Novel molluskan biomineralization proteins retrieved from proteomics: a case study with Upsalin

Ramos-Silva, P.; Benhamada, S.; Le Roy, N.; Marie, B.; Guichard, N.; Zannella-Cléon, I.; Plasseraud, L.; Corneillat, M.; Alcaraz, G.; Kaandorp, J.; Marin, F.

DOI
10.1002/cbic.201100708

Publication date
2012

Document Version
Final published version

Published in
ChemBioChem

DOI: 10.1002/cbic.201100708

Novel Molluskan Biomineralization Proteins Retrieved from Proteomics: A Case Study with Upsalin

Paula Ramos-Silva,*[a, b] Sana Benhamada,[a] Nathalie Le Roy,[a] Benjamin Marie,[a, c] Nathalie Guichard,[a] Isabelle Zanella-Cléon,[d] Laurent Plasseraud,[e] Marion Corneillat,[f] Gérard Alcaraz,[f] Jaap Kaandorp,[b] and Frédéric Marin*[a]

The formation of the molluskan shell is regulated by an array of extracellular proteins secreted by the calcifying epithelial cells of the mantle. These proteins remain occluded within the recently formed biominerals. To date, many shell proteins have been retrieved, but only a few of them, such as nacreins, have clearly identified functions. In this particular case, by combining molecular biology and biochemical approaches, we performed the molecular characterization of a novel protein that we named Upsalin, associated with the nacreous shell of the freshwater mussel *Unio pictorum*. The full sequence of the upsalin transcript was obtained by RT-PCR and 5′/3′ RACE, and the expression pattern of the transcript was studied by PCR and qPCR. Upsalin is a 12 kDa protein with a basic theoretical pI. The presence of Upsalin in the shell was demonstrated by extraction of the acetic-acid-soluble nacre matrix, purification of a shell protein fraction by mono-dimensional preparative SDS-PAGE, and by submitting this fraction, after trypsic digestion, to nano-LC-MS/MS. In vitro experiments with the purified protein showed that it interferes poorly with the precipitation of calcium carbonate. Homology searches also could not affiliate Upsalin to any other protein of known function, leaving open the question of its exact role in shell formation. An antibody raised against an immunogenic peptide of Upsalin was found to be specific to this protein and was subsequently assayed for immunogold localization of the target protein in the shell, revealing the ubiquitous presence of Upsalin in the nacreous and prismatic layers. Recently, with the application of high-throughput proteomic studies to shells, the number of candidate proteins without clear functions has been increasing exponentially. The Upsalin example highlights the crucial need, for the scientific community dealing with biomineralization in general, to dedicate the coming years to designing experimental approaches, such as gene silencing, that focus on the functions of mineral-associated proteins.

Introduction

Mollusks are known for their ability to synthesize their shells, calcified structures formed outside their living tissues. The mollusk shell, which is mainly composed of $CaCO_3$ crystals, plays an essential role in supporting the soft body and in protecting it against predation and desiccation. The shell formation process is initiated at the early stage of larval development (trochophore) and the shell continues to grow after metamorphosis, throughout the entire life of the animal.[1]

Like other biomineralization processes, shell formation is regulated by an extracellular organic matrix, which is a complex mixture of proteins, glycoproteins, polysaccharides, pigments, and presumably lipids, all secreted by the mantle tissue. This matrix controls crystal formation, nucleation, and growth and becomes occluded within the mineral during its growth. This process results in a stable and well-packed organomineral assembly,[2] in which proteins were recognized early as key components in shell formation[3] and in modulation of calcium carbonate crystal shapes.[4] Consequently, great effort has been put into their identification and the characterization of their primary structure,[1,5–7] in order to elucidate the mechanisms by which proteins interact in shell formation,[8] and to use these proteins to generate tailored composite biomaterials.[9]

Classically, shell proteins are retrieved by dissolving the mineral phase of the shell with weak acid.[10] This extraction gives

[a] *P. Ramos-Silva, S. Benhamada, Dr. N. Le Roy, Dr. B. Marie, N. Guichard, Dr. F. Marin*
UMR CNRS 6282 (ex 5561) CNRS Biogéosciences, Université de Bourgogne
6 boulevard Gabriel, 21000 Dijon (France)
E-mail: p.sequeiradosramossilva@uva.nl
frederic.marin@u-bourgogne.fr

[b] *P. Ramos-Silva, Dr. J. Kaandorp*
Section Computational Science, Faculty of Science
University of Amsterdam
Science Park 904, 1098 XH Amsterdam (The Netherlands)

[c] *Dr. B. Marie*
UMR 7245, CNRS, MCAM, Muséum National d'Histoire Naturelle (MNHN)
75005 Paris (France)

[d] *I. Zanella-Cléon*
Service de Spectrométrie de Masse, FR3302, UMS3444/US8
Institut de Biologie et Chimie des Protéines (IBCP), Université Lyon 1
7 Passage du Vercors, 69367 Lyon (France)

[e] *Dr. L. Plasseraud*
ICMUB, UMR CNRS 5260, Faculté des Sciences Mirande
Université de Bourgogne
21000 Dijon (France)

[f] *M. Corneillat, Prof. G. Alcaraz*
UPSP PROXISS, Département Agronomie Environnement, AgroSupDijon
26 boulevard Dr. Petitjean, B.P. 87999, 21079 Dijon Cedex (France)

rise to two organic fractions: the soluble matrix and the insoluble one.[1,10] Proteins of the soluble matrix can be further fractionated by electrophoresis or HPLC, and subjected to sequence analysis through mass spectrometry[11] or Edman degradation.[12] In order to obtain the full-length primary structures, degenerate primers are designed from partial protein sequences for amplification of shell protein-encoding transcripts, extracted from mantle tissues. This one-to-one protein approach enriched information on the primary structures of many shell proteins in the last decade.[1] Today, not all shell proteins have the same status: some of them have been fully characterized at both transcriptional and protein levels, and are firmly established as shell proteins (typical examples include prismalin-14[13] and pearlin[14]). Some proteins are true shell proteins, but because they have been sequenced directly, information at the transcript level is not available.[11,12] Conversely, some proteins have been characterized only at the transcript level, with demonstration of their presence in the shell remaining to be achieved.[15]

Lately, the introduction of high-throughput approaches to mollusks is changing this picture. Although genomic resources on mollusks are still scarce,[16] the last three years have seen the development of molluskan mantle transcriptome projects. The generated EST datasets were recently combined with proteomic data from shell extracts,[17] enabling the identification of biomineralization-related proteins at a record rate. So far, this methodology has been applied to a number of molluskan groups: clams,[18] mussels,[19] oysters,[17,20] and abalone.[21]

Thanks to proteomics, it has been possible to identify new homologous proteins belonging to the set of sequences that exhibit conserved domains such as N66 and nacrein-like homologous sequences with carbonic anhydrase domains.[17,19] Papilin homologues with Kunitz-like domains[21] or even perlucin, with a C-type lectin domain,[19] are just a few examples. In addition, proteins characterized by the predominance of one or two amino acids or by repeated short motifs in a domain have been identified, corroborating earlier studies on single proteins. Among the most popular examples are the aspartic acid-rich proteins, such as aspein,[17] which interact strongly with calcium carbonate crystals in vitro[22] and are usually considered high-capacity, low-affinity calcium-binding proteins.[23] Another example is to be found in proteins with basic domains, such as the KRMP and the glycine-rich shematrins,[17] which might play a role by interacting with negatively charged bicarbonate ions, or by anchoring acidic proteins.[15,24] In addition to the identification of homologous proteins, proteomics on shell extracts has revealed a set of novel unknown proteins for which structure–function relations are far from elucidation. Examples include MUSPs-2 and -3 (edible mussel),[19] MRNP34 (pearl oyster),[25] and IMSPs-1, -2, and -3 (manila clam).[18] The growing number of these "orphan shell proteins", without known functions, highlights the need for their complete characterization, in order to understand their roles in biomineralization. In this context, we report the characterization—both at transcriptional and protein level—of a novel biomineralization protein that does not exhibit any homology with previous known shell proteins. This protein, called Upsalin, is 125 residues long (≈12 kDa) and was retrieved from mantle tissues of the freshwater mussel *Unio pictorum*. The cDNA from Upsalin was identified by a classical molecular biology approach based on degenerate primers derived from sequences of peptides described in earlier proteomic work.[26] The occurrence of Upsalin in the shell was confirmed by proteomics, and a specific antibody showed that Upsalin was associated with the two shell layers of *U. pictorum*: prisms and nacre. In spite of several in vitro and functional characterization methods, we could not assign a clear function to Upsalin, in relation to shell biomineralization, and homology searches only revealed the tip of the iceberg. The example of Upsalin emphasizes the need for study of "orphan proteins" in order to identify functions in biomineralization not yet listed in previous works.

Results

Characterization of *Unio pictorum* shell

Unio pictorum belongs to the order Unionoida, one of the two orders of the subclass Palaeoheterodonta, in the Bivalvia class. The shell of *U. pictorum* is made of two fully mineralized layers of calcium carbonate (Figure 1 B). The outer mineralized layer is composed of prisms that develop perpendicularly to the outer shell surface. The internal layer is based on superimpositions of extremely thin flat tablets, which typifies the nacreous layer (Figure 1 C). The tablets are arranged in the "brickwall microstructure" pattern, a feature shared by most bivalvian nacres. Unlike in pteriomorphid bivalves, both nacre and prismatic layers of *U. pictorum* are fully aragonitic. On the external surface, a leathery olive-greenish thin organic sheet, the periostracum, covers the prismatic layer (Figure 1 A).

