Biochemical and molecular characterization of N66 from the shell of *Pinctada mazatlanica*

Crisalejandra Rivera-Perez 1, Catalina Magallanes-Dominguez 2, Rosa Virginia Dominguez-Beltran 3, Josafat Jehu Ojeda-Ramirez de Areyano 2, Norma Y. Hernandez-Saavedra Corresp. 2

1 Department of Fisheries Ecology, CONACyT-Centro de Investigaciones Biologicas del Noroeste (CIBNOR), La Paz, Baja California Sur, Mexico
2 Department of Fisheries Ecology, Molecular Genetics Laboratory, Centro de Investigaciones Biologicas del Noroeste (CIBNOR), La Paz, Baja California Sur, Mexico
3 Tecnológico Nacional de México, La Paz, Baja California Sur, Mexico

Corresponding Author: Norma Y. Hernandez-Saavedra
Email address: nhernan04@cibnor.mx

Mollusks shell mineralization is a tightly controlled process by shell matrix proteins. However, the study of shell matrix proteins has been limited to a few model species. In this study, the N66 mRNA of the pearl oyster *Pinctada mazatlanica* was cloned and functionally characterized. The full sequence of N66 mRNA consist of 1766 base pairs, and encodes one N66 protein. A sequence analysis revealed that N66 contained two carbonic anhydrase domains, a NG domain and several glycosylation sites. The sequence showed similarity to the carbonic anhydrase VII but also with its homolog protein nacrein. The native N66 protein was isolated from the shell and identified by mass spectrometry, the peptides sequenced matched to the nucleotide sequence obtained. Native N66 is a glycoprotein with a molecular mass of 60-66 kDa which displays carbonic anhydrase activity and calcium carbonate precipitation ability in presence of different salts. Also, a recombinant form of N66 was produced in *Escherichia coli*, and functionally characterized. The recombinant N66 displayed higher carbonic anhydrase activity and crystallization capability than the native N66, suggesting that the lack of posttranslational modifications in the recombinant N66 might modulate its activity.
Biochemical and molecular characterization of N66 from the shell of Pinctada mazatlanica

Crisalejandra Rivera-Perez¹, Catalina Magallanes-Dominguez², Rosa Virginia Dominguez-Beltran³, Josafat Jehu Ojeda-Ramirez de Areyano², Norma Y. Hernandez-Saavedra²*

¹Department of Fisheries Ecology, CONACYT-Centro de Investigaciones Biológicas del Noroeste (CIBNOR), La Paz, Baja California Sur, México.

²Department of Fisheries Ecology, Molecular Genetics Laboratory, Centro de Investigaciones Biológicas del Noroeste (CIBNOR), La Paz, 23096, Baja California Sur, México.

³Tecnológico Nacional de México, La Paz, 23080, Baja California Sur, México.

*Corresponding author: nhernan04@cibnor.mx
Abstract

Mollusks shell mineralization is a tightly controlled process by shell matrix proteins. However, the study of shell matrix proteins has been limited to a few model species. In this study, the N66 mRNA of the pearl oyster *Pinctada mazatlanica* was cloned and functionally characterized. The full sequence of N66 mRNA consist of 1766 base pairs, and encodes one N66 protein. A sequence analysis revealed that N66 contained two carbonic anhydrase domains, a NG domain and several glycosylation sites. The sequence showed similarity to the carbonic anhydrase VII but also with its homolog protein nacrein. The native N66 protein was isolated from the shell and identified by mass spectrometry, the peptides sequenced matched to the nucleotide sequence obtained. Native N66 is a glycoprotein with a molecular mass of 60-66 kDa which displays carbonic anhydrase activity and calcium carbonate precipitation ability in presence of different salts. Also, a recombinant form of N66 was produced in *Escherichia coli*, and functionally characterized. The recombinant N66 displayed higher carbonic anhydrase activity and crystallization capability than the native N66, suggesting that the lack of posttranslational modifications in the recombinant N66 might modulate its activity.
Introduction

Mollusks are known for their ability to build shells having a diversity of sizes, forms and structures. The mollusk shells are mainly composed of calcium carbonate (aragonite, calcite, and nacre), and synthesized by shell matrix proteins (SMPs) (Addadi et al., 2006). SMPs are synthesized in the mantle and secreted to the extrapalleal space, located between the shell and the mantle, where interact with bicarbonate ions (HCO$_3^-$), calcium (Ca$^{2+}$), polysaccharides and metal traces (Mg$^{2+}$) (Xie et al., 2016; Liu et al., 2011), leading to aragonite microstructures (nacre, crossed-lamellar and complex crossed-lamellar) and calcite microstructures (prismatic and foliated) (Zhang & Zhang, 2006). The nacre layer of the shell consists of organized crystal aragonite tablets forming a brick and mortar construction (Mutvei, 1980) which displays outstanding biomechanical properties, such as toughness, elasticity, light weight or softness (Checa, Macias-Sanchez & Ramirez-Rico, 2016).

Although, aragonite assembly process is not well understood, several SMPs involved in aragonite crystal growth have been identified in several marine mollusks (Zhang & Zhang, 2006) through transcriptomic and proteomic analysis (Miyamoto et al., 2013; Mann, Edsinger-Gonzalez & Mann, 2012). Nacrein and N66 proteins play a significant role in the biomineralization process, they are soluble enzymes involved in the calcium carbonate crystallization and participates in the formation of the nacreous layer (Miyamoto et al., 1996). Nacrein and N66 are metalloenzymes with one Gly-Xaa-Asn repeat domain (Xaa: Asp, Asn, or Glu, NG-repeat domain) at the C-terminal end (Song et al., 2014; Mann, Edsinger-Gonzalez & Mann, 2012; Norizuki & Samata, 2008; Kono, Hayashi & Samata, 2000; Miyamoto, Yano & Miyashita, 2003; Miyamoto et al., 1996) and two carbonic anhydrase (CA) domains which belong to the CA superfamily. All nacrein and N66 proteins described in mollusks belong to the
α-CA family, class II and VII, respectively (Mann, Edsinger-Gonzalez & Mann, 2012; Leggat et al., 2005; Kono, Hiyashi & Samata, 2000; Miyamoto et al., 1996). These proteins are known to contain a metal ion (Zn$^{2+}$) coordinated by three His residues at the active site (Ozensoy, Capasso & Supuran, 2016) and catalyze the reversible hydration of carbon dioxide to bicarbonate (Miyamoto et al., 1996).

These proteins are conserved in bivalves and gastropods and have a molecular weight ranging from 50 kDa to 66 kDa (Song et al., 2014; Mann, Edsinger-Gonzalez & Mann, 2012; Leggat et al., 2005; Miyamoto, Yano, Miyashita, 2003; Kono, Hayashi & Samata, 2000; Miyamoto et al., 1996). Previous studies have recognized that nacrein and N66 are glycoproteins (Leggat et al., 2005; Kono, Hayashi & Samata, 2000), where nacrein possess N-glycan containing sulfite and sialic acid at its terminus (Takakura et al., 2008), furthermore, they both have calcium binding ability (Miyamoto et al., 1996). Despite their similar biochemical properties, N66 has almost twice length of NG repeat domain than nacrein (Miyamoto, Yano & Miyashita, 2003; Kono, Hayashi & Samata, 2000; Miyamoto et al., 1996). This domain has been proved to function as negative regulator in shell formation in nacrein (Miyamoto, Miyoshi & Kohno, 2005), and also as a positive regulator, both as a Ca$^{2+}$ concentrator and an enzyme required for production of carbonate ions (Norizuki & Samata, 2008). The longer repeat NG domain have been related to stronger reaction of protein with Ca$^{2+}$ molecules, matrix components and crystals, resulting in an improved calcification capability (Samata et al., 1999).

