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.

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

4 Crisalejandra Rivera-Perez[1], Catalina Magallanes-Dominguez[2], Rosa Virginia Dominguez-

5 Beltran[3], Josafat Jehu Ojeda-Ramirez de Areyano[2], Norma Y. Hernandez-Saavedra[2*]

6 [1]Department of Fisheries Ecology, CONACYT-Centro de Investigaciones Biológicas del

7 Noroeste (CIBNOR), La Paz, Baja California Sur, México.

8 [2]Department of Fisheries Ecology, Molecular Genetics Laboratory, Centro de Investigaciones

9 Biológicas del Noroeste (CIBNOR), La Paz, 23096, Baja California Sur, México.

10 [3]Tecnológico Nacional de México, La Paz, 23080, Baja California Sur, México.

11 *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

66   α-CA family, class II and VII, respectively (Mann, Edsinger-Gonzalez & Mann, 2012; Leggat et

67   al., 2005; Kono, Hiyashi & Samata, 2000; Miyamoto et al., 1996). These proteins are known to

68   contain a metal ion (Zn^{2+}) coordinated by three His residues at the active site (Ozensoy, Capasso

69   & Supuran, 2016) and catalyze the reversible hydration of carbon dioxide to bicarbonate

70   (Miyamoto et al., 1996).

71   These proteins are conserved in bivalves and gastropods and have a molecular weight

72   ranging from 50 kDa to 66 kDa (Song et al., 2014; Mann, Edsinger-Gonzalez & Mann, 2012;

73   Leggat et al., 2005; Miyamoto, Yano, Miyashita, 2003; Kono, Hayashi & Samata, 2000;

74   Miyamoto et al., 1996). Previous studies have recognized that nacrein and N66 are glycoproteins

75   (Leggat et al., 2005; Kono, Hayashi & Samata, 2000), where nacrein possess N-glycan

76   containing sulfite and sialic acid at its terminus (Takakura et al., 2008), furthermore, they both

77   have calcium binding ability (Miyamoto et al., 1996). Despite their similar biochemical

78   properties, N66 has almost twice length of NG repeat domain than nacrein (Miyamoto, Yano &

79   Miyashita, 2003; Kono, Hayashi & Samata, 2000; Miyamoto et al., 1996). This domain has been

80   proved to function as negative regulator in shell formation in nacrein (Miyamoto, Miyoshi &

81   Kohno, 2005), and also as a positive regulator, both as a Ca^{2+} concentrator and an enzyme

82   required for production of carbonate ions (Norizuki & Samata, 2008). The longer repeat NG

83   domain have been related to stronger reaction of protein with Ca^{2+} molecules, matrix

84   components and crystals, resulting in an improved calcification capability (Samata et al., 1999).

85   The nacrein gene has been described in the bivalve *Pinctada maxima* (Kono, Hayashi &

86   Samata, 2000) and the gastropod *Turbo marmoratus* (Miyamoto, Yano & Miyashita, 2003),

87   while N66 gene has been only described in *P. maxima* (Kono, Hayashi & Samata, 2000).

88   Differential expression of SMP mRNAs in the mantle of mollusks has been related to their

89   crystal layer formation. Nacrein mRNA is highly expressed in the mantle edge and pallial

90   (Miyamoto, Miyoshi & Kohno, 2005; Miyamoto et al., 1996), while N66 mRNA is expressed in

91   the dorsal region of the mantle and the mantle edge (Kono, Hayashi & Samata, 2000), the former

92   has been associated for the nacreous layer function and the latter for the prismatic layer

93   formation (Sudo et al., 1997).

94   In order to gain insights into the biomineralization processes in mollusk and to increase the

95   knowledge of the nacre formation, a comprehensive study of the proteins involved in nacre

96   deposition is required. Pearl oyster, *Pinctada mazatlanica* is a good model species to study

97   biomineralization, since includes two layers, an inner nacreous layer, and an outer prismatic

98   calcite layer. In this study, a N66 from the shell was isolated and its coding sequence was

99   obtained from the mantle of the pearl oyster *P. mazatlanica* and overexpressed on a prokaryotic

100   system, the resulting native and recombinant proteins were functionally characterized.

Material & Methods

102   **Biological material.** Three adult female organisms were provided by Perlas del Cortez S. de

103   R.L. MI. located at Bahia de La Paz B.C.S. Organisms were transported to the Molecular

104   Genetics Laboratory at CIBNOR, and mantle tissue was dissected and stored at −80 °C until

105   used.

106

107   **Molecular characterization of N66 from mantle tissue**

108   **Total RNA extraction and RNA reverse transcription.** Total RNA was extracted with

109   TRIzol® (Invitrogen, U.S.A.) according to the manufacturer's instructions. Samples were

110   homogenized using a glass pestle; then, two consecutive extractions of each sample were made.

111   RNA purity and concentration were determined by spectrophotometry using a NanoDrop ND-

112   2000 (Thermo Scientific, U.S.A.) at 260/280 and 260/230 nm absorbance ratios (range, 1.9 –

113   2.0). The RNA integrity was assessed on 1% (w/v) agarose-synergel gel. To ensure complete

114   DNA absence, a direct PCR was performed using 1 µL (50 ng/µL) of each RNA preparation

115   with 28S ribosomal specific primer as a non-amplified control. After that, 1 µg of total RNA was

116   used from each verified RNA sample for cDNA synthesis using the cloned AMV First-Strand

117   cDNA Synthesis Reaction (Invitrogen, U.S.A.) and oligo-dT primer, afterwards cDNAs were

118   stored at −80 °C until use. Control reactions were performed without template or non-reverse

119   transcribed RNA to determine the presence of DNA.

120

121   **Amplification of coding sequence of N66 from mantle.** A search for sequences

122   encoding N66 from mollusks available at GenBank yielded three homologous sequences, N66

123   *Pinctada maxima* (GenBank: AB032613), N44 *P. maxima* (GenBank: FJ913472) and SP-S

124   *Mizuhopecten yessoensis* (GenBank: AB185328). Based on these sequences, specific primers

125   were designed to amplify the coding sequence of N66 (Table 1). PCR amplification of cDNA

126   fragments encoding N66 was done in a final volume of 12.5 µL containing 10 pmol of each

127   forward and reverse primer, 50 ng cDNA, 10 nmol dNTPs mixture, 1.25 µL 10X Taq Buffer, and

128   1 U Taq DNA polymerase (11615010A, Invitrogen). PCR amplification were carried out for 5

129   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,

130   followed of 35 cycles consisting 1 min at 94 °C, 1 min at 50 °C, and 2 min at 72 °C. In the last

131   cycle, the extension step at 72 °C lasted 10 min. PCR products were analyzed on 1 % (w/v)

132   agarose-synergel gels and visualized under UV light after staining with UView loading dye (166-

133   5112, Biorad). PCR products were cloned using the TOPO-TA cloning kit (K4500-01,

134   Invitrogen) and sequenced for both strands. The obtained sequence was submitted to the National

135   Center for Biotechnology Information (NCBI) database for BLAST searching and other

136   informatics analysis.

137

138   ***In silico* analyses.** Identification of the open reading frame was performed using the ORF

139   finder software (https://www.ncbi.nlm.nih.gov/orffinder/) and sequence alignment with

140   homologous sequences were performed by using CLUSTAL Omega. Identification of putative

141   protein motifs was performed using the MotifScan (Pfam HMMs global models database)

142   available at the Swiss Institute of Bioinformatic (http://myhits.isb-sib.ch/cgi-bin/motif_scan).

143   Identification of signal peptide was achieved by using SignalP 4.1 Server (Petersen et al., 2011).

144   Theoretical molecular weight, isoelectric point (pI), and amino acid composition of the protein

145   were calculated using the ProtParam software from ExPASy (Expert Protein Analysis System;

146   https://www.expasy.org/). Putative glycosylation and phosphorylation sites of N66 were

147   determined using NetOGly4.0 (Steentoft et al., 2013), NetNGlyc 1.0

148   (http://www.cbs.dtu.dk/services/NetNGlyc/) and NetPhos2.0 (Blom, Gammeltoft & Brunak,

149   1999), respectively.

150

151   **Vector construction and recombinant protein expression.** The coding region of N66

152   cDNA, without signal peptide, was synthesized by PCR with two primers (N66_HindF and

153   N66_HindR) to introduce the restriction enzyme sites (HindIII, AAGCTT) at the 5' and 3' ends

154   (Table 1). The PCR product was directly digested with HindIII (NEB, New England) and

155   isolated from gel electrophoresis, purified and ligated into a polyhistidine fusion protein

156   expression vector pTrcHis C (Invitrogen, Inc.) that was linearized with the same restriction

157   enzyme, yielding pTrcHis C/N66. *Escherichia coli* TOP10 cells were transformed with the

158   construct and cultured in 6 mL of LB medium containing 100 µg/mL ampicillin at 37 °C with

159   shaking at 150 rpm overnight. Five mL of the culture medium were transferred to a flask

160   containing 245 mL of the same medium and further incubated. When OD_{600} of the bacterial cells

161   reached ~0.7, isopropyl β-D-thiogalactosidase was added to a final concentration of 1 mM and

162   incubated at 37 °C with constant shaking at 150 rpm for 4 h. The bacterial cells were harvested

163   by centrifugation (7,500 rpm, 25 min at room temperature) and stored at −20 °C. The frozen

164   bacterial pellet was thawed and dispersed in ice-cold Tris-HCl buffer (50 mM, pH 7.4) and

165   disrupted with glass beads for 5 min of vigorous vortexing. The bacterial lysate was centrifuged

166   (12,000 rpm, 15 min, 4 °C), and the supernatant was recovered. Recombinant His-tagged

167   proteins were purified from the supernatant by using a Sepharose-Ni column (GE Healthcare).