Identification of a nucleotide sequence coding for a 12 kDa protein

In order to amplify a short sequence from the nacre of *U. pictorum*, a set of 16 degenerate primers was designed. These were

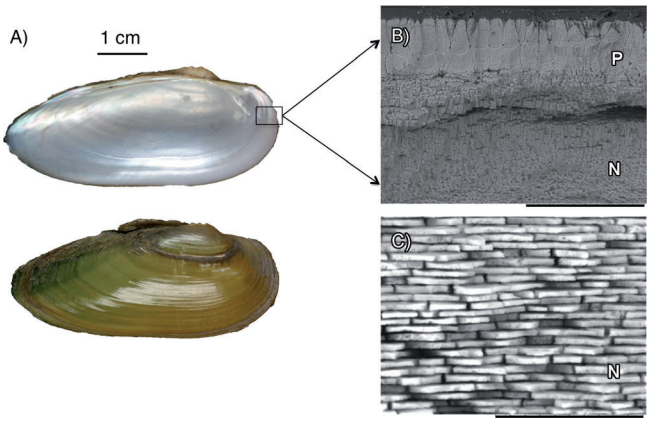


Figure 1. Scanning electron microscopy images showing the microstructure of the shell of *U. pictorum*. A) Inner and outer views of the shell. B) Shell broken in the transversal plane. P: prismatic layer; N: nacreous layer; scale bar = 200 μm. C) Closer view of the nacreous layer when broken in the transversal plane; scale bar = 20 μm.

based on eight peptide sequences corresponding to the 12 and 16 kDa protein bands previously analyzed by mass spectrometry.[26] One pair of primers corresponding to the peptides ECSDPEV-fwd and EFYGVCCPLR-rev gave a specific amplification product of 250 bp (Figure 2). Because the translation product of the sequenced fragment was consistent with four of the peptides identified for the 12 kDa protein band, the full-length sequence coding for this protein was further investigated. To this end, specific primers were designed from the amplified fragment, to provide the 5′- and 3′-ends by RACE PCR. The overlapping of the two amplified sides corresponded to a transcript sequence of 524 bp, which was confirmed by amplifying the full transcript with two primers situated at the 5′- and 3′-ends. This sequence consists of a start codon, a coding sequence of 375 bp, a TAA stop codon, the polyadenylation signal AATAAA, and the poly-A tail, 127 bp downstream of the stop codon (Figure 3 A).

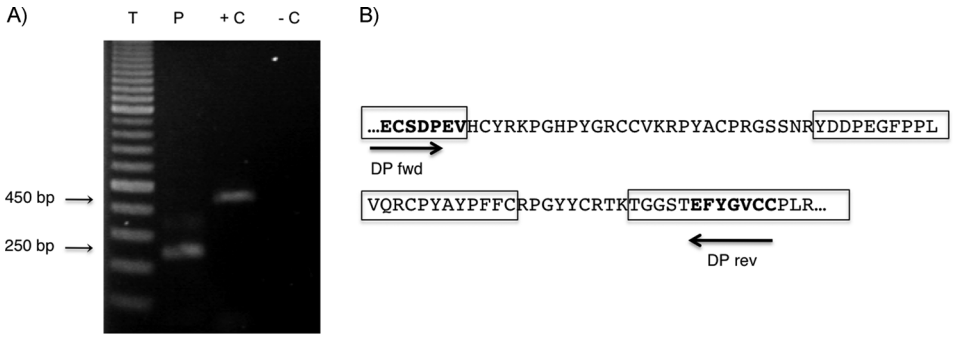


Figure 2. Amplification of a DNA fragment encoding for a shell protein. A) Electrophoresis in agarose gel (1.5%). T: DNA markers. P: 250 bp Upsalin amplified fragment. +C: positive control [amplified cDNA of *U. pictorum* β-actin (450 bp)]. −C: negative control. B) Predicted amino acid sequence for the 250 bp nucleotide sequence. Amino acids in bold were used as references to design the two degenerate primers: forward (DP fwd) and reverse (DP rev). Amino acids inside boxes correspond to the peptides previously identified from MS/MS spectra.[26]

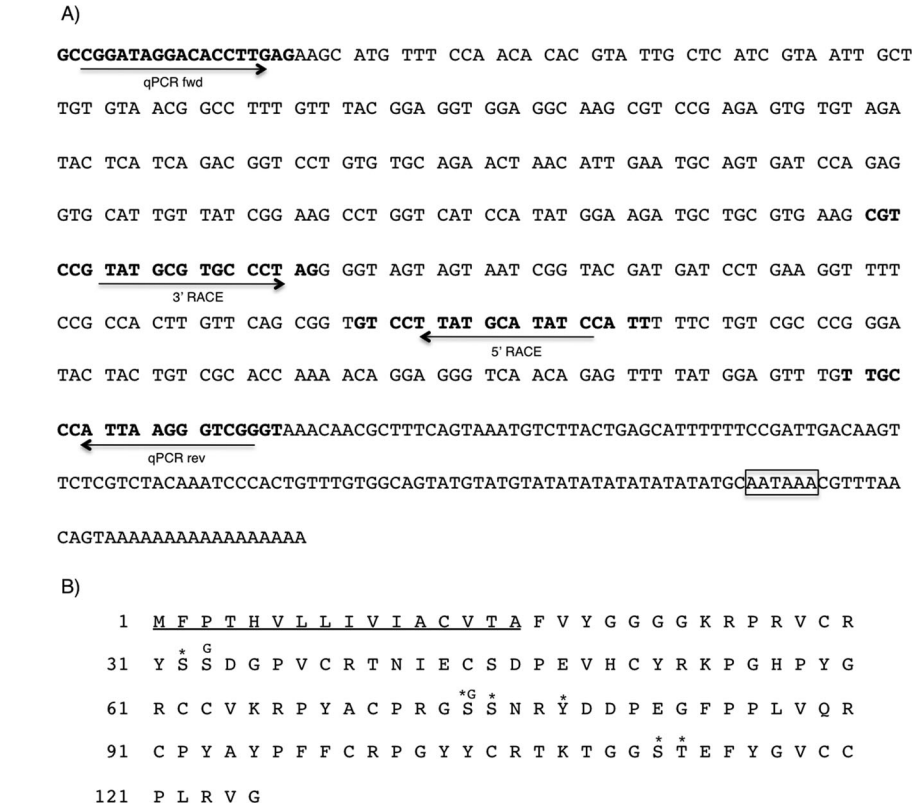


Figure 3. Nucleotide and predicted amino acid sequences of Upsalin precursor. A) Nucleotide sequence with the forward and reverse primers in bold and the polyadenylation signal boxed. B) Translated coding sequence region of Upsalin. The putative peptide signal is underlined and PTMs are highlighted with (*) for phosphorylations and (^{G}) for glycosylations.

Primary structure and molecular features of Upsalin

The open reading frame encodes a protein of 125 residues that we have named Upsalin (Figure 3 B). Upsalin has a predicted signal peptide with a cleavage site between the positions 16 and 17 and a putative transmembrane region (positions 5–23). The mature form of the protein consists of 109 aa residues and has a theoretical molecular weight of 12.3 kDa. This value is consistent with the 12 kDa band analyzed by MS/MS spectra analysis and from which the 16 degenerate primers were designed.[26] Five amino acid residues—Gly, Pro, Arg, Cys, and Tyr—constitute 59% of the Upsalin sequence, with individual percentages between 10 and 13% (Table 1). The frequencies of acidic amino acids (Asp, Glu) are low: less than 8% in total (Table 1). Consequently, Upsalin has a theoretical pI value of 9.07.

A computer search for putative post-translational modifications (PTMs) based on the NetPhos and YinOYang algorithms was performed. Six putative phosphorylation sites were found: four on Ser (positions 32, 74, 75, 112), one on Thr (position 109), and one on Tyr (position 78; Figure 3 B). Two putative glycosylation sites, based on Ser residues, were also detected at positions 33 and 74 (Figure 3 B). All these potential PTMs are located on the five peptides that were previously identified by

Table 1. Computed parameters for the mature Upsalin protein sequence (after the removal of the signal peptide): amino acid composition [%], isoelectric point (pI), and molecular weight (M_W).

Amino acid	[%]	Amino acid	[%]	Amino acid	[%]
Ala	1.8	**Arg**	**11.9**	Asn	1.8
Asp	3.7	**Cys**	**11.0**	Gln	0.9
Glu	3.7	**Gly**	**12.8**	His	1.8
Ile	0.9	Leu	1.8	Lys	3.7
Met	0.0	Phe	4.6	**Pro**	**12.8**
Ser	5.5	Thr	3.7	Trp	0.0
Tyr	**10.1**	Val	7.3		
pI	9.07	M_W [kDa]	12.3		

tandem mass spectrometry, but which did not actually exhibit PTMs. This finding implies two things: firstly, at least one Upsalin isoform does not bear PTMs in the nacre extract that was used for acquiring mass spectrometry data, and secondly that it cannot be ruled out that other PTM-bearing Upsalin peptides might also exist, but were not identified by our former proteomic analysis.

