# Structural Characterization of the Major Extrapallial Fluid Protein of the Mollusc *Mytilus edulis*: Implications for Function†

Yan Yin,‡,§ Jing Huang,‡,§ Michael L. Paine,|| Vernon N. Reinhold,§ and N. Dennis Chasteen*,§

*Department of Chemistry, University of New Hampshire, Durham, New Hampshire 03824, and School of Dentistry, University of Southern California, Los Angeles, California 90033*

*Received March 25, 2005; Revised Manuscript Received June 9, 2005*

ABSTRACT: The major protein component of the extrapallial fluid of the mollusc *Mytilus edulis* has been previously isolated and partially characterized. It was postulated to play a role in shell mineralization because of its intriguing property of $Ca^{2+}$-binding-induced self-assembling. However, it also binds other divalent ions, including $Cd^{2+}$, $Cu^{2+}$, $Mn^{2+}$, and $Mg^{2+}$. Herein is the initial report on the characterization of the primary structure of the extrapallial (EP) protein by RT-PCR and cDNA sequencing methods and by de novo peptide sequencing with mass spectrometry. The EP protein is comprised of 213 amino acids postcleavage of a signal peptide of 23 amino acids. The protein is rich in His, Glu, and Asp residues. The site of N-glycosylation, "NHTE", at amino acid positions 54−57 and the intramolecular disulfide bond between Cys 139 and Cys 171 of the protein have been characterized also. Sequence comparisons reveal that the EP protein possesses little homology to any presently known matrix proteins previously isolated from mollusc shells but rather it highly resembles a heavy metal binding protein and a histidine-rich glycoprotein, both from the hemolymph of *M. edulis*. The predicted domain profile and amino acid composition suggest that its N-terminus may be involved in calcium binding. The abundance of histidine residues of the protein may account for its heavy metal binding properties. Thus, the EP protein perhaps has multiple functions, serving as a $Ca^{2+}$-transport protein, a shell matrix protein, and a heavy metal detoxification protein.

Organisms are known to form more than 40 different minerals in over 30 phyla, from the nanoscale of magnetic compasses in magnetotactic bacteria (*1*, *2*) to the macroscale structures of seashells, bone, teeth, ivory, and corals, which fulfill a diversity of biological functions. Biomineralization is the study of the formation, structure, and properties of inorganic solids deposited in biological systems (*3*). It involves selective uptake of elements from surroundings and deposition of inorganic minerals to construct functional structures under strict biological control. The nucleation, growth, and cessation of mineralization typically take place in a liquid medium and are regulated by its content, mostly composed of matrix proteins and polysaccharides (*4*−*9*).

Molluscan shells are fascinating examples of functionalized inorganic/organic materials. They are composed of either one or both of the polymorphs of calcium carbonate, calcite and aragonite. The crystals are organized into different microarchitectural arrangement under the control of extracellular organic matrix frameworks. The formation of diverse molluscan shells is a favorable system for studying biomineralization, especially for investigating the role of organic matrix components (*10*). In the shell of the mollusc, *Mytilus edulis*, calcium carbonate is the dominant component, i.e., 95−99.9 wt %. The shell consists of two continuous layers that are ultrastructurally unique, an outer prismatic layer of calcite residing on top of the inner nacreous layer of aragonite along the long axis of the shell (*11*). Although the residual organic matrix components of the shell remaining after decalcification are minute, they are indispensable for providing nanoscale control over the shell fabrication and bring the shell tremendous strength as compared to the mineral alone.

A series of compartments are involved in the shell formation in the mollusc, the most important of which are the inner shell surface, the extrapallial cavity, and the outer mantle epithelium (*12*). The extrapallial (EP)[1] fluid is secreted by the outer mantle epithelium cells and fills the

† This work was supported by NIGMS Grants R01 GM20194 (N.D.C.) and R01 GM54045 (V.N.R.), NIDCR Grant R01 DE13404 (M.L.P.), and NCRR Grant RR018531 (V.N.R.) from the National Institutes of Health.

* To whom correspondence should be addressed. Phone: (603) 862-2520. Fax: (603) 862-4278. E-mail: ndc@cisunix.unh.edu.

‡ These individuals contributed equally to this work.

§ University of New Hampshire.

|| University of Southern California.