168

169   **Native N66 purification and characterization**

170   **Shell matrix extraction.** The organic matrix of prismatic and nacreous layers were

171   crushed to fine powder. The powdered matrix (20 g) were suspended in 100 mL cold acetic acid

172   (4 °C, 10% v/v) and incubated for 24 h with continuous stirring. Soluble (ASM) and insoluble

173   (AIM) matrix fractions were separated by centrifugation at 14,000 rpm for 20 minutes. The AIM

174   was rinsed with distilled water and lyophilized. The ASM was dialyzed 24 h against cold acetic

175   acid (4 °C, 1% v/v), afterwards the ASM was dialyzed other 24 h against distilled water. The

176   dialyzed ASM was concentrated by lyophilization. Protein concentration was determined by

177   Lowry method (Lowry et al., 1951).

178

179   **N66 purification by affinity chromatography.** Native and recombinant N66 were

180   isolated by affinity chromatography using a Sepharose-Ni column (GE Healthcare; 1.45 x 5.0

181 cm). Protein sample (1 mL) was loaded onto the column, previously equilibrated with 50 mM

182 Tris-HCl pH 7.4, 300 mM NaCl buffer, mixed and incubated for 1 h at 25 °C. The unbound

183 enzyme was washed with five bed volumes of 50 mM Tris-HCl pH 7.5, 300 mM NaCl buffer.

184 The bound enzyme was eluted by washing the column with elution buffer containing 50 mM

185 Tris-HCl pH 7.5, 300 mM NaCl, 300 mM Imidazole. Fractions of 1 mL were collected from the

186 elution and evaluated by electrophoresis. Fractions containing N66 were pooled and concentrated

187 through centrifugal filters (AMICON Ultra-10, Millipore) at 4,000 g for 20 min at 25 °C to a

188 final volume of 500 μL. Concentrated sample was desalted using PD-10 desalting column

189 (Amershan Pharmacia). Glycerol was added to the enzyme solution (final concentration 50%),

190 and the sample was stored at −20 °C until use. Protein content was quantified by Lowry method

191 (Lowry et al., 1951) and SDS-PAGE analysis of the purified protein was performed as described

192 above.

193

194 **SDS-PAGE and staining.** Protein bands, from native and recombinant proteins, were

195 analyzed by electrophoresis under reducing conditions using 10% SDS-PAGE (Mini-

196 PROTEAN, Bio-Rad), as described by Laemmli (1970) and stained with silver nitrate (Merril &

197 Washart, 1998) in 10% acrylamide gels. Samples were diluted (1:4) with sample buffer (0.125 M

198 Tris-HCl, 2% SDS, 20% v/v glycerol, 0.04% bromophenol blue, 5% β-mercaptoethanol at pH

199 6.8) and heated for 10 min at 100 °C. Electrophoresis separation was performed at 80V constant

200 current. To detect protein bands, gels were stained with 0.1% Coomassie Brilliant Blue R-250 in

201 7.5% acetic acid and 5% methanol at room temperature and distained in the 10% acetic acid and

202 40% methanol.

203   Glycosylation was analyzed by electrophoresis under reducing conditions (Thornton et al.,

204   1994). The gel was first incubated in fixing solution (5% TCA) for 5 min, then incubated in

205   oxidant solution (0.8% periodic acid in 0.3% sodium acetate), and rinsed with water. Then the

206   gel was incubated in Schiff's reagent for 10 min in the dark. Finally, the gel was washed three

207   times with 3% nitric acid for 3 min. The gel was stored in 7.5% acetic acid, 5% v/v methanol.

208   Ovoalbumin (SIGMA) was used as positive control.

209

210   **Mass spectrometry analyses.** To identify the mass spectrometric profile, 20 µg of the

211   native N66 protein was first separated by 10% SDS-PAGE. After electrophoresis, the gel was

212   stained with Coomassie Blue R-250 and bands were manually cut from the gel. After tryptic

213   digestion, the resulting peptides were subjected to Liquid Chromatography-Mass Spectrometry

214   (LC-MS). The LC-MS was performed at the Proteomic Unit Facility at Biotechnology Institute,

215   Cuernavaca. The resulting data were compared with the proteomic database at GenBank and

216   with the deduced amino acid sequence from cDNA of N66 from *P. mazatlanica*.

217

218   **Carbonic anhydrase assay.** Carbonic anhydrase activity from native and recombinant

219   N66 was assayed qualitatively and quantitatively. For qualitative assay, 1% agarose plates were

220   prepared with borate buffer (0.05 M boric acid, 0.025 M NaCl). Samples (10 µL of 0.2 mg

221   protein/ mL) were loaded on the agar plate, after sample diffusion 0.1% bromocresol purple

222   (dissolved in 0.1 M Tris-HCl pH 8.5) was added and incubated for 10 min at room temperature.

223   Afterwards, the solution was discarded and the agar plate was washed with distilled water.

224   Finally, saturated CO_2 water was added as substrate; the production of hydrogen ions during the

225   CO_2 hydration reaction lowers the pH of the solution until the formation of yellow halo zones

226   which indicate carbonic anhydrase activity while dark zone indicate lack of activity (Machenko,

227   2002). Bovine carbonic anhydrase (SIGMA) was used as positive control and boiled enzyme as

228   negative control.

229   The quantitative *in vitro* assay for CA activity is described in detail by Wilbur and

230   Anderson (1948). Briefly, CA activity was measured by the decrease of pH resulting from the

231   hydration of CO_2 to HCO_3^- and H^+ after the addition of substrate. All experiments were

232   performed at 4 °C. To run the assay, 5 mL of 0.012 M Tris-HCl buffer (pH 8.3) were transferred

233   to a small beaker and 1 mL of homogenate diluted in 0.012 M Tris-HCl buffer was added to

234   obtain 20 µg of protein. The mixture was constantly stirred with a magnetic stirring bar.

235   Afterwards, 4 mL of cold saturated CO_2 solution were added, and the decrease in pH was

236   recorded by a pH probe. CA activity was calculated as $(t_0-t)/t$, where t_0 is the time needed for the

237   non-catalyzed reaction and t is the time for the catalyzed reaction to obtain a pH decrease from

238   8.0 to 6.0. Units of enzyme activity (U) were normalized to the total protein content.

239

240   **Crystallization *in vitro* assay with native and recombinant N66.** Three crystallization

241   solutions were prepared, two induces calcite (calcitic crystallization solution) and one induces

242   aragonite (aragonite crystallization solution). Calcitic crystallization solutions (40 mM $CaCl_2$, pH

243   8.2, 100 mM $NaHCO_3$ and 100 mM $CaCO_3$) and aragonite crystallization solution (40 mM

244   $MgCl_2$, pH 8.2, 100 mM $NaHCO_3$) were prepared following the protocol of Weiss et al. 2000

245   and Hillner et al. 1992, respectively. Native and recombinant N66 (10 µL of 0.2 mg/ml) were

246   added to 50 µL of the saturated solution. As a control, buffer solution without protein was used.

247   Each solution experiment was made in triplicates and incubated at 4 °C for 21 days in a cover

248   slide situated at the bottom of a six-hole microplate which was sealed with parafilm to avoid

249   contamination. The morphology of the crystals produced were analyzed by scanning electron

250   microscopy (SEM) at the Electronic Microscopy Laboratory at CIBNOR.

251

252   **Statistical analysis**

253   Statistical analysis was performed using the GraphPad Prism Software. The results are expressed

254   as means ± s.d. Significant differences (P<0.05) were determined by ANOVA followed by a

255   pair-wise comparison of means (Tukey's test).

256

257   **Results**

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

259   Nucleotides and deduced amino acid sequence of N66 from the mantle of *P. mazatlanica* are

260   shown in Fig. 1. The open reading frame of N66 obtained from mantle cDNA was 1766 bp long

261   (GenBank: MH473230) and encoded a 576 amino acid protein with a calculated molecular mass

262   of 63.68 kDa before any post-translational modification and an isoelectric point (pI) of 7.38.

263   Removal of the signal peptide sequence (residues 1 to 22) resulted in a theoretical molecular

264   weight of 61.12 kDa and a theoretical pI of 7.21. It possess high proportions of N (19.6%) and G

265   (15.8%), which accounted for 35.4% of the total amino acid residues (Table 2). Analysis of the

266   N66 amino acid sequence showed homology with the carbonic anhydrase (CA) family,

267   exhibiting 34% identity with human CAVII (GenBank: NP_005173), with high homology to

268   amino acid residues involved in catalysis (Q_{125}, N_{154} and E_{168}) and three histidine residues

269   essential for the zinc cofactor binding (H_{156}, H_{158} and H_{181}) (Fig. 1).

270   N66 protein sequence possess in its modular structure a signal peptide (1-22 aa), two

271   carbonic anhydrase domains (G_{69}-G_{221} and D_{490}-K_{554}), and a NG rich sequence (N_{245}-G_{437}) (Fig.

272   S1), sharing 82.82% identity to N66 from *P. maxima*, ~92% to nacreins (B2, B3 and B4) from

273   *Pinctada margaritifera*, 72.62% to nacrein from *Pinctada fucata* and 78.70% to N45 from *P.*

274   *maxima* (Fig. 2). All the protein sequences contained the same protein structure, however the

275   length of the carbonic anhydrase (CA) and NG domain were variable among the proteins.

276   Further bioinformatics analysis of the N66 sequence suggests that some serine and threonine

277   residues in N66 may either be phosphorylated (36 putative sites) or N- and O-glycosylated (16

278   putative sites), as indicated in Fig. 1, which are common features related to proteins involved in

279   biomineralization (Nudelman et al., 2007). Finally, most of the O-glycosylation and

280   phosphorylation sites were found in the CA domain.

