPs accounted for approximately 80% of type 3, and this group was dominated by the MSI60 family proteins. In *Pinctada fucata*, the previously reported conventional MSI60 corresponds to v4.1_scaf_13a.g31190.t1 (Sudo et al. 1997). This protein is characterized by an exceptionally high content of alanine and glycine residues (Osuna-Mascaró et al. 2015), a compositional bias resembling that of silk and spider-silk fibroins (Marie et al. 2017; Sarashina and Endo 2006). Because proteins enriched in low-complexity and hydrophobic sequences are known to self-assemble into insoluble fibrous structures through hydrophobic interactions (Furuhashi et al. 2009; Liu et al. 2017), These features support the idea that MSI60 family proteins are major candidates for components contributing to the core of the interlamellar membrane.

In the present dataset, we additionally identified two proteins with MSI60-like sequence characteristics, v4.1_scaf_13a.g31191.t1 and v4.1_scaf_13a.g30686.t1. These proteins share the characteristic features of MSI60, namely an overall alanine/glycine-rich sequence with inserted aspartic acid-rich acidic regions. Among them, v4.1_scaf_13a.g31191.t1 is similar to conventional MSI60, whereas v4.1_scaf_13a.g30686.t1 was found to contain an N-terminal region showing high structural similarity to the chitin-binding fold of human chitotriosidase CHIT1 (Fadel et al. 2016) (Fig. S8; Fig. S9A). We therefore refer to this protein as chitin-binding MSI60 (CB-MSI60). Its predicted chitin-binding region suggests that CB-MSI60 may mediate interactions between protein aggregates and β-chitin fibers within the interlamellar membrane. Furthermore, the chitin-binding region of CB-MSI60 also showed similarity to the chitin-binding region of Pif-family proteins (Suzuki et al. 2013) (Fig. S9B, C), raising the possibility that Pif and MSI60 either diverged from a common ancestral module or convergently acquired similar chitin-binding motifs during the evolution of nacre formation.

Another shared feature of the MSI60 family is the presence of inserted aspartic acid-rich acidic regions (Fig. S10). Highly acidic peptides have long been considered to participate in biomineralization through interactions with Ca^{2+} (Rivera-Perez et al. 2020; Verch et al. 2011; Wildt et al. 2022; Schuitemaker et al. 2024). From this perspective, the acidic segments of MSI60 family proteins may influence the local ionic environment at

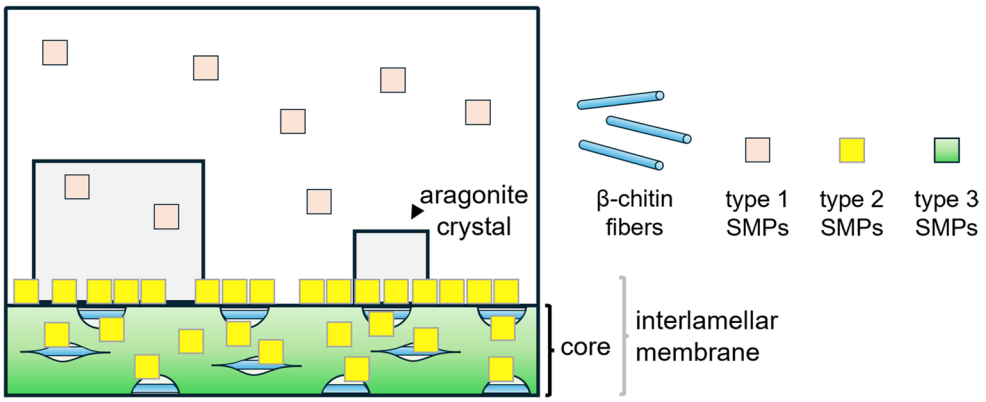

Fig. 7 Proposed extraction-defined three-compartment working model of nacreous layer formation. Type 3 SMPs are interpreted as insoluble core-associated components of the interlamellar membrane, together with β-chitin fibers. Type 2 SMPs are interpreted as relatively extractable components associated with the interlamellar membrane surface and the mineral–organic interface. Type 1 SMPs are interpreted as soluble matrix components potentially involved in regulating the mineralization environment. The exact ultrastructural positional relationship between β-chitin fibers and type 3 SMPs is not directly resolved by the present data. The arrangement shown here is therefore intended as a provisional working interpretation, drawn with reference to the ultrastructural model proposed by Osuna-Mascaró et al. (2015) for the nacre of *Pteria hirundo*. In that model, a chitin is embedded in a protein matrix

the interlamellar membrane during crystal growth by imparting a net negative charge to the membrane core and providing potential binding sites for cations such as Ca^{2+}.

