DOI: 10.1002/cbic.200900009

Evolution of Nacre: Biochemistry and Proteomics of the Shell Organic Matrix of the Cephalopod *Nautilus macromphalus*

Benjamin Marie,*[a] Frédéric Marin,*[a] Arul Marie,[b] Laurent Bédouet,[c] Lionel Dubost,[b] Gérard Alcaraz,[d] Christian Milet,[c] and Gilles Luquet*[a]

In mollusks, one of the most widely studied shell textures is nacre, the lustrous aragonitic layer that constitutes the internal components of the shells of several bivalves, a few gastropods, and one cephalopod: the nautilus. Nacre contains a minor organic fraction, which displays a wide range of functions in relation to the biomineralization process. Here, we have biochemically characterized the nacre matrix of the cephalopod *Nautilus macromphalus*. The acid-soluble matrix contains a mixture of polydisperse and discrete proteins and glycoproteins, which interact with the formation of calcite crystals. In addition, a few bind calcium ions. Furthermore, we have used a proteomic approach, which was applied to the acetic acid-soluble and -insoluble shell matrices, as well as to spots obtained after 2D gel electrophoresis. Our data demonstrate that the insoluble and soluble matrices, although different in their bulk monosaccharide and amino acid compositions, contain numerous shared peptides. Strikingly, most of the obtained partial sequences are entirely new. A few only partly match with bivalvian nacre proteins. Our findings have implications for knowledge of the long-term evolution of molluskan nacre matrices.

Introduction

The calcification of the molluskan shell is a genetically controlled process based on subtle interactions between an extracellular matrix, which is synthesized by the epithelial cells of the calcifying mantle, and the mineral ion precursors of calcification: calcium, bicarbonate, and minor elements. The proteins, glycoproteins, and polysaccharides that are incorporated within the forming shell constitute the shell matrix. They are believed to have several key functions:[1] to provide an organic template for mineral growth, to create local physicochemical conditions for promoting nucleation, to determine the calcium carbonate polymorph (calcite vs. aragonite), to guide the mineral growth in privileged directions, to organize the crystal architecture in well-defined microstructures, and to inhibit the crystal growth when and where necessary.[2] The matrix is also thought to play a role in signaling to the mantle epithelial cells.[3]

For several reasons related to its remarkable mechanical properties and to its high value in jewelry, nacre is by far the most widely studied molluskan shell microstructure. In addition, nacre is a biocompatible material, which may be used in bone surgery because of its osteoinductive and osteogenic properties.[4] From a structural viewpoint, it consists of a regular superimposition of polygonal flat aragonitic tablets of 0.5 μm in thickness, embedded in a peripheral thin organic matrix.[5,6] Recent ultrastructural advances have shown that each nacre tablet has a hierarchical organization[7] and is composed of nanodomains.[8] In parallel, biochemical characterizations have shown that the nacre matrix comprises a large set of macromolecular components, such as chitin,[9] hydrophobic "framework" proteins,[10] proteins with elastomeric properties,[11] and several soluble proteins and glycoproteins.[12,13] For a few years, new dynamic models of nacre formation have been used in attempts to conciliate both biochemical and ultrastructural data.[1] Particular emphasis has been placed on the prominent role played by chitin in spatially structuring the organic framework,[1,14] the distribution of key biochemical functions at the surfaces of nacre tablets,[15] the existence of a transient precursor amorphous phase,[16] and the growth of nacre tablets in a hydrophobic gel, which hardens and becomes insoluble when nacre tablets coalesce.[17]

Nacre also constitutes a fascinating model of study from an evolutionary viewpoint. Firstly, nacre appears early in the fossil record,[18] somewhere in the Cambrian period, and since then has been remarkably stable throughout the Phanerozoic eon.

[a] *Dr. B. Marie, Dr. F. Marin, Dr. G. Luquet*
UMR CNRS 5561 Biogéosciences, Université de Bourgogne
6, bd Gabriel, 21000 Dijon (France)
Fax: (+33) 3-80-39-63-87
E-mail: benjamin.marie@u-bourgogne.fr
gilles.luquet@u-bourgogne.fr
frederic.marin@u-bourgogne.fr

[b] *A. Marie, L. Dubost*
Muséum National d'Histoire Naturelle
Plateforme de Spectrométrie de Masse et de Protéomique
Département Régulation Développement et Diversité Moléculaire
FRE 3206 CNRS, Molécules de Communication
et Adaptations des Micro-Organismes
75005 Paris (France)

[c] *Dr. L. Bédouet, Dr. C. Milet*
UMR CNRS 5178, Biologie des Organismes Marins et Ecosystèmes
MNHN, 75231 Paris (France)

[d] *Prof. G. Alcaraz*
UPSP PROXISS, Département Agronomie Environnement
ENESAD, 21000 Dijon (France)

Secondly, nacre is usually considered a rather primitive microstructure.[19,20] Thirdly, nacre is restricted to the phylum Mollusca, where it occurs in four of the five classes that constitute the conchiferan mollusks: monoplacophorans, bivalves, gastropods, and cephalopods.[19] Taken together, these elements give rise to one fundamental question: are all nacre microstructures constructed from similar matrix assemblages? Giving an answer to this question should shed light on the process of recruitment of nacre matrix proteins in the Cambrian, and on the evolutionary constraints exerted on these proteins during the Phanerozoic.

To this end, we have investigated the biochemistry of the nacre matrix of the cephalopod *Nautilus macromphalus*. The choice of this model is dictated by the basal position of the genus *Nautilus* in the diverse phylogenetic trees of the class Cephalopoda.[21] It is the representative of the order Nautiloida, the most ancient lineage of the extant cephalopods[22] and the only one to have conserved an external calcified shell with a nacroprismatic texture, truly homologous to that of bivalve or gastropod. Furthermore, although the nautilus has been the focus of several ultrastructural[5,23–26] and biochemical investigations into the bulk shell matrix,[9,27–29] only a few papers have dealt with the detailed characterization of its shell proteins,[30–32] and only three short N-terminal sequences are so far available.[33] The aim of this work was to characterize the shell nacre matrix of the nautilus by combining the biochemical and proteomic approaches. It emphasizes the key role displayed by saccharide moieties in the modulation of calcium carbonate crystals. Finally, it sheds new light on the macroevolution of nacre matrix proteins.

Results

Extraction of organic matrix

Like those of other nautiloids, the outer wall of the shell of *N. macromphalus* has a bilayered structure, made up of an outer prismatic and an inner nacreous layer, and the septal inner shell is only constituted of nacre. The whole shell is aragonitic.

We carefully removed the periostracum, the septa, and the prismatic layer from the shell preparation and subsequently extracted the matrix associated exclusively with nacre. The acid-insoluble matrix (AIM) represents around 4% by weight of the nacre powder. The acid-soluble matrix (ASM) represents only 0.1% by weight of the nacre powder and consists of 70–75% proteins as determined by the bicinchoninic acid (BCA) protein assay.

FTIR profiles of AIM and ASM

Figure 1A shows the FTIR profiles of nacre AIM and ASM, which also testify that the decalcification was complete because no aragonite band was observed.[34,35] Both AIM and ASM exhibit characteristic bands of proteinaceous and/or gly-

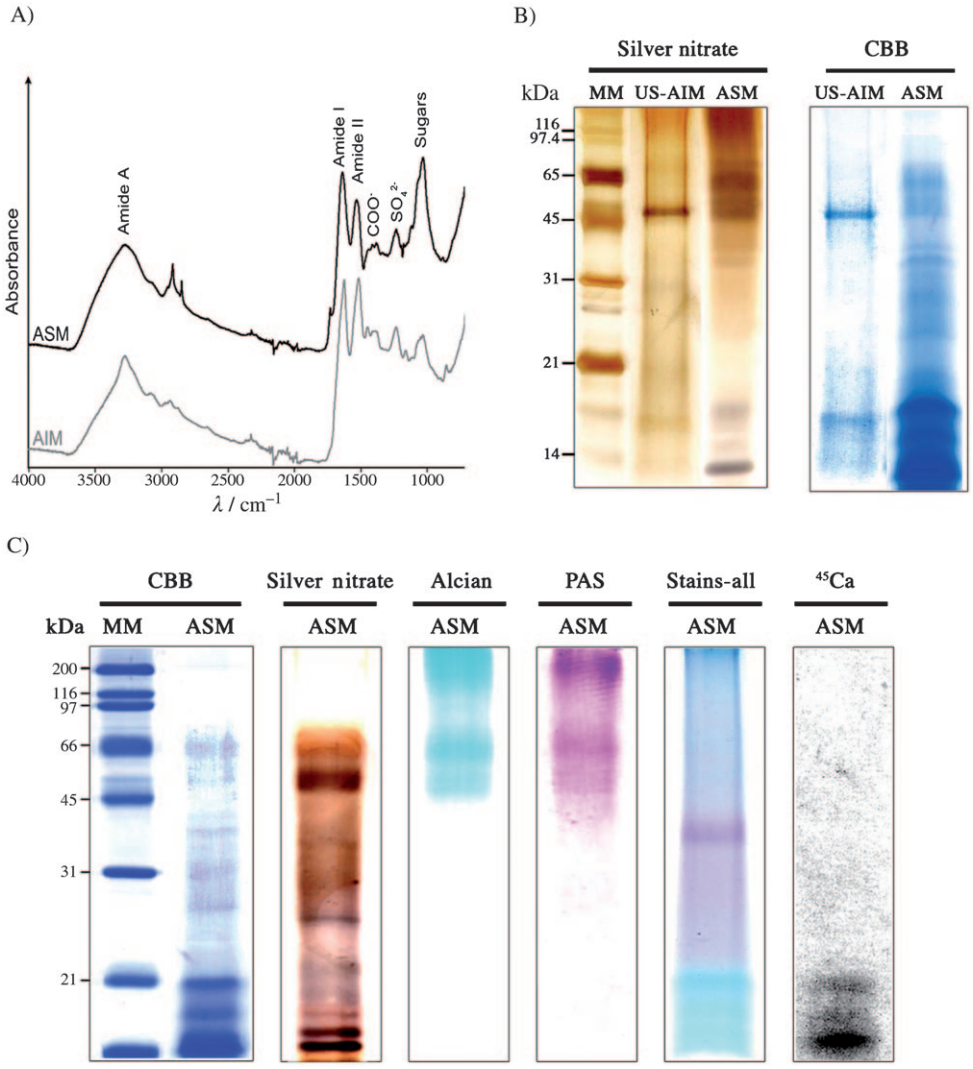


Figure 1. Macromolecular and functional compositions of the organic matrix of *Nautilus macromphalus* nacreous shell layer. A) Infrared spectra of the acid soluble (ASM, black line) and the acid insoluble (AIM, gray line) matrices obtained by complete decalcification of the nacre powder in acetic acid (5%, 4 °C). B) Electrophoretic analysis of the urea-soluble extracts of the AIM (US-AIM) and of the ASM on 12% SDS-PAGE, under denaturing conditions. The AIM was suspended in urea (8 M, 60 °C, 2 h) before electrophoresis. The gels were stained with silver (left) or Coomassie brilliant blue (CBB; right). C) SDS-PAGE analysis of the ASM of *Nautilus macromphalus* nacre. Acrylamide gels (12%) were stained with CBB, silver nitrate, Alcian blue, periodic acid Schiff (PAS), Stains-all. The result of the calcium overlay test (^{45}Ca) is shown on the far right. A similar amount of shell matrix (40 μg) was loaded on each lane; MM: molecular mass markers.