Tissue-specific gene expression of Upsalin

In order to confirm the full sequence of Upsalin predicted by the RACE approach and to analyze its expression in other tissues, two specific primers from the 3′- and 5′-ends were synthesized and tested on cDNA produced from the mantle, gills, foot, and adductor muscles. Figure 4A shows the results from one of the two specimens tested, after RNA extraction and DNase treatment under reproducible conditions. After PCR, all four products were purified, cloned, and sequenced. In the four amplifications, the nucleotide sequence was coincident with a size of 400 bp, confirming the presence of Upsalin in the four tested tissues of the freshwater mussel. However, much stronger bands appeared for the mantle and the adductor muscles, while the corresponding bands from gills and foot were fainter, suggesting lower expression levels of Upsalin in these tissues. A quantitative analysis of the transcription levels was carried out on three specimens with use of the same set of primers, as shown in Figure 4B. Upsalin expression ratios were normalized with β-actin and the expression in the mantle was used as the control conditions with the relative expression of 1. The qPCR results confirmed what had been observed on gel with standard PCR: transcript levels of the three tested samples were the highest in the mantle, slightly lower in the adductor muscles, but negligible in the gills and foot (Figure 4B). These results suggest that Upsalin is not mantle-specific but might also be involved in other functions beyond shell formation. One should keep in mind, however, that adductor muscles are firmly attached to the shell's internal side, meaning in other words that they are located closely to calcification sites.

Homology search

In order to check whether the sequence of Upsalin was unique or whether it could be affiliated to other proteins or protein families, homology searches were performed with the Upsalin nucleotide sequence and use of the tblastx program from NCBI. The most significant result (*E* value of $2e^{-40}$) was obtained against the EST database of all organisms with a sequence encoding a putative protein of the mantle of the Chinese pearl mussel *Hyriopsis cumingii* (GI: 312832200). Figure 5 shows the optimal alignment for the two sequences, with 94 exact matches (in black). Interestingly, *U. pictorum* and *H. cumingii* both belong to the same small Unionidae family of freshwater mussels in the Palaeoheterodonta subclass. At a lower taxonomic level, *U. pictorum* and *H. cumingii* are classified into different subfamilies: Unioninae and Ambleminae, respectively. This clearly suggests that Upsalin or related proteins could be present in the mantle tissues of other members of the Unionidae family, in particular genera that are closely related to *Unio*.

Blast searches were also performed with the predicted amino acid sequence of Upsalin against the UniprotKB. The

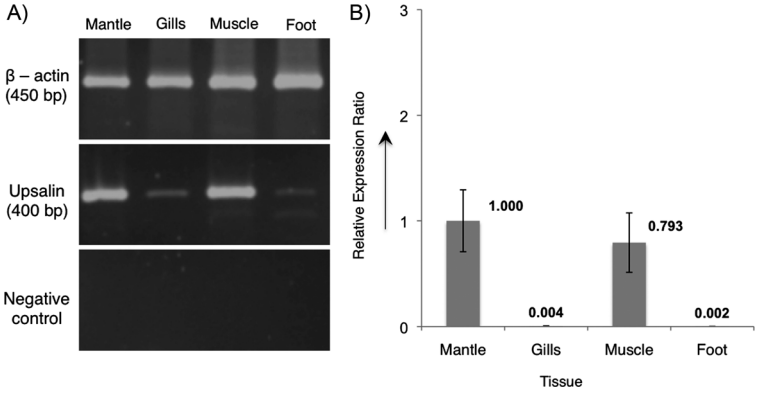


Figure 4. Analysis of the expression of upsalin transcript in four different tissues: mantle, gills, adductor muscles, and foot. A) Electrophoresis (agarose, 1.5%) of the PCR products resulting from a standard PCR carried out with specific primers designed from the 5′- and 3′-ends. The positive control was performed by amplification of a housekeeping gene and the negative control was prepared without addition of cDNA. B) Histogram of the real-time PCR. Each column represents the relative expression ratio of upsalin in one specific tissue by a mean expression ratio ($n = 6$). The transcription of upsalin in the mantle is used as reference.

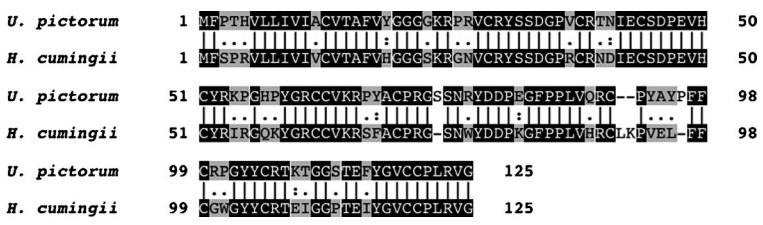


Figure 5. Pairwise alignment of the predicted amino acid sequences of Upsalin from *Unio pictorum* and *Hyriopsis cumingii* (produced by EMBOSS Needle): "|" denotes identical, ":" denotes conserved substitution, and "·" denotes semi-conserved substitution.

best matches with *E* values <1 were obtained with proteins from fruit flies and the yellow fever mosquito. The former proteins are predicted to be on the extracellular region (GO: 0005576) and to have a serine-type endopeptidase inhibitor activity (GO: 0004867), including several annotated domains along their sequences.

We decided to select the non-redundant protein segments of the best blastp hits and to align them with the sequences of Upsalin from *U. pictorum* and its homologue from *H. cumingii*. The optimal alignment was obtained with the tool T-Coffee and visualization was performed with EsPript 2.2. Figure 6 shows several similar groups along the alignment that ends at a highly conserved motif of six residues (YGVCCP). This motif is located at the C-terminal side of Upsalin, but is positioned further upstream for all the insect sequences. All the selected protein segments that gave high-scoring pairs with Upsalin constitute part of an EGF-like (epidermal growth factor) domain (IPR003645) and of one cysteine-rich repeat (IPR006150). However, in SMART or InterPro searches with Upsalin, no domains within the query were found aside from the signal peptide. The domains discussed above were still detected but with scores above the required threshold.

Purification and characterization of Upsalin by SDS-PAGE from shell extracts

Although peptide fragments of Upsalin had been obtained by proteomics in a previous study,[26] a thorough demonstration of the presence of Upsalin in the shell of *U. pictorum* was required. To this end, we purified the protein directly from *U. pictorum* shell extracts. We first extracted the acetic-acid-soluble matrix (ASM) of the nacre of *U. pictorum*. Quantification of both ASM and acetic-acid-insoluble matrix (AIM) indicated that ASM represents about 0.04% of the weight of the nacre shell powder and AIM represents 0.5%. In subsequent analyses, only ASM was used, and tested on 1D gels, as shown in Figure 7A (lane 2).

ASM is characterized by the abundance of polydisperse and discrete macromolecules. A polyclonal antibody was prepared from a peptide of Upsalin and, when tested against the nacre ASM on ELISA (Figure 7E), showed a strong reactivity, suggesting the presence of the protein in the extract. In subsequent experiments with this antibody, the second bleed (not the last one) was used, because of its higher reactivity. ASM was fractionated by preparative SDS-PAGE coupled to a fraction collector, and the Upsalin fraction was detected by dot-blotting of each fraction and incubation of the membrane with the antibody, as shown in Figure 7B. A pure fraction, eluted between tubes 17 and 21, was obtained and tested on gel (Figure 7A, lane 3) and on Western blot (Figure 7D, lane 3), together with ASM (Figure 7A, lane 2 and Figure 7D, lane 2). The purified fraction has an apparent molecular weight of 12 kDa, which is consistent with the calculated molecular weight of the mature Upsalin. Furthermore, the Western blot results unambiguously show that the antibody is specific for the purified fraction and does not cross-react with other macromolecules of the ASM (Figure 7D, lane 2).

In parallel, the purified fraction of Figure 7A (lane 3) was analyzed by nano-LC-MS/MS after trypsic digestion. Three peptides of 31, 25, and 15 aas, covering 65% of the sequence of mature Upsalin, were obtained (Figure 7C). These peptides, flanked by the basic residues Arg and Lys upstream and Arg downstream, completely match with the sequence of Upsalin. This result established that the purified fraction is Upsalin,

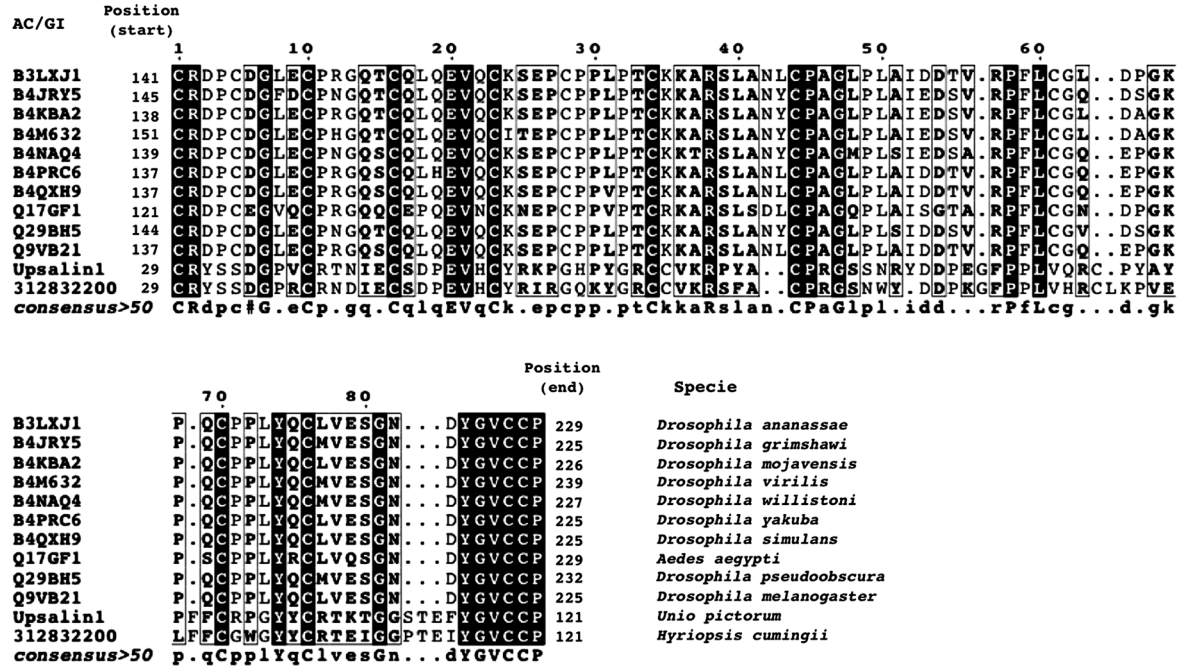


Figure 6. Multiple sequence alignment of Upsalin with segments of proteins from nine different species of fruit fly and from one mosquito. UniprotKB accession numbers are listed on the left and species name on the right. A: strict identity. **A**: similarity in a group. A: similarity across groups.