Following the MSI60 family, shematrin was also a major component of the type 3 insoluble framework category. Shematrin is characterized by hydrophobic low-complexity segments and often contains glycine/tyrosine-rich motifs (Yano et al. 2006). These features are consistent with assembly into robust insoluble organic networks observed in mollusks (Chen et al. 2018). In the present dataset, shematrin was classified predominantly as type 3, supporting the view that shematrin is a major candidate component of the insoluble core of the interlamellar membrane, together with β-chitin and MSI60 family proteins.

Because tyrosinase was enriched in type 2, it may act as a relatively extractable, surface-associated enzyme rather than as a component deeply embedded within the insoluble core of the interlamellar membrane. Tyrosinase catalyzes the oxidation of phenolic residues and can generate quinone intermediates that promote covalent cross-linking among tyrosine-containing proteins (Miglioli et al. 2019). This reaction is known to contribute to hardening and stabilization of organic layers in invertebrate biomaterials (Miserez et al. 2010). In the nacreous layer, such oxidative cross-linking may be particularly important for glycine/tyrosine-rich proteins, including shematrin-like framework components (Kinoshita et al. 2011). Thus, tyrosinase may participate in oxidative cross-linking of tyrosine-rich matrix proteins, potentially contributing to stabilization of the organic interface associated with nacre formation. This possibility remains to be tested by direct analyses of tyrosinase activity, cross-linked products, and their spatial distribution in the nacreous layer.

A Working Three-Compartment Model of Nacre Formation

Based on these findings, our solubility-based type classification supports an extraction-defined three-compartment framework for the nacreous layer: (i) type 3 SMPs are interpreted as core-associated insoluble components of the interlamellar membrane together with β-chitin fibers; (ii) type 2 SMPs are interpreted as relatively extractable components associated with the membrane surface and organic–mineral interface; and (iii) type 1 SMPs are interpreted as soluble matrix components potentially involved in ionic regulation and protein maturation (Fig. 7).

By contrast, type 4 proteins were not incorporated as a distinct component in the present model, because they showed mixed or ambiguous partitioning patterns rather than a clear predominant association with a single extraction-defined compartment. In addition, such patterns may partly reflect differences in apparent solubility associated with proteolytic processing, cross-linking, or interactions with other matrix components. Because only 12 of the 327 identified SMPs were classified as type 4, and all were of very low abundance, this group may also include proteins whose quantitative ratios fell near the classification thresholds.

This three-compartment model provides a working hypothesis for understanding the spatial organization and division of labor of SMPs during nacre formation. However, because this model is based on extraction behavior and quantitative partitioning, it does not directly demonstrate the in situ localization or molecular activity of individual

SMPs. Future studies combining localization studies with detailed functional characterization are expected to connect the extraction-based framework established here to a more precise mechanistic understanding of nacreous layer formation.

Conclusion

In shell biomineralization, SMPs have long been discussed within a qualitative binary framework consisting of soluble "intracrystalline" components and insoluble "framework" components (Suzuki and Nagasawa 2013; Paleček et al. 2024). Although this distinction has been useful as a first approximation, substantial overlap between these pools indicates that SMPs are distributed more continuously across extraction fractions than this simple classification implies (Bédouet et al. 2007; Mann et al. 2012). Consequently, qualitative labels alone are insufficient to describe the predominant fraction association and quantitative significance of individual SMPs.

In the present study, nacre SMPs were clearly separated into three fractions, and a solubility-based type classification was introduced on the basis of their quantitative partitioning among fractions. This framework quantitatively refines the conventional soluble/insoluble classification and provides a reproducible way to relate biochemical extractability to inferred function. Furthermore, the quantitative analysis of SMPs may provide a new organizing axis for comparative study across molluscan shell proteomes. Comparisons of SMPs among species have often relied on the presence or absence of particular domains or protein families. By contrast, the present framework extends such comparisons to the quantitative evaluation of the relative abundance of those molecular groups within a shell proteome. Thus, it may improve the resolution of comparative shell proteomics and provide a basis for understanding the molecular systems underlying distinct shell microstructures.