281

282   **N66 protein characterization**

283   **Native N66 from *P. mazatlanica* shells.** The organic matrix of the shell from *P.*

284   *mazatlanica* was separated by SDS-PAGE revealing a wide range of abundant molecular weight

285   proteins in the ASM fraction (Fig. 3A). The ASM fraction was used to isolate the N66 by affinity

286   chromatography using a Sepharose-Ni matrix (Fig. S2) since the active site of the CA domain

287   contains three histidines involved in metal ion binding which has been shown to be conserved in

288   N66 from *P. mazatlanica* (this study) and *P. maxima* (Kono, Hayashi & Samata, 2000). The

289   fractions containing the eluted N66 (Fig. S2) were pooled and desalted. Two bands in SDS-

290   PAGE and silver stained gels were observed (Fig. 3B) with molecular weights of ~60 and 64

291   kDa. A flow sheet of the purification process for N66 is presented in Table 3. The enzyme was

292   enriched 3.39-fold and the yield was 56.52%. Electrophoretic analysis of N66 with PAS stain,

293   demonstrated that N66 is glycosylated (Fig. 3B), however, the type of carbohydrate associated

294   was not analyzed. The results of LC-MS of the bands matched the deduced amino acid sequence

295   of the N66 of the mantle of *P. mazatlanica* (Table 4), as described before (Fig. 1), and allowed

296   us to identify it as a N66, a shell matrix protein from the shell of *P. mazatlanica*.

297

298   **Recombinant N66 from *P. mazatlanica.*** Recombinant N66 protein was expressed as a

299   soluble form at 37 °C with 1 mM IPTG in *E. coli* after 4h induction, and the bands

300   corresponding to the protein in SDS-PAGE analysis agreed with the native N66 of ASM from *P.*

301   *mazatlanica* (Fig. 3B). The recombinant N66 was successfully purified to homogeneity by

302   affinity chromatography using Ni-Sepharose columns (Fig. S3). Electrophoretic analysis showed

303   two protein bands of ~60 and 64 kDa which agreed with that of the native N66 of ASM from *P.*

304   *mazatlanica* (Fig. 3B). Recombinant N66 did not show glycosylation after staining with PAS

305   (data not shown).

306

307   **Carbonic anhydrase activity of native and recombinant N66.** The carbonic anhydrase

308   activity of the native and recombinant N66 was confirmed qualitatively (Fig. 4) and

309   quantitatively. Figure 4 showed comparable yellow halos obtained with the native and

310   recombinant N66, indicating CA activity. The commercial bovine CA was used as positive

311   control. The comparable CA activity of the native and recombinant N66 proteins was confirmed

312   with the Wilbur and Anderson method. The results showed significant differences between

313   native (161 ± 12.5 U/mg protein) and recombinant protein (197 ± 7.5 U/mg protein), being the

314   recombinant protein the one with higher activity.

315   ***In vitro* crystallization by native and recombinant N66.**   To elucidate the effect of

316   native and recombinant N66 on the growth of calcium carbonate crystals, three individual *in*

317   *vitro* crystallization assays testing the formation of aragonite and calcite were established and

318   analyzed by SEM (Fig. 5). First, when no protein was added (negative control) to the three salts

319   tested (NaHCO$_3$+MgCl$_2$, NaHCO$_3$+CaCl$_2$ and CaCO$_3$), different crystal morphology was

320   observed. The crystals formed when sodium bicarbonate in presence of magnesium chloride

321   were unorganized, while in presence of calcium chloride crystals become compact and thicker,

322   and in presence of calcium carbonate the formed crystals exhibited the typical rhombohedra of

323   calcite (Fig. 5A-C). On the other hand, in the presence of 0.2 mg protein/ mL of native N66 to

324   the three salts preparations showed polycrystalline aggregates of aragonite and calcite in

325   presence of MgCl$_2$ and CaCO$_3$ respectively, however in presence of CaCl$_2$, crystals were small

326   structures of calcite and in lesser quantity per area (Fig. 5D-F). In contrast to native N66,

327   recombinant form produced bigger and organized crystal structures forming plates (Fig. 5G-I).

328   **Discussion**

329   The mollusk shell is a complex structure made of organic and mineral components. The

330   shells are composed by layers formed from calcium carbonate polymorphs as aragonite or calcite

331   crystals (Takeuchi et al., 2008; Zhang & Zhang, 2006). The calcium carbonate polymorphs are

332   formed by shell matrix proteins (SMPs), which have been classified as soluble and insoluble

333   based on the two main extraction methods, EDTA and acetic acid (Marie et al., 2007; Kono,

334   Hayashi & Samata, 2000). The insoluble proteins, are chitin-proteinaceous complexes rich in

335   aliphatic amino acids, such as Gly and Ala, while the soluble proteins have been found to be

336   polyanionic, enriched mainly in Asp; among the soluble proteins, functional enzymes such as

337   carbonic anhydrases have been described (Kono, Hayashi & Samata, 2000), which seems

338   contribute to the control of biomineralization.

339   Most of the information about genes encoding enzymes with CA activity to form nacre in

340   mollusks is available for nacreins (Mann, Edsinger-Gonzalez & Mann, 2012; Leggat et al., 2005;

341   Miyamoto et al., 1996). Nacrein gene has been described in *Crassostrea gigas* (Song et al.,

342   2014), *Patella vulgata*, *Mytilus californianus*, *P. maxima* and *Pinctada magaritifera* (Joubert et

343   al., 2010), while N66 mRNA has been only described in *P. maxima* (Kono, Hayashi & Samata,

344   2000) and *P. margaritifera* (Joubert et al., 2010). This homologous genes have been found to be

345   expressed in the mantle (Miyamoto, Myoshi & Kohno, 2005; Kono, Hayashi & Samata, 2000),

346   which is a conserved organ involved in shell formation throughout mollusks. We followed this

347   lead and proved that *P. mazatlanica* express the N66 mRNA in the mantle and secrete the

348   enzyme into the extrapalleal space to participate in the shell growth.

349   Currently, only six SMPs with carbonic anhydrase activity have been isolated and

350   characterized from pearl oysters, nacrein from the Akoya pearl oyster *Pinctada fucata*, turban

351   shell *Turbo marmoratus,* the edible Iwagaki oyster *Crassostrea nippona*, Yesso scallop

352   *Patinopecten yessoensis* and giant clam *Tridacna gigas* (Norizuki & Samata; Leggat et al., 2005;

353   Miyamoto et al., 2003; Miyamoto et al., 1996), and only one N66 from *P. maxima* (Miyamoto et

354   al., 1996). Carbonic anhydrases (CAs) are usually purified via the agarose-sulfonamide affinity

355   column chromatographic procedure (Ozensoy, Capasso & Supuran, 2016), however, in this study

356   a Sepharose-Ni column was successfully used to capture the N66 protein from the shell of *P.*

357   *mazatlanica*, since it is well known that the CA domain possess three conserved histidine

358   residues at the active site and His exhibits the strongest interaction to metal ion matrices, as

359   electron donors groups on the imidazole ring of His form coordinated bonds with the

360   immobilized metal (Bornhorst & Falke, 2000). The isolated protein was identified by mass

361   spectrometry as a N66, due to its similarity to N66 from *P. maxima* and its homolog protein,

362   nacrein; native N66 and its recombinant form showed two bands with a molecular weight of ~60

363   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,

370   migrate similar on SDS-PAGE, the formation of oligomers of N66 is suggested in this study.

371   However, further experiments are required to establish if the double band of N66 from the shell

372   of *P. sterna* is due to glycosylations or oligomerization of the protein.

373   Structurally, N66 in *P. mazatlanica* is similar to N66 from *P. maxima* (Kono, Hayashi &

374   Samata, 2000) and its homologous proteins, nacreins and N45 (Miyamoto et al., 1996), they

375   share in its modular structure a signal peptide, two CA domains and a NG domain. α-CAs

376   contain a domain which is composed by three amino acid residues (Gln, Asn and Glu) and three

377   His (Vulo et al., 2008), according to the deduced primary structure of N66 from *P. mazatlanica*,

378   the protein contains two CA domains composed by Gln_{125}, Asn_{154}, Glu_{168} and three His residues

379   which act as zinc ligands (His_{156}, His_{158} and His_{181}), suggesting that N66 possess a functional CA

380   domain. CA activity has been only proved in nacrein from *P. fucata* (Miyamoto et al., 1996) and

381   N66 from *P. mazatlanica* and its recombinant protein, in this study. The recombinant N66

382   showed higher CA activity than the native N66 and the bovine CA. The NG domain found in

383   N66 from *P. mazatlanica* differs in length and composition to its homologues, accounting for

384   over ~30% of their sequences. This domain has been proved to function as negative regulator in

385   shell formation in nacrein (Miyamoto, Miyoshi & Kohno, 2005), but also as a positive regulator,

386   both as a Ca^{2+} concentrator and an enzyme required for production of carbonate ions (Norizuki

387   & Samata, 2008). It is hypothesized that the longer repeat NG domain might have a stronger

388   interaction of protein with Ca^{2+} molecules, matrix components and crystals, leading to better

389   calcification ability (Samata et al., 1999). This suggests, that N66 might have better calcification

390   ability than its homologs, although, further experiments are required to prove this hypothesis.