coproteinaceous components: the thick band around 3270 cm^{-1} is attributable to the amide A group (N–H bonds), the two small bands at 2916 cm^{-1} and 2850 cm^{-1} were assigned to the C–H bonds, and the two noteworthy bands near 1640 cm^{-1} and 1530 cm^{-1} were ascribed to the amide I (C–O bond) and the amide II (C–N bond) groups, respectively, commonly associated with proteins. This is in agreement with previous works on organic compounds in other mollusk shells.[29,32,36–38] Carboxylate (COO^-) and sulfate (SO_4^{2-}) absorption bands are also present in both extracts, at around 1420 cm^{-1} and 1234 cm^{-1}, respectively.[38] On the other hand, ASM exhibits a strong carbohydrate absorption band around 1060 cm^{-1}, whereas for AIM this band is weaker. This suggests that ASM contains a more important saccharide fraction.

Amino acid compositions

The total amounts of amino acids, as determined from amino acid quantification, are 16 and 66% *w/w* of the shell nacre AIM and ASM, respectively. We consider that the weak solubility of the AIM, even in HCl (5.7 N), produces a significant underestimation of its proteinaceous content. Table 1A indicates the amino acid compositions (in mol%) of the AIM and ASM. Because of deamidation of Asn and Gln during the acid hydrolysis, Asx and Glx residues represent Asn + Asp and Gln + Glu, respectively. In the AIM extract, Gly, Ala, and Glx are prominent amino acids, followed by Ser and Asx. The first three constitute almost half of the total amino acids. This result is consistent with those previously obtained with other *Nautilus* species,[31] because the AIM exhibits the signature of hydrophobic extracts, which characterizes the "silk fibroin-like" proteins[10,27] found in the nacre of various mollusks.[37,39,40] The composition of the ASM is characterized by large amounts of Gly and Asx, which represent more than two-fifths of the total amino acids. If most of the Asx residues are in their acidic form, as previously suggested,[28] this implies that the ASM of *Nautilus macromphalus* nacre is acidic, an assumption that is further confirmed on 2D gels.

Table 1. Compositions of the AIM and ASM extracts of *Nautilus macromphalus* nacre.

Amino acids[a]	% of the total amino acids		Monosaccharides[b]	ng per μg of matrix (% of the total)	
	AIM	ASM		AIM	ASM
Asx	9.5	20.8	fucose	0.8 (1)	2.5 (4)
Glx	10.3	7.5	rhamnose	ND	2.5 (4)
Ser	9.6	6.8	arabinose	1.7 (3)	1.0 (2)
His	0	2.1	galactose	10 (16)	8.0 (14)
Gly	24.0	21.4	glucose	9.5 (15)	9.3 (17)
Thr	2.8	4.8	mannose	ND	1.3 (2)
Ala	13.3	7.3	xylose	ND	ND
Arg	5.6	3.1	galactosamine	4.6 (8)	6.5 (12)
Tyr	2.8	2.9	glucosamine	34.2 (56)	9.7 (17)
Cys	0	1.1	galacturonic acid	ND	3.0 (5)
Val	3.4	3.1	glucuronic acid	ND	3.8 (7)
Met	0.9	0.7	Neu5Ac	TR	TR
Phe	4.8	2.7	Neu5Gc	ND	ND
Ile	2.9	2.3	sulfated sugars	0.5 (1)	8.3 (15)
Leu	5.9	3.2	total	61.3 (100)	55.9 (100)
Lys	2.2	3.9			
Pro	2.0	6.4			
Trp	ND	ND			

[a] Amino acid composition: data are presented as molar percentage of total amino acids for each extract. Note that Asx = Asn + Asp and Glx = Gln + Glu. Cysteine residues were quantified after oxidation. Tryptophan residues were not detected due to the hydrolysis conditions. [b] Monosaccharide composition: the neutral sugar compositions were obtained by high-performance anion exchange-pulsed amperometric detection (HPAE-PAD). Data are represented in ng per μg of the total matrix and as percentages of the total identified carbohydrate compounds; ND: not detected; TR: trace amounts.

Monosaccharide composition

The total amounts of neutral, acidic, and amino sugars obtained after TFA (2 M) hydrolysis represent 6.1% and 5.2% of the AIM and the ASM, respectively (Table 1B). However, for the AIM, the standard hydrolysis for monosaccharide determination does not release the total sugar content, as can be seen in the presence of insoluble residues after hydrolysis. The AIM yields only six monosaccharide residues, out of which glucosamine represents 56%, galactose 16%, and glucose 15% of the composition. The technique used does not allow glucosamine to be distinguished from its N-acetylated form, which is the monomer of chitin. Chitin has been detected in the insoluble nacre matrix of several mollusks[41,42] including *Nautilus macromphalus*.[9,23] If it is assumed that a large fraction of the glucosamine residues originate from the hydrolysis of chitin, this macromolecular constituent could represent an important fraction of the AIM.

The ASM monosaccharide composition appears to be more diversified than that of the AIM, because it contains six of the seven standard neutral sugars (with the exception of xylose), the two amine sugars, and also noticeable amounts of glucuronic and galacturonic acids. As in the case of the AIM, in which glucosamine, glucose, galactose, and galactosamine represent 95% of the monosaccharide composition, these four residues are also dominant in the ASM, with 17, 17, 14, and 12%, respectively. Traces of sialic acids (Neu5Ac only) were detected in both matrices.

Because previous qualitative methods had determined the presence of acidic sulfated sugars in the nacre of *Nautilus* sp.,[15,24,32] an accurate quantification of these sugars in the AIM and in the ASM was performed by spectrophotometry (Table 1B). Significant amounts, representing 15% of the total sugar content, were quantified in the ASM. The AIM contains only traces of sulfated sugars.

Characterization of matrices by SDS-PAGE

The treatment of the AIM with urea dissolves the insoluble components only partly. The urea-treated fraction is called urea-soluble AIM. When resolved on mini-SDS-PAGE and stained with silver or Coomassie brilliant blue (CBB), the urea-soluble AIM (US-AIM) and the ASM were found to be composed of various distinct macromolecular components (Fig-

ure 1B). The US-AIM shows two main bands migrating around 50 and 18 kDa. A few minor bands around 30 and 20 kDa are also visible, although only after silver staining. Silver and CBB stain the nacre ASM in different ways. The silver staining reveals high-molecular-weight components between 200 and 100 kDa, two thick bands around 63 and 52 kDa, and small discrete bands between 20 and 12 kDa. Although silver staining is much more sensitive than CBB, its reproducibility is poorer: the 100–200 kDa components sometimes appear negatively stained, and the other components exhibit variations in their staining intensities.

CBB does not reveal the high-molecular-weight components between 200 and 100 kDa, suggesting that these components are either heavily glycosylated or are strongly negatively charged. CBB does stain the two thick bands around 63 and 52 kDa and a few compounds between 40 and 22 kDa fairly strongly, and it heavily stains the small discrete bands between 20 and 12 kDa. Previous observations of CBB-stained electrophoretic profiles obtained from EDTA- or acetic acid-soluble matrices of *Nautilus pompilius* shell[30,31] revealed thick bands above 45 kDa and a few discrete compounds between 20 and 10 kDa.

Because silver-nitrate-stained profiles of the ASM can show negative staining in the 100–200 kDa range (Figure 1C), we investigated the biochemical properties of this extract further by using different gel and blot staining techniques, as shown in Figure 1C. In particular, Alcian and periodic acid Schiff (PAS) procedures heavily stain the high-molecular-weight compounds, as well as the two discrete bands at 63 and 52 kDa. The Alcian blue staining method performed at pH 1 is specific for the very acidic sulfated sugars, while PAS reveals vicinal diol groups on peripheral sugars with a specific red/purple color. Taken together, these results suggest that the high-molecular-weight compounds, together with the 63 and 52 kDa bands, are acidic sulfated glycoproteins.

The carbocyanine dye stains the 37 kDa band in dark blue, whereas the proteins below 20 kDa exhibit a particular light blue color, different from the metachromatic blue characteristic of calcium-binding proteins. Nevertheless, the ability of the low-molecular-weight proteins to bind calcium ions was unambiguously demonstrated by the ^{45}Ca overlay test under denaturing conditions.