It should be emphasized, however, that the solubility-based type classification and the three-compartment model derived from it do not directly measure the actual spatial localization of SMPs, but rather represent hypotheses based on extraction behavior and inferred function. Because the spatial interpretation proposed here is based on the integration of inferred functions and previous findings, the spatial validity of the three-compartment model will need to be tested in future studies that directly resolve in situ localization.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10126-026-10631-4.

Author Contributions K.O. and M.S. conceived the study. T.A. provided the live P. fucata specimens and gave advice of sampling. K.O. took scanning electron microscopy images, isolated proteins, carried out LC-MS/MS sample preparation, and analysis of LC-MS/MS data. L.N. and H.K. supplied the technology of LC-MS/MS measurement. K.O. and M.S. prepared the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding Open Access funding provided by The University of Tokyo. This research was partially funded by two Grants-in-Aid for Scientific Research B and A (JP19H03045 to MS and JP23H00339 to MS) and Scientific Research on Innovative Areas (JP19H05771 to MS, JP25H01449 to MS, JP25H01447 to MS), This study was partially based on results obtained from project JPNP22100161-0, which was subsidized by the New Energy and Industrial Technology Development Organization (NEDO) (to MS). This study was partially funded by the Environment Research and Technology Development Fund (grant number CN2201) of the Environmental Restoration and Conservation Agency of Japan (to MS). This work was supported by JST Support for Pioneering Research Initiated by the Next Generation, Grant Number JPMJSP0603 (to KO). This research was supported by Research Support Project for Life Science and Drug Discovery (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP25ama121009 (to HK).

Data Availability The raw LC–MS/MS datasets, the protein sequence database (FASTA file), and the MaxQuant output file proteinGroups.txt used for downstream analysis have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository under the accession number PXD071717. Protein sequences for all accessions beginning with "v4.1_scaf_" mentioned in the main text are freely available from the OIST Marine Genomics Unit Genome Browser (https:/marinegenomics.oist.jp/pearl_4_1A/viewer/info? project_id=110).

Declarations

Competing interests The authors declare no competing interests.

References

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung CC, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Žídek A, Bapst V, Kohli P, Jaderberg M, Hass abis D, Jumper JM

(2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630(8016):493–500. https://doi.org/10.1038/s41586-024-07487-w

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389

Alves G, Ogurtsov AY, Kwok S, Wu WW, Wang G, Shen RF, Yu YK (2008) Detection of co-eluted peptides using database search methods. Biol Direct 3:27. https://doi.org/10.1186/1745-6150-3-27

Bahn SY, Jo BH, Choi YS, Cha HJ (2017) Control of nacre biomineralization by Pif80 in pearl oyster. Sci Adv 3(8):e1700765. https://doi.org/10.1126/sciadv.1700765

Bai Y, Liu S, Hu Y, Yu H, Kong L, Xu C, Li Q (2023) Multi-omic insights into the formation and evolution of a novel shell microstructure in oysters. BMC Biol 21(1):204. https://doi.org/10.1186/s12915-023-01706-y

Barkovits K, Pacharra S, Pfeiffer K, Steinbach S, Eisenacher M, Marcus K, Uszkoreit J (2020) Reproducibility, specificity and accuracy of relative quantification using spectral library-based data-independent acquisition. Mol Cell Proteom 19(1):181–197. https://doi.org/10.1074/mcp.RA119.001714

Batzel GO, Moreno BK, Lopez LS, Nguyen CK, Livingston BT, Joester D, Lyons DC (2022) Proteomic and transcriptomic analyses in the slipper snail *Crepidula fornicata* uncover shell matrix genes expressed during adult and larval biomineralization. Integr Org Biol 4(1):obac023. https://doi.org/10.1093/iob/obac023

Bédouet L, Marie A, Dubost L, Péduzzi J, Duplat D, Berland S, Puisségur M, Boulzaguet H, Rousseau M, Milet C, Lopez E (2007) Proteomics analysis of the nacre soluble and insoluble proteins from the oyster *Pinctada margaritifera*. Mar Biotechnol 9(5):638–649. https://doi.org/10.1007/s10126-007-9017-1