391   Post-translational modifications (PTM) of SMPs are a common feature that have been

392   suggested to play an important role in biomineralization. Phosphorylation and glycosylation are

393   the major players in many of the protein functions (Saraswathy & Ramalingam, 2011). In silico

394   analysis of N66 from *P. mazatlanica* showed 36 putative sites of phosphorylation and 16

395   putative sites of glycosylation. The biochemical analysis confirmed that N66 is glycosylated

396   (Fig. 3B), a modification that has been observed in other matrix proteins (Montagnani et al.,

397   2011; Suzuki et al., 2011). Glycosylation participates in the folding of the protein and/or

398   interactions with Ca^{2+} (Suzuki et al., 2011; Borbas, Wheeler & Sikes, 1991) and provide sites of

399   high anionic charge through addition of sialic acid or sulfated oligosaccharides (Michenfelder et

400   al., 2003) which have been related to promote the uptake of Ca^{2+} (Song et al., 2014). The

401   saccharide chains of N66 in *P. mazatlanica* probably have similar effects, however, the chemical

402   identity of the oligosaccharides moieties, the number and the location of actual glycosylation

403   sites along the protein of N66 remain to be determined.

404   Experiments involving *in vitro* calcium carbonate crystallization have been used to prove

405   functional activity of SMPs since most of the described SMPs does not possess catalytic

406   domains. Crystal formation comprise two main process, nucleation and growth. It begins with an

407   amorphous calcium carbonate as an intermediate of the crystalline carbonate polymorph such as

408   calcite or aragonite (Wang et al., 2009). The presence of Mg^{2+} ions foster metastable aragonite

409   growth that make up the mineral phase of nacre (Evans, 2017; Su et al., 2016; Samata et al.,

410   1999), and when Mg^{2+} ion are low induces the formation of calcite (Raz, Weiner & Addadi,

411   2000). Morphological observation of calcite and aragonite by SEM has been made by several

412   authors, showing that crystals of calcite display a rombohedral form while aragonite has an

413   orthorhombic form (Zigovecki, Posilovic & Bermanee, 2009), although the morphology of each

414   polymorph can be changed due to the crystallization condition, e.g. aragonite is needle-like but it

415   changes to flake-like or cauliflower-like (Chakrabarty & Mahapatra, 1999). In this study, N66

416   and its recombinant protein have differential CA activity and lead to calcium carbonate

417   precipitation, displaying structures such as aragonite and calcite when crystallization was made

418   in presence of $MgCl_2$ and $CaCl_2$, respectively. Similar results were obtained by Kono et al.

419   (2000), for N66 from *P. maxima*, which induces aragonite layers, however it requires the

420   presence of N14 and Mg^{2+} ions. The recombinant N66 foster higher crystal growth than native

421   N66, this evidence raise the possibility that the presence of PTMs (e.g. glycosylations and

422   phosphorylations) affect the growth and shape of calcite and aragonite crystals, which is in

423   agreement to previous evidences in *Unio pictorum*, that showed that glycosylations might change

424   the crystal morphology in *in vitro* crystallization experiments (Marie et al., 2007).

425   **Conclusion**

426   In this study, a N66 mRNA from the mantle of *P. mazatlanica* was identified. The

427   transcript is translated to one single protein named N66. The N66 is a glycoprotein with carbonic

428   anhydrase activity and calcium carbonate precipitation ability *in vitro*, suggesting that N66 is

429   involved in the mineralization process in *P. mazatlanica*. Experiments with the recombinant N66

430   suggest that posttranslational modifications might play an important role in the modulation of the

431   CA activity. The identification of the carbohydrates associated to N66, as well as the

432   determination of other posttranslational modification (e.g. phosphorylation, sulfation, etc.), and

433   their role in the calcium carbonate precipitation (e.g. crystal growth, type of polymorph) remains

434   to be tested. Also, to understand the potential oligomerization process of the N66 and its role in

435   the mineralization process remains unclear.

436

437   **References**

438   Addadi L, Joester D, Nudelman F, Weiner S. 2006. Mollusk shell formation: A source of new

439   concepts for understanding biomineralization processes. *Chemistry European Journal* 12:980-

440   987. DOI: 10.1002/chem.200500980.

441   Blom N, Gammeltoft S, Brunak S. 1999. Sequence- and structure-based prediction of eukaryotic

442   protein phosphorylation sites. *Journal of Molecular Biology* 294:1351-1362. DOI:

443   10.1006/jmbi.1999.3310.

444   Borbas JE, Wheeler AP, Sikes S. 1991. Molluscan shell matrix phosphoproteins: Correlation of

445   degree of phosphorylation to shell mineral microstructure and to in vitro regulation of

446   mineralization. *Journal of Experimental Zoology* 258:1-13. DOI:10.1002/jez.1402580102.

447   Bornhorst JA, Falke JJ. 2000. Purification of proteins using polyhistidine affinity tags. *Methods*

448   *in Enzymol* 326:245-254.

449   Buren S, Ortega-Villasante C, Blanco-Rivero A, Martinez-Bernardini A, Shutova T, 2011.

450   Importance of post-translational modifications for functionality of a chloroplast-localized

451   carbonic anhydrase (CAH1) in *Arabidopsis thaliana. PLOS One* 6(6): e21021. DOI:

452   10.1371/journal.pone.0021021

453   Chakrabarty D, Mahapatra S. 1999. Aragonite crystals with unconventional

454   morphologies. *Journal of Materials Chemistry* 9:2953–2957. DOI: 10.1039/a905407c

455 Checa AG, Macías-Sánchez E, Ramírez-Rico J. 2016. Biological strategy for the fabrication of

456 highly ordered aragonite helices: the microstructure of the cavolinioidean gastropods. *Scientific*

457 *Reports* 6:25989. DOI: 10.1038/srep25989.

458 De Luca V, Del Prete S, Vullo D, Carginale V, Di Fonzo P, Osman SM, AlOthman Z, Supuran

459 CT, Capasso C. 2016. Expression and characterization of a recombinant psychrophilic carbonic

460 anhydrase (NcoCA) identified in the genome of the Antarctic cyanobacteria belonging to the

461 genus Nostoc. *Journal of Enzyme Inhibition and Medical Chemistry*, 31(5): 810-817. DOI:

462 10.3109/14756366.2015.1069289.

463 Evans JS. 2017. Polymorphs, proteins, and nucleation theory: A critical analysis. *Minerals.* 7: 6.

464 DOI: 10.3390/min7040062.

465 Hillner PE, Manne S, Gratz AJ, Hansma PK. 1992. AFM images of dissolution and growth on a

466 calcite growth crystal. *Ultramicroscopy* 42:1387-1393. DOI: 10.1016/0304-3991(92)90454-R.

467 Hilvo M, Baranauskiene L, Salzano AM, Sacloni A, Matulis D, Innocenti A, Scozzafava A, Monti

468 SM, Di Fiore A, De Simone G, Lindfors M, Janis J, Valjakka J, Pastorekova S, Pastorek J,

469 Kulomaa MS, Nordlund HR, Supuran CT, Parkkila S. 2008. Biochemical characterization of

470 CAIX, one of the most active carbonic anhydrase isozymes. *Journal of Biological Chemistry* 41:

471 27799-27809. DOI: 10.1074/jbc.M800938200.

472 Joubert C, Piquemal D, Marie B, Manchon L, Pierrat F, Zanella-Cleon I, Cochennec-Laureau N,

473 Gueguen Y, Montagnani C. 2010. Transcriptome and proteome analysis of *Pinctada*

474 *margaritifera* calcifying and shell: focus on biomineralization. *BMC Genomics* 11:613. DOI:

475 10.1186/1471-2164-11-613.

476   Kono M, Hayashi N, Samata T. 2000. Molecular Mechanism of the Nacreous Layer Formation

477   in *Pinctada maxima*. *Biochemical and Biophysical Research Communications* 269:213-218.

478   DOI: 10.1006/bbrc.2000.2274.

479   Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of

480   bacteriophage T4. *Nature* 227:680-685. DOI: 10.1038/227680a0.

481   Leggat W, Dixon R, Saleh S, Yellowlees D. 2005. A novel carbonic anhydrase from the giant

482   clam *Tridacna gigas* contains two carbonic anhydrase domains. *FEBS Journal* 272:3297-3305.

483   DOI: 10.1111/j.1742-4658.2005.04742.x.

484   Liu X, Liu C, Chen L, Sun J. 2011. A new method to extract matrix proteins directly from the

485   secretion of the mollusk mantle and the role of these proteins in shell biomineralization. *Marine*

486   *Biotechnology* 13:981-991. DOI: 10.1007/s10126-011-9362-y.

487   Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. 1951. Protein measurement with the folin

488   phenol reagent. *Journal of Biological Chemistry* 193:265.

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

490   CRC Press.

491   Mann K, Edsinger-Gonzalez E, Mann M. 2012. In depth proteomic analysis of a molluscan shell:

492   acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*. *Proteome Science* 10:28.

493   DOI: 10.1186/1477-5956-10-28.

494   Marie B, Luquet G, Pais de Barros JP, Guichard N, Morel S, Alcaraz G, Bollache L, Marin F.

495   2007. The shell matrix of the freshwater mussel *Unio pictorum* (Paleoheterodonta, Unionoida).

496   *FEBS Journal* 274:2933-2945. DOI: 10.1111/j.1742-4658.2007.05825.x.

497   Merril CR, Washart KM. 1998. Gel electrophoresis of proteins: a practical approach. IN Press,

498   O.U. (Ed.) New York.

499   Michenfelder M, Fu G, Lawrence C, Weeaver JC, Wustman BA, Taranto L, Evans JS, Morse

500   DE. 2003. Characterization of two molluscan crystal-modulating biomineralization proteins and

501   identification of putative mineral binding domains. *Biopolymers* 70:522-533. DOI:

502   10.1002/bip.10536.

503   Miyamoto H, Miyashita T, Okushima M, Nakano S, Morita T, Matsushiro A. 1996. A carbonic

504   anhydrase from the nacreous layer in oyster pearls. *Proceedings of the National Academy of*

505   *Sciences* 93:9657-9660.