Interaction of the ASM with in vitro calcite growth

The results of this test are shown in Figure 2. In the control blank experiment, the crystals exhibited the typical habitus of calcite (Figure 2A, B). The ASM was tested at concentrations varying from 0.1 to 20 $\mu g\,mL^{-1}$. In that range, we checked two parameters: the crystal size and shape, and the density of crystals per surface unit. Significant effects were observed above a threshold concentration of 0.5 $\mu g\,mL^{-1}$. At 0.5 $\mu g\,mL^{-1}$, most of the crystals were polycrystalline and exhibited smooth faces (Figure 2C, D). At 1 $\mu g\,mL^{-1}$, the number of crystals increased (Figure 2E) and all of them were polycrystalline, with rounded edges and microsteps (Figure 2F). At 2 $\mu g\,mL^{-1}$, the density of crystals decreased moderately (Figure 2G) and the polycrystalline aggregates exhibited foliated faces (Figure 2H). At 5 $\mu g\,mL^{-1}$, the number of crystals did not vary (Figure 2I), but the angles of crystals appeared more rounded, with slightly curved faces (Figure 2J). At 10 $\mu g\,mL^{-1}$, the density of crystals decreased (Figure 2K) and they exhibited rounded shapes (Figure 2L). At 20 $\mu g\,mL^{-1}$, the size distribution of the crystals was heterogeneous (Figure 2M), but the majority were smaller than those produced at lower ASM concentrations. This decrease in size was probably due to the inhibiting effect of the ASM at high concentration. The formed polycrystalline aggregates had curved faces, edges, and corners (Figure 2N). At a higher mag-

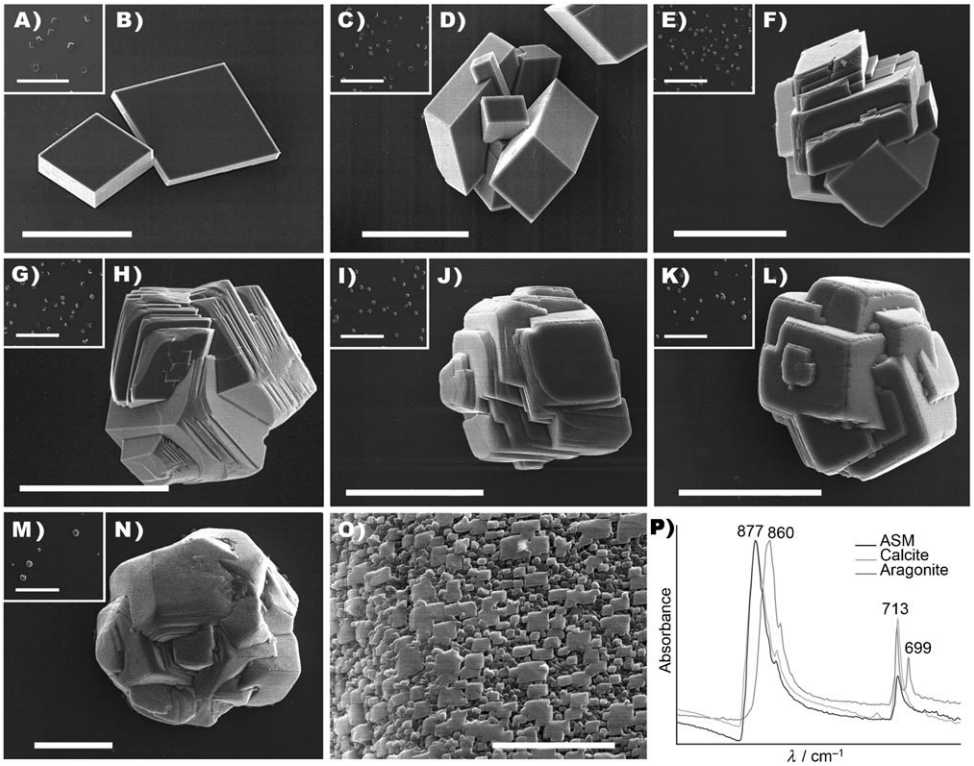


Figure 2. Effects of *Nautilus macromphalus* nacre ASM on $CaCO_3$ crystal growth. SEM micrographs of calcite crystals grown in vitro, with different concentrations of ASM in the 0.1–20 $\mu g\,mL^{-1}$ range. A), B) Negative control without matrix; C), D) 0.5 $\mu g\,mL^{-1}$ ASM; E), F) 1 $\mu g\,mL^{-1}$ ASM; G), H) 2 $\mu g\,mL^{-1}$ ASM; I), J) 5 $\mu g\,mL^{-1}$ ASM; K), L) 10 $\mu g\,mL^{-1}$ ASM; M), N) 20 $\mu g\,mL^{-1}$ ASM; O) detail of N). Scale bars: 600 μm for A), C), E), G), I), K), and M); 30 μm for B), D), F), H), J), L), and N); 2 μm for O). P) FTIR profiles of the crystals obtained with nacre ASM, in comparison with those of aragonite and calcite standards. Under these conditions, we only observed the formation of calcite crystals.

nification, the faces appeared to be formed by submicron crystallites of regular quadratic shape (Figure 2O). We checked the polymorph of the newly formed $CaCO_3$ crystals by FTIR and observed the presence of bands specific to calcite (877 and 713 cm^{-1}) whether or not in the presence of nacre ASM (Figure 2P). No aragonite bands (860 and 699 cm^{-1}) were detected.

Separation of ASM proteins by 2D gel electrophoresis (2DE) and mass spectrometry analysis

The results of the two-dimensional gel analysis of nacre ASM are shown in Figure 3. The bands previously observed by 1D-electrophoresis can be retrieved. Most of the ASM components are concentrated either at very acidic or very basic pI values.

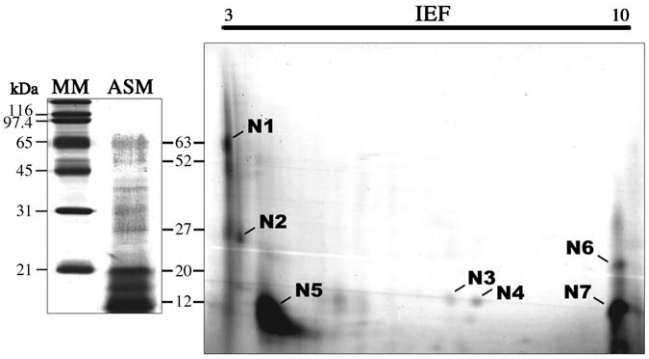


Figure 3. 2D gel electrophoresis (2DE) analysis of *Nautilus macromphalus* nacre ASM. The monodimensional gel (left) with the ASM extract shows the correspondence between the protein bands and the spots observed on the 2DE (right) after CBB staining, by using a pH 3–10 immobilized pH gradient (IPG) strip in the first dimension; MM: molecular mass markers. The seven most intense spots (N1–N7) were excised from the gel and further analyzed by MALDI-TOF mass spectrometry after reduction, alkylation, and digestion with trypsin.

The two prominent 63 kDa (spot N1) and 52 kDa glycoproteins are clustered at pI 3; this suggests that these proteins each exhibit a pI of 3 or less. On the other hand, the set of low-molecular-weight proteins migrating between 20 and 12 kDa appears to be composed variously of very acidic (N5), neutral (N3 and N4), and very basic components (N6 and N7). Interestingly, the 12 kDa band is not homogenous and divides into two large spots around pI 3.5–4 and pI 9–10 (N5 and N7, respectively).

Seven spots were excised from the gel for protein characterization by MALDI-TOF/TOF with peptide mass fingerprinting (PMF) data search after digestion with trypsin (see one example with spot N1 in Figure 4A). For the main peptides, the MS/MS spectra were acquired for de novo sequencing and results were subjected to a BLAST data search. The peptides obtained after trypsin digestion from the 2DE spots were searched against the NCBI nonredundant database with use of MASCOT software and no significant result was found. Because of the absence of genomic/transcriptomic database on *Nautilus* sp., it is not surprising that this approach could not provide a reliable result for identifying proteins. Interestingly, we observed that many peptides were common to various spots (Table 2); this suggests that the different ASM proteins can share similar peptide domains. In spite of the possibility that peptides of different sequences could share the same *m/z* value, the numerous cases of similar *m/z* values shown in Table 2 indicate that the peptides are more likely truly identical.

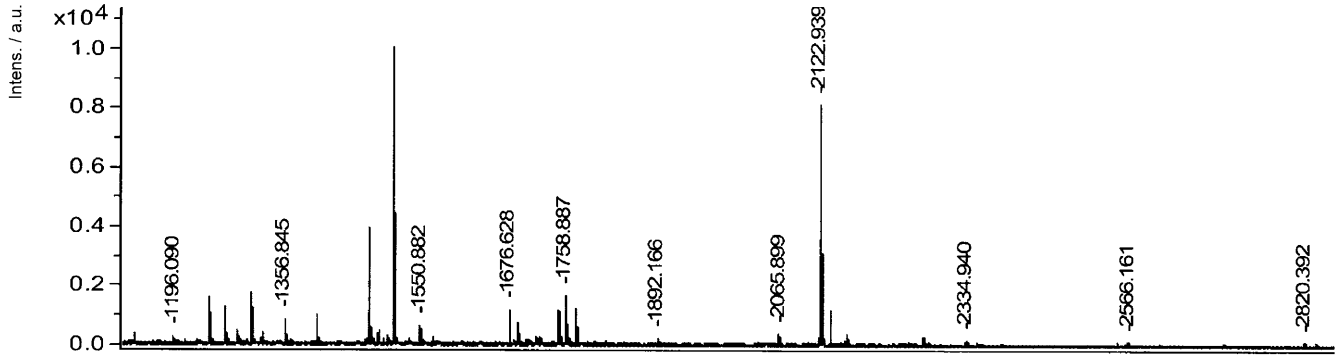
We performed LC/MS-MS analysis both of the ASM and of the AIM after hydrolysis with trypsin. This approach provided an important set of MS and MS/MS data for both extracts. When searched against MASCOT, no protein was identified. Strikingly, the ASM and the AIM possess many peptides of the same *m/z* value. By comparing the ASM and AIM MS/MS spectra two by two, we determined that 33 different peptides of the two matrices strictly share the same MS/MS-spectra patterns, as exemplified in Figure 5. Our results could have two implications: proteins of the AIM, on one hand, and the ASM, on the other, might be made up of similar peptide blocks, or alternatively, some proteins of the ASM might also be present in the AIM extract.

Partial and full internal sequences were obtained by interpretation of MS/MS spectra for de novo sequencing (see, for example, Figure 4A–C). The interpretation of the mass differences between ions was assisted by use of PEAKS or Analyst QS1.1 software, and sequences were further confirmed manually. We only considered amino acid positions that were unambiguously determined with the *y*-ion series and confirmed with the *b*-ion series. Only a few peptides analyzed in MS/MS mode could be fractionated or gave good MS/MS spectra for sequence interpretations. It is very likely that some of the peptides digested with trypsin bear glycosyl moieties (for example, for N1) or other post-translational modifications, which restrain the MS/MS fractionation. With the 2DE coupled with MALDI-TOF/TOF approach, we obtained four peptide sequences for the N1, N5, and N7 spots (Table 3). From LC-MS-MS/MS analysis, we deduced 14 and 20 internal sequences for the ASM and AIM extracts, respectively. Two peptides of 675.35 and 675.85 *m/z* from the AIM, share the same amino acid sequence, and differ only in one residue—N or D—at the fifth position. This one-residue polymorphism could be explained by the presence of two isoforms of a particular domain in different matrix proteins or the occurrence of a repeat within the same protein as already observed for mucoperlin.[43] We also noticed that the 492.26 and 729.38 *m/z* peptides share identical internal sequences (DSSVLT).