Bruderer R, Sondermann J, Tsou CC, Barrantes-Freer A, Stadelmann C, Nesvizhskii AI, Schmidt M, Reiter L, Gomez-Varela D (2017) New targeted approaches for the quantification of data-independent acquisition mass spectrometry. Proteomics 17(9):1700021. https://doi.org/10.1002/pmic.201700021

Checa AG (2018) Physical and biological determinants of the fabrication of molluscan shell microstructures. Front Mar Sci 5:353. https://doi.org/10.3389/fmars.2018.00353

Chen Y, Gao J, Xie J, Liang J, Zheng G, Xie L, Zhang R (2018) Transcriptional regulation of the matrix protein Shematrin-2 during shell formation in pearl oyster. J Biol Chem 293(46):17803–17816. https://doi.org/10.1074/jbc.RA118.005281

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372. https://doi.org/10.1038/nbt.1511

Cusack M, Freer A (2008) Biomineralization: Elemental and organic influence in carbonate systems. Chem Rev 108(11):4433–4454. https://doi.org/10.1021/cr078270o

Du X, Fan G, Jiao Y, Zhang H, Guo X, Huang R, Zheng Z, Bian C, Deng Y, Wang Q, Wang Z, Liang X, Liang H, Shi C, Zhao X, Sun F, Hao R, Bai J, Liu J, Chen W, Liang J, Liu W, Xu Z, Shi Q, Xu X, Zhang G, Liu X (2017) The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. Gigascience 6(8):1–12. https://doi.org/10.1093/gigascience/gix059

Du YP, Chang HH, Yang SY, Huang SJ, Tsai YJ, Huang JJ, Chan JC (2016) Study of binding interaction between Pif80 protein fragment and aragonite. Sci Rep 6:30883. https://doi.org/10.1038/srep30883

Evans JS (2019) The biomineralization proteome: Protein complexity for a complex bioceramic assembly process. Proteomics 19(16):e1900036. https://doi.org/10.1002/pmic.201900036

Fadel F, Zhao Y, Cousido-Siah A, Ruiz FX, Mitschler A, Podjarny A (2016) X-ray crystal structure of the full length human chitotriosidase (CHIT1) reveals features of its chitin binding domain. PLoS One 11(4):e0154190. https://doi.org/10.1371/journal.pone.0154190

Furuhashi T, Schwarzinger C, Miksik I, Smrz M, Beran A (2009) Molluscan shell evolution with review of shell calcification hypothesis. Comp Biochem Physiol B Biochem Mol Biol 154(3):351–371. https://doi.org/10.1016/j.cbpb.2009.07.011

Gerhard EM, Wang W, Li C, Guo J, Ozbolat IT, Rahn KM, Armstrong AD, Xia J, Qian G, Yang J (2017) Design strategies and applications of nacre-based biomaterials. Acta Biomater 54:21–34. https://doi.org/10.1016/j.actbio.2017.03.003

Gong N, Shangguan J, Liu X, Yan Z, Ma Z, Xie L, Zhang R (2008) Immunolocalization of matrix proteins in nacre lamellae and their in vivo effects on aragonitic tablet growth. J Struct Biol 164(1):33–40. https://doi.org/10.1016/j.jsb.2008.05.009

Griss J, Côté RG, Gerner C, Hermjakob H, Vizcaíno JA (2011) Published and perished? The influence of the searched protein database on the long-term storage of proteomics data. Mol Cell Proteom 10(9): M111.008490. https://doi.org/10.1074/mcp.M111.008490

Guo Q, Li D, Zhai Y, Gu Z (2020) CCPRD: a novel analytical framework for the comprehensive proteomic reference database construction of non-model organisms. ACS Omega 5(25):15370–15384. https://doi.org/10.1021/acsomega.0c01278

Hanif MA, Han JD, Kim SC, Hossen S, Kho KH (2023) EF-Hand-binding secreted protein Hdh-SMP5 regulates shell biomineralization and responses to stress in Pacific Abalone, *Haliotis discus hannai*. Curr Issues Mol Biol 45(12):10079–10096. https://doi.org/10.3390/cimb45120629

Heck M, Neely BA (2020) Proteomics in non-model organisms: a new analytical frontier. J Proteome Res 19(9):3595–3606. https://doi.org/10.1021/acs.jproteome.0c00448