[1] Abbreviations: 2D GE, two-dimensional gel electrophoresis; ACN, acetonitrile; C1q, complement 1q; CAF, chemical-assisted fragmentation; cDNA, complementary DNA; CHAPS, 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate; CHCA, α-cyano-4-hydroxycinnamic acid; CID, collision-induced dissociation; DHB, 2,5-dihydroxybenzoic acid; DTT, dithiothreitol; EP, extrapallial; EPR, electron paramagnetic resonance; ESI, electrospray ionization; HIP, heavy metal binding protein; HRG, histidine-rich glycoprotein; MALDI, matrix-assisted laser desorption ionization; PCR, polymerase chain reaction; PSD, postsource decay; PMF, peptide mass fingerprint; QIT, quadrupole ion trap; QTOF, quadrupole time of flight; RACE, rapid amplification of cDNA ends; RT, reverse transcription; SOD, superoxide dismutase; SDS, sodium dodecyl sulfate; SMART, simple modular architecture research tool; TFA, trifluoroacetic acid; TOF, time of flight.

10.1021/bi0505565 CCC: $30.25 
Published on Web 07/13/2005

extrapallial cavity, the space between the shell and the outermost visceral organ, the mantle. It is the medium from which the nacre layer growth occurs and the prismatic layer of the shell thickens (*13*−*15*). The EP fluid contains proteins, glycoproteins, carbohydrates, and amino acids and is also believed to be supersaturated with respect to the shell minerals. The composition of its ion content is found to differ from that of the animal blood and surrounding seawater (16−18). Both the anatomical location and the biomolecular content of the EP fluid imply that it plays an important role in shell formation in vivo (*19*, *20*). Despite this fact, surprisingly little study has been carried out on the protein components of the EP fluid, while the organic matrix from shells, in particular proteins, has received extensive investigation (*10*, *21*−*28*).

Only recently has the major protein component of the EP fluid, hereafter termed the EP protein, been isolated and partially characterized (*29*). It is an acidic glycoprotein with a reported isoelectric point ranging from 4.08 to 4.67 for various isoforms. The protein is a homodimer of 28.3 kDa monomers, composed of 14.3 wt % carbohydrate and is rich in His, Asx, and Glx residues. The most intriguing property of the EP protein in vitro is that it binds $Ca^{2+}$, which induces the aggregation of monomers to form a series of multimeric species of increasing molecular mass. The binding causes a major reduction in $\beta$-sheet accompanied with an increase in $\alpha$-helical content and is reversible (*29*). All of these findings suggest that the EP protein is possibly a precursor or building block to the soluble organic matrix of the shell. However, the direct role of the EP protein in mollusc shell formation still remains undefined.

One of the most crucial questions, yet poorly studied, is whether the EP protein is transported from the EP fluid and becomes part of the shell matrix framework, possibly in some modified form, where it participates in the process of shell formation together with other organic matrix components. Additionally, the elucidation of its calcium binding site(s) demands the determination of the three-dimensional structure by single crystal X-ray diffraction. The primary structure of the EP protein is a prerequisite for providing insights into both areas. However, in the initial study only the amino acid composition and a partial N-terminal sequence of 20 amino acids were determined (*29*). In this paper, we report the cloning and characterization of the complete primary structure of the EP protein. The mature EP protein is comprised of 213 amino acids, whose sequence is deduced from its full-length complementary DNA (cDNA) and confirmed by mass spectrometry. The site of N-glycosylation and the intramolecular disulfide of the EP protein have been characterized also. Sequence comparisons reveal that it greatly resembles a heavy metal binding protein and a histidine-rich glycoprotein from the hemolymph of *M. edulis* rather than any known matrix proteins previously isolated from mollusc shells. The predicted domain profile and amino acid composition suggest that its N-terminus is possibly involved in calcium binding. The abundant His residues of the protein may account for its heavy metal binding properties. The EP protein is thus proposed to transport $Ca^{2+}$ between the EP fluid and the plasma for mineralization, possibly functioning as a matrix protein and serving as a detoxification protein through heavy metal binding.