Sequence alignments

Only the sequences longer than seven amino acids were further considered for SIM alignment or BLAST. We performed an alignment two by two with all the 65 known molluskan shell proteins,[13] using the SIM tool from the Expasy server (http://www.expasy.org/tools/sim.html). In no case did we obtain a perfect match. In some cases, however, some alignments might be significant (Figure 6A–E). Three peptides match partially with nacrein and nacrein-like proteins (Figure 6A)—two nacre-associated proteins that are known to exhibit carbonic

A)



B) De novo results

Search parameter: MS Tol.: 50.0 ppm, MSMS Tol.: 0.4 Da

Tree hierarchy	Calcd $[M+H]^+$	Meas. M_w	Calcd M_w	Dev.(Da)	Score	Sequence (N ter – C ter)
MSMS 1	2122.976	2121.922	2121.968	-0.046	16411	(LGSSDP)SDCACLHALGHVAR 9:&11: Carbam (C)

C)

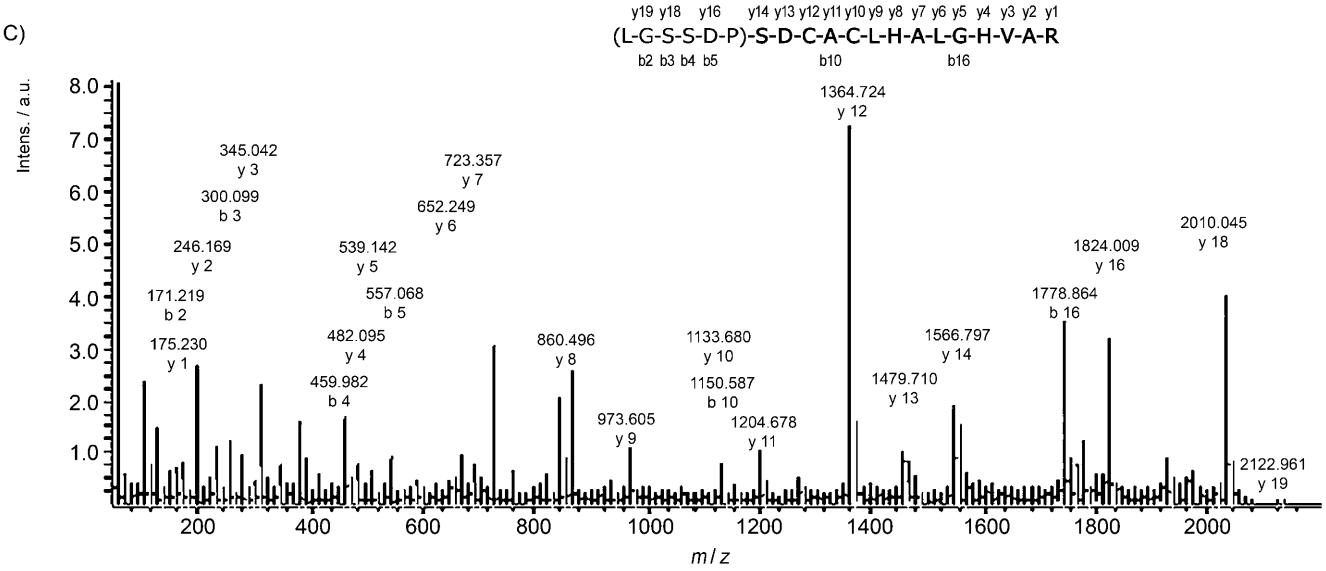


Figure 4. Proteomic approach performed on *Nautilus macromphalus* nacre ASM proteins after separation by 2DE. Example of the analysis performed on the Naut_smpp1 peptide (*m/z* 2122.94, monocharged) from the N1 spot. A) Tryptic PMF of the N1 spot protein acquired in the linear mode by MALDI-TOF with HCCA matrix. The most intensive peaks were manually selected for MS/MS fractionation. The MASCOT search engine was used for protein identification by comparison of the mass spectra of the fragmented peptides. B) Result of the RapiDeNovo software interrogation for MS/MS spectra interpretation for the 2122.94 *m/z* peptide fractionation. A high score indicates good correspondence between MS/MS spectra of a fractionated peptide and the sequence proposed by the software. The sequences were further validated by manual de novo sequencing. C) MS/MS spectrum of the 2122.94 *m/z* peptide and manual interpretation of the spectrum for sequence identification. We only consider amino acid positions that were unambiguously determined. Uncertain amino acids are mentioned in brackets but not validated, and peptide sequences with uncertain amino acids in an internal position were not further considered. In this example, we validated the sequence SDCACLHALGHVAR for further BLAST, alignment with the SIM tool, and sequence registration on the Swiss-Prot database.

anhydrase (CA) activity. In spite of this sequence similarity, we did not detect any CA activity in the ASM extract of *N. macromphalus* (data not shown). Similarly, three other peptides match significantly with N14, N16, and pearlin (Figure 6B). Tyrosinase-like proteins 1 and 2 are also similar to two peptides from *N. macromphalus* shell proteins (Figure 6C). In mollusks, tyrosinase-like proteins are suggested to play an essential role in quinone-tanning processes, by converting tyrosine residues into DOPA. In the shell, they may be directly involved in the phase transformation of a gel or of soluble proteins into insoluble "framework" proteins. Three peptides match with mucoperlin (Figure 6D), a mucin-like protein from *Pinna nobilis* nacre.[43] Furthermore, weak similarities are observed between two peptides of the AIM and low-complexity sequences from MSI31 and MSI60 (Figure 6E), but these similarities might be fortuitous as a result of the bias due to the dominance of a single amino acid. Surprisingly, all of the few similarities detected are with shell proteins of bivalve, and not gastropod, origin.

Table 2. Tryptic PMF of the spot proteins from the *Nautilus macromphalus* shell nacre recorded by MALDI-TOF.[a]

Spot name	*m/z* values (monocharged ions)[b]
N1	1196.09, 1356.85*, 1394.79, 1489.71*, 1505,68, 1512.94*, 1550.88*, 1567.78, 1676.63*, 1690.63, 1692.63, 1723.43, 1748.93*, 1758.89*, 1774.89, 1892.17, 2065.90, **2122.94***, 2334.94, 2566.19, 2820.39, 2835.37
N2	1197.25, 1361.81*, 1433.65, 1459.63, 1490.78, 1512.92*, 1645.95, 1722.93*, 1723.41, 1760.86, 1765.76*, 1774.87, 1791.74*, 1837.98, 1876.72, 1892.15, 1900.89, 1909.02, 2122.91, 2705.18, 3162.28*, 3211.46, 3312.31
N3	1998.94*, 2711.17*
N4	1584.73*, 1941.94, 1998.95*, 2653.72, 2695.19, 2711.20*, 4017.69, 4034.76
N5	1003.37*, 1013.54*, 1029.52*, 1286.32*, 1310.57*, 1320.62, 1482.66*, 1513.77*, 1523.61, 1524.59*, 1531.52, 1580.63, 1655.65, 1701.55, 1707.78, 1722.91, 1740.85, 1744.88, 1758.87, 1765.75, 1791.77, 1811.88, 1836.96*, 1858.58, 1873.77, **1874.76***, 1912.71, 1931.77, 1975.59, 2017.02, 2033.02*, 2312.14, 2399.00, 2689.40, 2740.14, 2817.49, 3312.35, 3621.65
N6	1140.55, 1169.69, 1184.68, 1188.64*, 1260.54*, 1274.55*, 1286.29, 1325.78*, 1364.67, 1401.63, 1457.71*, 1482.63*, 1490.73, 1508.74, 1555.80, 1568.73, 1585.80*, 1645.90, 1660.79, 1688.11, 1689.82, 1722.90, 1738.01*, 1745.90, 1761.78, 1774.83, 1805.75, 1876.75, 1901.88, 1943.92, 2105.94, 2192.97, 2309.04*, 2465.12, 2576.09, 2732.16, 3211.42, 3273.43, 3429.41, 4407.92
N7	1003.38, 1286.32*, 1374.60, 1391.65, 1425.64*, 1438.67, 1465.64, 1467.64, **1482.67***, 1502.47, 1520.62, 1531.55, 1602.67, 1625.71, 1627.70, 1669.69, 1671.68, 1696.83, 1702.47*, 1743.83*, 1758.85, 1759.84, 1760.83*, 1796.79, 1819.76*, 1836.74, 1859.75, **1876.79***, 1893.79*, 1914.74*, 1952.58, 2672.15, 3079.37, 3085.83, 3127.38*, 3170.36, 3255.44, 3270.43, 3314.38, 3553.56, 3695.60

[a] Numerous spots were composed of many peptides with the same *m/z* values. For example, the spot protein N5 shares three peptide ions with N6 (*m/z* 1286.32, 1482.66, 1722.91) and five peptide ions with N7 (*m/z* 1003.37, 1286.32, 1482.66, 1531.52, 1758.87). This suggests that the N5 protein (pI ~ 3.5–4) and N6 or N7 proteins (pI 9–10) could exhibit similar domains. On the other hand, the N2 spot protein has five peptide ions in common with N1 (*m/z* 1512.92, 1723.41, 1774.87, 1892.15, 2122.9), three peptide ions in common with N5 (*m/z* 1722.93, 1791.74, 3312.31), and six peptide ions in common with N6 (*m/z* 1490.78, 1645.95, 1722.93, 1774.87, 1876.72, 3211.46). This suggests that N2 proteins share similar amino acid sequences with N1, N5, and N6. [b] Underlined values: peptides with the same *m/z* in different protein spots; bold values: peptides from which sequences were obtained by MS/MS analysis; *: peptides analyzed in MS/MS mode.