Heinemann B, Künzler P, Eubel H, Braun HP, Hildebrandt TM (2021) Estimating the number of protein molecules in a plant cell: protein and amino acid homeostasis during drought. Plant Physiol 185(2):385–404. https://doi.org/10.1093/plphys/kiaa050

Holm L (2022) Dali server: structural unification of protein families. Nucleic Acids Res 50(W1):W210–W215. https://doi.org/10.1093/nar/gkac387

Hu A, Noble WS, Wolf-Yadlin A (2016) Technical advances in proteomics: new developments in data-independent acquisition. F1000Res. 5:F1000 Faculty Rev–419. https://doi.org/10.12688/f1000research.7042.1

Jin C, Zhang Y, Cheng K, Jiang R, Jiang S, Shi Y, Ren G, Luo W (2023) Hcprismatin-14 of *Hyriopsis cumingii*, a novel matrix protein is crucial for framework recognition and crystal deposition during prismatic layer formation. Front Mar Sci 10:1154968. https://doi.org/10.3389/fmars.2023.1154968

Keene EC, Evans JS, Estroff LA (2010) Matrix interactions in biomineralization: aragonite nucleation by an intrinsically disordered nacre polypeptide, n16N, associated with a β-Chitin substrate. Cryst Growth Des 10:1383–1389. https://doi.org/10.1021/cg901389v

Kinoshita S, Wang N, Inoue H, Maeyama K, Okamoto K, Nagai K, Kondo H, Hirono I, Asakawa S, Watabe S (2011) Deep sequencing of ESTs from nacreous and prismatic layer producing tissues and a screen for novel shell formation-related genes in the pearl oyster. PLoS One 6(6):e21238. https://doi.org/10.1371/journal.pone.0021238

Kintsu H, Nishimura R, Negishi L, Kuriyama I, Tsuchihashi Y, Zhu L, Nagata K, Suzuki M (2020) Identification of methionine-rich insoluble proteins in the shell of the pearl oyster, *Pinctada fucata*. Sci Rep 10:18335. https://doi.org/10.1038/s41598-020-75444-4

Kong J, Liu C, Yang D, Yan Y, Chen Y, Liu Y, Zheng G, Xie L, Zhang R (2019) A novel basic matrix protein of *Pinctada fucata*, PNU9, functions as inhibitor during crystallization of aragonite. CrystEngComm 21:1250–1261. https://doi.org/10.1039/C8CE02194E

Lao Y, Zhang X, Zhou J, Su W, Chen R, Wang Y, Zhou W, Xu ZF (2007) Characterization and in vitro mineralization function of a soluble protein complex P60 from the nacre of *Pinctada fucata*. Comp Biochem Physiol B Biochem Mol Biol 148(2):201–208. https://doi.org/10.1016/j.cbpb.2007.05.010

Li J, Li J, Zhao WG, Sun HD, Guo ZG, Liu XY, Tang XY, She ZF, Yuan T, Liu SN (2020) Comprehensive proteomics and functional annotation of mouse brown adipose tissue. PLoS One 15(12):e0232084. https://doi.org/10.1371/journal.pone.0232084

Li J, Smith LS, Zhu HJ (2021) Data-independent acquisition (DIA): an emerging proteomics technology for analysis of drug-metabolizing enzymes and transporters. Drug Discov Today Technol 39:49–56. https://doi.org/10.1016/j.ddtec.2021.06.006

Liang J, Xu G, Xie J, Lee I, Xiang L, Wang H, Zhang G, Xie L, Zhang R (2015) Dual roles of the lysine-rich matrix protein (KRMP)-3 in shell formation of pearl oyster, *Pinctada fucata*. PLoS One 10(7):e0131868. https://doi.org/10.1371/journal.pone.0131868

Liu C, Li S, Kong J, Liu Y, Wang T, Xie L, Zhang R (2015) In-depth proteomic analysis of shell matrix proteins of *Pinctada fucata*. Sci Rep 5:17269. https://doi.org/10.1038/srep17269

Liu C, Yuan Y, Zhang W, Huang J (2024) Proteomic analysis of shell matrix proteins from the chiton *Acanthopleura loochooana*. Comp Biochem Physiol D Genomics Proteomics 49:101176. https://doi.org/10.1016/j.cbd.2023.101176