506   Miyamoto H, Yano M, Miyashita T. 2003. Similarities in the structure of nacrein, the shell-

507   matrix protein, in a bivalve and a gastropod. *Journal of Molluscan Studies* 69:87-89. DOI:

508   10.1093/mollus/69.1.87.

509   Miyamoto H, Miyoshi F, Kohno J. 2005. The carbonic anhydrase domain protein nacrein is

510   expressed in the epithelial cells of the mantle and acts as a negative regulator in calcification in

511   the mollusc *Pinctada fucata*. Zoological Science 22:311-315. DOI: 10.2108/zsj.22.311.

512   Miyamoto H, Endo H, Hashimoto N, Limura K, Isowa Y, Kinoshita S, Kotaki T, Masaoka T,

513   Miki T, Nakayama S, Nogawa C, Notazawa A, Ohmori F, Sarashina I, Suzuki M, Takagi R,

514   Tkahashi J, Takeuchi T, Yokoo N, Satoh N, Toyohara H, Miyashita T, Wada H, Samata T, Endo

515   K, Nagasawa H, Watabe S. 2013. The diversity of shell matrix proteins: genome-wide

516   investigation of the pearl oyster, *Pinctada fucata*. *Zoological Science* 30:801-816. DOI:

517   10.2108/zsj.30.801.

518   Montagnani C, Marie B, Marin F, Belliard C, Riquet F, Tayalé A, Zanella-Cleon I, Fleury E,

519   Gueguen Y, Piquemal D, Cochennec-Laureau N. 2011. Pmarg-Pearlin is a matrix protein

520   involved in nacre framework formation in the pearl oyster *Pinctada margaritifera*.

521   *ChemBioChem.* 2011; 12:2033-2043. DOI: 10.1002/cbic.201100216.

522 Mutvei H. 1980. The nacreous layer in molluscan shells. In: The mechanisms of

523 biomineralization in animals and plants. Proceedings of the third international biomineralization

524 symposium, pp. 49-56. Tokai University Press, Tokyo.

525 Norizuki M, Samata T. 2008. Distribution and function of the nacrein-related proteins inferred

526 from structural analysis. *Marine Biotechnology* 10:234-241. DOI: 10.1007/s10126-007-9061-x.

527 Nudelman F, Chen HH, Goldberg SW, Addadi L. 2007. Memorial lecture lessons from

528 biomineralization: comparing the growth strategies of mollusc shell prismatic layers in *Atrina*

529 *rigida*. *Faraday Discussions* 136:9-25.

530 Ozensoy GO, Capasso C, Supuran CT. 2016. A magnificent enzyme superfamily: carbonic

531 anhydrase, their purification and characterization. *Journal of Enzyme Inhibition and Medicinal*

532 *Chemistry* 31:5, 689-694. DOI: 10.3109/14756366.2015.1059333.

533 Pastorek J, Pastorekova S, Callebaut I, Mornon JP, Zelnik V, Opavsky R, Zatovicova M, Liao S,

534 Portetelle D, Stanbridge EJ. 1994. Cloning and characterization of MN, a human tumor-associated

535 a protein with a domain homologous to carbonic anhydrase and a putative helix-loop-helix DNA

536 binding segment. Oncogene, 9(10): 2877-2888.

537 Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal

538 peptides from transmembrane regions. *Nature Methods* 8:785-786. DOI: 10.1038/nmeth.1701.

539 Raz S., Weiner S., Addadi L. 2000. Formation of high magnesian calcites via an amorphous

540 precursor phase. Possible biological Implications. *Advanced Materials* 12, 38–42. DOI:

541 10.1002/(SICI)1521-4095(200001)12:1<38::AID-ADMA38>3.0.CO;2-I.

542 Samata T, Hayashi N, Kono M, Hasegawa K, Horita C, Akera S. 1999. A new matrix protein

543 family related to the nacreous layer formation of *Pinctada fucata*. *FEBS Letters* 462:225-229.

544   Saraswathy N, Ramalingam P. Phosphoproteomics 2011. In: Concepts and Techniques in

545   Genomics and Proteomics 203-211.

546   Song X, Wang X, Li L, Zhang G. 2014. Identification two novel nacrein-like proteins involved

547   in the shell formation of the Pacific oyster *Crassostrea gigas*. *Molecular and Biological Reports*

548   41:4273-4278.

549   Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT,

550   Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U,

551   Brunak S, Wandall HH, Levery SB, Clausen H. 2013. Precision mapping of the human O-

552   GalNAc glycoproteome through SimpleCell technology. *EMBO Journal* 32**:**1478-1488. DOI:

553   10.1038/emboj.2013.79.

554   Su J, Zhu F, Zhang G, Wang H, Xie L, Zhang R. 2016. Transformation of amorphous calcium

555   carbonate nanoparticles into aragonite controlled by ACCBP. *Cryst Eng Comm* 18:2125–2134.

556   DOI: 10.1039/C5CE02288F.

557   Sudo S, Fujikawa T, Nagakura T, Ohkubo T, Sakaguchi K, Tanaka M, Nakasima K, Takahashi

558   T. 1997. Structure of mollusc shell framework proteins. *Nature* 387:563-564. DOI:

559   10.1038/42391.

560   Suzuki M, Iwashima A, Tsutsui N, Ohira T, Kogure T, Nagasawa H. 2011. Identification and

561   characterization of a calcium carbonate-binding protein, blue mussel shell protein (BMSP), from

562   the nacreous layer. *ChemBioChem* 12:2478-2487. DOI: 10.1002/cbic.201100317.

563   Takakura D, Norizuki M, Ishikawa F, Samata T. 2008. Isolation and characterization of the N-

564   linked oligosaccharides in nacrein from *Pinctada fucata*. *Marine Biotechnology* 10:290-296.

565   DOI: 10.1007/s10126-007-9063-8.

566 Takeuchi T, Sarashina I, Iijima M, Endo K. 2008. In vitro regulation of CaCO₃ crystal

567 polymorphism by the highly acidic molluscan shell protein Aspein. *FEBS Letters* 582:591-596.

568 DOI: 10.1016/j.febslet.2008.01.026.

569 Thornton DJ, Carlstedt I, Sheehan JK. 1994. Identification of glycoproteins on nitrocellulose

570 membranes and gels. IN Jm, W. (Ed.) *Method Molecular Biology.* Totowa, NJ: Human Press.

571 Ulmasov B, Waheed A, Shah GN, Grubb JH, Sly WS, Tu C, Silverman DN. 2000. Purification

572 and kinetic analysis of recombinant CAXII, a membrane carbonic anhydrase overexpressed in

573 certain cancers. *Proceedings of the National Academy of Sciences* 97(26): 14212-14217.

574 DOI: 10.1073/pnas.97.26.14212

575 Vulo D, Nishimori I, Scozzafava A, Supuran CT. 2008. Carbonic anhydrase activators:

576 activation of the human cytosolic isozyme III and membrane-associated isoform IV with amino

577 acids and amines. *Bioorganic & Medicinal Chemistry Letters* 18:4303-4307. DOI:

578 10.1016/j.bmcl.2008.06.075.

579 Wang D, Wallace AF, De Yoreo JJ, Dove PM. 2009. Carboxylated molecules regulate

580 magnesium content of amorphous calcium carbonates during calcification. 106(51):21511-6.

581 DOI: 10.1073/pnas.0906741106.

582 Wilbur KM, Anderson NG. 1948. Electrometric and colorimetric determination of carbonic

583 anhydrase. *Journal of Biological Chemistry* 176:147-154.

584 Weiss IM, Kaufmann S, Mann K, Fritz M. 2000. Purification and Characterization of Perlucin

585 and Perlustrin, Two New Proteins from the Shell of the Mollusc *Haliotis laevigata. Biochemical*

586 *and Biophysical Research Communications* 267**:**17-21. DOI: 10.1006/bbrc.1999.1907.

587   Xie J, Liang J, Sun J, Gao J, Zhang S, Liu Y, Xie L, Zhang R. 2016. Influence of the extrapallial

588   fluid of *Pinctada fucata* on the crystallization of calcium carbonate and shell biomineralization.

589   *Crystal Growth & Design* 16:672-680. DOI: 10.1021/acs.cgd.5b01203.

590   Yarra T, Gharbi K, Blaxter M, Peck LS, Clark MS. 2016. Characterization of the mantle

591   transcriptome in bivalves: *Pecten maximus*, *Mytilus edulis* and *Crassostrea gigas*. *Marine*

592   *Genomics* 27:9-15. DOI: 10.1016/j.margen.2016.04.003.

593   Zhang C, Zhang R. 2006. Matrix proteins in the outer shells of molluscs. *Marine Biotechnology*

594   8:572-586. DOI: 10.1007/s10126-005-6029-6.

595   Zigovecki GZ, Posilovic H, Bermanee V. 2009. Identification of biogenetic calcite and aragonite

596   using SEM. *Geologia Croatia* 62(3): 201-206. DOI: 10.4154/gc.2009.14.

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

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*.