We did not find any homology with the few sequences published by Zhao and co-workers on *Nautilus pompilius*.[33]

Discussion

We have investigated the biochemical properties of the nacre-associated matrix extracted from the shell of the cephalopod *Nautilus macromphalus*. In classical views on molluskan shell biomineralization, the shell matrix can be divided into soluble and insoluble fractions: after mineral dissolution, the insoluble macromolecular complexes are recovered by simple centrifugation, whereas the soluble components require additional concentration/dialysis/purification steps. The proportion of insoluble to soluble varies widely,[44] but nacre is always characterized by a high proportion of insoluble matrix,[45] whatever the extraction technique used. In the present case, with a matrix content above 4 wt%, the nacre of *Nautilus macromphalus* probably represents the molluskan shell microstructure with the highest organic content, and its AIM/ASM ratio is about 40. Our analyses performed on the bulk matrices make a clear distinction between these two fractions. The urea soluble parts of the AIM and ASM exhibit different electrophoretic patterns. The AIM is enriched in aliphatic residues, mainly Gly and Ala, and in glucosamine. Gly and Ala are characteristic of "framework" proteins,[10] whereas large amounts of glucosamine indicate the likely presence of chitin.[29] The ASM is more hydrophilic, and enriched in Asx residues, whereas its sugar composition is shared out between different monosaccharide residues.

This clear dichotomy between the AIM and the ASM is largely attenuated by our MS and MS/MS data, which show that several tryptic peptides are present in both matrices. Our finding is not an isolated case in calcium carbonate biomineralization: recently, Bédouet and co-workers[46] have noted that the insoluble and water-soluble matrices of the pearl oyster *Pinctada margaritifera* contain identical MSI31 peptides; Mann et al. have made a similar observation for the eggshell matrix of the hen *Gallus gallus*.[47] This fact may be explained in different manners. Firstly, soluble proteins of the shell matrix can become insoluble correlatively to their ability to form homopolymers of increasing size. Self-polymerization of a single shell protein has indeed been observed for P20[48] and for caspartin.[49] Secondly, soluble proteins can be components of complex insoluble heteropolymers, which result from cross-linking between matrix macromolecules. Such a phenomenon, called quinone-tanning, has been observed during the formation of the periostracum.[50,51] It may occur as well in the formation of nacre tissues, in particular for hardening of the chitino–protein framework.[17] Thirdly, from recent reviews on shell proteins,[12,13] the primary structures of several shell matrix proteins exhibit organization in modules. Some of our recent data suggest that these modules may serve as "functional blocks" in different proteins, soluble or not.[52,53] Finally, we cannot exclude a mechanism opposite to polymerization: protein cleavage. Some proteins of the AIM may indeed maturate and be cleaved into short peptides, which would be easily solubilized. The degradation of a long precursor into shorter peptides of well-defined size is largely documented for enamel proteins, in particular for amelogenin and MEPE.[54,55]

The proteomic approach that we used allowed 38 amino acid sequences, of six to 18 residues long, to be obtained. We are fully aware that this sequence pool represents only a small part of the whole sequences of the shell matrix proteins, thus giving a rather partial picture of the complete shell proteome of *Nautilus macromphalus*. Furthermore, the de novo sequencing by mass spectrometry presents some technical limitations: it does not work on proteins, which lack appropriate cleavage sites (arginine and lysine in the case of trypsic digestion) and the resulting peptides, if not ionized, are not detectable. The technique might thus introduce a bias in the representativeness of the analyzed peptides. Finally, our approach focuses

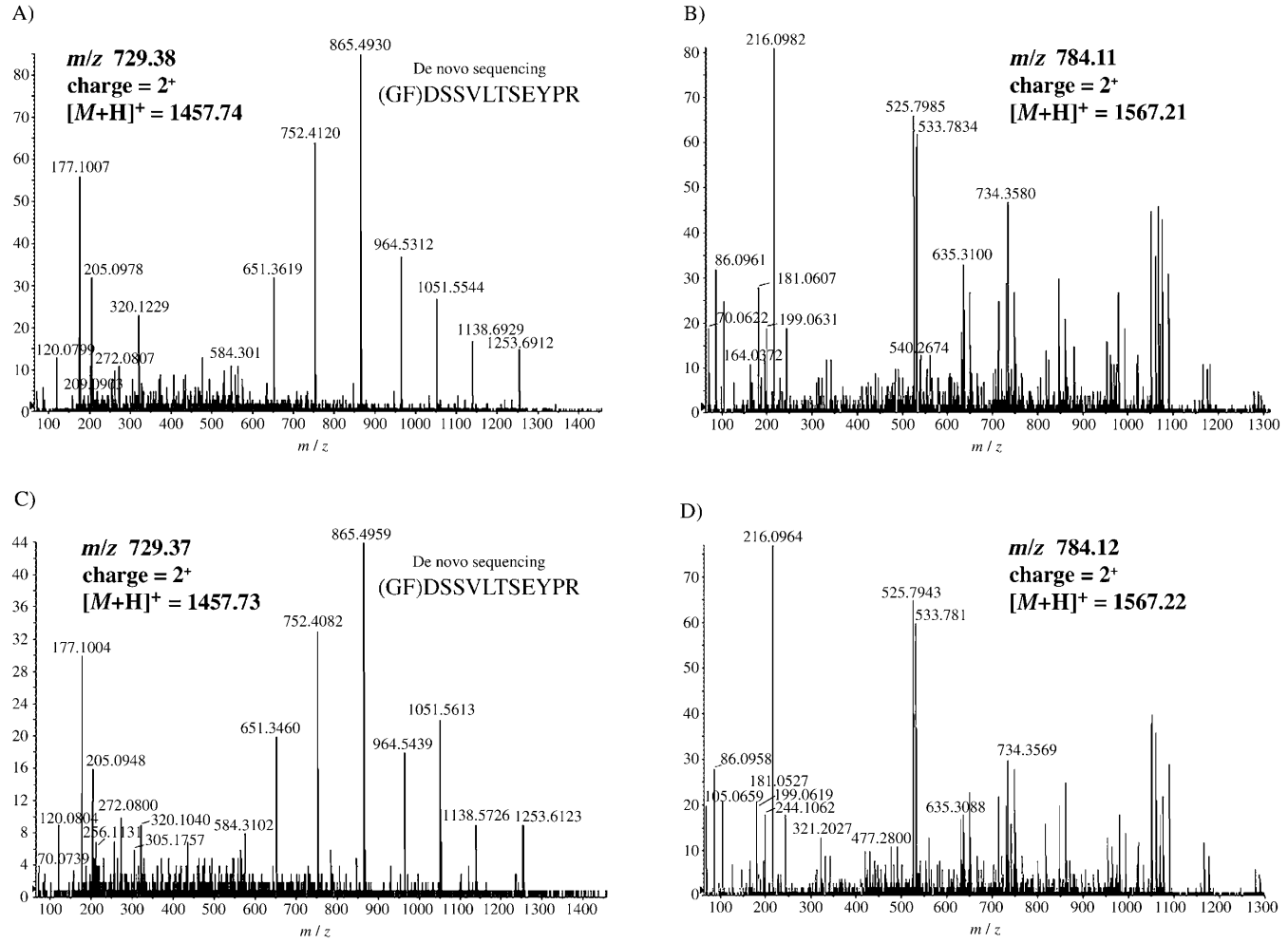


Figure 5. The *Nautilus macromphalus* nacre ASM and AIM share common proteins. Examples of MS/MS spectra of peptides that exhibit similar *m/z* values, acquired by ESI-QqTOF/TOF from trypsin digestion of: A), B) ASM, and C), D) AIM. The minimum ion intensity was set to ten counts. The ion-spray potential and the declustering potential were 5200 and 50 V, respectively. The collision energies for the gas-phase fragmentations of the precursor ions were determined automatically by IDA on the basis of their *m/z* values. A), C) MS/MS spectra of 729.4 *m/z* peptides from ASM and AIM tryptic digests, respectively. For these two peptides, de novo sequencing gave the same sequence, (GF)DSSVLTSEYPR. B), D) MS/MS spectra of 784.1 *m/z* peptides from ASM and AIM tryptic peptides, respectively. These two peptides share the same MS/MS profiles, but because of the complexity of the spectrum sequencing was not possible for this example. By using this comparative approach for peptides of similar *m/z* value from the two extracts, we were able to show that the ASM and AIM share at least 33 common peptides that exhibit the same MS/MS profiles and the same amino acid sequences.

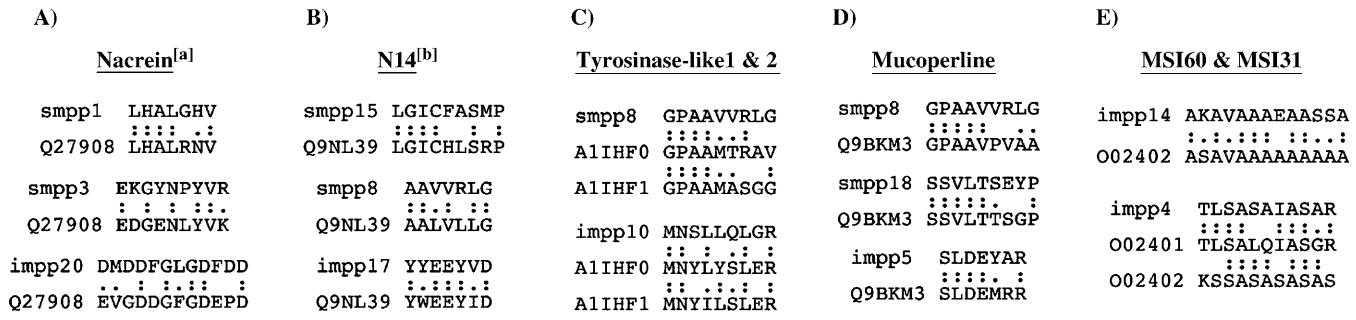


Figure 6. Alignment of *Nautilus macromphalus* shell matrix sequences with previously characterized molluskan shell proteins. A) Multiple alignments with *P. fucata* nacrein (accession number Q27908). Note that similar alignments are observed with the six nacrein-like proteins (A0ZSF2, A0ZSF3, A0ZSF4, A0ZSF5, A0ZSF6, and A0ZSF7). B) Multiple alignments with N14 (Q9L39). Similar alignments are observed with N16 and perlin (O97048, Q14A6). C) Multiple alignments with tyrosinase-like proteins 1 and 2 (A1IHF0, A1IHF1). D) Multiple alignments with mucoperlin (Q9BKM3). E) Alignments with MSI60 (O02402) and MSI31 (O02401). Comparisons of sequences were performed two-by-two with the new sequence of *N. macromphalus* shell matrix and the 65 full or partial sequences of mollusk shell proteins (reviewed in ref. [18]), with the SIM tool from the Expasy software (http://www.expasy.org/tools/sim.html). [a] Similar conserved domains exhibiting partial alignment with *N. macromphalus* shell proteins are also observed in narcein-like proteins. [b] Similar conserved domains exhibiting partial alignment with *N. macromphalus* shell proteins are also observed in N16 and perline.