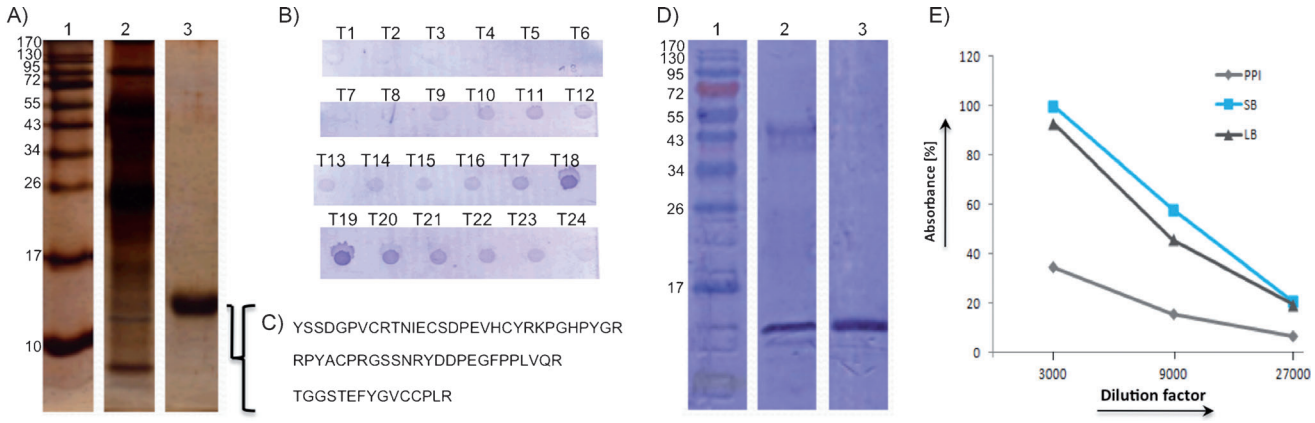


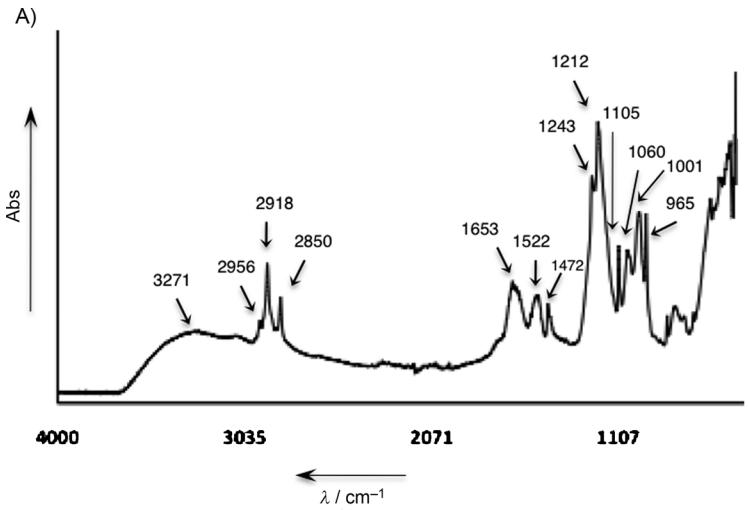
Figure 7. Purification of Upsalin, proteomic analysis of the purified protein fraction, and testing of the specificity of the anti-Upsalin antibody. A) SDS-PAGE of the ASM (lane 2) and of the purified Upsalin (lane 3), gel stained with silver nitrate; lane 1: molecular weight markers. B) Nacre ASM fractions generated by preparative electrophoresis, tested on dot-blot with the antibody elicited against the Upsalin peptide. Upsalin was eluted between tubes 17 and 22 and tested (lane 3). C) Sequences of the three peptides identified after nano-LC-MS/MS analysis of the purified protein fraction of lane 3. This analysis confirms that the sequence obtained by PCR and 3′-/5′-RACE is that of Upsalin. D) Western blot of the ASM (lane 2) and of the purified Upsalin (lane 3); the signal is specific of Upsalin (lane 1: molecular weight markers). E) ELISA of the polyclonal anti-Upsalin peptide antibody against the nacre ASM of *U. pictorum*. PPI: pre-immune serum. SB: second bleed. LB: last bleed. SB was used in subsequent experiments.

which in other words means that Upsalin is a protein from the ASM fraction associated with the shell of *U. pictorum*.

Glycosylation of Upsalin

We investigated the glycosylation patterns of the purified Upsalin fraction. FTIR (ATR) spectra were acquired from Upsalin lyophilizates, as shown in Figure 8A. The characteristic protein peaks are observed at 1653 cm^{-1} (amide I), at 1522 and 1472 cm^{-1} (amide II), and at 1212 cm^{-1} (amide III).[27] In the 4000–2500 cm^{-1} range, the broad band absorption at 3271 cm^{-1} and the sharp bands located at 2956, 2918, and 2850 cm^{-1} are attributable to the presence of NH and aliphatic CH groups, respectively.[28] Peaks characteristic of sugar moieties are observed in the 1200–900 cm^{-1} range and at 1105, 1060, 1001, and 965 cm^{-1}, and can be attributed in particular to C–C–O and C–O–C stretchings.[7,29] Furthermore, the weak absorption at 1243 cm^{-1} might reflect the presence of sulfated groups.[30]

The monosaccharide composition of Upsalin was determined after hydrolysis and is shown in Figure 8B. Of a list of eleven classical monosaccharides, including acidic (GlcA, GalA), neutral (Fuc, Rha, Ara, Xyl, Gal, Man, Glc), and amino sugars (GlcN, GalN), only seven were detected, representing a total of 5 ng per µg of Upsalin lyophilizate. This indicates that Upsalin is only weakly glycosylated (0.5 *w*% glycosylation). The two amino sugars, GlcN and GalN, together represent more than 60% of the composition, whereas Glc, Gal, and Man constitute 14, 9, and 8% respectively. Interestingly, the acetylated forms of GlcN and GalN (which convert to GlcN and GalN upon hy-



B)

Sugars	ng/µg	%
arabinose	0,05	1
galactosamine	1,35	26,9
galactose	0,46	9,2
glucosamine	1,83	36,5
glucose	0,71	14,1
mannose	0,41	8,2
xylose	0,21	4,2
Total	5,02	

Figure 8. Qualitative and quantitative analysis of saccharidic moieties of purified Upsalin. A) FTIR (ATR) spectrum of Upsalin showing in particular the amide I and II vibration modes of the protein, located at 1653 and 1522 cm^{-1}, respectively, and, in the 1200–1000 cm^{-1} range, the characteristic absorptions of sugar moieties (C–C–O and C–O–C stretchings) resulting from glycosylation. B) Monosaccharide composition of purified Upsalin, as determined by HPAE-PAD chromatography.

drolysis) are mostly the monomers, which link the saccharidic moieties to Ser or Thr (O-linked sugars), or to Asn (N-linked sugars). No traces of Fuc, Rha, or the acidic sugars GalA and GlcA were detected. We cannot exclude the possibility that the monosaccharide analysis after hydrolysis underestimates the amount of covalently linked sugars, if these sugars are mainly of the sialic acid type. If so, the FTIR spectrum does not give conclusive evidence of the presence of sialic acids because they are characterized by a FTIR peak at 1608 cm^{-1},[31] which in this case could be completely masked by the amide I peak at 1653 cm^{-1}.