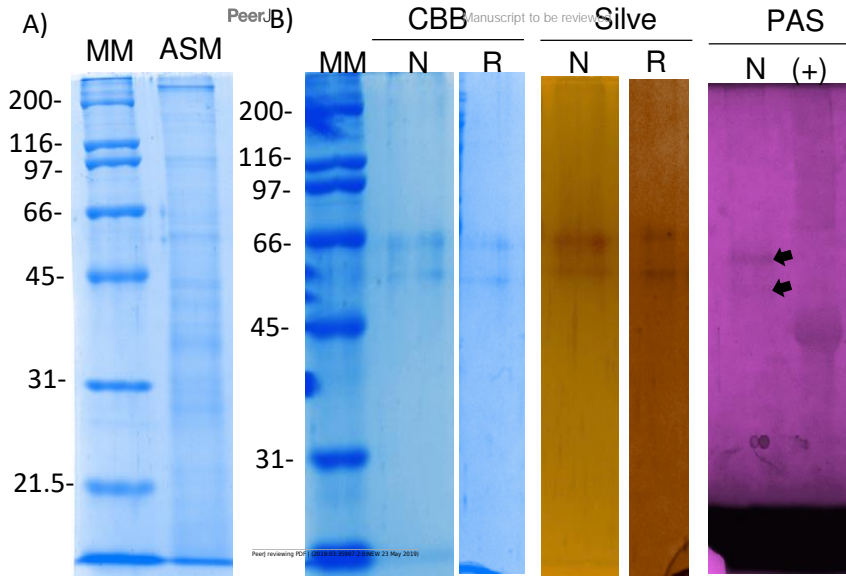
```
   1  tatgtggagaatgacgacgcttcttcacttgactgctctgcttgttcttgattccattatgt
   1  M  W  R  M  T  T  L  L  H  L  T  A  L  L  V  L  I  P  L  C
  62  cattgtgcctccatgcacaggcatgaccattatatggacatggatcaaacctaccgtgat
  21  H  C  A  S  M  H  R  H  D  H  Y  M  D  M  D  Q  T  Y  R  D
 121  catgtgggaaactgtcattattcagggggaagtagctgtgacgccgggttagctacaat
  41  H  V  G  N  C  H  Y  S  G  G  S  S  C  V  D  C  A  G  F  S  Y  N  D
 181  agggaacaaaatgaggaacaatgccacgggccgtatgactggcacactatatctagttg
  61  R  E  Q  N  E  E  Q  C  H  G  P  Y  D  W  H  T  I  S  S  C
 241  tttaaggcatgtggaagtaaagagagacaatcaccaatcaacatttggtcacatagagca
  81  F  K  A  C  G  S  K  E  R  Q  S  P  I  N  I  W  S  H  R  A
 301  cttttccgaaaactgccaagactgaaattcaagccacatatgaaatcattggatacgaaa
 101  L  F  R  K  L  P  R  L  K  F  K  P  H  M  K  S  L  D  T  K
 361  gtgtcaaatcaccaaaatcatgccctgaattcgattctgaggacgaaaaacttcatgtt
 121  V  S  N  H  A  N  H  A  P  E  F  D  S  E  D  E  K  L  H  V
 421  aaactgaagaaacttgttgatgacattataaattccgcaatctccatattcacattggc
 141  K  L  K  K  L  V  D  G  H  Y  K  F  R  A  I  L  A  I  L  I  G
 481  aaagtagacgaaagggctccgaacacagcgttgacagacatttcacctatgaggactt
 161  K  S  R  R  K  G  S  A  H  S  V  D  R  H  F  T  P  M  E  A
 541  catttagtgttccatcatgatgagaaaaagaaatcaaacctcctaggattccgttagga
 181  A  L  V  F  H  H  D  E  K  K  E  I  K  P  P  R  I  P  L  G
 601  agaaatttcagtggtattaatgaatttgttgtcgttggggttttttctagaggatggtgat
 201  R  N  F  S  G  I  N  E  F  V  V  V  G  V  F  L  E  D  G  D
 661  gaaggatacggtgatgaaccggacgactatgaatgtaagcgcatattaaagggtcattac
 221  E  G  Y  G  D  E  P  D  D  Y  E  C  K  R  I  L  K  G  H  Y
 721  gatcattgcgacaacaatggagacaacggttacaactgtgataacggcaacaatgaaac
 241  D  H  C  D  N  N  G  D  N  G  Y  N  C  D  N  G  N  N  E
 781  aacggaaacaatggtaatggtaacaacggctataacggtaacaacggttataacggtaat
 261  N  G  N  N  G  N  G  N  G  Y  N  G  N  N  G  N  N  G  Y  N  G  N
 841  aacggtgacaatggcaacagtgggaaacaatggtaatggtaacaacggttataacggtgac
 281  N  G  D  N  G  N  S  G  N  N  G  N  N  G  Y  N  G  D
 901  aatggcaacaacggaaacaatggtaatggtaacaacggtaataacggtaataacggtggc
 301  N  G  N  S  G  N  N  G  N  N  G  N  N  G  N  N  G  G
 961  aacggaaacaacaatggtaatggtaatggtaacaacggttataacggaataacggtggt
 321  N  G  N  N  R  N  N  G  N  G  N  Y  N  G  N  N  G  G
1021  aatggcaacaacggaaacaatggtaatggtaacaacggaaataatggtaatggtaacaac
 341  N  G  N  N  G  N  N  G  N  G  N  N  G  N  N  G  N  N
1081  ggaaataacggtggcaatggcaacaacggaaacaatggtaatggtaacaacggaaataat
 361  G  N  N  G  G  N  G  N  N  G  N  N  G  N  N  G  N  N
1141  ggtaatggtaacaacggaaataacggtggcaatggcaacaatggtaatggtaacaatgga
 381  G  N  G  N  N  G  N  N  G  G  N  G  N  N  G  N  N  G
1201  ggtaatggtaacggtgactacggtagtaatggtaacaatggtggaaacgggaacaatggt
 401  G  N  G  N  G  D  Y  G  S  N  G  N  N  G  G  N  G  N  G
1261  tcgaacggtgataacggtaatggcgacaatggttataacggtgataatggtaacaatgac
 421  S  N  G  D  N  G  N  G  D  N  G  Y  N  G  D  N  G  N  S  D
1321  gggcgactcagacgtttgggacttggaaaatgtccgccgcattcataccgagcgatcac
 441  G  R  L  R  R  W  D  L  E  N  V  R  R  I  H  T  E  R  Y  H
1381  atcagcggaagatgtattgtcaaaaaagcaaaacgcctcaacgaggattctcgaatgcgca
 461  I  S  G  R  C  I  V  K  K  A  K  R  L  S  R  I  L  E  C  A
1441  tatagacacaaaaagtcagagaattcaaaaggaatggagaacaacaaaaagtcttgatgtt
 481  Y  R  H  K  K  V  R  E  F  K  R  N  G  E  H  K  S  L  D  V
1501  gaaattacacccgaaaattggttctaccgccaataaaatacgagcaattatacctatgaa
 501  E  I  T  P  E  M  V  L  P  P  I  K  Y  R  Q  Y  Y  T  Y  E
1561  ggatctttgacaaccccttccttctaccagagagagcgtccgttgggttgtagaaaaatgccac
 521  G  S  L  T  T  P  P  C  T  E  S  V  R  W  V  V  E  K  C  H
1621  gtgcaagtatccgacaagaggttgtgatgcattgcggaggataatatgatgatggt
 541  V  Q  V  S  R  R  V  L  D  A  L  R  K  V  E  G  Y  D  D  G
1681  accacgtttgacaagtatggaaacaagacgtcccacacagagaaacataaaacctgtac
 561  T  T  F  E  Q  V  K  Q  D  V  P  H  R  E  T  -
1741  ctgtgtacaaaaacttatatgaga
```

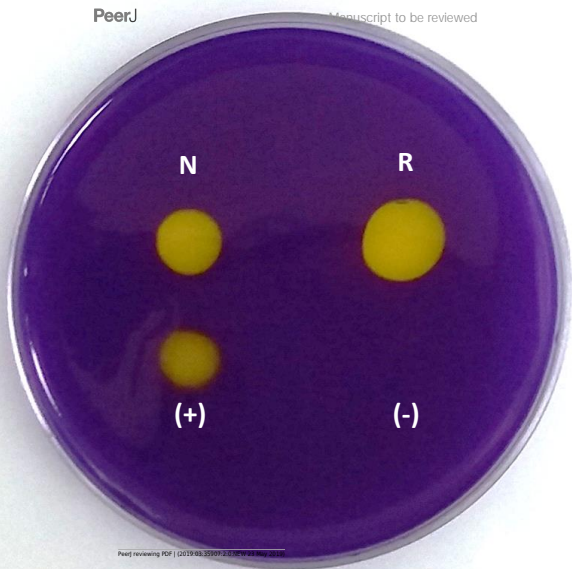
Amino acid sequence alignment of N66 and related shell matrix proteins.