Table 3. Full and partial (−) peptide sequences of *Nautilus macromphalus* nacre ASM and AIM extracts. Sequences were deduced from MS/MS spectra interpreted for de novo sequencing from 2DE + MALDI-TOF/TOF or LC ESI-QqTOF/TOF analysis. These protein sequence data appear in the UniProt knowledgebase under the accession numbers P85364–P85401.

Peptide	*m/z*	Charge	$[M+H]^+$	Partial (−) or full sequence	Accession No.
MALDI-TOF/TOF analysis					
Spot N1					
SMPP1	2122.94	1^+	2123.95	−SDCACLHALGHVAR	P85390
Spot N5					
SMPP2	1874.76	1^+	1875.76	DGDDGDGD−	P85398
Spot N7					
SMPP3	1482.67	1^+	1483.68	EKGYNPYVR	P85388
SMPP4	1876.79	1^+	1877.79	DMYSDNLGLCDN−	P85391
ESI-QqTOF/TOF analysis					
ASM					
SMPP5[a]	360.19	2^+	719.37	FAPCPK	P85372
SMPP6	367.72	2^+	734.43	ADLFLR	P85373
SMPP7[a]	483.78	2^+	966.56	LYSLLTEK	P85374
SMPP8	484.31	2^+	967.62	GPAAVVRLLGK	P85375
SMPP9	492.26	2^+	983.51	SFDSSVLTK	P85376
SMPP10[a]	516.79	2^+	1032.56	QLFQPPFR	P85364
SMPP11[a]	517.78	2^+	1034.55	TSFTLYMR	P85392
SMPP12[a]	528.32	2^+	1055.63	ALAQLLSLAR	P85377
SMPP13[a]	549.28	2^+	1097.54	TSTVFSDGQR	P85378
SMPP14	554.75	2^+	1108.49	DFFFDFDR	P85379
SMPP15[a]	584.31	2^+	1167.61	LGICFASMPR	P85380
SMPP16	593.82	2^+	1186.64	−FIAGLNGLR	P85381
SMPP17	681.38	2^+	1361.76	LSLQEFLGLWR	P85382
SMPP18	729.38	2^+	1457.74	−DSSVLTSEYPR	P85383
AIM					
IMPP1	372.71	2^+	744.41	FVSTYK	P85389
IMPP2	465.78	2^+	930.55	VSVVGTVR	P85384
IMPP3	470.26	2^+	939.51	YLTCNLR	P85393
IMPP4	480.78	2^+	960.54	TLASAIASAR	P85385
IMPP5	512.26	2^+	1023.52	LGSLDEYAR	P85386
IMPP6	519.77	2^+	1038.54	SFSESLAAAR	P85387
IMPP7	582.81	2^+	1164.61	TFVSSQVSGPR	P85394
IMPP8	600.83	2^+	1200.65	AAALSQAADNLR	P85395
IMPP9	625.83	2^+	1250.66	FTAASQTLEAGR	P85396
IMPP10	675.35	2^+	1349.69	NSTMNSLLQLGR	P85397
IMPP11	675.85	2^+	1350.70	NSTMDSLLQLGR	P85370
IMPP12	682.85	2^+	1364.70	LSVNYQWLVR	P85399
IMPP13	697.82	2^+	1394.64	ALVSADGDSWFAR	P85368
IMPP14	737.39	2^+	1473.78	VTAKAVAAAEAASSAR	P85371
IMPP15	762.88	2^+	1524.76	LGSPFGGFDTLGSNR	P85369
IMPP16	842.46	2^+	1683.92	PGGGVSGALSSLVQSLVR	P85365
IMPP17	862.4	2^+	1723.79	YYEEYVDGQGMVVR	P85366
IMPP18	890.10	3^+	2669.29	−MFLGPGGAGGQAFDQAR	P85400
IMPP19	978.96	2^+	1956.91	−GGGSNFGQFAGGLLDR	P85401
IMPP20	1057.4	2^+	2113.79	−DMDDFGLGDFDDGR	P85367

[a] Peptide also present in the AIM extract.

strictly on secreted proteins, which are incorporated in the shell. It does not take into consideration the "silent" secreted proteins, such as extrapallial fluid proteins, nor the proteins of the mantle epithelium, which may contribute to the shell elaboration without being constitutive of the shell matrix.[56] In spite of these technical limitations, it is puzzling to observe that most of the peptide sequences that we obtained do not exhibit any similarity with known molluskan shell proteins. This obviously points to the fundamental question of the diversity of molluskan shell proteins. Since 1996, the number of fully sequenced proteins/genes has exponentially increased, forming a wide range of protein families.[12,13] However, these characterizations have been performed on a limited number of biological models, the two most studied being the pearl oyster *Pinctada* sp. and the abalone *Haliotis* sp. Through investigation of new biological models on one hand, and through the use of new approaches on the other, such as "mantle transcriptomics"[56,57] and "shell proteomics" as here, the information on molluskan shell proteins increases at different levels: protein families, domains, and functions. Ultimately, it should allow more accurate definition of the basic "protein equipment" required for making a shell.

Another remarkable finding of our proteomic approach on *Nautilus* is that the few peptides that exhibit sequence homology with known shell proteins match with those of bivalve origin. So far, we have not found any sequence homology with the nacre proteins characterized from the archeogastropod *Haliotis*, such as lustrin A, perlustrin, perlucin, perlwapin, perlinhibin, AP7, or AP24 (reviewed in ref. [13]) or the set of proteins determined by EST.[56] In the most consensual phylogenetic reconstructions of the phylum Mollusca, cephalopods and gastropods are grouped together in the Visceroconch (also called Cyrtosoma) clade,[58] which represents the sister group of diasomes (bivalves + scaphopods). The putative phylogenetic proximity of cephalopods and gastropods would be reflected in their respective nacre textures, which exhibit similarities, in contrast to bivalvian nacre.[59] The first are columnar, and constituted of superimposed tablets;[26] the second represent the "brickwall-type", also called "sheet nacre", and nacre tablets grow "in terraces".[6] It is then surprising to observe that the morphological resemblance between nautilus and abalone nacres is not translated in a "sampling" of their protein contents. This discrepancy is puzzling, in particular because nacre textures are usually considered to be plesiomorphic[19,60] and one should expect their constitutive proteins to be highly conserved during evolution. Our findings suggest, but do not demonstrate, that nacre proteins might be less evolutionarily constrained than expected, and/or that similar types of nacre might be constructed through different biochemical pathways, by use of different "macromolecular tools".

Experimental Section

Shell material and matrix extraction: Fresh shells of 20–25 mm in diameter from the cephalopod *Nautilus macromphalus* were collected on the coast of New Caledonia (Pacific). The external prismatic layers were removed by abrasion under cold water. Whole shells were mechanically crushed and fragments of the siphon were removed. The shell fragments were immersed in NaOCl (1%, *v/v*) for 24 h to remove superficial contaminants, and were then thoroughly rinsed with water. All the following extraction procedure was performed at 4 °C. The shell powder (< 200 μm) was suspended in MilliQ water and decalcified, overnight, in cold dilute acetic acid (5%, *v/v*), which was added from an automated burette (Titronic Universal, Schott, Mainz, Germany) at a flow rate of 100 μL every 5 s. The solution (final pH around 4.2) was centrifuged at

3900 *g* (30 min). The pellet, corresponding to the acid-insoluble matrix (AIM), was rinsed six times with MilliQ water, freeze-dried, and weighed. The supernatant containing the acid-soluble matrix (ASM) was filtered (5 μm) and concentrated with an Amicon ultrafiltration system on a Millipore® membrane (YM10; 10 kDa cut-off). The solution (about 5–10 mL) was extensively dialyzed against MilliQ water (three days, several water changes) before being freeze-dried and weighed.

Infrared spectra analysis: Infrared spectra were directly recorded from dry lyophilized samples at a 2 cm^{-1} resolution on a Fourier transform infrared (FTIR) spectrometer (Ucker Vector 22) fitted with a Specac Golden Gate™ ATR device in the 4000–500 cm^{-1} wavenumber range. Each extract generated several spectra with high reproducibility.

Biochemical analysis of shell matrices with 1D gels: The separation of matrix components was performed under denaturing conditions by 1D SDS-PAGE containing polyacrylamide (12%, Mini-protean 3; Bio-Rad). The protein concentration of the ASM was estimated with the aid of a BCA-200 Protein Assay kit (Pierce, Rockford, US). The ASM was directly suspended in Laemmli sample buffer containing β-mercaptoethanol and heat denatured.[61] ASM samples (50 μg) were loaded in each well. AIM (1 mg) was partly dissolved in urea (8 M, 400 μL) at 60 °C (2 h). The urea-soluble AIM (US-AIM; 20 μL) was loaded onto the gel after heat-denaturation with Laemmli sample buffer. Proteins were visualized on gel both with silver nitrate[62] and with Coomassie brilliant blue (CBB) R-250. For checking the reproducibility of the extraction and the variability of matrix composition, we performed different decalcifications from single shells.

Glycosylations of ASM macromolecules were studied qualitatively. Saccharide moieties were investigated on denaturing mini-gels by use of periodic acid Schiff (PAS)[63] and Alcian Blue 8GX staining,[64] which at pH 1 is specific for sulfated sugars.[65]

The calcium-binding ability of the ASM was investigated by "Stains-All" (cationic carbocyanine dye) staining[66] and by the ^{45}Ca overlay test.[67] For the second of these, the gel was transferred to a PVDF Immobilon-P membrane (Millipore Corp.), with a mini-Trans Blot module (Bio-Rad) for 90 min at 120 mA. The membrane was subsequently incubated three times for 20 min in an overlay buffer (60 mM KCl, 5 mM $MgCl_2$, 10 mM imidazole·HCl, pH 6.8), and then for 30 min in the same buffer containing $^{45}CaCl_2$ (40 MBq L^{-1}, 1.4 μM). After a brief washing (50% ethanol, 2×2 min), the membranes were air-dried and exposed to a film (X-Omat, Kodak) for 1 week.