In vitro crystallization assay with purified Upsalin

The purified Upsalin was assayed to check its ability to interact with the formation of calcium carbonate in vitro by the diffusion method. The results are shown in Figure 9. Increasing concentrations of Upsalin were tested, from 0 $\mu g\,mL^{-1}$ (control experiments) to 20 $\mu g\,mL^{-1}$. In the blank experiment, typical rhombohedrons of calcite were produced. Upsalin did not modify the shape of the crystals even at high concentrations. In particular, we did not observe polycrystalline aggregates, as often observed when intermediate concentrations are tested (5–10 $\mu g\,mL^{-1}$). We did not record inhibition of the crystal formation, even at higher concentrations. It seems that Upsalin has no interfering effect on the precipitation of calcium carbonate. In contrast, the ASM of *U. pictorum* exhibits a dose-dependent effect, as it has been shown earlier.[7]

In situ localization of Upsalin in the shell

Localization of Upsalin directly in the shell was performed, with the aid of the antibody raised against the Upsalin peptide. Six experiments were conducted; the results shown in Figure 10 are the best obtained from the six experiments. All the experiments were performed on cross-sections of shell fresh fractures, in order to examine the distribution of the signal in the shell thickness. We observed that Upsalin is local-

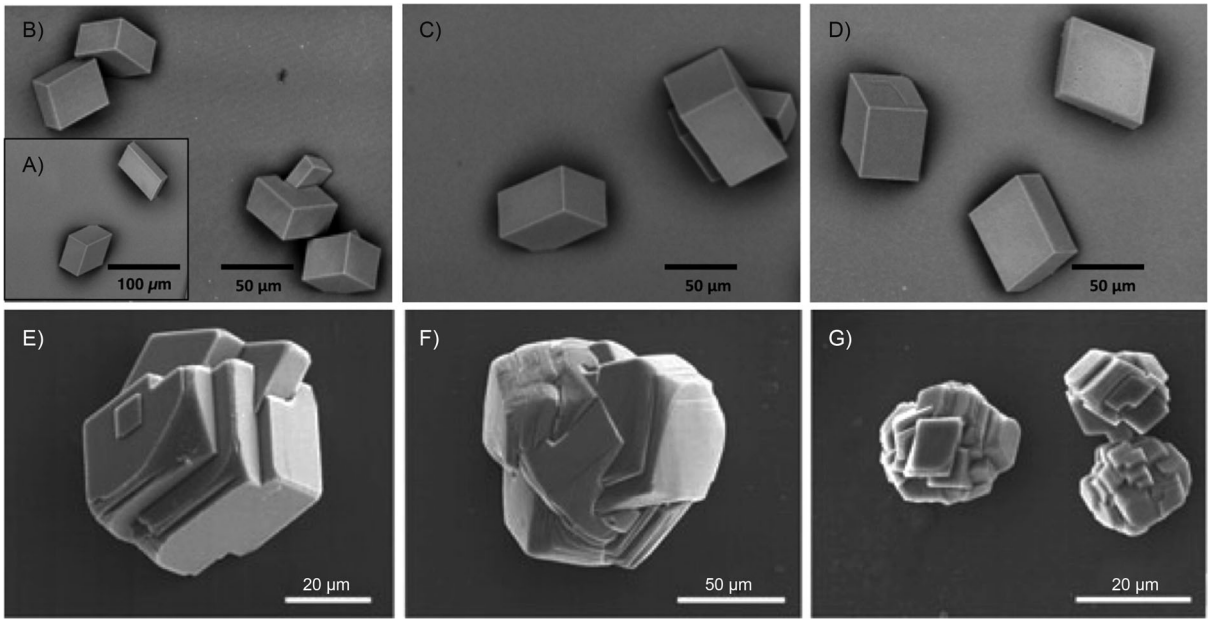


Figure 9. In vitro crystallization of calcium carbonate in the presence of Upsalin. A) Blank experiment (no Upsalin). B) [Upsalin] = 5 $\mu g\,mL^{-1}$. C) [Upsalin] = 10 $\mu g\,mL^{-1}$. D) [Upsalin] = 20 $\mu g\,mL^{-1}$. Almost no effect is recorded. E)–G) Effect of the ASM at similar concentrations.[7] E) [ASM] = 1 $\mu g\,mL^{-1}$. F) [ASM] = 5 $\mu g\,mL^{-1}$. G) [ASM] = 20 $\mu g\,mL^{-1}$.

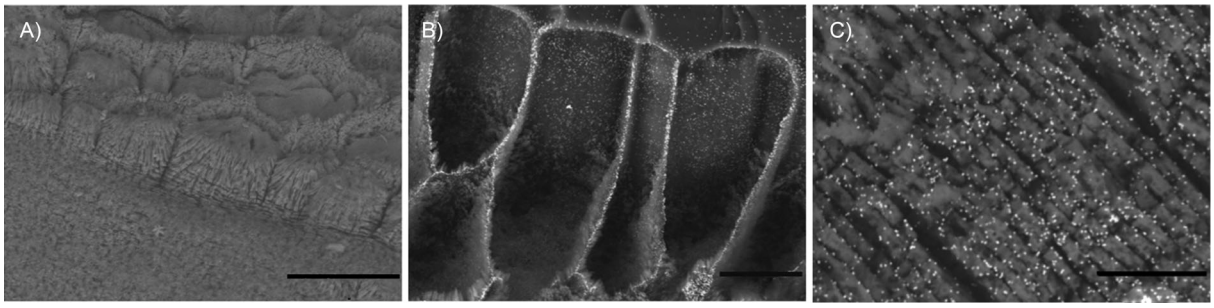


Figure 10. Immunogold staining of *U. pictorum* shell fragments with the polyclonal anti-Upsalin peptide antibody. A) Negative control. Scale bar: 50 μm. B) Results for the prismatic layer. Scale bar: 15 μm. C) Results for the nacreous layer. Upsalin is present in both layers. Scale bar: 10 μm.

ized in the nacreous shell layer. The signal seems to be concentrated in the interlamellar interface, but it is mostly discontinuous. In addition, a clear signal was also observed in the outer prismatic layer of the shell, in this case localized along the periprismatic sheaths that surround individual prisms. These results suggest that Upsalin is localized in both mineralized layers of the shell of *U. pictorum*.

Discussion

We have described the identification of a new protein, associated with the shell biomineralization process of the freshwater mussel *Unio pictorum*. We have characterized Upsalin both at the transcriptional and at the protein levels. To this end, we used the data set of previously published[26] peptides from the same species, in order to design degenerate primers that were subsequently tested on mantle cDNAs. The full sequence of the transcript was obtained by 5′- and 3′-RACE. This sequence encodes a 125-residue-long basic soluble protein, Upsalin, which possesses a signal peptide. In addition to being secreted, Upsalin is incorporated in the shell: it was indeed purified from the ASM of the nacreous layer and the sequence of the purified protein was confirmed by proteomics. Upsalin is weakly glycosylated. An antibody elicited against an immunogenic peptide of Upsalin showed that it is present in both layers (nacreous and prismatic) of the shell of *U. pictorum*. Upsalin, together with prismalin-14,[13] mucoperlin,[32] N14/N16/pearlin,[33–35] and BMSP, thus represents one of the few examples of a shell protein to be characterized both at the transcriptional and the protein levels, and for which in vitro characterization has been performed after purification from the shell.

Unlike other biomineralization transcripts that are exclusively expressed in the mantle,[13,36] the expression of Upsalin is particularly high in the mantle and adductor muscles and it is also faintly detectable in the gills and foot. A similar situation has been observed previously with some members of the shematrin family and with the methionine-rich protein MRNP-34,[25] which are thought to provide frameworks for calcification. Shematrins-1, -2, -5, and -6 showed expression levels that were not restricted to the mantle but were also present in other tissues such as the adductor muscles.[24]

What might be the function of Upsalin in biomineralization? Firstly, a blast search at the nucleotide level indicates that the highest similarity found so far is with a hypothetical mantle protein of the Chinese freshwater pearl mussel *Hyriopsis cumingii*, the function of which is totally unknown. Upsalin and this protein may be considered true homologues. This suggests that Upsalin or Upsalin-like proteins might also be found in other genera of the Unionidae, but further experimental RT-PCR investigations will be required, in different representatives of this family, and, beyond, of the whole Palaeoheterodonta subclass. At present, because of the scarcity of molecular data for this bivalve group in the NCBI or SwissProt databases, in silico analyses for retrieval of Upsalin homologues are irrelevant. Additionally, we were not able to detect any Upsalin homologue even in the large 270 000 EST database from *Pinctada margaritifera* mantle cDNA library (data not shown),[17] suggesting that Upsalin is not present in some other mollusks with nacreous shells, and could be taxon-specific.