The nacrein gene has been described in the bivalve *Pinctada maxima* (Kono, Hayashi & Samata, 2000) and the gastropod *Turbo marmoratus* (Miyamoto, Yano & Miyashita, 2003), while N66 gene has been only described in *P. maxima* (Kono, Hayashi & Samata, 2000). Differential expression of SMP mRNAs in the mantle of mollusks has been related to their
crystal layer formation. Nacrein mRNA is highly expressed in the mantle edge and pallial (Miyamoto, Miyoshi & Kohno, 2005; Miyamoto et al., 1996), while N66 mRNA is expressed in the dorsal region of the mantle and the mantle edge (Kono, Hayashi & Samata, 2000), the former has been associated for the nacreous layer function and the latter for the prismatic layer formation (Sudo et al., 1997).

In order to gain insights into the biomineralization processes in mollusk and to increase the knowledge of the nacre formation, a comprehensive study of the proteins involved in nacre deposition is required. Pearl oyster, *Pinctada mazatlanica* is a good model species to study biomineralization, since includes two layers, an inner nacreous layer, and an outer prismatic calcite layer. In this study, a N66 from the shell was isolated and its coding sequence was obtained from the mantle of the pearl oyster *P. mazatlanica* and overexpressed on a prokaryotic system, the resulting native and recombinant proteins were functionally characterized.

**Material & Methods**

**Biological material.** Three adult female organisms were provided by Perlas del Cortez S. de R.L. MI. located at Bahia de La Paz B.C.S. Organisms were transported to the Molecular Genetics Laboratory at CIBNOR, and mantle tissue was dissected and stored at −80 °C until used.

**Molecular characterization of N66 from mantle tissue**

**Total RNA extraction and RNA reverse transcription.** Total RNA was extracted with TRIzol® (Invitrogen, U.S.A.) according to the manufacturer’s instructions. Samples were homogenized using a glass pestle; then, two consecutive extractions of each sample were made. RNA purity and concentration were determined by spectrophotometry using a NanoDrop ND-
2000 (Thermo Scientific, U.S.A.) at 260/280 and 260/230 nm absorbance ratios (range, 1.9 –
2.0). The RNA integrity was assessed on 1% (w/v) agarose-synergel gel. To ensure complete
DNA absence, a direct PCR was performed using 1 µL (50 ng/µL) of each RNA preparation
with 28S ribosomal specific primer as a non-amplified control. After that, 1 µg of total RNA was
used from each verified RNA sample for cDNA synthesis using the cloned AMV First-Strand
cDNA Synthesis Reaction (Invitrogen, U.S.A.) and oligo-dT primer, afterwards cDNAs were
stored at −80 °C until use. Control reactions were performed without template or non-reverse
transcribed RNA to determine the presence of DNA.

**Amplification of coding sequence of N66 from mantle.** A search for sequences
encoding N66 from mollusks available at GenBank yielded three homologous sequences, N66
*Pinctada maxima* (GenBank: AB032613), N44 *P. maxima* (GenBank: FJ913472) and SP-S
*Mizuhopecten yessoensis* (GenBank: AB185328). Based on these sequences, specific primers
were designed to amplify the coding sequence of N66 (Table 1). PCR amplification of cDNA
fragments encoding N66 was done in a final volume of 12.5 µL containing 10 pmol of each
forward and reverse primer, 50 ng cDNA, 10 nmol dNTPs mixture, 1.25 µL 10X Taq Buffer, and
1 U Taq DNA polymerase (11615010A, Invitrogen). PCR amplification were carried out for 5
min at 94 °C, followed by 5 cycles consisting 1 min at 94 °C, 1 min at 40 °C, and 2 min at 72 °C,
followed of 35 cycles consisting 1 min at 94 °C, 1 min at 50 °C, and 2 min at 72 °C. In the last
cycle, the extension step at 72 °C lasted 10 min. PCR products were analyzed on 1 % (w/v)
agarose-synergel gels and visualized under UV light after staining with UView loading dye (166-
5112, Biorad). PCR products were cloned using the TOPO-TA cloning kit (K4500-01,
Invitrogen) and sequenced for both strands. The obtained sequence was submitted to the National
Center for Biotechnology Information (NCBI) database for BLAST searching and other informatics analysis.

**In silico analyses.** Identification of the open reading frame was performed using the ORF finder software (https://www.ncbi.nlm.nih.gov/orffinder/) and sequence alignment with homologous sequences were performed by using CLUSTAL Omega. Identification of putative protein motifs was performed using the MotifScan (Pfam HMMs global models database) available at the Swiss Institute of Bioinformatic (http://myhits.isb-sib.ch/cgi-bin/motif_scan). Identification of signal peptide was achieved by using SignalP 4.1 Server (Petersen et al., 2011). Theoretical molecular weight, isoelectric point (pI), and amino acid composition of the protein were calculated using the ProtParam software from ExPASy (Expert Protein Analysis System; https://www.expasy.org/). Putative glycosylation and phosphorylation sites of N66 were determined using NetOGly4.0 (Steentoft et al., 2013), NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) and NetPhos2.0 (Blom, Gammeltoft & Brunak, 1999), respectively.

**Vector construction and recombinant protein expression.** The coding region of N66 cDNA, without signal peptide, was synthesized by PCR with two primers (N66_HindF and N66_HindR) to introduce the restriction enzyme sites (HindIII, AAGCTT) at the 5’ and 3’ ends (Table 1). The PCR product was directly digested with HindIII (NEB, New England) and isolated from gel electrophoresis, purified and ligated into a polyhistidine fusion protein expression vector pTrcHis C (Invitrogen, Inc.) that was linearized with the same restriction enzyme, yielding pTrcHis C/N66. *Escherichia coli* TOP10 cells were transformed with the
Construct and cultured in 6 mL of LB medium containing 100 µg/mL ampicillin at 37 °C with shaking at 150 rpm overnight. Five mL of the culture medium were transferred to a flask containing 245 mL of the same medium and further incubated. When OD$_{600}$ of the bacterial cells reached ~0.7, isopropyl β-D-thiogalactosidase was added to a final concentration of 1 mM and incubated at 37 °C with constant shaking at 150 rpm for 4 h. The bacterial cells were harvested by centrifugation (7,500 rpm, 25 min at room temperature) and stored at −20 °C. The frozen bacterial pellet was thawed and dispersed in ice-cold Tris-HCl buffer (50 mM, pH 7.4) and disrupted with glass beads for 5 min of vigorous vortexing. The bacterial lysate was centrifuged (12,000 rpm, 15 min, 4 °C), and the supernatant was recovered. Recombinant His-tagged proteins were purified from the supernatant by using a Sepharose-Ni column (GE Healthcare).

Native N66 purification and characterization

**Shell matrix extraction.** The organic matrix of prismatic and nacreous layers were crushed to fine powder. The powdered matrix (20 g) were suspended in 100 mL cold acetic acid (4 °C, 10% v/v) and incubated for 24 h with continuous stirring. Soluble (ASM) and insoluble (AIM) matrix fractions were separated by centrifugation at 14,000 rpm for 20 minutes. The AIM was rinsed with distilled water and lyophilized. The ASM was dialyzed 24 h against cold acetic acid (4 °C, 1% v/v), afterwards the ASM was dialyzed other 24 h against distilled water. The dialyzed ASM was concentrated by lyophilization. Protein concentration was determined by Lowry method (Lowry et al., 1951).