## EXPERIMENTAL PROCEDURES

*Materials.* Blue mussels, *M. edulis*, from the Great Eastern Mussel Farms (Tenants Harbor, ME) were purchased locally or collected from the waters off the northeast coast of the United States (Dover Point, NH), and the EP protein was extracted and purified following procedures as described elsewhere (*29*). The enzyme PNGase F and associated glycan release reagents (10× denaturing buffer, NP-40, G7 buffer) were purchased from New England Biolabs (Beverly, MA); Montage in-gel digestion kit and C-18 Sep-Pak and Ziptip C-18 columns were from Millipore (Bedford, MA); water-$^{18}O$ (normalized 95 atom % $^{18}O$), dithiothreitol (DTT), iodoacetamide, tributylphosphine, proteomic grade trypsin, endoproteinase Glu-C, and MALDI matrices, including 2,5-dihydroxybenzoic (DHB), $\alpha$-cyano-4-hydroxycinnamic (CHCA), and sinapinic acids, were from Sigma-Aldrich (St. Louis, MO); acetonitrile and methanol were from E. M. Science (Gibbstown, NJ), and formic acid and acetic acid were from VWR Scientific Product (West Chester, PA). The Ettan CAF MALDI sequencing kit was donated by Amersham Biosciences (Piscataway, NJ). Chemicals were of ACS or HPLC grade. All water used was obtained from a Milli-Q system (Millipore, Bedford, MA).

*RNA Purification.* The mantle and gill tissues from several live mussels were excised and immediately stabilized in the RNA*Later* RNA stabilization reagent (Qiagen). The tissues were disrupted with a mortar and pestle and then homogenized with QiaShredder columns (Qiagen). Total RNA was isolated with an RNeasy mini kit (Qiagen). The average yield was about 10 $\mu$g of total RNA from ~25 mg of tissues. The RNA purity was determined by the $A_{260nm}/A_{280nm}$ ratio spectrophotometrically. The integrity of RNA was analyzed on a 1.2% denaturing agarose gel with ethidium bromide staining. All reagents used were RNase and DNase free.

*RT-PCR and Cloning.* Reverse transcription (RT) and polymerase chain reaction (PCR) were carried out with a ThermoScript RT-PCR system (Invitrogen). The PCR reactions were performed in a GeneAmp PCR system 9600 (Applied Biosystems). The degenerate oligonucleotide primer DP1 (5′-CAY GAY GAY CAY CAY GAY GC-3′) (Figure 1) encoding the amino acid sequence of HDDHHDA near the N-terminus of the EP protein for PCR amplification was obtained from Integrated DNA Technologies (Coralville, IA). In the primer sequence, Y codes for either T or C. About 2.2 $\mu$g of total RNA was used for RT with 50 pmol of oligo-[$d(T)_{20}$] primer in a 20 $\mu$L reaction, catalyzed by the enclosed ThermoScript RT reverse transcriptase. The cDNA sequence of the EP protein was then amplified by PCR using DP1 as the sense primer and oligo[$d(T)_{20}$] as the antisense primer. A 50 $\mu$L PCR reaction mixture consisted of 2.5 $\mu$L of first-strand cDNA, 0.2 $\mu$M primer, 200 mM dNTP mix, 2.5 units of Platinum *Taq* DNA polymerase, and 1× PCR buffer containing 1.5 mM $MgCl_2$. The thermal cycling program used for PCR amplification included an initial step at 95 °C for 3.5 min followed by 30 cycles composed of 95 °C for 30 s, 50 °C for 1 min, and 72 °C for 2 min and then ended with an additional step of 72 °C for 3 min. RT-PCR reactions omitting total RNA were incubated in parallel as negative controls.

The PCR products were analyzed by electrophoresis on a 1% agarose gel. The major PCR fragment of the expected size ~1 kb was extracted from the gel and purified with a MinElute gel extraction kit (Qiagen). Then it was ligated into the pCR 2.1 vector and transformed into *E. coli* TOP10 cells using a TA cloning kit (Invitrogen). The presence of inserts was confirmed by restriction enzyme *Eco*RI digestion in addition to blue/white screening. Plasmid DNA from three positive clones was sequenced by the dideoxynucleotide chain termination method (*30*) using M13 forward and reverse primers at the Norris Comprehensive Cancer Center Core Facility, University of Southern California.