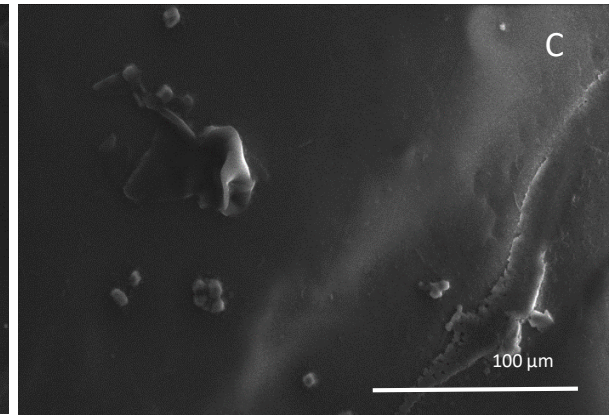
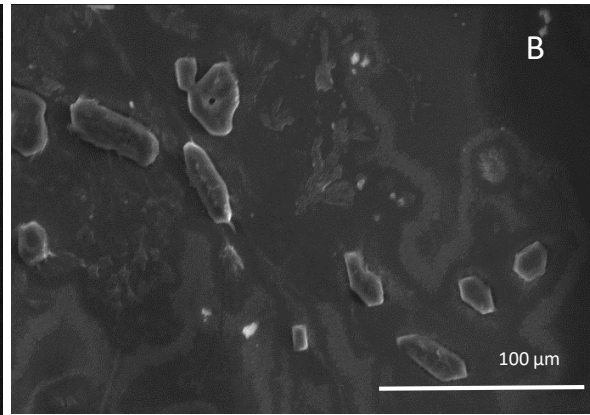
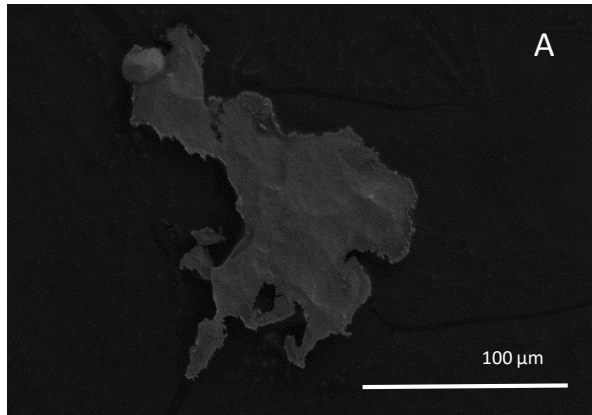
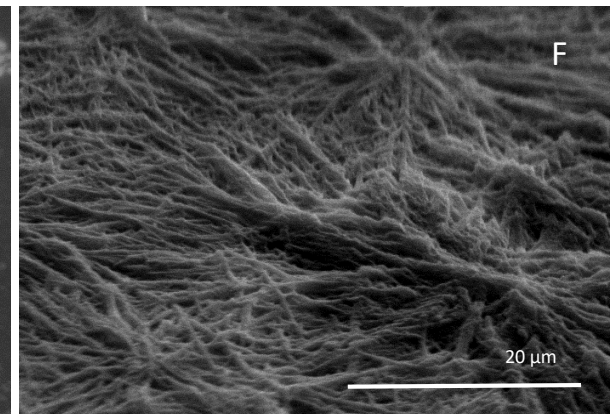
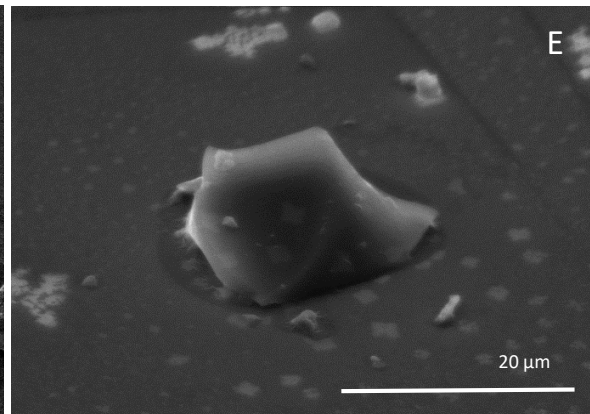
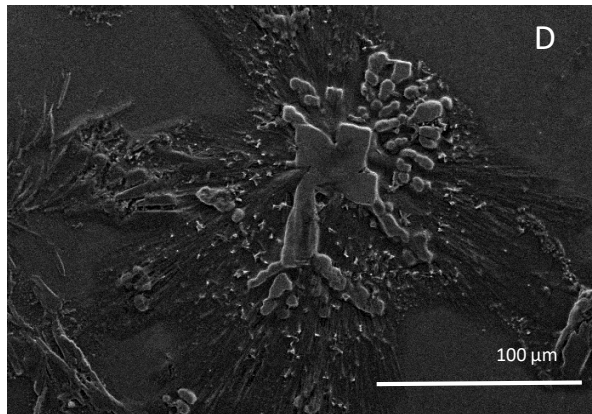
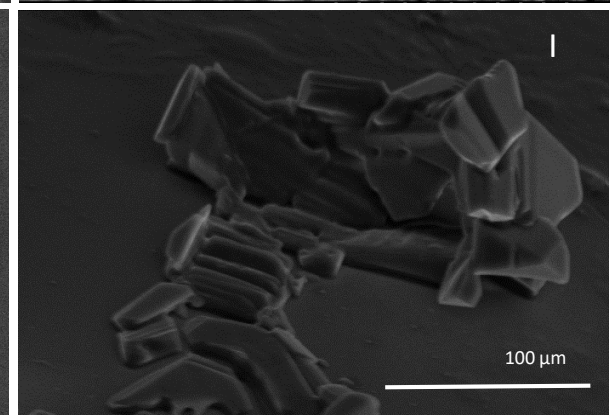
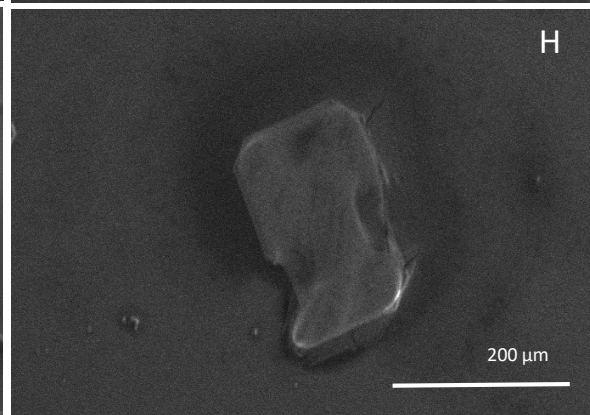
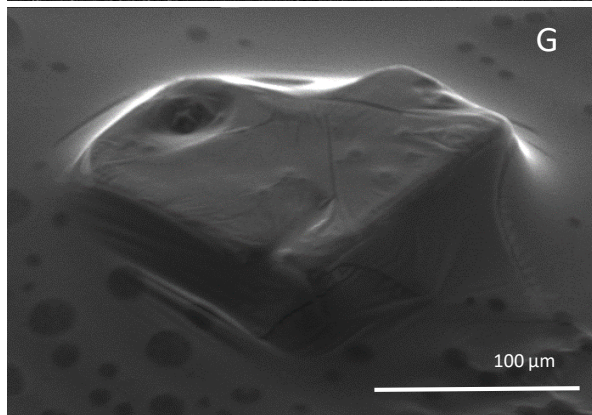
Accession numbers: *P. mazatlanica* N66 (this study, GenBank: MH473230), *P. maxima* N66 (GenBank: BAA90540.1), *P. fucata* Nacrein (GenBank: BAA119401.1), *P. margaritifera* nacrein B2 (GenBank: ADY69618.1), B3 (GenBank: AEC03971.1), B4 (GenBank: AEC03972.1) and C5 (GenBank: AEC03973.1), and *P. maxima* 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.

```
      N66 P.mazatlanica   MIRWRTMTLLLHLTALLVLIFLPHCASMHRHDHYMDMDQTYRDRWQKHYSGGSSCCDAGFSYN   60
      N66 P.maxima        MIRWRTMTLLLHLTSLLVLIFLPHCASMHRHDHYMDMDQTYPNGLGYCEPSGESSCKAGFSYN   60
      N45 P.maxima        MIRWRTMTLLLHLTSLLVLIFLPHCASMHRHDHYMDMDQTYPNGLGYCEPSGESSCKAGFSYN   60
      NacreinB4           MIRWRTMTLLLHLTSLLVLIFLPHCASMHRHDHYMDMDQTYRNRWQKCHSGGNSCCDAGFSYN   60
      NacreinB3           MIRWRTMTLLLHLTSLLVLIFLPHCASMHRHDHYMDMDQTYRNRWQKCHS-GGNSCCKAGFSYN   59
      NacreinB2           MIRWRTMTLLLHLTSLLVLIFLPHCASMHRHDHYMDMDQTYRNRWQKCHSGGSSCCKAGFSYN   60
      NacreinC5           MIRWRTMTLLLHLTSLLVLIFLPHCASMHRHDHYMDMDQTYRDGSSCKAGFSYQ            60
                          *****:*** ** :*:***:*****:******:*******   *     : .*.****:**:

      N66 P.mazatlanica   REQNEQCHGPYDWHTISBCFKACGSRKERQSPINIWSHRALVFR-KLPRLRFKPHMKSLDT   119
      N66 P.maxima        REQNEQCHGPYDWHTISBCFKACGSRKERQSPINIWSHRALVFR-RLPRLRFKPHMKSLDT   119
      N45 P.maxima        RD----ICQPYDWHTISBCFKACGSRKERQSPINIWSHRALVFR-RLPRLRFKPHMKSLDT   116
      NacreinB4           REQSEQCHGPYDWHTISBCFKACGSRKRQSPINIWSHRALVFRHRHRLPRLRFKPHMKSLDT   120
      NacreinB3           REQSEQCHGPYDWHTISBCFKACGSRKRQSPINIWSHRALVFR-ILPRLRFKPHMKSLDT   117
      NacreinB2           REQSEQCHGPYDWHTISBCFKACGSRKRQSPINIWSHRALVFR-TLPRLRFKPHMKSLDT   117
      NacreinC5           R----DICQPYDWHTISBCFKACASKKRQSPINIWSHRAVYFRVPDLPRLRFKPHMKSLDT   116
                          *.  : * **.****:***:**.*:::::****:***.* :**       ***:***:***

      N66 P.mazatlanica   DVTHNQNRAPEFEPEDCGRELKLVKVKLKNLUGHYKFRNLHIHGKSRRRKGSEHSVDRHFTPPM   178
      N66 P.maxima        DVTHNQNRAPEFEPEDCGRELKLVKVKLKNLUGHYKFRNLHIHNGKSRRRKGSEHSVNRHFTPPM   176
      N45 P.maxima        DVTHNQNRAPEFEPEDCGRELKLVKVKLKNLUGHYKFRNLHIHNGKSRRRKGSEHSVNRHFTPPM   176
      NacreinB4           EVSHNQNRAPEFDSED-EKLHVKLKNLUGHYKFRNLHITGKSRRRKGSEHSVDRHFTPPM   179
      NacreinB3           EVSHNQNRAPEFDSED-EKLHVKLKNLUGHYKFRNLHITGKSRRRKGSEHSVDRHFTPPM   177
      NacreinB2           EVSHNQNRAPEFDSED-EKLHVKLKNLUGHYKFRNLHITGKSRRRKGSEHSVDRHFTPPM   177
      NacreinC5           EVSHNQNHAPEFDSED-EKPHVKLKNLUGHYKFRNLHITGKSRRRKGSRHSVDRHFTPPM   175
                          :*:****:***** :*  :*  :*:********:**.****:** :* ***:***.:*.***.*

      N66 P.mazatlanica   EAHLVFHHDDKKEIKPPRKLGRNFSGINEKFVVVGFVELEVGDEGYGDEPDDECKRILRG   238
      N66 P.maxima        EAHLVFHHDDKKEIKPPRKLGGVYAGRNEKFVVVGFVELEVGDEGYGDEPDDECKRILRG   238
      N45 P.maxima        EAHLVFHHDDKKEIKPPVKLGGVYAGRNEKFVVVGFVELEVGDEGYGDEPDDECKRILRG   238
      NacreinB4           EAHLMFRPDDEKKEIKPSRIWLGGKNVNGINEKFVVVGFVELEVGDEGYGDEPDDECKRILRG   237
      NacreinB3           EAHLMFRPDDEKKEIKPSRIWLGGKNVNGINEKFVVVGFVELEVGDEGYGDEPDDECKRILRG   237
      NacreinB2           EAHLMFRPDDEKKEIKPSRIWLGGKNVNGINEKFVVVGFVELEVGDEGYGDEPDDECKRILRG   235
      NacreinC5           
                          ****:.*  :*:* * :  .    . . **:**:*** * ***********************

      N66 P.mazatlanica   HYDHCDNNGCGNYCDNGNNGNNGNNGN------------GNNGYNGNNGYNGGNNGN   284
      N66 P.maxima        ---HCENNGDNGNNCDNGNNGNNGNNGN---------NNGNNGNNGNNGYNGGNN---   279
      N45 P.maxima        HYDHCDNNGCGNYCDNG--------------------------------------   251
      NacreinB4           HYDHCDNNGCGNYCDNGNNGNNGNNGN------------GNNGYNGNNGYNGGNNGN   299
      NacreinB3           HYDHCDNNGCGNYCDNGNNGNNGNNGN------------GNNGYNGNNGYNGGNNGN   283
      NacreinB2           HYDHCDNNGCGNYCDNG-------------------GNNGYNGNNGYNGGNNGN   275
      NacreinC5           HYDYCDNNGGN------------------------------------   249
                          :.:*******

      N66 P.mazatlanica   GNSGNNGNGNNGYNG---------------------DNGNSGNNGNGNNGYNGGNNGN   321
      N66 P.maxima        GNNG------------------HGNGNNGNGYNGNNGNGYNGGNNGN   312