**N66 purification by affinity chromatography.** Native and recombinant N66 were isolated by affinity chromatography using a Sepharose-Ni column (GE Healthcare; 1.45 x 5.0
Protein sample (1 mL) was loaded onto the column, previously equilibrated with 50 mM Tris-HCl pH 7.4, 300 mM NaCl buffer, mixed and incubated for 1 h at 25 °C. The unbound enzyme was washed with five bed volumes of 50 mM Tris-HCl pH 7.5, 300 mM NaCl buffer. The bound enzyme was eluted by washing the column with elution buffer containing 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 300 mM Imidazole. Fractions of 1 mL were collected from the elution and evaluated by electrophoresis. Fractions containing N66 were pooled and concentrated through centrifugal filters (AMICON Ultra-10, Millipore) at 4,000 g for 20 min at 25 °C to a final volume of 500 µL. Concentrated sample was desalted using PD-10 desalting column (Amershan Pharmacia). Glycerol was added to the enzyme solution (final concentration 50%), and the sample was stored at −20 °C until use. Protein content was quantified by Lowry method (Lowry et al., 1951) and SDS-PAGE analysis of the purified protein was performed as described above.

**SDS-PAGE and staining.** Protein bands, from native and recombinant proteins, were analyzed by electrophoresis under reducing conditions using 10% SDS-PAGE (Mini-PROTEAN, Bio-Rad), as described by Laemmli (1970) and stained with silver nitrate (Merril & Washart, 1998) in 10% acrylamide gels. Samples were diluted (1:4) with sample buffer (0.125 M Tris-HCl, 2% SDS, 20% v/v glycerol, 0.04% bromophenol blue, 5% β-mercaptoethanol at pH 6.8) and heated for 10 min at 100 °C. Electrophoresis separation was performed at 80V constant current. To detect protein bands, gels were stained with 0.1% Coomassie Brilliant Blue R-250 in 7.5% acetic acid and 5% methanol at room temperature and distained in the 10% acetic acid and 40% methanol.
Glycosylation was analyzed by electrophoresis under reducing conditions (Thornton et al., 1994). The gel was first incubated in fixing solution (5% TCA) for 5 min, then incubated in oxidant solution (0.8% periodic acid in 0.3% sodium acetate), and rinsed with water. Then the gel was incubated in Schiff’s reagent for 10 min in the dark. Finally, the gel was washed three times with 3% nitric acid for 3 min. The gel was stored in 7.5% acetic acid, 5% v/v methanol. Ovoalbumin (SIGMA) was used as positive control.

Mass spectrometry analyses. To identify the mass spectrometric profile, 20 µg of the native N66 protein was first separated by 10% SDS-PAGE. After electrophoresis, the gel was stained with Coomassie Blue R-250 and bands were manually cut from the gel. After tryptic digestion, the resulting peptides were subjected to Liquid Chromatography-Mass Spectrometry (LC-MS). The LC-MS was performed at the Proteomic Unit Facility at Biotechnology Institute, Cuernavaca. The resulting data were compared with the proteomic database at GenBank and with the deduced amino acid sequence from cDNA of N66 from *P. mazatlanica*.

Carbonic anhydrase assay. Carbonic anhydrase activity from native and recombinant N66 was assayed qualitatively and quantitatively. For qualitative assay, 1% agarose plates were prepared with borate buffer (0.05 M boric acid, 0.025 M NaCl). Samples (10 µL of 0.2 mg protein/mL) were loaded on the agar plate, after sample diffusion 0.1% bromocresol purple (dissolved in 0.1 M Tris-HCl pH 8.5) was added and incubated for 10 min at room temperature. Afterwards, the solution was discarded and the agar plate was washed with distilled water. Finally, saturated CO₂ water was added as substrate; the production of hydrogen ions during the CO₂ hydration reaction lowers the pH of the solution until the formation of yellow halo zones.
which indicate carbonic anhydrase activity while dark zone indicate lack of activity (Machenko, 2002). Bovine carbonic anhydrase (SIGMA) was used as positive control and boiled enzyme as negative control.

The quantitative in vitro assay for CA activity is described in detail by Wilbur and Anderson (1948). Briefly, CA activity was measured by the decrease of pH resulting from the hydration of CO$_2$ to HCO$_3^-$ and H$^+$ after the addition of substrate. All experiments were performed at 4 °C. To run the assay, 5 mL of 0.012 M Tris-HCl buffer (pH 8.3) were transferred to a small beaker and 1 mL of homogenate diluted in 0.012 M Tris-HCl buffer was added to obtain 20 µg of protein. The mixture was constantly stirred with a magnetic stirring bar. Afterwards, 4 mL of cold saturated CO$_2$ solution were added, and the decrease in pH was recorded by a pH probe. CA activity was calculated as $(t_0-t)/t$, where $t_0$ is the time needed for the non-catalyzed reaction and $t$ is the time for the catalyzed reaction to obtain a pH decrease from 8.0 to 6.0. Units of enzyme activity (U) were normalized to the total protein content.

Crystallization in vitro assay with native and recombinant N66. Three crystallization solutions were prepared, two induces calcite (calcitic crystallization solution) and one induces aragonite (aragonite crystallization solution). Calcitic crystallization solutions (40 mM CaCl$_2$, pH 8.2, 100 mM NaHCO$_3$ and 100 mM CaCO$_3$) and aragonite crystallization solution (40 mM MgCl$_2$, pH 8.2, 100 mM NaHCO$_3$) were prepared following the protocol of Weiss et al. 2000 and Hillner et al. 1992, respectively. Native and recombinant N66 (10 µL of 0.2 mg/ml) were added to 50 µL of the saturated solution. As a control, buffer solution without protein was used. Each solution experiment was made in triplicates and incubated at 4 °C for 21 days in a cover slide situated at the bottom of a six-hole microplate which was sealed with parafilm to avoid
contamination. The morphology of the crystals produced were analyzed by scanning electron microscopy (SEM) at the Electronic Microscopy Laboratory at CIBNOR.

Statistical analysis

Statistical analysis was performed using the GraphPad Prism Software. The results are expressed as means ± s.d. Significant differences (P<0.05) were determined by ANOVA followed by a pair-wise comparison of means (Tukey’s test).

Results

Sequence analysis of cDNA from mantle of *Pinctada mazatlanica* encoding a N66.

Nucleotides and deduced amino acid sequence of N66 from the mantle of *P. mazatlanica* are shown in Fig. 1. The open reading frame of N66 obtained from mantle cDNA was 1766 bp long (GenBank: MH473230) and encoded a 576 amino acid protein with a calculated molecular mass of 63.68 kDa before any post-translational modification and an isoelectric point (pI) of 7.38. Removal of the signal peptide sequence (residues 1 to 22) resulted in a theoretical molecular weight of 61.12 kDa and a theoretical pI of 7.21. It possess high proportions of N (19.6%) and G (15.8%), which accounted for 35.4% of the total amino acid residues (Table 2). Analysis of the N66 amino acid sequence showed homology with the carbonic anhydrase (CA) family, exhibiting 34% identity with human CAVII (GenBank: NP_005173), with high homology to amino acid residues involved in catalysis (Q_{125}, N_{154} and E_{168}) and three histidine residues essential for the zinc cofactor binding (H_{156}, H_{158} and H_{181}) (Fig. 1).

N66 protein sequence possess in its modular structure a signal peptide (1-22 aa), two carbonic anhydrase domains (G_{69}-G_{221} and D_{490}-K_{554}), and a NG rich sequence (N_{245}-G_{437}) (Fig.
S1), sharing 82.82% identity to N66 from *P. maxima*, ~92% to nacreins (B2, B3 and B4) from *Pinctada margaritifera*, 72.62% to nacrein from *Pinctada fucata* and 78.70% to N45 from *P. maxima* (Fig. 2). All the protein sequences contained the same protein structure, however the length of the carbonic anhydrase (CA) and NG domain were variable among the proteins. Further bioinformatics analysis of the N66 sequence suggests that some serine and threonine residues in N66 may either be phosphorylated (36 putative sites) or N- and O-glycosylated (16 putative sites), as indicated in Fig. 1, which are common features related to proteins involved in biomineralization (Nudelman et al., 2007). Finally, most of the O-glycosylation and phosphorylation sites were found in the CA domain.