*3′ and 5′ RACE.* To obtain the full cDNA sequence, rapid amplification of cDNA ends (RACE) was carried out using a FirstChoice RLM-RACE kit (Ambion). All of the gene-specific primers were synthesized by Integrated DNA Technologies (Corallville, IA). For 3′ RACE, the sense primer 3P (5′-GAA ATG ATT ATT CAC GCA GAC GCA GAG C-3′), as shown in Figure 1, was designed on the basis of the sequence determined from RT-PCR. The downstream antisense primer was supplied with the kit. The reaction condition was similar to those described above, except that it was catalyzed by SuperTaq DNA polymerase (Ambion). The PCR settings were an initial step at 94 °C for 3 min and then 35 cycles of 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 1 min, followed by a final extension step of 72 °C for 7 min. Two gene-specific antisense primers were used in the nested PCR for 5′ RACE. They were $5P_{outer}$ (5′-GTT CGT CAA TCT CAT GCT TGT TAT GTT CTG T-3′) for the outer PCR and $5P_{inner}$ (5′-AAT TTC CTT TTC GAT TTC GTG GTG GAT GAA C-3′) for the inner PCR, respectively (Figure 1). Both sense primers were provided with the kit. The optimum PCR conditions for both inner and outer 5′ RACE were similar to that of 3′ RACE except that the annealing temperatures were both 60 °C. Both 3′ and 5′ RACE products were purified, cloned, and sequenced following the similar aforementioned procedures.

*Enzymatic N-Linked Glycan Release.* The EP protein was denatured in 10 $\mu$L of 1× denaturing buffer (5% SDS and 10% $\beta$-mercaptoethanol) at 100 °C for 10 min. The sample was cooled to room temperature followed by the addition of 1 $\mu$L each of G7 buffer (0.5 M sodium phosphate, pH 7.5), NP-40, and 0.1 unit of PNGase F. The deglycosylation experiment was incubated overnight at 37 °C.

*Two-Dimensional Gel Electrophoresis (2D GE).* Two-dimensional gel electrophoresis was carried out on ElectrophoretIQ 2000 equipment (Proteome Systems). The EP protein was precipitated by cold acetone and resuspended in 400 $\mu$L of universal 2D sample buffer consisting of 40 mM Tris, 10 mM acrylamide, 2 M thiourea, 7 M urea, 2% CHAPS [3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate], 0.8% pH 3−10 carrier ampholytes, 2% SB3-10 detergent, and 0.01% bromophenol blue. Then a final concentration of 5 mM tributylphosphine was added before rehydrating 24 cm IPG strips with pH 4−7 gradient (Amersham) overnight at room temperature.

Isoelectric focusing was performed at 18 °C using a voltage step gradient from 400 to 6000 V with 120000 total V·h. The strips were cut in half and equilibrated twice for 10 min in 10 mL of equilibration buffer (50 mM Tris/acetate, pH 7.0, 3 M urea, 2.5% SDS, and 0.01% bromophenol blue). The equilibrated strips were applied to a 6−15% polyacrylamide gel and run in Tris/Tricine/SDS running buffer at 15 °C with an applied current of 50 mA per gel for 90 min. After electrophoresis, the gels were put in a fixative solution (25% methanol and 10% acetic acid) for 60 min and stained with Coomassie blue overnight. Deglycosylated EP protein was analyzed by 2D GE following the same procedures described above.

*In-Gel Trypsin Digestion.* Protein spots or bands (2−3 mm diameter) were cut from the gel and transferred to Eppendorf tubes. To each tube was added 100 $\mu$L of destaining solution of 50 mM $NH_4HCO_3$ in 50% acetonitrile (ACN), and the mixture was incubated for 20 min. The solution was discarded, and this process was repeated three times. Then 200 $\mu$L of 100% ACN was added, and the solution was incubated for 10 min. The solution was taken out, and the gel pieces were dried in the SpeedVac for 15 min. After that, 5 $\mu$L of trypsin solution (0.0033 $\mu$g/$\mu$L in 25 mM $NH_4HCO_3$) was added to each tube containing dry gel pieces, and the resultant solution was incubated at 37 °C overnight. The peptides were recovered from the gel pieces by incubating with 50 $\mu$L of an extraction solution of 50% ACN/0.1% TFA (trifluoroacetic acid) for 30 min. The extracted peptide solution was dried in the SpeedVac to approximately 10 $\mu$L and stored at −20 °C. Peptides were desalted and concentrated using Ziptip C-18 columns before MS analysis.

*Enzymatic Digest with Isotope Labeling.* The EP protein was dissolved in 100 mM $NH_4HCO_3$ and 50% $H_2{}^{16}O/H_2{}^{18}O$ buffer (pH 8.5), and 1−2 $\mu$g of trypsin was then added. The sample was incubated at 37 °C overnight, and the peptide mixture was desalted before MS analysis.