Amino acid compositions of ASM and AIM: The amino acid compositions of the two fractions were determined by Eurosequence (Groningen, The Netherlands). Freeze-dried samples were hydrolyzed with HCl (5.7 M) in the gas phase for 1.5 h at 150 °C. The resulting hydrolysates were analyzed by using a HP 1090 Aminoquant[68] instrument by an automated two-step precolumn derivatization with *o*-phthalaldehyde (OPA) for primary and *N*-(9-fluorenyl)-methoxycarbonyl (FMOC) for secondary amino acids. Cysteine residues were quantified after oxidation. The hydrolysis procedure does not allow the quantification of tryptophan residues.

Saccharide analysis of AIM and ASM: The monosaccharide compositions were determined as follows. Lyophilized sample matrices (100 μg) were hydrolyzed in trifluoroacetic acid (TFA, 2 M, 100 μL) at 105 °C for 4 h. Samples were evaporated to dryness before being resuspended with NaOH (20 mM, 100 μL). The acidic, neutral, and amino sugar contents of the hydrolysates were determined by high-performance anion exchange-pulsed amperometric detection (HPAE-PAD) by using a CarboPac PA100 column (Dionex P/N043055), according to the Dionex instructions. Nonhydrolyzed samples were analyzed similarly, in order to detect free monosaccharides that could have contaminated the sample during dialysis.

Sialic acid determination was performed by HPAE-PAD after a moderate hydrolysis (acetic acid, 0.5 M, 3 h, 80 °C), as described by Rohrer.[69] The solution containing the released sialic acids was centrifuged and filtered (Centricon, 3 kDa cut-off), and was then dried under vacuum. Samples were re-dissolved in MilliQ water for measurement.

The direct 1,9-dimethylmethylene blue (DMB) method for quantifying sulfated polysaccharides[70] was adapted for microplate reading at 655 nm by use of the Blyscan™ kit provided by Biocolor (Newtownabbey, UK). Chondroitin 4-sulfate was used as a standard for the quantification of matrix sulfated sugars.

Growth of calcite crystals in the presence of ASM: $CaCO_3$ precipitation was performed in vitro by slow diffusion of ammonium carbonate vapor in a calcium chloride solution.[71] This test has been performed with soluble matrices extracted from both calcite and aragonite.[72–75] The test was adapted as follows: different amounts of ASM (0.1 μg mL^{-1} to 20 μg mL^{-1}) in $CaCl_2$ solution (10 mM, 500 μL) were introduced onto an 8-well culture slide (BD Falcon). The culture slide was closed with a plastic cover, which had been pierced with 1 mm holes. It was incubated for 48 h at 4 °C in a closed vessel containing crystals of ammonium bicarbonate. After incubation, the slides were gently dried by capillarity and then by rapid incubation at 50 °C. Blank controls were performed under the same conditions but without any sample. The polymorph identity of the calcium carbonate crystals was determined by direct observation by FTIR spectroscopy as described above. Samples were subsequently carbon-sputtered and observed in the secondary electron mode at 15 keV by scanning electron microscopy (SEM, JEOL 6400) at the LRS (Laboratoire de Réactivité des Solides, Dijon).

ASM analysis by 2D gel electrophoresis: After quantification by using the micro-BCA assay, the ASM was analyzed by 2D gel electrophoresis. Isoelectric focusing (IEF) was carried out with a Protean IEF cell (Bio-Rad). Precast strips (7 cm linear, pH 3–10 immobilized-pH gradient (IPG)) were rehydrated, overnight, at 50 V (25 °C) with buffer (150 μL) containing ASM (80 μg) in urea (6 M), thiourea (2 M), CHAPS (4%, *w/v*), DTT (20 mM), ampholytes (0.1%), and bromophenol blue (0.001%). Immediately afterwards IEF was carried out at 250 V for 15 min in linear mode. Voltage was then raised up to 8000 V in rapid mode and maintained at 8000 V until 10000 Vh had been accumulated. The IPG strips were subsequently transferred for 10 min to equilibration buffer (6 M urea, 2% SDS, 375 mM Tris-HCl pH 8.8, 20% glycerol, 130 mM DTT; 2 mL) and for 10 min to the same buffer (without DTT) but containing iodoacetamide (135 mM). Strips were rinsed in Tris-glycine SDS buffer (TGS: 25 mM Tris, 192 mM glycine, 0.1% SDS), and were positioned on top of precast NuPAGE® BisTris Novex SDS-polyacrylamide gels (4–10%) covered with agarose in TGS (0.5%, *w/v*). Electrophoresis was then performed at 200 V for 40 min.

Spot protein digestion and MALDI-TOF/TOF analysis: Excised gel fractions were washed for 10 min in NH_4HCO_3 (100 mM), and incubated in acetonitrile (ACN, 100%, 20 μL) for 10 min, with gentle stirring (the supernatants were discarded). Samples were subsequently reduced with DTT (20 mM, 30 min, 56 °C in 0.1 M NH_4HCO_3) and alkylated with iodoacetamide (0.1 M; 20 min, 25 °C in 0.1 M NH_4HCO_3). Proteins were digested in gel with trypsin (10 μg mL^{-1}, 20 μL; Roche, France) in TFA (0.01%, 30 min at 4 °C), supernatant

(17 μL) was then discarded, and pieces of gels were covered with NH_4HCO_3 (25 mM, 7 μL) for incubation (2 h, 37 °C). Peptide-containing fractions were extracted by addition of TFA (0.5%, 1 μL) and the supernatant was then recovered. The gel was rinsed with ACN (8 μL) and the supernatants were pooled. Tryptic peptides were analyzed with the mass spectrometer (MS) and the MS/MS modes on a matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF/TOF) Ultraflex II (Bruker Daltonics) instrument with a HCCA matrix.

LC ESI-QqTOF/TOF analysis of whole ASM and AIM: ASM (1 mg) was dissolved in NH_4HCO_3 (100 mM, 20 μL) and proteins were reduced at 60 °C for 15 min with DTT (5 mM), followed by alkylation with iodoacetamide (20 mM), in the dark at room temperature for 25 min. Samples were diluted in NH_4HCO_3 buffer (400 μL) with acetonitrile (20 μL) and incubated for 18 h at 37 °C with trypsin (10 μg, Proteomics grade, Sigma). After centrifugation, the supernatant was lyophilized and stored at −20 °C. AIM (100 mg) was suspended in NH_4HCO_3 (100 mM, 5 mL) and treated as described above for the ASM. Then the sample was lyophilized. The dry residue (1 mg) was dissolved in NH_4HCO_3 buffer (100 μL) containing acetonitrile (5%) and incubated with trypsin (10 μg) for 18 h at 37 °C. After centrifugation, the supernatant was lyophilized and stored at −20 °C. The peptide digests were dissolved in MilliQ water (50 μL) and aliquots (5 μL) were used for proteomic analysis. High-performance liquid chromatography (HPLC) was performed on the tryptic peptides with a 0.32 mm diameter capillary C_{18} column (Discovery® Bio wide pore 5 μm, Sigma) at a 50 $\mu L\,min^{-1}$ flow rate with a gradient of acetonitrile in 1% formic acid (5 to 80% in 60 min). The fractionated peptides were analyzed with an electrospray ionization hybrid quadripole time-of-flight (ESI-QqTOF) hybrid mass spectrometer (pulsar I, Applied Biosystems) by information-dependent acquisition (IDA), which allows switching between MS and MS/MS experiments. The data were acquired and analyzed with the Analyst QS software (Version 1.1). After 1 s acquisition of the MS spectrum, the two most intense multiple charged precursor ions (+2 to +4) could be selected for 2 s-MS/MS spectral acquisitions. The mass-to-charge ratios of the selected precursor ions were excluded for 60 s to avoid re-analysis. The minimum threshold intensity of the ion was set to ten counts. The ion-spray potential and declustering potential were 5200 and 50 V, respectively. The collision energies for the gas-phase fragmentation of the precursor ions were determined automatically by the IDA on the basis of their mass-to-charge ratio (*m/z*) values.

Mass spectrometry data and database searching: The mass spectra data were searched against the NCBI nonredundant database (October 2007), with use of the MASCOT search engine (http://www.matrixscience.com). The mass tolerance was 50 ppm for MALDI-TOF MS or 0.5 Da for ESI-QqTOF MS and was 0.2 Da for both MS/MS experiments with one missed cleavage. Other search parameters were carbamidomethyl as fixed modification and oxidation of methionine as variable modification. Because no identification was achieved by MASCOT searching, MS/MS spectra were analyzed and interpreted de novo manually with the assistance of the Analyst QS1.1 software or the PEAKS software;[76] L indicates both Ile and Leu residues. All the candidate sequences were merged in a MS BLAST search and were compared and aligned two by two with the other sequences of mollusk shell proteins, by use of the SIM tool from the Expasy server (http://www.expasy.ch/tools/sim.html).

Acknowledgements

This work is supported by an ANR project for the period 2007–2010 (ACCRO-Earth, ref. BLAN06–2_159971, coordinator Gilles Ramstein, LSCE, Gif/Yvette). B.M. was financed by a PhD Fellowship from the Ministère Délégué à la Recherche et aux Nouvelles Technologies, (No. 15351) accompanying the ACI JC3049 (F.M.). The "Conseil Régional de Bourgogne" (Dijon, France) provided additional support for the acquisition of new equipment in the Biogeosciences research unit (UMR CNRS 5561). Jean-Paul Pais de Barros (U198, Métabolisme des Lipoprotéines Humaines et Interactions Vasculaires, Université de Bourgogne) is acknowledged for his help in performing the 2D gel electrophoresis. B.M. and F.M. thank Claudie Josse (Laboratoire de Réactivités des Solides, Université de Bourgogne) for helping with the SEM, and Danielle Ballivet-Tkatchenko and Laurent Plasseraud (UMR 5188, LSEO, Université de Bourgogne) for the infrared measurements. They also thank Olivier Mathieu and Jean Lévèque for elemental analysis (UMR 5561, Biogéosciences, Université de Bourgogne) and "l'aquarium des lagons" de Nouméa (Nouvelle-Calédonie) for providing fresh shells of Nautilus macromphalus*. The protein sequence data appear in the UniProt knowledgebase under the accession numbers P85364–P85401.*

Keywords: biomineralization · evolution · mollusk shell nacre · *Nautilus macromphalus* · proteomics