**N66 protein characterization**

**Native N66 from *P. mazatlanica* shells.** The organic matrix of the shell from *P. mazatlanica* was separated by SDS-PAGE revealing a wide range of abundant molecular weight proteins in the ASM fraction (Fig. 3A). The ASM fraction was used to isolate the N66 by affinity chromatography using a Sepharose-Ni matrix (Fig. S2) since the active site of the CA domain contains three histidines involved in metal ion binding which has been shown to be conserved in N66 from *P. mazatlanica* (this study) and *P. maxima* (Kono, Hayashi & Samata, 2000). The fractions containing the eluted N66 (Fig. S2) were pooled and desalted. Two bands in SDS-PAGE and silver stained gels were observed (Fig. 3B) with molecular weights of ~60 and 64 kDa. A flow sheet of the purification process for N66 is presented in Table 3. The enzyme was enriched 3.39-fold and the yield was 56.52%. Electrophoretic analysis of N66 with PAS stain, demonstrated that N66 is glycosylated (Fig. 3B), however, the type of carbohydrate associated was not analyzed. The results of LC-MS of the bands matched the deduced amino acid sequence
of the N66 of the mantle of *P. mazatlanica* (Table 4), as described before (Fig. 1), and allowed
us to identify it as a N66, a shell matrix protein from the shell of *P. mazatlanica*.

**Recombinant N66 from *P. mazatlanica***. Recombinant N66 protein was expressed as a
soluble form at 37 °C with 1 mM IPTG in *E. coli* after 4h induction, and the bands
corresponding to the protein in SDS-PAGE analysis agreed with the native N66 of ASM from *P.
mazatlanica* (Fig. 3B). The recombinant N66 was successfully purified to homogeneity by
affinity chromatography using Ni-Sepharose columns (Fig. S3). Electrophoretic analysis showed
two protein bands of ~60 and 64 kDa which agreed with that of the native N66 of ASM from *P.
mazatlanica* (Fig. 3B). Recombinant N66 did not show glycosylation after staining with PAS
(data not shown).

**Carbonic anhydrase activity of native and recombinant N66**. The carbonic anhydrase
activity of the native and recombinant N66 was confirmed qualitatively (Fig. 4) and
quantitatively. Figure 4 showed comparable yellow halos obtained with the native and
recombinant N66, indicating CA activity. The commercial bovine CA was used as positive
control. The comparable CA activity of the native and recombinant N66 proteins was confirmed
with the Wilbur and Anderson method. The results showed significant differences between
native (161 ± 12.5 U/mg protein) and recombinant protein (197 ± 7.5 U/mg protein), being the
recombinant protein the one with higher activity.

**In vitro crystallization by native and recombinant N66**. To elucidate the effect of
native and recombinant N66 on the growth of calcium carbonate crystals, three individual *in
vitro* crystallization assays testing the formation of aragonite and calcite were established and
analyzed by SEM (Fig. 5). First, when no protein was added (negative control) to the three salts tested (NaHCO$_3$+MgCl$_2$, NaHCO$_3$+CaCl$_2$ and CaCO$_3$), different crystal morphology was observed. The crystals formed when sodium bicarbonate in presence of magnesium chloride were unorganized, while in presence of calcium chloride crystals become compact and thicker, and in presence of calcium carbonate the formed crystals exhibited the typical rhombohedra of calcite (Fig. 5A-C). On the other hand, in the presence of 0.2 mg protein/ mL of native N66 to the three salts preparations showed polycrystalline aggregates of aragonite and calcite in presence of MgCl$_2$ and CaCO$_3$ respectively, however in presence of CaCl$_2$, crystals were small structures of calcite and in lesser quantity per area (Fig. 5D-F). In contrast to native N66, recombinant form produced bigger and organized crystal structures forming plates (Fig. 5G-I).

**Discussion**

The mollusk shell is a complex structure made of organic and mineral components. The shells are composed by layers formed from calcium carbonate polymorphs as aragonite or calcite crystals (Takeuchi et al., 2008; Zhang & Zhang, 2006). The calcium carbonate polymorphs are formed by shell matrix proteins (SMPs), which have been classified as soluble and insoluble based on the two main extraction methods, EDTA and acetic acid (Marie et al., 2007; Kono, Hayashi & Samata, 2000). The insoluble proteins, are chitin-proteinaceous complexes rich in aliphatic amino acids, such as Gly and Ala, while the soluble proteins have been found to be polyanionic, enriched mainly in Asp; among the soluble proteins, functional enzymes such as carbonic anhydrases have been described (Kono, Hayashi & Samata, 2000), which seems contribute to the control of biomineralization.

Most of the information about genes encoding enzymes with CA activity to form nacre in mollusks is available for nacreins (Mann, Edsinger-Gonzalez & Mann, 2012; Leggat et al., 2005;
Miyamoto et al., 1996). Nacrein gene has been described in *Crassostrea gigas* (Song et al., 2014), *Patella vulgata, Mytilus californianus, P. maxima* and *Pinctada margaritifera* (Joubert et al., 2010), while N66 mRNA has been only described in *P. maxima* (Kono, Hayashi & Samata, 2000) and *P. margaritifera* (Joubert et al., 2010). This homologous genes have been found to be expressed in the mantle (Miyamoto, Myoshi & Kohno, 2005; Kono, Hayashi & Samata, 2000), which is a conserved organ involved in shell formation throughout mollusks. We followed this lead and proved that *P. mazatlanica* express the N66 mRNA in the mantle and secrete the enzyme into the extrapallial space to participate in the shell growth.

Currently, only six SMPs with carbonic anhydrase activity have been isolated and characterized from pearl oysters, nacrein from the Akoya pearl oyster *Pinctada fucata*, turban shell *Turbo marmoratus*, the edible Iwagaki oyster *Crassostrea nippona*, Yesso scallop *Patinopecten yessoensis* and giant clam *Tridacna gigas* (Norizuki & Samata; Leggat et al., 2005; Miyamoto et al., 2003; Miyamoto et al., 1996), and only one N66 from *P. maxima* (Miyamoto et al., 1996). Carbonic anhydrases (CAs) are usually purified via the agarose-sulfonamide affinity column chromatographic procedure (Ozensoy, Capasso & Supuran, 2016), however, in this study a Sepharose-Ni column was successfully used to capture the N66 protein from the shell of *P. mazatlanica*, since it is well known that the CA domain possess three conserved histidine residues at the active site and His exhibits the strongest interaction to metal ion matrices, as electron donors groups on the imidazole ring of His form coordinated bonds with the immobilized metal (Bornhorst & Falke, 2000). The isolated protein was identified by mass spectrometry as a N66, due to its similarity to N66 from *P. maxima* and its homolog protein, nacrein; native N66 and its recombinant form showed two bands with a molecular weight of ~60 and ~66 kDa, respectively. The double bands of the proteins observed on SDS-PAGE under
364 reducing conditions have been also observed in other CAs such as CAH1, CAXII and CAIX, the 
365 last one migrates as a double band between 54/58 kDa under reducing conditions. This effect is 
366 most likely due to glycosylations (Buren et al., 2011; Ulmasov et al., 2000) and/or equivalent 
367 populations of mature and pro-forms of the enzyme with both possessing a propensity to 
368 oligomerize in non-reducing conditions (Hilvo et al., 2008; Ulmasov et al., 2000; Pastorek et al., 
369 1994). Since the recombinant N66 do not possess PTM and both, native and recombinant, 
369 migrate similar on SDS-PAGE, the formation of oligomers of N66 is suggested in this study. 
370 However, further experiments are required to establish if the double band of N66 from the shell 
371 of *P. sterna* is due to glycosylations or oligomerization of the protein.