*Deglycosylation of the Glycopeptide and Isotope Labeling of the Glycosylation Site.* The EP protein was digested either by trypsin or by Glu-C as described before in 100% $H_2{}^{16}O$ digestion buffer. The peptide mixture was then dried in the SpeedVac. Ten microliters of denaturing buffer (5% SDS, 10% $\beta$-mercaptoethanol) was added, and the mixture was boiled at 100 °C for 5 min to deactivate the trypsin/Glu-C. After being cooled to room temperature, 0.5 $\mu$L of PNGase F was added, and the sample was incubated at 37 °C overnight. For the isotope labeling sample, the denaturing buffer was composed of 50% $H_2{}^{16}O/H_2{}^{18}O$.

The glycan and peptides were separated on a C-18 Sep-Pak column (Millipore). The peptides were eluted with 3 mL of 50% ACN/0.1% TFA followed by drying down in the SpeedVac. Samples were then reconstituted in 0.1% TFA for MALDI analysis.

*Chemical-Assisted Fragmentation (CAF).* All of the derivatization steps were performed on the Ziptip C-18 resin bed. The resin was wetted with 50% ACN/0.5% TFA and equilibrated with 0.1% TFA. The sample was loaded and treated with *o*-methylisourea hydrogen sulfate overnight. The resin tip was washed by Milli-Q $H_2O$ followed by the addition of the CAF reagent (NHS ester of 3-sulfopropionic acid anhydride). The reaction was run for at least 3 min before the stopping solution (50% hydroxylamine) was added to quench the reaction. The derivatized sample was eluted in 5 $\mu$L of 80% ACN/0.5% TFA and analyzed by MALDI-TOF directly.

*Protein Reduction and Alkylation.* The EP protein was dissolved in 50 $\mu$L of 100 mM $NH_4HCO_3$ buffer (pH 8.5). Five microliters of 45 mM DTT was added, and the reaction was incubated at 60−65 °C for 1 h. After being cooled to

room temperature, 5 $\mu$L of 100 mM iodoacetamide was added. The alkylation reaction was allowed to occur in the dark for 1 h before the solution was diluted to 250 $\mu$L by adding 200 $\mu$L of 100 mM $NH_4HCO_3$. Then, 2 $\mu$g of trypsin was added, and the sample was incubated at 37 °C overnight.

*MALDI-TOF Mass Spectrometry.* MALDI-TOF mass spectra were recorded using an Axima-Curved Field Reflectron (CFR) mass spectrometer (Shimadzu/Kratos Analytical, Manchester, U.K.) equipped with a nitrogen laser ($\lambda = 337$ nm). The instrument was calibrated externally using two protein standards (cytochrome *c* and apomyoglobin) for the linear mode and two peptide standards (angiotensin II and $P_{14}R$) for the reflectron mode. The EP protein was desalted with a spin column (Pierce) before MALDI-TOF MS analysis. Approximately 0.5 $\mu$L of sample and 0.5 $\mu$L of matrix were spotted onto a stainless steel target plate and allowed to air-dry before analysis. The samples were ablated using ~100 laser shots fired in 10 shot packets while the laser rastered over the target surface. All of the spectra were acquired in the positive-ion reflectron mode using either 2,5-DHB (2,5-dihydroxybenzoic acid) or α-CHCA (α-cyno-4-hydroxycinnamic acid) (10 mg/mL in 50% ACN/0.1% TFA) as the matrix. Sinapinic acid (10 mg/mL in 70% ACN/0.1% TFA) was used as the matrix for protein analysis in the linear mode. PSD spectra were acquired by increasing 15−20% laser power. Spectra processing was done with Kratos Launchpad software.

*MALDI-IT-TOF Mass Spectrometry.* MALDI-IT-TOF mass spectrometry was performed on the Axima QIT (Shimadzu/Kratos Analytical, Manchester, U.K.). The sample plate was prepared in an identical way. Spectra were collected using 200 laser shots fired in 2 shot units while manually moving the laser around the crystal edge. $MS^n$ experiments were achieved by introducing helium gas into the ion trap and increasing the cell collision energy.