In addition, Upsalin shows similarity with a group of proteins of insect origin (Diptera). The similarity region involves about 80 residues located in the first halves of the sequences of extracellular proteins with serine-type endopetidase inhibitor activity (GO: 0004867). The best hits include proteins from nine species of *Drosophila*. The alignment of Upsalin and of its *Hyriopsis* homologue with these proteins revealed a high degree of sequence conservation with ten conserved cysteine residues and eleven other residues, eight of which are hydrophobic (G, L, V, P, Y). In addition, a short, fully conserved YGVCCP motif is present at the C-terminal side of the alignment. It is known that cysteine residues play key role in the 3D structures of proteins, because they form intra- and intermolecular disulfide bridges, constraining the proteins to proper folding and stabilization in the extracellular domain. In the dipterian proteins, the aligned regions are part of one EGF-like domain and of one cysteine-rich-like domain. According to InterPro resources on the domain organization of the cysteine-rich repeat IPR006150, it is commonly associated with Kunitz domains that might function as serine peptidase inhibitors. This is consistent with our previous proteomic analysis on the ASM of *U. pictorum* and of its 12/16 kDa SDS-PAGE bands, which also showed peptides with similar putative serine protease inhibitor domains.[26]

Finally, the amino acid sequence of Upsalin does not exhibit any homology with those of previously discovered shell matrix proteins, in particular with proteins associated with nacre microstructures, whatever their origin: abalone,[21] pearl oysters,[17,20] or blue mussel.[19] This finding is striking, not to say remarkable, and once more highlights the unexpected diversity of molluskan shell proteins, one fact among others that suggests that mollusks have "invented" more than one single "biochemical" pathway for synthesizing nacre.[19,37]

Unlike in the cases of many other molluskan shell-associated proteins,[13,38] no internal repeats or low-complexity regions are observed in the Upsalin primary sequence. Furthermore, Upsalin is only weakly glycosylated, and it is very unlikely that its sugar moieties exert a role in concentrating calcium ions in the vicinity of the nucleation sites[39] as has been proposed for other proteins.[11,32] Previous calcium-binding studies on the ASM of *Unio pictorum*, by staining with Stains All and by autoradiography with ^{45}Ca, suggest weak calcium binding ability at molecular weights below 14 kDa.[7] Consistently with this, Upsalin does not interfere in vitro with the precipitation of calcium carbonate. These considerations suggest that Upsalin does not have a direct interaction with calcium or calcium carbonate crystals. Furthermore, Upsalin does not exhibit any chitin-binding capacity (data not shown). Taken together, these data suggest that Upsalin is involved in other functions, which are indirectly related to biomineralization. Putative roles include, among others, protection of the matrix against enzymatic degradation, organization of the spatial arrangement of other matrix macromolecules, or signaling activity towards the epithelial cells of the calcifying mantle, but these hypotheses will need to be tested experimentally. In addition, the fact that

Upsalin has a considerable amount of proline units (12.8%) in its sequence suggests a rigid backbone with good potential for binding other proteins.[40]

Eight years ago, Gotliv and co-workers[41] published a paper that pointed out the fact that several proteins retrieved from shell tissues did not have a clear function in mineralization. Since that publication, although our knowledge about shell proteins has immensely progressed, we feel compelled to note that this observation is still valid. Upsalin, like several other proteins, enters the growing collection of orphan proteins with no association to known functional traits or protein families. This clearly indicates that the coming years should be urgently dedicated to setting up approaches such as gene knock-down, as has recently been performed on the pearl oyster *Pinctada fucata*,[42] or the use of two-hybrid systems, that should be extremely useful for deciphering the functions of molluskan shell proteins in biomineral deposition.

Experimental Section

Sample collection and characterization: Living adult *Unio pictorum* specimens with shells between 30–80 mm in length were collected in the stream of La Varaude (Izeure, Côte d'Or, France). The soft tissues were removed from the shell. The mantle, foot, gills, and muscles were sampled under RNAse-free conditions and transferred to liquid nitrogen, and were stored at −80 °C until further experiments. The fresh shells were cleaned, fragmented, and etched with EDTA (1%, *w*/*v*) for direct microscopic observations with a Hitachi TM-1000 table-top scanning electron microscope (SEM) with an acceleration voltage of 15 kV, under a back-scattered electron mode.

General strategy for the identification of Upsalin: A previous proteomic study on the bulk nacre shell matrix of *U. pictorum* and on SDS-fractionated protein bands allowed the identification of several peptides. Twelve of them were identified by de novo sequencing from a 12 kDa band. They were used for the design of degenerate primers, in order to obtain the full sequence of the corresponding transcript. Once this was complete, we designed a polyclonal antibody from an immunogenic peptide of the found sequence, to purify the protein from nacre extracts and to localize it in shell tissues. The predicted protein sequence was confirmed by proteomic analysis of the purified extract, which was subsequently used for in vitro functional assays.

Total RNA extraction and cDNA synthesis: The total RNA was extracted from the mantle, foot, gills, and adductor muscles (ca. 80 mg tissue) by grinding the tissue with a Teflon pestle until complete homogenization in TRI Reagent (Euromedex), according to the manufacturer's instructions. RNA integrity was checked by fractionation in a non-denaturing agarose gel (TBE 1.0X, 1.2%) that was subsequently stained with ethidium bromide. Complementary DNA synthesis was performed with the iScript cDNA synthesis kit (BioRad, USA). Quantifications both of the total RNA and of cDNA were determined by measuring the $OD_{260\,nm}$ with a SmartSpec Plus Spectrophotometer (BioRad, USA).

Identification of a cDNA fragment: The cDNA from the mantle was first screened by PCR with degenerate primers. Amplification of cDNA was performed with GoTaq Flexi DNA Polymerase (1.25 U, Promega, USA) and an automated MJ Mini Gradient Thermal Thermocycler (USA, BioRad) by the following program: 5 min of initial denaturation at 95 °C, 1 min of denaturation at 95 °C (35×), 1 min of annealing at 55 °C (35×), 5 min of extension at 72 °C (35×) and a final extension of 5 min at 72 °C. A set of 16 degenerate primers based on eight peptide sequences (sense and antisense) obtained from our previous study was designed. The peptides were chosen because of their exclusive presence in the protein bands of 12 and 16 kDa obtained by SDS-PAGE analysis of the ASM. From the set of degenerate primers, one pair gave a highly specific PCR product of 250 bp (sense, 5′-GARTG YWSNG AYCCN GARGT-3′ and antisense, 5′-GGRCA RCANA CNCCR TARAA YTC-3′).

Rapid amplification of cDNA ends (5′- and 3′-RACE): The cDNA ends were identified by use of the 3′-RACE (ref. 18373–019) and 5′-RACE (ref. 18374–058) Systems for Rapid Amplification of cDNA ends (Invitrogen). The procedure was carried out according to the manufacturer's instructions with gene-specific primers based on the nucleotide sequence of the Upsalin cDNA fragment. The 3′-end product was obtained by PCR amplification of the reverse transcribed cDNA sequence with the gene-specific primer (GSP) 5′-CGTCC GTATG CGTGC CCTAG-3′, with an annealing temperature of 52 °C. To obtain the 5′-end product, the total RNA was reverse-transcribed by use of a first antisense gene-specific primer (GSP1) 5′-AATGG ATATG CATAA GGAC-3′ and amplification of the cDNA was performed by use of a nested antisense specific primer (GSP2) 5′-GTGGC GGAAA ACCTT CAGGA-3′ with an annealing temperature of 50.4 °C.

Purification, amplification, and sequencing: All PCR products were run on agarose gel (1.5%, TBE 1.0×), purified with the Wizard SV gel and PCR clean-up system (Promega, ref. A9281, USA), and then cloned in a pGEM-T Easy vector system I (Promega, USA) with JM109 Competent Cells (Promega, USA). Transformed colonies containing the cloning vector were selected on LB agar plates with ampicillin (100 $\mu g\,mL^{-1}$) at 37 °C for 12 h. From each single colony, a small portion was sampled and scattered in new numbered Petri dishes and grown under the same previously described conditions for 20 h. For selection of the positive clones, a small portion of each grown colony was taken and put in nuclease-free water (10 μL). The solution was heated at 99 °C for 20 min to release the plasmids and a PCR reaction was performed with use of the pGEM-T Easy vector-specific primers T7 (5′-TAATA CGACT CACTA TAGGG-3′) and SP6 (5′-CATTT AGGTG ACACT ATAG-3′). This allowed identification of the positive recombinant plasmids, which were subsequently purified by use of the QIAprep miniprep kit (Qiagen, ref. 27106) and sequenced by the Eurofins MWG Operon sequencing service (Eurofins MWG Operon sequencing service, Ebersberg, Germany).

Amplification of the full nucleotide sequence and quantitative real-time PCR: To confirm that the overlap of the sequences obtained by RACE corresponded to the predicted complete nucleotide sequence, the full-length cDNA was amplified with use of two specific primers based on the 5′- and 3′-ends. In order to check its presence in different tissues, RNA was extracted from mantle, gills, foot, and adductor muscles. RNA was treated with DNAse (Invitrogen) to remove DNA contaminants and then converted into cDNA. PCR and quantitative PCR were performed with the primers sense 5′-GCCGG ATAGG ACACC TTGAG-3′, upstream of the coding region, and anti-sense 5′-ACCCG ACCCT TAATG GGCAA-3′, coding for the C terminus of the predicted protein. Standard PCR was performed with the following program: 2 min of initial denaturation at 95 °C, 1 min of denaturation at 95 °C (35×), 1 min of annealing at 60 °C (35×), 1 min of extension at 72 °C (35×) and a final extension of 4 min at 72 °C. Quantitative PCR was carried out with an iQ SYBR Green Supermix kit (BioRad, USA) and an iCycler iQ Real Time PCR Detection System (Bio-Rad, USA). Firstly, fusion curve and primers

efficiency were tested with sequentially diluted cDNAs (1, 1:10, 1:100, 1:1000) from the four tissues. Consequently, each real-time PCR reaction was performed in triplicate with non-diluted cDNA (1 μL) and the mean of three independent biological replicates was calculated. All results were normalized to the β-actin mRNA levels and calculated by the $2^{-\Delta\Delta Ct}$ method.[43]