372 Structurally, N66 in *P. mazatlanica* is similar to N66 from *P. maxima* (Kono, Hayashi & 
373 Samata, 2000) and its homologous proteins, nacreins and N45 (Miyamoto et al., 1996), they 
374 share in its modular structure a signal peptide, two CA domains and a NG domain. α-CAs 
375 contain a domain which is composed by three amino acid residues (Gln, Asn and Glu) and three 
376 His (Vulo et al., 2008), according to the deduced primary structure of N66 from *P. mazatlanica*, 
377 the protein contains two CA domains composed by Gln_{125}, Asn_{154}, Glu_{168} and three His residues 
378 which act as zinc ligands (His_{156}, His_{158} and His_{181}), suggesting that N66 possess a functional CA 
379 domain. CA activity has been only proved in nacrein from *P. fucata* (Miyamoto et al., 1996) and 
380 N66 from *P. mazatlanica* and its recombinant protein, in this study. The recombinant N66 
381 showed higher CA activity than the native N66 and the bovine CA. The NG domain found in 
382 N66 from *P. mazatlanica* differs in length and composition to its homologues, accounting for 
383 over ~30% of their sequences. This domain has been proved to function as negative regulator in 
384 shell formation in nacrein (Miyamoto, Miyoshi & Kohno, 2005), but also as a positive regulator, 
385 both as a Ca^{2+} concentrator and an enzyme required for production of carbonate ions (Norizuki
It is hypothesized that the longer repeat NG domain might have a stronger interaction of protein with Ca\(^{2+}\) molecules, matrix components and crystals, leading to better calcification ability (Samata et al., 1999). This suggests, that N66 might have better calcification ability than its homologs, although, further experiments are required to prove this hypothesis.

Post-translational modifications (PTM) of SMPs are a common feature that have been suggested to play an important role in biomineralization. Phosphorylation and glycosylation are the major players in many of the protein functions (Saraswathy & Ramalingam, 2011). In silico analysis of N66 from *P. mazatlanica* showed 36 putative sites of phosphorylation and 16 putative sites of glycosylation. The biochemical analysis confirmed that N66 is glycosylated (Fig. 3B), a modification that has been observed in other matrix proteins (Montagnani et al., 2011; Suzuki et al., 2011). Glycosylation participates in the folding of the protein and/or interactions with Ca\(^{2+}\) (Suzuki et al., 2011; Borbas, Wheeler & Sikes, 1991) and provide sites of high anionic charge through addition of sialic acid or sulfated oligosaccharides (Michenfelder et al., 2003) which have been related to promote the uptake of Ca\(^{2+}\) (Song et al., 2014). The saccharide chains of N66 in *P. mazatlanica* probably have similar effects, however, the chemical identity of the oligosaccharides moieties, the number and the location of actual glycosylation sites along the protein of N66 remain to be determined.

Experiments involving *in vitro* calcium carbonate crystallization have been used to prove functional activity of SMPs since most of the described SMPs does not possess catalytic domains. Crystal formation comprise two main process, nucleation and growth. It begins with an amorphous calcium carbonate as an intermediate of the crystalline carbonate polymorph such as calcite or aragonite (Wang et al., 2009). The presence of Mg\(^{2+}\) ions foster metastable aragonite growth that make up the mineral phase of nacre (Evans, 2017; Su et al., 2016; Samata et al.,
and when Mg\(^{2+}\) ion are low induces the formation of calcite (Raz, Weiner & Addadi, 2000). Morphological observation of calcite and aragonite by SEM has been made by several authors, showing that crystals of calcite display a rombohedral form while aragonite has an orthorhombic form (Zigovecki, Posilovic & Bermanee, 2009), although the morphology of each polymorph can be changed due to the crystallization condition, e.g. aragonite is needle-like but it changes to flake-like or cauliflower-like (Chakrabarty & Mahapatra, 1999). In this study, N66 and its recombinant protein have differential CA activity and lead to calcium carbonate precipitation, displaying structures such as aragonite and calcite when crystallization was made in presence of MgCl\(_2\) and CaCl\(_2\), respectively. Similar results were obtained by Kono et al. (2000), for N66 from \textit{P. maxima}, which induces aragonite layers, however it requires the presence of N14 and Mg\(^{2+}\) ions. The recombinant N66 foster higher crystal growth than native N66, this evidence raise the possibility that the presence of PTMs (e.g. glycosylations and phosphorylations) affect the growth and shape of calcite and aragonite crystals, which is in agreement to previous evidences in \textit{Unio pictorum}, that showed that glycosylations might change the crystal morphology in \textit{in vitro} crystallization experiments (Marie et al., 2007).

\textbf{Conclusion}

In this study, a N66 mRNA from the mantle of \textit{P. mazatlanica} was identified. The transcript is translated to one single protein named N66. The N66 is a glycoprotein with carbonic anhydrase activity and calcium carbonate precipitation ability \textit{in vitro}, suggesting that N66 is involved in the mineralization process in \textit{P. mazatlanica}. Experiments with the recombinant N66 suggest that posttranslational modifications might play an important role in the modulation of the CA activity. The identification of the carbohydrates associated to N66, as well as the determination of other posttranslational modification (e.g. phosphorylation, sulfation, etc.), and
their role in the calcium carbonate precipitation (e.g. crystal growth, type of polymorph) remains to be tested. Also, to understand the potential oligomerization process of the N66 and its role in the mineralization process remains unclear.

References

Addadi L, Joester D, Nudelman F, Weiner S. 2006. Mollusk shell formation: A source of new concepts for understanding biomineralization processes. *Chemistry European Journal* 12:980-987. DOI: 10.1002/chem.200500980.

Blom N, Gammeltoft S, Brunak S. 1999. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology* 294:1351-1362. DOI: 10.1006/jmbi.1999.3310.

Borbas JE, Wheeler AP, Sikes S. 1991. Molluscan shell matrix phosphoproteins: Correlation of degree of phosphorylation to shell mineral microstructure and to in vitro regulation of mineralization. *Journal of Experimental Zoology* 258:1-13. DOI:10.1002/jez.1402580102.

Bornhorst JA, Falke JJ. 2000. Purification of proteins using polyhistidine affinity tags. *Methods in Enzymol* 326:245-254.

Buren S, Ortega-Villasante C, Blanco-Rivero A, Martinez-Bernardini A, Shutova T, 2011. Importance of post-translational modifications for functionality of a chloroplast-localized carbonic anhydrase (CAH1) in *Arabidopsis thaliana*. *PLOS One* 6(6): e21021. DOI: 10.1371/journal.pone.0021021

Chakrabarty D, Mahapatra S. 1999. Aragonite crystals with unconventional morphologies. *Journal of Materials Chemistry* 9:2953–2957. DOI: 10.1039/a905407c
Checa AG, Macías-Sánchez E, Ramírez-Rico J. 2016. Biological strategy for the fabrication of highly ordered aragonite helices: the microstructure of the cavolinioidean gastropods. *Scientific Reports* 6:25989. DOI: 10.1038/srep25989.

De Luca V, Del Prete S, Vullo D, Carginale V, Di Fonzo P, Osman SM, AlOthman Z, Supuran CT, Capasso C. 2016. Expression and characterization of a recombinant psychrophilic carbonic anhydrase (NcoCA) identified in the genome of the Antarctic cyanobacteria belonging to the genus Nostoc. *Journal of Enzyme Inhibition and Medical Chemistry*, 31(5): 810-817. DOI: 10.3109/14756366.2015.1069289.

Evans JS. 2017. Polymorphs, proteins, and nucleation theory: A critical analysis. *Minerals*. 7: 6. DOI: 10.3390/min7040062.

Hillner PE, Manne S, Gratz AJ, Hansma PK. 1992. AFM images of dissolution and growth on a calcite growth crystal. *Ultramicroscopy* 42:1387-1393. DOI: 10.1016/0304-3991(92)90454-R.