[1] L. Addadi, D. Joester, F. Nudelman, S. Weiner, *Chem. Eur. J.* **2006**, *12*, 980–987.
[2] S. Mann, *Nature* **1988**, *332*, 119–124.
[3] D. Sud, D. Doumenc, E. Lopez, C. Milet, *Tissue Cell* **2001**, *33*, 154–160.
[4] P. Westbroek, F. Marin, *Nature* **1998**, *392*, 861–862.
[5] H. Mutvei, *The Mechanisms of Biomineralization in Animals and Plants*, Tokai University Press, Tokyo, **1990**.
[6] H. Nakahara, *Mechanisms and Phylogeny of Mineralization in Biological Systems*, Springer, New York, **1991**.
[7] Y. Oaki, H. Imai, *Angew. Chem. Int. Ed.* **2005**, *44*, 6571–6575.
[8] M. Rousseau, E. Lopez, P. Stempflé, M. Brendlé, L. Franke, A. Guette, R. Naslain, X. Bourrat, *Biomaterials* **2005**, *26*, 6254–6262.
[9] G. Goffinet, C. Jeuniaux, *Cahiers Biologie Mar.* **1979**, *20*, 341–349.
[10] S. Sudo, T. Fujikawa, T. Nagakura, T. Ohkubo, K. Sakaguchi, M. Tanaka, K. Nakashima, T. Takahashi, *Nature* **1997**, *387*, 563–564.
[11] X. Shen, A. Belcher, P. Hansma, G. Stucky, D. Morse, *J. Biol. Chem.* **1997**, *272*, 32472–32481.
[12] C. Zhang, R. Zhang, *Mar. Biotechnol.* **2006**, *8*, 572–586.
[13] F. Marin, G. Luquet, B. Marie, D. Medakovic, *Curr. Top. Dev. Biol.* **2007**, *80*, 209–276.
[14] G. Falini, S. Weiner, L. Addadi, *Calcif. Tissue Int.* **2003**, *72*, 548–554.
[15] F. Nudelman, B. Gotliv, L. Addadi, S. Weiner, *J. Struct. Biol.* **2006**, *153*, 176–187.
[16] I. Weiss, N. Tuross, L. Addadi, S. Weiner, *J. Exp. Zool.* **2002**, *293*, 478–491.
[17] Y. Levi-Kalisman, G. Falini, L. Addadi, S. Weiner, *J. Struct. Biol.* **2001**, *135*, 8–17.
[18] J. Cope, *Origin and Evolutionary Radiation of the Mollusca*, Oxford University Press, Oxford, **1996**.
[19] J. Carter, G. Clark, *Mollusk, Note for a Short Course, Studies in Geology*, University of Tennessee Press, Knoxville, **1985**.
[20] A. Palmer, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1379–1382.
[21] A. Lindgren, G. Giribet, M. Nishiguchi, *Cladistics* **2004**, *20*, 454–486.
[22] L. Bonnaud, C. Ozouf-Costaz, R. Boucher-Rodoni, *C. R. Biol.* **2004**, *327*, 133–138.
[23] G. Goffinet, *Comp. Biochem. Physiol.* **1969**, *29*, 835–839.

[24] M. Crenshaw, H. Ristedt, *The Mechanisms of Mineralization in the Invertebrates and Plants*, University of South Carolina Press, Columbia, **1976**.
[25] R. Velazquez-Castillo, J. Reyes-Gasga, D. Garcia-Gutierrez, M. Jose-Yacaman, *Biomaterials* **2006**, *27*, 4508–4517.
[26] R. Velazquez-Castillo, J. Reyes-Gasga, D. Garcia-Gutierrez, M. Jose-Yacaman, *J. Mater. Res.* **2006**, *21*, 1484–1489.
[27] C. Grégoire, *Chemical Zoology, Vol. 7*, Academic Press, New York, **1972**.
[28] S. Weiner, *Calcif. Tissue Int.* **1979**, *29*, 163–167.
[29] Y. Dauphin, F. Marin, *Experientia* **1995**, *51*, 278–283.
[30] S. Weiner, H. Lowenstam, L. Hood, *J. Exp. Mar. Biol. Ecol.* **1977**, *30*, 45–51.
[31] J. Keith, S. Stockwell, D. Ball, K. Remillard, D. Kaplan, T. Thannhauser, R. Sherwood, *Comp. Biochem. Physiol. Part B* **1993**, *105*, 487–496.
[32] Y. Dauphin, *Zoology* **2006**, *109*, 85–95.
[33] H. Zhao, T. Samata, D. Takakura, R. Hashimoto, Y. Miyazaki, T. Nozawa, Y. Hikita, *Biomineralization (BIOM2001): Formation, Diversity, Evolution and Application*, Tokai University Press, Kanagawa, **2003**.
[34] J. Balmain, B. Hannoyer, E. Lopez, *J. Biomed. Mater. Res.* **1999**, *48*, 749–754.
[35] D. Worms, S. Weiner, *J. Exp. Zool.* **1986**, *237*, 11–20.
[36] J. Marxen, W. Becker, *Comp. Biochem. Physiol. Part B* **1997**, *118*, 23–33.
[37] B. Marie, G. Luquet, J. P. Païs de Barros, N. Guichard, S. Morel, G. Alcaraz, L. Bollache, F. Marin, *FEBS J.* **2007**, *274*, 2933–2945.
[38] J. Marxen, M. Hammer, T. Gehrke, W. Becker, *Biol. Bull.* **1998**, *194*, 231–240.
[39] H. Nakahara, G. Bevelander, M. Kakei, *Venus* **1982**, *41*, 33–46.
[40] L. Pereira-Mouriès, M. Almeida, C. Ribeiro, J. Peduzzi, M. Barthélemy, C. Milet, E. Lopez, *Eur. J. Biochem.* **2002**, *269*, 4994–5003.
[41] S. Weiner, W. Traub, *FEBS Lett.* **1980**, *111*, 311–316.
[42] F. Zentz, L. Bédouet, M. Schuller, C. Milet, E. Lopez, M. Giraud, *Mar. Biotechnol.* **2001**, *3*, 36–44.
[43] F. Marin, P. Corstjens, B. Gaulejac, E. de Vrind-De Jong, P. Westbroek, *J. Biol. Chem.* **2000**, *275*, 20667–20675.
[44] S. Weiner, W. Traub, H. Lowenstam, *Biomineralization and Biological Metal Accumulation*, Reidel, Dordrecht, **1983**.
[45] H. Kasai, N. Ohta, *Studies of Molluscan Paleobiology. Professor Masae Omori Memorial Volume*, Niigata University Press, Niigata, Japan **1981**.
[46] L. Bédouet, A. Marie, L. Dubost, J. Peduzzi, D. Duplat, S. Berland, M. Puisségur, H. Boulzaguet, M. Rousseau, C. Milet, E. Lopez, *Mar. Biotechnol.* **2007**, *9*, 638–649.
[47] K. Mann, B. Macek, J. Olsen, *Proteomics* **2006**, *6*, 3801–3810.
[48] L. Bédouet, M. Schuller, F. Marin, C. Milet, E. Lopez, M. Giraud, *Comp. Biochem. Physiol. Part B* **2001**, *128*, 389–400.
[49] 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.
[50] J. Waite, A. Saleuddin, S. Andersen, *J. Comp. Physiol.* **1979**, *130*, 301–307.
[51] K. Nagai, N. Yano, K. Morimoto, H. Miyamoto, *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **2007**, *146*, 207–214.
[52] F. Marin, B. Pokroy, G. Luquet, P. Layrolle, K. De Groot, *Biomaterials* **2007**, *28*, 2368–2377.
[53] B. Marie, G. Luquet, L. Bédouet, C. Milet, N. Guichard, D. Medakovic, F. Marin, *ChemBioChem* **2008**, *9*, 2515–2523.
[54] J. Bartlett, J. Simmer, *Crit. Rev. Oral Biol. Med.* **1999**, *1*, 425–441.
[55] A. Martin, V. David, J. Laurence, P. Schwartz, E. Lafer, A. Hedge, P. Rowe, *Endocrinology* **2008**, *149*, 1757–1772.
[56] D. Jackson, C. McDougall, K. Green, F. Simpson, G. Wörheide, B. Degnan, *BMC Biol.* **2006**, *4*, 40–49.
[57] D. Jackson, G. Wörheide, B. Degnan, *BMC Evol. Biol.* **2007**, *7*, 160–177.
[58] G. Haszprunar, C. Schander, K. Halanych, *Phylogeny and Evolution of the Mollusca*, University of California Press, Berkeley, **2008**.
[59] D. Chateigner, C. Hedegaard, H. Wenk, *J. Struct. Geol.* **2000**, *22*, 1723–1735.
[60] A. Checa, T. Okamoto, J. Ramirez, *Proc. R. Soc. London Ser. B* **2006**, *273*, 1329–1337.
[61] U. Laemmli, *Nature* **1970**, *227*, 680–685.
[62] J. Morrissey, *Anal. Biochem.* **1981**, *117*, 307–310.
[63] R. Zacharius, T. Zell, J. Morrison, J. Woodlock, *Anal. Biochem.* **1969**, *30*, 148–152.
[64] R. Wall, T. Gyi, *Anal. Biochem.* **1988**, *175*, 298–299.
[65] R. Lev, S. Spicer, *J. Histochem. Cytochem.* **1964**, *12*, 309.
[66] K. Campbell, D. MacLennan, A. Jorgensen, *J. Biol. Chem.* **1983**, *258*, 11 267–11 273.
[67] K. Maruyama, T. Mikawa, S. Ebashi, *J. Biochem.* **1984**, *95*, 511–519.
[68] R. Schuster, *J. Chromatogr.* **1988**, *431*, 271–284.
[69] J. Rohrer, *Anal. Biochem.* **2000**, *283*, 3–9.
[70] I. Barbosa, S. Garcia, V. Barbier-Chassefière, J. P. Caruelle, I. Martelly, D. Papy-Garcia, *Glycobiology* **2003**, *13*, 647–653.
[71] L. Addadi, S. Weiner, *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 4110–4114.
[72] S. Albeck, L. Addadi, S. Weiner, *Connect. Tissue Res.* **1996**, *35*, 365–370.
[73] Q. Feng, G. Pu, Y. Pei, F. Cui, H. Li, T. Kim, *J. Cryst. Growth* **2000**, *216*, 459–465.
[74] G. Fu, S. Qiu, C. Orme, D. Morse, J. De Yoreo, *Adv. Mater.* **2005**, *17*, 2678–2683.
[75] L. Treccani, K. Mann, F. Heinemann, M. Fritz, *Biophys. J.* **2006**, *91*, 2601–2608.
[76] B. Ma, K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby, G. Lajoie, *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337–2342.

Received: January 5, 2009
Published online on May 26, 2009