In silico analysis of the deduced amino acid sequence: Homology searches with the Upsalin nucleotide sequence were performed with the program tblastx, available from NCBI, against the nucleotide databases nr and est. The default algorithm parameters were used, with the exception of the *E* value, which was set to 10^{-3}. The predicted amino acid sequence was also used as query against UniprotKB with use of the blastp program. The putative presence of a signal peptide was determined by use of SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/).[44] The theoretical molecular weight (M_W), isoelectric point (pI), and amino acid composition were determined with the tools available from the Expasy website (http://web.expasy.org/protparam/).[45] Putative phosphorylation and glycosylation sites were checked with the NetPhos 2.0 server (http://www.cbs.dtu.dk/services/NetPhos/)[46] and the YinOYang 1.2 server (http://www.cbs.dtu.dk/services/YinOYang/),[47,48] respectively. In order to identify potential protein domains and functional sites, protein sequences were queried against the InterPro database (http://www.ebi.ac.uk/Tools/pfa/iprscan/)[49] and the SMART tool (http://smart.embl-heidelberg.de/).[50] Sequence alignments were performed with EMBOSS Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/),[51] T-Coffee (http://www.ebi.ac.uk/Tools/msa/tcoffee/),[52] and EsPript 2.2 (http://espript.ibcp.fr/ESPript/ESPript/).[53]

Extraction of the shell organic matrix: Shells were rinsed in tap water and brushed, and were then immersed for 24 h in a solution of sodium hypochlorite (dilution 1:20, 0.13% active chlorine) to remove the organic contaminants. They were subsequently rinsed in milli-Q water and dried. The external prismatic layer was removed by mechanical polishing. The internal nacreous layer was crushed and sieved (< 200 μm) to afford a fine calcium carbonate powder, which was decalcified overnight with cold dilute acetic acid (5%, *v*/*v*), by our standard procedure.[10] The resulting solution was centrifuged (30 min, 3900 *g*, 4 °C) to separate the acetic acid-soluble fraction (ASM, supernatant) from the acetic acid-insoluble one (AIM, pellet). The ASM was filtered on cellulose membrane (5 μm) and then ultrafiltered on a 10 kDa-cutoff membrane (Amicon) to concentrate the macromolecules, which were subsequently dialyzed at 4 °C for a few days against milli-Q water (with several water changes). The solution was finally lyophilized. The AIM was resuspended in milli-Q water, the suspension was centrifuged, and the supernatant was put aside. After several suspension/centrifugation cycles, the final pellet was freeze-dried, while the pooled supernatants were added to the ASM. Both ASM and AIM were quantified by direct weighing of the lyophilizates with a precision balance.

Protein purification and characterization on mono-dimensional gel and on Western blots: The ASM was tested by conventional mono-dimensional SDS-PAGE (Bio-Rad, miniProtean III). After complete denaturation with Laemmli buffer,[54] gels were stained with silver nitrate.[55] Alternately, gels were electro-transferred onto PVDF membranes and the membranes were subsequently exposed to a specific antibody elicited against the target protein, by a standard Western blot procedure.[56] To detect the antigen-antibody complex, we used a secondary antibody coupled to alkaline phosphatase (Sigma A3687, 1:30 000), and the complex was stained with NBT/BCIP (Sigma B5655).

The ASM was used for purification of the protein of interest by a procedure that we developed:[10] in brief, this involves a blind fractionation of the ASM on a preparative SDS-PAGE (Bio Rad, Prep Cell model 491) equipped downstream with a fraction collector, followed by a test of the eighty collected fractions on dot-blot, with an antibody elicited against a peptide of the protein deduced from the DNA sequence (see below). The purity of the protein fraction was checked with an acrylamide gel (16%), which was stained with silver nitrate.

Proteomic analysis of the purified fraction: A proteomic analysis was performed in order to check the identity of the protein fraction purified by preparative electrophoresis. Briefly, the fraction was enzymatically digested with trypsin[57] and then purified with a Vivapure C18 micro membrane (Vivascience). The purified samples were analyzed by nano-LC-MS/MS with a nano-Liquid Chromatography system (LC Packings, Dionex) and a nano-ESI-qQ-TOF mass spectrometry system (QSTAR XL, AB Sciex). Mass spectrometry data were acquired with use of Analyst QS 1.1 software (AB Sciex) operated in IDA mode as previously described.[57] Protein identification was performed with ProteinPilot 3.0 software (Applied Biosystems) with use of the Paragon database search algorithm and the sequence determined from the DNA sequencing.

Antibody production and ELISA testing: A peptide, corresponding to the sequence ACPRGSSNRYDDPEGF, was designed, synthesized, and coupled with a carrier (KLH) for eliciting polyclonal antibodies in two white rabbits, according to the 28 days Speedy program developed by Eurogentec: immunizations at days 0, 7, 10, and 18, and blood sampling at days 0 (pre-immune serum), 21 (intermediate), and 28 (last bleeding). Polyclonal antibodies were tested by ELISA, by a standard procedure.[58] This experimental procedure allowed the determination of the reactivity and the titer of the antibody elicited against its target peptide. In this case, the tested antigen was the nacre ASM of *Unio pictorum*.

Glycosylation studies: The qualitative and quantitative characterization of post-translational modifications (PTMs)—that is, glycosylation—of the purified protein was investigated by two techniques: FTIR and monosaccharide analysis. FTIR spectra were recorded from dry lyophilized purified protein samples with a Bruker Vector 22 instrument fitted with a Specac Golden Gate ATR device (Specac Ltd., Orpington, UK) in the 4000–500 cm^{-1} wavenumber range (ten scans at a spectral resolution of 2 cm^{-1}). The assignment of absorption bands was performed by comparison with previous spectra descriptions available in the bibliography.

For quantifying the monosaccharide content of the purified protein, lyophilized samples were hydrolyzed in trifluoroacetic acid (TFA, 2 M, 100 μL) at 105 °C for 4 h. These hydrolytic conditions do not allow the quantification of GalNAc and GlcNac, which are converted into GalN (galactosamine) and GlcN (glucosamine), respectively. Sialic acids, such as *N*-acetylneuraminic acid, are destroyed during the hydrolytic procedure. Samples were evaporated to dryness before being dissolved with NaOH (100 μL, 20 mM). The sugar contents of the hydrolysates were determined by high-performance anion exchange with pulsed amperometric detection (HPAE-PAD) on a CarboPac PA100 column (Dionex Corp., Sunnyvale, CA, USA). Carbohydrate standard (Sigma) was injected at 16, 8, and 4 ppm. Non-hydrolyzed samples were analyzed similarly in order to detect free monosaccharides that could have contaminated the sample during dialysis. For the reasons indicated above, this technique does not allow quantification of sialic acids.

In vitro interaction of the purified protein with calcium carbonate: $CaCO_3$ precipitation was performed in vitro by slow diffusion

of ammonium carbonate vapor into calcium chloride solution.[4] The test was adapted as follows: solutions of $CaCl_2$ (10 mM, 200 μL) containing different amounts of the purified protein (0 to 20 $\mu g\,mL^{-1}$) were applied to 16-well culture slides (BD Falcon, Becton Dickinson, Franklin Lakes, NJ, USA). Blank controls were performed without any sample. They were incubated for 48 or 72 h at room temperature or at 4 °C in a closed desiccator containing crystals of ammonium bicarbonate. They were dried and directly observed with a tabletop scanning electron microscope (Hitachi TM-1000).

Immunogold localization of the purified protein on shell fragments: Immunogold labeling was performed with shell fragments as described previously[59] by use of the antibody raised against the purified protein (diluted 1:5000) and a secondary antibody (goat anti-rabbit, dilution 1:400), coupled with 5 nm gold particles (British Biocell International, Cardiff, UK, ref. EM.GAR5). The size of the gold particles was increased further by incubating the shell fragments in a silver-enhancing solution (BBI, ref. EKL15) for 15 min. The samples were then rinsed, dried, and directly observed with a Hitachi TM1000 SEM. Control experiments were performed similarly, variously without the first antibody step (but with the secondary antibody and the silver enhancement incubations) or with the pre-immune serum (with all the incubation steps), or with the silver enhancement step alone.

Acknowledgements

The work of P.R.S. was supported by the Marie Curie ITN (Initial Training Network), BIOMINTEC, piloted by the Johannes Gutenberg University, Mainz, Germany (Prof. H. C. Schröder, see http://www.biomintec.de/). The work of S.B. was supported by UMR CNRS 5561, Biogéosciences. Additional financial support includes the ACCRO-Earth ANR program 2007–2011 (LCSE, G. Ramstein), an INTERRVIE program from INSU (2010), and finally a BQR project between UMR CNRS 5561 and UPSP PROXISS, AgroSup. The authors thank Stéphane Fraichard (UMR CNRS 5548) for his kind help and advice in qPCR experiments.

P.R.-S. performed all the molecular biology work, with the assistance of N.L.R., and all the in silico analyses, S.B. did all the biochemical characterization, together with N.G., L.P. performed the FTIR measurements, I.Z.-C. the proteomic analysis, and M.C. and G.A. the monosaccharide analysis. The study was designed by F.M. and P.R.-S. Technical advice was given by B.M. and J.K. The manuscript was written by P.R.-S. and F.M., with input from S.B.