Hilvo M, Baranauskiene L, Salzano AM, Sacloni A, Matulis D, Innocenti A, Scozzafava A, Monti SM, Di Fiore A, De Simone G, Lindfors M, Janis J, Valjakka J, Pastorekova S, Pastorek J, Kulomaa MS, Nordlund HR, Supuran CT, Parkkila S. 2008. Biochemical characterization of CAIX, one of the most active carbonic anhydrase isozymes. *Journal of Biological Chemistry* 41: 27799-27809. DOI: 10.1074/jbc.M800938200.

Joubert C, Piquemal D, Marie B, Manchon L, Pierrat F, Zanella-Cleon I, Cochenneuc-Laureau N, Gueguen Y, Montagnani C. 2010. Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying and shell: focus on biomineralization. *BMC Genomics* 11:613. DOI: 10.1186/1471-2164-11-613.
Kono M, Hayashi N, Samata T. 2000. Molecular Mechanism of the Nacreous Layer Formation in *Pinctada maxima*. *Biochemical and Biophysical Research Communications* 269:213-218. DOI: 10.1006/bbrc.2000.2274.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685. DOI: 10.1038/227680a0.

Leggat W, Dixon R, Saleh S, Yellowlees D. 2005. A novel carbonic anhydrase from the giant clam *Tridacna gigas* contains two carbonic anhydrase domains. *FEBS Journal* 272:3297-3305. DOI: 10.1111/j.1742-4658.2005.04742.x.

Liu X, Liu C, Chen L, Sun J. 2011. A new method to extract matrix proteins directly from the secretion of the mollusk mantle and the role of these proteins in shell biomineralization. *Marine Biotechnology* 13:981-991. DOI: 10.1007/s10126-011-9362-y.

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. 1951. Protein measurement with the folin phenol reagent. *Journal of Biological Chemistry* 193:265.

Manchenko GP. 2002. *Handbook of detection of enzymes on electrophoretic gels*, Boca Raton: CRC Press.

Mann K, Edsinger-Gonzalez E, Mann M. 2012. In depth proteomic analysis of a molluscan shell: acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*. *Proteome Science* 10:28. DOI: 10.1186/1477-5956-10-28.

Marie B, Luquet G, Pais de Barros JP, Guichard N, Morel S, Alcaraz G, Bollache L, Marin F. 2007. The shell matrix of the freshwater mussel *Unio pictorum* (Paleoheterodonta, Unionoida). *FEBS Journal* 274:2933-2945. DOI: 10.1111/j.1742-4658.2007.05825.x.

Merril CR, Washart KM. 1998. Gel electrophoresis of proteins: a practical approach. IN Press, O.U. (Ed.) New York.
Michenfelder M, Fu G, Lawrence C, Weeaver JC, Wustman BA, Taranto L, Evans JS, Morse DE. 2003. Characterization of two molluscan crystal-modulating biomineralization proteins and identification of putative mineral binding domains. *Biopolymers* 70:522-533. DOI: 10.1002/bip.10536.

Miyamoto H, Miyashita T, Okushima M, Nakano S, Morita T, Matsushiro A. 1996. A carbonic anhydrase from the nacreous layer in oyster pearls. *Proceedings of the National Academy of Sciences* 93:9657-9660.

Miyamoto H, Yano M, Miyashita T. 2003. Similarities in the structure of nacrein, the shell-matrix protein, in a bivalve and a gastropod. *Journal of Molluscan Studies* 69:87-89. DOI: 10.1093/mollus/69.1.87.

Miyamoto H, Miyoshi F, Kohno J. 2005. The carbonic anhydrase domain protein nacrein is expressed in the epithelial cells of the mantle and acts as a negative regulator in calcification in the mollusc *Pinctada fucata*. Zoological Science 22:311-315. DOI: 10.2108/zsj.22.311.

Miyamoto H, Endo H, Hashimoto N, Limura K, Isowa Y, Kinoshita S, Kotaki T, Masaoka T, Miki T, Nakayama S, Nogawa C, Notazawa A, Ohmori F, Sarashina I, Suzuki M, Takagi R, Tkahashi J, Takeuchi T, Yokoo N, Satoh N, Toyohara H, Miyashita T, Wada H, Samata T, Endo K, Nagasawa H, Watabe S. 2013. The diversity of shell matrix proteins: genome-wide investigation of the pearl oyster, *Pinctada fucata*. *Zoological Science* 30:801-816. DOI: 10.2108/zsj.30.801.

Montagnani C, Marie B, Marin F, Belliard C, Riquet F, Tayalé A, Zanella-Cleon I, Fleury E, Gueguen Y, Piquemal D, Cochennec-Laureau N. 2011. Pmarg-Pearlin is a matrix protein involved in nacre framework formation in the pearl oyster *Pinctada margaritifera*. *ChemBioChem*. 2011; 12:2033-2043. DOI: 10.1002/cbic.201100216.
Mutvei H. 1980. The nacreous layer in molluscan shells. In: The mechanisms of biomineralization in animals and plants. Proceedings of the third international biomineralization symposium, pp. 49-56. Tokai University Press, Tokyo.

Norizuki M, Samata T. 2008. Distribution and function of the nacrein-related proteins inferred from structural analysis. *Marine Biotechnology* 10:234-241. DOI: 10.1007/s10126-007-9061-x.

Nudelman F, Chen HH, Goldberg SW, Addadi L. 2007. Memorial lecture lessons from biomineralization: comparing the growth strategies of mollusc shell prismatic layers in *Atrina rigida*. *Faraday Discussions* 136:9-25.

Ozensoy GO, Capasso C, Supuran CT. 2016. A magnificent enzyme superfamily: carbonic anhydrase, their purification and characterization. *Journal of Enzyme Inhibition and Medicinal Chemistry* 31:5, 689-694. DOI: 10.3109/14756366.2015.1059333.

Pastorek J, Pastorekova S, Callebaut I, Mormon JP, Zelnik V, Opavsky R, Zatovicova M, Liao S, Portetelle D, Stanbridge EJ. 1994. Cloning and characterization of MN, a human tumor-associated a protein with a domain homologous to carbonic anhydrase and a putative helix-loop-helix DNA binding segment. *Oncogene*, 9(10): 2877-2888.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 8:785-786. DOI: 10.1038/nmeth.1701.

Raz S., Weiner S., Addadi L. 2000. Formation of high magnesian calcites via an amorphous precursor phase. Possible biological Implications. *Advanced Materials* 12, 38–42. DOI: 10.1002/(SICI)1521-4095(200001)12:1<38::AID-ADMA38>3.0.CO;2-I.

Samata T, Hayashi N, Kono M, Hasegawa K, Horita C, Akera S. 1999. A new matrix protein family related to the nacreous layer formation of *Pinctada fucata*. *FEBS Letters* 462:225-229.
Saraswathy N, Ramalingam P. Phosphoproteomics 2011. In: Concepts and Techniques in Genomics and Proteomics 203-211.

Song X, Wang X, Li L, Zhang G. 2014. Identification two novel nacrein-like proteins involved in the shell formation of the Pacific oyster *Crassostrea gigas*. *Molecular and Biological Reports* 41:4273-4278.

Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U, Brunak S, Wandall HH, Levery SB, Clausen H. 2013. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO Journal* 32:1478-1488. DOI: 10.1038/emboj.2013.79.

Su J, Zhu F, Zhang G, Wang H, Xie L, Zhang R. 2016. Transformation of amorphous calcium carbonate nanoparticles into aragonite controlled by ACCBP. *Cryst Eng Comm* 18:2125–2134. DOI: 10.1039/C5CE02288F.

Sudo S, Fujikawa T, Nagakura T, Ohkubo T, Sakaguchi K, Tanaka M, Nakasima K, Takahashi T. 1997. Structure of mollusc shell framework proteins. *Nature* 387:563-564. DOI: 10.1038/42391.