      N45 P.maxima        ------------------------------------------   251
      NacreinB4           GNSGNNGNGNNGYNGGNNGDNGNGYNGNNGYNGGNNGNRNNGNGNGYNGGNNGN   359
      NacreinB3           GNSGNNGN------------GNNGYNGNNGYNGGNNGNGYNGGNNGN   325
      NacreinB2           ----------------------------GNNGYNGNGNNGYNGGNNGN   298
      NacreinC5           ------------------------------------------   249

      N66 P.mazatlanica   GNNRNNGNGNNGYNGNNGDNGNNGNNGNGNNGNNGNNGNNG--------NNGG------   365
      N66 P.maxima        ------GNDNN---GNN-------GGNGNNGNNGNGNNGNGNNGNGNNGGNGNN----   351
      N45 P.maxima        ------------------------------------------   251
      NacreinB4           GNNRNNGNGNNGYNGNNGDDDGNGNNGNGNGNNGNGNNGNNGNNGNGNGNGGNGNNGNNG   405
      NacreinB3           GNNRNNGNGNNGYNGNNGDNGNNGNNGNNGNGNNGNNGNNGNGNNGNNGGNGNN   389
      NacreinB2           GNNRNNGNGNNGYNGNNGDNGNNGNNGNNGNGNN--------NGNNGNNGGNGNN   350
      NacreinC5           ------------------------------------------   249

      N66 P.mazatlanica   NGNNGNNGNGNNGNNGNNGGGNNGGNGNGNNGNGGNGNGNSNGNGDYGSNGNNGGNGNNGNNGGN   425
      N66 P.maxima        ----GNNGNNGNNGNGNNGNGNNGNNGNNGNNGNNGNNGNGNNGGNNGGGNGNNGGNGGN   407
      N45 P.maxima        ------------------------------------------   251
      NacreinB4           NGNNGNNGNGNNGNNGNNGNNGNGNGNNGNNGGGNGNNGYGSNGNNGNGNNGGNGNNGGN   465
      NacreinB3           NGNNGNNGNGNNGNNGNNGNNGNGNGNNGNNGGGNGNNGNGNNGNNGGNGNNGGN   449
      NacreinB2           NGNNGNNGNGNNGNNGNNGNNGNGNGNNGNNGGGNGNNGGNNGNNGGN   410
      NacreinC5           ----------------GNNGNNGNGGNGNNGNNGNHGGNGNN   294

      N66 P.mazatlanica   GNGGNGYNGDNGNSDGRIRRWDLLENVRRHTTKHYHHSGRCIYKKAKRLSRILECAYRHKK   485
      N66 P.maxima        GNGGNGYNGDNGNSDGRIRRWDLLENVRRHTTKHYHFSGRCIYKKAKRLSRILECAYRHKK   467
      N45 P.maxima        -NGGNGYNGDNGNSDGRPRRWDLLENVRRHTTEHYHFSGRCIYKKAKRLSRILECAYRHKK   310
      NacreinB4           GNGGNGYNGDNGNSDGPRRWDLLENVRRHTTEHYHFSGRCIYKKAKRLSRILECAYRHKK   525
      NacreinB3           GNGGNGYNGDNGNSDGRIRRWDLLENVRRHTTEHYHFSGRCIYKKAKRLSRILECAYRHKK   509
      NacreinB2           GNGGNGYNGDNGNSDGRIRRWDLLENVRRHTTEHYHFSGRCIYKKAKRLSRILECAYRHKK   470
      NacreinC5           GNGGNGYNGDNGNSDGRIRRWDLLENVRRHTTEHYHFSGRCIYKKAKRLSRILECAYRHKK   354
                          .******:***** :*  **:****:*** ***** .::   .*:**. *:*:*** **:**

      N66 P.mazatlanica   VRKFRHHGEHKSLDVEITPEMVLPPHYIKRQYYTYEGSLSTPPPCSVHWVVEKCHVQVSR   545
      N66 P.maxima        VRKFRHHGEHKSLDVEITPEMVLPPHYIKRQYYTYEGSLTTPPPCSVHWVVEKCHVQVSR   527
      N45 P.maxima        VRKFRHHGEHKSLDVEITPEMVLPPHKYKRQYYTYEGSLTTPPPCNTVLWVVEKCHVQVSR   370
      NacreinB4           VRKFRHHGEHKSLDVEITPEMVLPPHYIKRQYYTYEGSLTTPPPCNTVLWVVEKCHVQVSR   585
      NacreinB3           VRKFRHHGEHKSLDVEITPEMVLPPHYIKRQYYTYEGSLTTPPPCNTVLWVVEKCHVQVSR   569
      NacreinB2           VRKFRHHGEHKSLDVEITPEMVLPPTKYKRQYYTYEGSLTTPPPCNTVLWVVEKCHVQVSR   530
      NacreinC5           VRKFRHHGEHKSLDVEITPEMVLPPTKYKRQYYTYEGSLTTPPPCNTVLWVVEKCHVQVSR   414
                          ****:****:*:*:*,****:******   *****::.::*:***** .: .:.****:**.

      N66 P.mazatlanica   -----------VPHRET----------   576
      N66 P.maxima        RVLDALRRMREGYEDGTTLSKFGTRRPTQRNKPLRVVYKNFI   568
      N45 P.maxima        GCLNMPCTDKDM---KMCPR------------   387
      NacreinB4           RVLDALRRMREGYEDGTTLSKFGTRRPTQRNKIPVPVVYKNFI   626
      NacreinB3           RVLDALRRMREGYEDGTTLSKFGTRRPTQRNKIPVPVVYKNFI   610
      NacreinB2           RVLDALRRMREGYEDGTTLSKFGTRRPTQRNIKPVTVYKNFI   571
      NacreinC5           RVLDALRRMREGYEDGTTLSKFGTRRPTQRNIKPVTVYKNFI   455
                          *  .:  . . . :
```

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.

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.

Primer sequences and expected amplicon size for N66 amplification.

Primers	Primer sequence (5'-3')	Amplicon size (pb)
N66_1F	ATGTGGAGAATGACGACGCTT	425
N66_425R	TGTGAATATGGAGATTGCGG	
N66_1F	ATGTGGAGAATGACGACGCTT	680
N66_680R	CCTTTAATATGCGCTTACATTC	
N66_680Fw	GAATGTAAGCGCATATTAAAGG	962
N66_1642Rv	GGGACGTCTGGTTCCATAC	
N66_1F	ATGTGGAGAATGACGACGCTT	1766
N66_1766R	CTCATATAAAGTTTTTGTACACAGG	
N66_HindF	GTG**AAGCTT**TGCCTCCATGCACAGGCATG	1642
N66_HindR	CAC**AAGCTT**GGGACGTCTTGTTTCCATAC	

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	

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

2 Protein concentration was estimated by Lowry method. The experiments were conducted three
3 times. CA activity was estimated by Wilbur and Anderson method.

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