Keywords: biomineralization · freshwater bivalves · mollusks · proteins · proteomics

[1] F. Marin, G. Luquet, B. Marie, D. Medakovic, *Curr. Top. Dev. Biol.* **2007**, *80*, 209 – 276.
[2] F. Marin, G. Luquet, *C. R. Palevol.* **2004**, *3*, 469 – 492.
[3] K. Simkiss, *Comp. Biochem. Physiol.* **1965**, *16*, 427 – 435.
[4] L. Addadi, S. Weiner, *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 4110 – 4114.
[5] J. Keith, S. Stockwell, D. Ball, K. Remillard, D. Kaplan, T. Thannhauser, R. Sherwood, *Comp. Biochem. Physiol. B* **1993**, *105*, 487 – 496.
[6] H. Miyamoto, T. Miyashita, M. Okushima, S. Nakano, T. Morita, A. Matsushiro, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 9657 – 9660.
[7] B. Marie, G. Luquet, J.-P. Pais De Barros, N. Guichard, S. Morel, G. Alcaraz, L. Bollache, F. Marin, *FEBS J.* **2007**, *274*, 2933 – 2945.
[8] C. Zhang, R. Zhang, *Mar. Biotechnol.* **2006**, *8*, 572 – 586.
[9] J. S. Evans, *Chem. Rev.* **2008**, *108*, 4455 – 4462.
[10] F. Marin, L. Pereira, P. Westbroek, *Protein Expression Purif.* **2001**, *23*, 175 – 179.
[11] B. Marie, G. Luquet, L. Bédouet, C. Milet, N. Guichard, D. Medakovic, F. Marin, *ChemBioChem* **2008**, *9*, 2515 – 2523.
[12] F. Marin, R. Amons, N. Guichard, M. Stigter, A. Hecker, G. Luquet, P. Layrolle, G. Alcaraz, C. Riondet, P. Westbroek, *J Biol Chem.* **2005**, *280*, 33895 – 33908.
[13] M. Suzuki, E. Murayama, H. Inoue, N. Ozaki, H. Tohse, T. Kogure, H. Nagasawa, *Biochem. J.* **2004**, *382*, 205 – 213.
[14] C. Montagnani, B. Marie, F. Marin, C. Belliard, F. Riquet, A. Tayalé, I. Zanella-Cléon, E. Fleury, Y. Gueguen, D. Piquemal, N. Cochennec-Laureau, *ChemBioChem* **2011**, *12*, 2033 – 2043.
[15] C. Zhang, L. Xie, J. Huang, X. Liu, R. Zhang, *Biochem. Biophys. Res. Commun.* **2006**, *344*, 735 – 740.
[16] W. Simison, J. L. Boore in *Phylogeny and Evolution of the Mollusca* (Eds.: W. Ponder, D. R. Lindberg), University of California Press, Berkeley, **2008**, pp. 447 – 462.
[17] C. Joubert, D. Piquemal, B. Marie, L. Manchon, F. Pierrat, I. Zanella-Cléon, N. Cochennec-Laureau, Y. Gueguen, C. Montagnani, *BMC Genomics* **2010**, *11*, 613.
[18] B. Marie, N. Trinkler, I. Zanella-Cleon, N. Guichard, M. Becchi, C. Paillard, F. Marin, *Mar. Biotechnol.* **2011**, *13*, 955 – 962.
[19] B. Marie, N. Le Roy, I. Zanella-Cléon, M. Becchi, F. Marin, *J. Mol. Evol.* **2011**, *72*, 531 – 546.
[20] B. Marie, I. Zanella-Cléon, N. Guichard, M. Becchi, F. Marin, *Mar. Biotechnol.* **2011**, *13*, 1159 – 1168.
[21] B. Marie, A. Marie, D. J. Jackson, L. Dubost, B. M. Degnan, C. Milet, F. Marin, *Proteome Sci.* **2010**, *8*, 54.
[22] F. Marin, G. Luquet in *Handbook of Biomineralization: Biological Aspects and Structure Formation* (Ed.: E. Bauerlein), Wiley-VCH, Weinheim, **2007**, pp. 273 – 290.
[23] P. Maurer, E. Hohenester, J. Engel, *Curr Opin Cell Biol* **1996**, *8*, 609 – 617.
[24] M. Yano, K. Nagai, K. Morimoto, H. Miyamoto, *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2006**, *144*, 254 – 262.
[25] B. Marie, C. Joubert, C. Belliard, A. Tayale, I. Zanella-Cléon, F. Marin, Y. Gueguen, C. Montagnani, *Amino Acids* **2011**; DOI: 10.1007/s00726-011-0932-0.
[26] B. Marie, I. Zanella-Cléon, N. Le Roy, M. Becchi, G. Luquet, F. Marin, *ChemBioChem* **2010**, *11*, 2138 – 2147.
[27] A. Barth, *Biochim. Biophys. Acta* **2007**, *1767*, 1073 – 1101.
[28] R. Silverstein, G. Bassler, T. Morrill, *Spectrometric Identification of Organic Compounds*, Wiley, New York, **1981**.
[29] J. C. Marxen, M. Hammer, T. Gehrke, W. Becker, *Biol. Bull.* **1998**, *194*, 231 – 240.
[30] S. Albeck, S. Weiner, L. Addadi, *Chem. Eur. J.* **1996**, *2*, 278 – 284.
[31] R. Knörle, P. Schnierle, A. Koch, N. P. Buchholz, F. Hering, H. Seiler, T. Ackermann, G. Rutishauser, *Clin. Chem.* **1994**, *40*, 1739 – 1743.
[32] F. Marin, P. Corstjens, B. de Gaulejac, E. de Vrind-De Jong, P. Westbroek, *J. Biol. Chem.* **2000**, *275*, 20667 – 20675.
[33] M. Kono, N. Hayashi, T. Samata, *Biochem. Biophys. Res. Commun.* **2000**, *269*, 213 – 218.
[34] T. Samata, N. Hayashi, M. Kono, K. Hasegawa, C. Horita, S. Akera, *FEBS Lett.* **1999**, *462*, 225 – 229.
[35] T. Miyashita, R. Takagi, M. Okushima, S. Nakano, H. Miyamoto, E. Nishikawa, A. Matsushiro, *Protein J.* **2000**, 409 – 418.
[36] Y. Kong, G. Jing, Z. Yan, C. Li, N. Gong, F. Zhu, D. Li, Y. Zhang, G. Zheng, H. Wang, L. Xie, R. Zhang, *J. Biol. Chem.* **2009**, *284*, 10841 – 10854.
[37] D. J. Jackson, C. McDougall, B. Woodcroft, P. Moase, R. A. Rose, M. Kube, R. Reinhardt, D. S. Rokhsar, C. Montagnani, C. Joubert, D. Piquemal, B. Degnan, *Mol. Biol. Evol.* **2010**, *27*, 591 – 608.
[38] X. Shen, A. M. Belcher, P. K. Hansma, G. D. Stucky, D. E. Morse, *J. Biol. Chem.* **1997**, *272*, 32472 – 32481.
[39] L. Addadi, J. Moradian, E. Shay, N. G. Maroudas, S. Weiner, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 2732 – 2736.
[40] M. P. Williamson, *Biochem. J.* **1994**, *297*, 249 – 260.
[41] B.-A. Gotliv, L. Addadi, S. Weiner, *ChemBioChem* **2003**, *4*, 522 – 529.

[42] M. Suzuki, K. Saruwatari, T. Kogure, Y. Yamamoto, T. Nishimura, T. Kato, H. Nagasawa, *Science* **2009**, *325*, 1388 – 1390.
[43] K. J. Livak, T. D. Schmittgen, *Methods* **2001**, *25*, 402 – 408.
[44] J. D. Bendtsen, H. Nielsen, G. V. Heijne, S. Brunak, *J Mol Biol* **2004**, *340*, 783 – 795.
[45] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch in *The Proteomics Protocols Handbook* (Ed.: J. M. Walker), Humana, Totowa, **2005**, pp. 571 – 608.
[46] N. Blom, S. Gammeltoft, S. Brunak, *J Mol Biol* **1999**, *294*, 1351 – 1362.
[47] R. Gupta, S. Brunak, *Pac. Symp. Biocomput.* **2002**, *7*, 310 – 322.
[48] Y. Sakaidani, T. Nomura, A. Matsuura, M. Ito, E. Suzuki, K. Murakami, D. Nadano, T. Matsuda, K. Furukawa, T. Okajima, *Nat. Commun.* **2011**, *2*, 583.
[49] E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, R. Lopez, P. Bucher, L. Cerutti, A. Res, *Nucleic Acids Res.* **2005**, *33*, 116 – 120.
[50] I. Letunic, T. Doerks, P. Bork, *Nucleic Acids Res.* **2011**, *39*, 302 – 305.
[51] P. Rice, I. Longden, A. Bleasby, *Trends Genet. Trends Gen* **2000**, *16*, 276 – 277.
[52] C. Notredame, D. Higgins, J. Heringa, *J. Mol. Biol.* **2000**, *302*, 205 – 217.
[53] P. Gouet, E. Courcelle, D. I. Stuart, F. Métoz, *Bioinformatics* **1999**, *15*, 305 – 308.
[54] U. K. Laemmli, *Nature* **1970**, *227*, 680 – 685.
[55] J. H. Morrissey, *Anal. Biochem.* **1981**, *117*, 307 – 310.
[56] H. Towbin, T. Staehelin, J. Gordon, *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 4350 – 4354.
[57] B. Marie, F. Marin, A. Marie, L. Bédouet, L. Dubost, G. Alcaraz, C. Milet, G. Luquet, *ChemBioChem* **2009**, *10*, 1495 – 1506.
[58] M. F. Clark, A. N. Adams, *J. Gen. Virol.* **1977**, *34*, 475 – 483.
[59] F. Marin, B. Pokroy, G. Luquet, P. Layrolle, K. De Groot, *Biomaterials* **2007**, *28*, 2368 – 2377.

Received: November 11, 2011
Published online on April 4, 2012