Suzuki M, Iwashima A, Tsutsui N, Ohira T, Kogure T, Nagasawa H. 2011. Identification and characterization of a calcium carbonate-binding protein, blue mussel shell protein (BMSP), from the nacreous layer. *ChemBioChem* 12:2478-2487. DOI: 10.1002/cbic.201100317.

Takakura D, Norizuki M, Ishikawa F, Samata T. 2008. Isolation and characterization of the N-linked oligosaccharides in nacrein from *Pinctada fucata*. *Marine Biotechnology* 10:290-296. DOI: 10.1007/s10126-007-9063-8.
Takeuchi T, Sarashina I, Iijima M, Endo K. 2008. In vitro regulation of CaCO$_3$ crystal polymorphism by the highly acidic molluscan shell protein Aspein. *FEBS Letters* 582:591-596.

DOI: 10.1016/j.febslet.2008.01.026.

Thornton DJ, Carlstedt I, Sheehan JK. 1994. Identification of glycoproteins on nitrocellulose membranes and gels. IN Jm, W. (Ed.) *Method Molecular Biology*. Totowa, NJ: Human Press.

Ulmason B, Waheed A, Shah GN, Grubb JH, Sly WS, Tu C, Silverman DN. 2000. Purification and kinetic analysis of recombinant CAXII, a membrane carbonic anhydrase overexpressed in certain cancers. *Proceedings of the National Academy of Sciences* 97(26):14212-14217.

DOI: 10.1073/pnas.97.26.14212

Vulo D, Nishimori I, Scozzafava A, Supuran CT. 2008. Carbonic anhydrase activators: activation of the human cytosolic isozyme III and membrane-associated isoform IV with amino acids and amines. *Bioorganic & Medicinal Chemistry Letters* 18:4303-4307. DOI: 10.1016/j.bmcl.2008.06.075.

Wang D, Wallace AF, De Yoreo JJ, Dove PM. 2009. Carboxylated molecules regulate magnesium content of amorphous calcium carbonates during calcification. 106(51):21511-6.

DOI: 10.1073/pnas.0906741106.

Wilbur KM, Anderson NG. 1948. Electrometric and colorimetric determination of carbonic anhydrase. *Journal of Biological Chemistry* 176:147-154.

Weiss IM, Kaufmann S, Mann K, Fritz M. 2000. Purification and Characterization of Perlucin and Perlustrin, Two New Proteins from the Shell of the Mollusc *Haliotis laevigata*. *Biochemical and Biophysical Research Communications* 267:17-21. DOI: 10.1006/bbrc.1999.1907.
Xie J, Liang J, Sun J, Gao J, Zhang S, Liu Y, Xie L, Zhang R. 2016. Influence of the extrapallial fluid of *Pinctada fucata* on the crystallization of calcium carbonate and shell biomineralization. *Crystal Growth & Design* 16:672-680. DOI: 10.1021/acs.cgd.5b01203.

Yarra T, Gharbi K, Blaxter M, Peck LS, Clark MS. 2016. Characterization of the mantle transcriptome in bivalves: *Pecten maximus*, *Mytilus edulis* and *Crassostrea gigas*. *Marine Genomics* 27:9-15. DOI: 10.1016/j.margen.2016.04.003.

Zhang C, Zhang R. 2006. Matrix proteins in the outer shells of molluscs. *Marine Biotechnology* 8:572-586. DOI: 10.1007/s10126-005-6029-6.

Zigovecki GZ, Posilovic H, Bermanee V. 2009. Identification of biogenetic calcite and aragonite using SEM. *Geologia Croatia* 62(3): 201-206. DOI: 10.4154/gc.2009.14.
Nucleotide and deduced amino acid sequence in *P. mazatlanica* N66 cDNA (GenBank: MH473230).

The putative signal peptide is *underlined*. Conserved anhydrase carbonic domains are highlighted by *shaded boxes*. *Triangles* indicate amino acid residues involved in catalysis. Putative phosphorylation sites are shown in *bold letters*, and amino acids representing N- and O-glycosylation are *circle* and *boxed* respectively. The putative polyadenylation signal (AATAAA) is indicated in *italics*. The peptide sequences obtained by mass spectrometry analyses are *boxed with dotted line*.
Amino acid sequence alignment of N66 and related shell matrix proteins.

Accession numbers: *P. mazatlanica* N66 (this study, GenBank: MH473230), *P. maxima* N66 (GenBank: BAA90540.1), *P. fucata* Nacrein (GenBank: BAA119401.1), *P. margaritifera* nacrein B2 (GenBank: ADY69618.1), B3 (GenBank: AEC03971.1), B4 (GenBank: AEC03972.1) and C5 (GenBank: AEC03973.1), and *P. maxima* N45 (GenBank: ACT55367.1). Signal peptide is underlined; conserved carbonic anhydrase domains are highlighted by shaded boxes. Triangles indicate amino acid residues involved in catalysis, open triangles indicates zinc binding histidines. Alpha helix (cylinders) and beta sheets (arrows) are shown along the amino acid sequence. Below the protein sequences is a key denoting conserved sequence (*), conservative (:) and semi-conservative (.) modifications.
Manuscript to be reviewed

NacreinB2              GNGDNGYNGDNGNSDGRLRRWDL
NacreinB3              GNGDNGYNGDNGNSDWRLRRWDL
N45_P.maxima           -NGDNGYNGDNGNSDGRLRRWDL
N66_P.maxima           GNGDNGYNGDNGNSDGRLRRWDL

NacreinB2              NGNNGNNGNGNNGNNGNGNNGNNGGNGNNGNNGDYGSNDNNGGNGNNGNNGDN 410
N45_P.maxima           ------------------------------------------------------------ 251
NacreinB2              GNNRNNGNGNNGYNGNNGDNGNNGNNGNGNNGNNGNGNNGNNGYNGNNGNNGNNG 350
NacreinB3              GNNRNNGNGNNGYNGNNGDNGNNGNNGNGNNGNNGNGNNGNNGGNDNNGNNG 389
NacreinB4              GNNRNNGNGNNGYNGNNGDDGNNGNNGNGNNGNNGYGNNGNN--------------GNNG 405
N45_P.maxima           ------------------------------------------------------------ 251
N66_P.mazatlanica      GNNRNNGNGNNGYNGNNGDNGNNGNNGNGNNGNNGNGNNG--------NNGG-------- 365

NacreinC5              ------------------------------------------------------------ 249
NacreinB2              -------------------------------------GDNGNSGNNGNGNNGNNGNNGGN 298
NacreinB3              GNSGNNGN--------------GNNGYNGNNGYNGNNGDNGNSGNNGNGNNGYNGNNGGN 329
NacreinB4              GNSGNNGNGNNGYNGNNGDNGNGNNGYNGNNGYNGNNGDNGNSRNNGNGNNGYNGNNGDN 359
N45_P.maxima           ------------------------------------------------------------ 251
N66_P.mazatlanica      HYDHCDNNGDNGYNCDNGNNENNGNNGNNGNNGYNGNNGDNGNGNNGYNGNNGYNGNNGDN 284

NacreinC5              EAHLVFRHDEKKEIEPPRIWLGRNFNGSNEFVVVGVFLEVGDEGYGDEPDDDECKRILKG 235
NacreinB2              EAHLMFRPDEKKEIKPSRIWLGKNYNGSNEFVVVGVFLEVGDEGYGDEPDDDECKRILKG 238
NacreinB3              EAHLVFRHDEKKEIRPPRILLGRNFSGINEFVVVGVFLEVGDEGYGDEPDDDECKRILKG 237
N45_P.maxima           EAHLVFHHDDKKEIKPPRVKLGGVYAGRNKFVVVGVFLEVGDEGYGDEPDDDECKRILKG 236
N66_P.mazatlanica      EAHLVFHHDEKKEIKPPRIPLGRNFSGINEFVVVGVFLEDGDEGYGDEPDDYECKRILKG 238