Liu C, Zhang R (2021) Biomineral proteomics: a tool for multiple disciplinary studies. J Proteomics 238(30):104171. https://doi.org/10.1016/j.jprot.2021.104171

Liu X, Jin C, Wu L, Dong S, Zeng S, Li J (2017) Hic74, a novel alanine and glycine rich matrix protein related to nacreous layer formation in the mollusc *Hyriopsis cumingii*. Aquac Fish 2(3):119–123. https://doi.org/10.1016/j.aaf.2017.03.005

Mann K, Edsinger E (2014) The *Lottia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. Proteome Sci 12:28. https://doi.org/10.1186/1477-5956-12-28

Mann K, Edsinger-Gonzales E, Mann M (2012) In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*. Proteome Sci 10(1):28. https://doi.org/10.1186/1477-5956-10-28

Marie B, Arivalagan J, Mathéron L, Bolbach G, Berland S, Marie A, Marin F (2017) Deep conservation of bivalve nacre proteins highlighted by shell matrix proteomics of the Unionoida *Elliptio complanata* and *Villosa lienosa*. J R Soc Interface 14(126):20160846. https://doi.org/10.1098/rsif.2016.0846

Marin F, Le Roy N, Marie B (2012) The formation and mineralization of mollusk shell. Front Biosci 4(3):1099–125. https://doi.org/10.2741/s321

Mehta D, Scandola S, Uhrig RG (2022) BoxCar and library-free data-independent acquisition substantially improve the depth, range, and completeness of label-free quantitative proteomics. Anal Chem 94(2):793–802. https://doi.org/10.1021/acs.analchem.1c03338

Miglioli A, Dumollard R, Balbi T, Besnardeau L, Canesi L (2019) Characterization of the main steps in first shell formation in *Mytilus galloprovincialis*: possible role of tyrosinase. Proc Biol Sci 286(1916):20192043. https://doi.org/10.1098/rspb.2019.2043

Miserez A, Rubin D, Waite JH (2010) Cross-linking chemistry of squid beak. J Biol Chem 285(49):38115–38124. https://doi.org/10.1074/jbc.M110.161174

Miyamoto H, Miyashita T, Okushima M, Nakano S, Morita T, Matsushiro A (1996) A carbonic anhydrase from the nacreous layer in oyster pearls. Proc Natl Acad Sci U S A 93(18):9657–9660. https://doi.org/10.1073/pnas.93.18.9657

Miyamoto H, Miyoshi F, Kohno J (2005) The carbonic anhydrase domain protein nacrein is expressed in the epithelial cells of the mantle and acts as a negative regulator in calcification in the mollusc *Pinctada fucata*. Zool Sci 22(3):311–315. https://doi.org/10.2108/zsj.22.311

Miyashita T, Takagi R (2011) Tyrosinase causes the blue shade of an abnormal pearl. J Molluscan Stud 77(3):312–314. https://doi.org/10.1093/mollus/eyr013

Muntel J, Xuan Y, Berger ST, Reiter L, Bachur R, Kentsis A, Steen H (2015) Advancing urinary protein biomarker discovery by data-independent acquisition on a quadrupole-Orbitrap mass spectrometer. J Proteome Res 14(11):4752–4762. https://doi.org/10.1021/acs.jproteome.5b00826

Namikawa Y, Zhu L, Lu P, Nagata K, Suzuki M (2025) Calcium dissociation with carbonate ions from Pf-SCP, sarcoplasmic calcium-binding protein in *Pinctada fucata*, contributes to calcium mineralization for shell formation. Protein Sci 34(11):e70336. https://doi.org/10.1002/pro.70336

Nielsen H, Tsirigos KD, Brunak S, von Heijne G (2019) A brief history of protein sorting prediction. Protein J 38:200–216. https://doi.org/10.1007/s10930-019-09838-3

Nudelman F, Gotliv BA, Addadi L, Weiner S (2006) Mollusk shell formation: mapping the distribution of organic matrix components underlying a single aragonitic tablet in nacre. J Struct Biol 153(2):176–187. https://doi.org/10.1016/j.jsb.2005.09.009

Osuna-Mascaró AJ, Cruz-Bustos T, Marin F, Checa AG (2015) Ultrastructure of the interlamellar membranes of the nacre of the bivalve *Pteria hirundo*, determined by immunolabelling. PLoS One 10(4):e0122934. https://doi.org/10.1371/journal.pone.0122934