NacreinB4              KVSNHQNHAPEFDSED-EKL
N45_P.maxima           DVTNHQNRAPEFEPEDGDKL

************** *  **** *.**:*:****:*  ******:**************

NacreinB4              R----DICQGPYYWHT
NacreinB3              REQSEEQCHGPYDWHT
NacreinB4              REQSEEQCHGPYDWHT
N45_P.maxima           RD----ICQGPYHWHT
N66_P.maxima           RD----ICQGPYHWHT

NacreinB4              M
NacreinB2              M
NacreinB4              M
N66_P.maxima           M
N66_P.mazatlanica      M

NacreinB4              TESVLWVVEKCHVQVSR
N66_P.maxima           NETVLWVVEKCHVQVSR

NacreinC5              M
NacreinB2              M
NacreinB4              M
N66_P.maxima           M
N66_P.mazatlanica      M

NacreinB4              RCIVKKAKRLSRILECAYRHKK
N66_P.maxima           GCIVKKAKRLSRILECAYRHKK

NacreinB4              PKNSSCKAG
NacreinC5              PKNSSCKAG

NacreinB4              FSY
NacreinC5              FSY

VPR

LCHCASMHRHDHYMDMDQTYPNGLGY

ENVRRIHTERY

ENVRRMHTERY

ENARRMHTERY

ANVRRMHAERY

LKN

CGSKERQSPI

HIH

RCIVKKAKRLSRILECAYRHKK

IGKSRRKGSEHSVDRHFTPM 175

NGKSRRKGSEHSVNRHFTPM 176

NGKSRRKGSEHSVNRHFTPM 176

HCIVKKAKRLSRILECAYRHKK

VFDPRLPKFPRMLRKSLDT 116

RGFRPRLPKFPRMLRKSLDT 118

RHKLPRLKFKPHMKSLDT 120

LPYLPRLKFKPHMKSLDT 117

R-KLPRLKFKPHMKSLDT 119

TESVLWVVEKCHVQVSR

NETVLWVVEKCHVQVSR

CHY

CEP

PKNSSCKAG

SGGSSCDAG

SGESSCKAG

SGESSCKAG

SGGSSCDAG

FSY

Q 60

530

545

354

525

485

470

430

425

410

298

284

283

275

269

251

251

289

251

249

235

230

226

222

218

214

210

196

192

188

184

180

176

172

168

164

160

156

152

148

144

140

136

132

128

124

120

116

112

108

104

100

96

92

88

84

80

76

72

68

64

60

56

52

48

44

40

36

32

28

24

20
Biochemical characterization of the native and recombinant N66 from *P. mazatlanica*.

Analysis by 10% SDS-PAGE. A) Acetic soluble matrix protein extract; B) CBB: Coomassie Brilliant Blue; Silver nitrate stain; and PAS: Periodic acid/Schiff staining. MM: molecular marker; N: Native protein, R: Recombinant protein and Ovoalbumin as positive control (+) for PAS.
Figure 4 (on next page)

Carbonic anhydrase activity of native and recombinant N66 from *P. mazatlanica*.

Carbonic anhydrase activity was visualized by soaking the agarose plate in bromocresol purple, and exposing the plate to saturated CO$_2$ water. Commercial bovine carbonic anhydrase was used as positive control and water as negative control. N: native N66 and R: Recombinant N66 from *P. mazatlanica*. 
Effect of native and recombinant N66 from *P. mazatlanica* on CaCO$_3$ crystal growth.

Effect of native and recombinant N66 from *P. mazatlanica* on CaCO$_3$ crystal growth. SEM micrographs of calcium carbonate crystal growth in the presence of 0.2 mg/mL of N66 (native or recombinant) incubated at 4 °C for 21 days. Controls (A-C): Salts without N66; Native N66 (D-F): Salts in the presence of native N66 and, Recombinant (G-I): Salts in the presence of recombinant N66. Pictures are representative of three independent experiments.
\[ \text{NaHCO}_3 + \text{MgCl}_2 \]

\[ \text{NaHCO}_3 + \text{CaCl}_2 \]

\[ \text{CaCO}_3 \]
Table 1 (on next page)

Primer sequences and expected amplicon size for N66 amplification.
| Primers          | Primer sequence (5’-3’)                   | Amplicon size (pb) |
|------------------|-------------------------------------------|-------------------|
| N66_1F           | ATGTGGAGAATGACGACGCTT                     | 425               |
| N66_425R         | TGTAATATGGAGATTTGGG                      |                   |
| N66_1F           | ATGTGGAGAATGACGACGCTT                     | 680               |
| N66_680R         | CTTTAATATGCGCTTACATTC                    |                   |
| N66_680Fw        | GAATGTAAGCGCATATTTAAAAGG                 | 962               |
| N66_1642Rv       | GGGACGTCTGGTTCCACATAC                    |                   |
| N66_1F           | ATGTGGAGAATGACGACGCTT                     | 1766              |
| N66_1766R        | CTCATATAAGTTTTGTGACACAGG                 |                   |
| N66_HindF        | GTGAAGCTTTTGCTCCATGCACAGGCATG           | 1642              |
| N66_HindR        | CACAAAGCTTGGGACGTCTTTGTTCCACATAC        |                   |
Table 2 (on next page)

Amino acid composition of N66 from mantle of *P. mazatlanica*. 
| Amino acid residue | Residues | Mol%  |
|--------------------|----------|-------|
| Ala                | 10       | 1.7%  |
| Arg                | 35       | 6.1%  |
| Asn                | 113      | 9.6%  |
| Asp                | 35       | 6.1%  |
| Cys                | 14       | 2.4%  |
| Gln                | 9        | 1.6%  |
| Glu                | 30       | 5.2%  |
| Gly                | 91       | 5.8%  |
| His                | 29       | 5.0%  |
| Ile                | 6        | 2.8%  |
| Leu                | 28       | 4.9%  |
| Lys                | 30       | 5.2%  |
| Met                | 8        | 1.4%  |
| Phe                | 13       | 2.3%  |
| Pro                | 17       | 3.0%  |
| Ser                | 28       | 4.9%  |
| Thr                | 16       | 2.8%  |
| Trp                | 7        | 1.2%  |
| Tyr                | 23       | 4.0%  |
| Val                | 24       | 4.2%  |
| Mr                 | 63.68    |       |
| pI                 | 7.38     |       |
Table 3 (on next page)

Summary of the native N66 purification from the shell of *P. mazatlanica*.
| Step                        | Protein (mg/ml) | Total protein (mg) | Total activity (U) | Specific activity (U/mg) | Purification (fold) | Yield (%) |
|-----------------------------|-----------------|--------------------|--------------------|--------------------------|--------------------|------------|
| Acetic acid extraction-dialysis | 19.20           | 19.20              | 990.72             | 51.60                    | 1                  | 100        |
| Sepharose-Ni                | 0.32            | 3.20               | 560.00             | 175.0                    | 3.39               | 56.52      |

Protein concentration was estimated by Lowry method. The experiments were conducted three times. CA activity was estimated by Wilbur and Anderson method.
Table 4 (on next page)

Native N66 peptide sequences matching with mass spectrometry and deduced amino acid sequences from the N66 gene.

*Peptides obtained only for the 60 kDa band.
| Peptide       | Calculated mass (Da) | Observed mass (Da) |
|---------------|----------------------|--------------------|
| QSPINIWSHR    | 1237.38              | 1237.64            |
| NLHIHIGK      | 931.11               | 931.54             |
| HFTPMEAHLVFHH | 1465.69              | 1500.73            |
| RWDLENVR      | 1087.20              | 1087.56            |
| SLDVEITPEMVLPPIK* | 1781.14          | 1780.97            |
| RVLDALR*      | 842.01               | 842.52             |