Otter LM, Eder K, Kilburn MR, Yang L, O'Reilly P, Nowak DB, Cairney JM, Jacob DE (2023) Growth dynamics and amorphous-to-crystalline phase transformation in natural nacre. Nat Commun 14(1):2254. https://doi.org/10.1038/s41467-023-37814-0

Paleček D, Milano S, Gutiérrez-Zugasti I, Talamo S (2024) Stable isotopes in the shell organic matrix for (paleo)environmental reconstructions. Commun Chem 7(1):16. https://doi.org/10.1038/s42004-023-01076-0

R Development Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://doi.org/10.32614/R.manuals

Rivera-Perez C, Flores-Sánchez IA, Ojeda Ramírez de Areyano JJ, Rojas Posadas DI, Hernández-Saavedra NY (2020) A shell matrix protein of *Pinctada mazatlanica* produces nacre platelets *in vitro*. Sci Rep 10(1):20201. https://doi.org/10.1038/s41598-020-77320-7

Samata T, Hayashi N, Kono M, Hasegawa K, Horita C, Akera S (1999) A new matrix protein family related to the nacreous layer formation of *Pinctada fucata*. FEBS Lett 462(1–2):225–229. https://doi.org/10.1016/S0014-5793(99)01387-3

Sarashina I, Endo K (2006) Skeletal matrix proteins of invertebrate animals: comparative analysis of their amino acid sequences. Paleontol Res 10(4):311–336. https://doi.org/10.2517/prpsj.10.311

Saruwatari K, Matsui T, Mukai H, Nagasawa H, Kogure T (2009) Nucleation and growth of aragonite crystals at the growth front of nacres in pearl oyster, *Pinctada fucata*. Biomaterials 30(16):3028–3034. https://doi.org/10.1016/j.biomaterials.2009.03.011

Schuitemaker A, Koziara KB, Raiteri P, Gale JD, Demichelis R (2024) New model for aspartic acid species in aqueous calcium carbonate growth environments: challenges and perspectives. Phys Chem Chem Phys 26(6):4909–4921. https://doi.org/10.1039/d3cp04674e

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian

gene expression control. Nature 473:337–342. https://doi.org/10.1038/nature11848
Shin JB, Krey JF, Hassan A, Metlagel Z, Tauscher AN, Pagana JM, Sherman NE, Jeffery ED, Spinelli KJ, Zhao H, Wilmarth PA, Choi D, David LL, Auer M, Barr-Gillespie PG (2013) Molecular architecture of the chick vestibular hair bundle. Nat Neurosci 16:365–374. https://doi.org/10.1038/nn.3312
Sudo S, Fujikawa T, Nagakura T, Ohkubo T, Sakaguchi K, Tanaka M, Nakashima K, Takahashi T (1997) Structures of mollusc shell framework proteins. Nature 387:563–564. https://doi.org/10.1038/42391
Suzuki M, Iwashima A, Kimura M, Kogure T, Nagasawa H (2013) The molecular evolution of the Pif family proteins in various species of mollusks. Mar Biotechnol 15:145–158. https://doi.org/10.1007/s10126-012-9471-2
Suzuki M, Nagasawa H (2013) Mollusk shell structures and their formation mechanism. Can J Zool 91(6):349–366. https://doi.org/10.1139/cjz-2012-0333
Suzuki M, Saruwatari K, Kogure T, Yamamoto Y, Nishimura T, Kato T, Nagasawa H (2009) An acidic matrix protein, Pif, is a key macromolecule for nacre formation. Science 325(5946):1388–1390. https://doi.org/10.1126/science.1173793
Takeuchi T, Kawashima T, Koyanagi R, Gyoja F, Tanaka M, Ikuta T, Shoguchi E, Fujiwara M, Shinzato C, Hisata K, Fujie M, Usami T, Nagai K, Maeyama K, Okamoto K, Aoki H, Ishikawa T, Masaoka T, Fujiwara A, Endo K, Endo H, Nagasawa H, Kinoshita S, Asakawa S, Watabe S, Satoh N (2012) Draft genome of the pearl oyster *Pinctada fucata*: a platform for understanding bivalve biology. DNA Res 19(2):117–130. https://doi.org/10.1093/dnares/dss005
Takeuchi T, Koyanagi R, Gyoja F, Kanda M, Hisata K, Fujie M, Goto H, Yamasaki S, Nagai K, Morino Y, Miyamoto H, Endo K, Endo H, Nagasawa H, Kinoshita S, Asakawa S, Watabe S, Satoh N, Kawashima T (2016) Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. Zool Lett 2:3. https://doi.org/10.1186/s40851-016-0039-2
Takeuchi T, Suzuki Y, Watabe S, Nagai K, Masaoka T, Fujie M, Kawamitsu M, Satoh N, Myers EW (2022) A high-quality, haplotype-phased genome reconstruction reveals unexpected haplotype diversity in a pearl oyster. DNA Res 29(6):dsac035. https://doi.org/10.1093/dnares/dsac035
Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673–4680. https://doi.org/10.1093/nar/22.22.4673
Tsou CC, Tsai CF, Teo GC, Chen YJ, Nesvizhskii AI (2016) Untargeted, spectral library-free analysis of data-independent acquisition proteomics data generated using Orbitrap mass spectrometers. Proteomics 16(15–16):2257–2271. https://doi.org/10.1002/pmic.201500526
Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 13(9):731–740. https://doi.org/10.1038/nmeth.3901
Venable JD, Dong MQ, Wohlschlegel J, Dillin A, Yates JR (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nat Methods 1:39–45. https://doi.org/10.1038/nmeth705
Verch A, Gebauer D, Antonietti M, Cölfen H (2011) How to control the scaling of CaCO3: a fingerprinting technique to classify additives. Phys Chem Chem Phys 13(37):16811–16820. https://doi.org/10.1039/c1cp21328h
Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A (2023) The conserved domain database in 2023. Nucleic Acids Res 51(D1):D384–D388. https://doi.org/10.1093/nar/gkac1096
Wildt BWM, van der Meijden R, Bartels PAA, Sommerdijk NAJM, Akiva A, Ito K, Hofmann S (2022) Bioinspired silk fibroin mineralization for advanced in vitro bone remodeling models. Adv Funct Mater 32:2206992. https://doi.org/10.1002/adfm.202206992
Wong WC, Kwan YH, He X, Chen C, Xiang S, Xiao Y, Long L, Gao K, Wang N, Wu L, Qian PY, Sun J (2025) Proteomic analyses reveal the key role of gene co-option in the evolution of the scaly-foot snail scleritome. Commun Biol 8:337. https://doi.org/10.1038/s42003-025-07785-7
Xu Y, Galloway JM, Hasselt LJ, Meldrum FC (2025) The role of confinement in biomineralization. Chem Rev 125(24):12128–12197. https://doi.org/10.1021/acs.chemrev.5c00659
Yano M, Nagai K, Morimoto K, Miyamoto H (2006) Shematrin: a family of glycine-rich structural proteins in the shell of the pearl oyster *Pinctada fucata*. Comp Biochem Physiol B Biochem Mol Biol 144(2):254–262. https://doi.org/10.1016/j.cbpb.2006.03.004
Yano M, Nagai K, Morimoto K, Miyamoto H (2007) A novel nacre protein N19 in the pearl oyster *Pinctada fucata*. Biochem Biophys Res Commun 362(1):158–163. https://doi.org/10.1016/j.bbrc.2007.07.172
Yarra T, Blaxter M, Clark MS (2021) A bivalve biomineralization toolbox. Mol Biol Evol 38(9):4043–4055. https://doi.org/10.1093/molbev/msab153
Zheng M, Evdokimov AG, Moshiri F, Lowder C, Haas J (2020) Crystal structure of a Vip3B family insecticidal protein reveals a new fold and a unique tetrameric assembly. Protein Sci 29(4):824–829. https://doi.org/10.1002/pro.3803
Zhu M, Mu H, Dai X (2024) Integrated control of bacterial growth and stress response by (p)ppGpp in *Escherichia coli*: a seesaw fashion. iScience 27(2):108818. https://doi.org/10.1016/j.isci.2024.108818
Zou X, Wang L, Chen Y, Fu H, Gao Y, Liu B, Tan M, Zhai L (2025) In-depth analysis of data characteristics and comparative evaluation of DDA and DIA accuracy in label-free quantitative proteomics of biological samples. Clin Proteomics 23(1):2. https://doi.org/10.1186/s12014-025-09572-2