*ESI-QTOF Mass Spectrometry.* ESI-QTOF mass spectrometry was performed in the positive ion mode on a Q-TOF Ultima API instrument (Micromass, Manchester, U.K.). Peptide samples were dissolved in 50% methanol/0.1% formic acid and analyzed by a nanospray capillary tip with an estimated flow rate of ~200 nL/min and a spray voltage of 1.0−1.2 kV. Samples were desolvated by a heated capillary at a temperature of 100 °C. Extraction cone voltage was 35 V for MS profile spectra and 45 V for MS/CID-MS experiments. The collision energy for CID studies was adjusted between 10 and 80 to get optimal fragments. In general, spectra represent the summation of 50−350 scans (30 scan/min) for MS and MS/MS experiments. Spectra processing was done with MassLynx v3.5 software.

*Shell Matrix Protein Identification by Proteomic Methods.* All of the tissues attached to the shell were removed using a scalpel. The shells were soaked in 10% (v/v) ammonia solution overnight and then rinsed with deionized water. Both the external and inner shell surfaces were scrubbed with sand paper to remove residual tissues. The shells were washed thoroughly with deionized water before being pulverized in a ball mill grinder. The EDTA-soluble matrix proteins were extracted by decalcification of 70 g of shell powder in 600 mL of 0.5 M ethylenediaminetetraacetic acid (EDTA), pH 8.0, solution with continuous stirring for 8 days. After centrifugation, the supernatant was filtered through a 0.2 $\mu$m syringe filter. The EDTA-soluble matrix proteins were then desalted and concentrated by ultrafiltration using a membrane disk with a molecular weight cutoff of 3000 (Millipore, MA). Shell powder (40 g) was added slowly to 400 mL of 10% (v/v) acetic acid solution with continuous stirring for 7 days. The mixture was centrifuged and filtered as aforementioned. The insoluble materials obtained from acetic acid decalcification were washed with deionized water, air-dried, and extracted with 200 mL of 1% SDS and 10 mM DTT in 50 mM Tris-HCl (pH. 8.0) at 80−100 °C for 4 h with continuous stirring. The acetic acid-insoluble matrix proteins were then collected after centrifugation, filtration, desalting, and concentrating procedures as aforementioned.

The extracted matrix protein mixtures were separated by 2D GEs following the same procedure as for the EP protein, except that 18 cm IPG strips with a pH 3−10 gradient were used instead. Individual protein spots were excised and digested following the in-gel trypsin digestion protocol. The tryptic peptides of the protein spot were analyzed by MALDI-TOF/PSD using α-CHCA as the matrix. The acquired PMFs were searched in protein databases using MASCOT to identify each protein component.

## RESULTS

*cDNA Sequencing and Derived Amino Acid Sequence of the EP Protein.* By using RT-PCR strategy, the partial cDNA sequence of the EP protein was amplified. A prominent ~1 kbp product was generated from reverse transcription using the primer oligo[$d(T)_{20}$] and a degenerate primer, DP1, that corresponds to an internal peptide sequence of the EP protein (see Experimental Procedures). The deduced sequence of the obtained PCR product is comprised of 191 amino acids, and its N-terminal sequence matches perfectly to the peptide sequence used to design the degenerate primer DP1 used to amplify the EP protein. Three gene-specific primers, 3P, $5P_{inner}$, and $5P_{outer}$, were then designed on the basis of the sequence of this partial cDNA sequence to establish the C- and N-terminal sequences by 3′ and 5′ RACE reactions, respectively (see Experimental Procedures). The ~400 and ~350 bp nucleotide sequences obtained by 3′ and 5′ RACE, respectively, were aligned with the partial cDNA sequence obtained from RT-PCR. Excluding the nucleotide sequence contributed to the RT-PCR product from the degenerate primer DP1, the overlapping sequences for the RT-PCR, 3′ and 5′ RACE products matched exactly. This equated to a 128 bp overlap for the 3′ RACE product and a 105 bp overlap for the 5′ RACE product. The complete cDNA sequence (982 bp) of the EP protein was thus constructed and its amino acid sequence derived (Figure 1).

Sequence analysis reveals an open reading frame of 236 amino acids with the translation initiation codon ATG at nucleotide position 61. At position +4 from this initiation codon there is a guanine nucleotide, and at position −3 there exists an adenine nucleotide, representing a Kozak initiation sequence, the optimal sequence for initiation by eukaryotic ribosomes (*31*). A stop codon TAA is located at position 769. A putative polyadenylation signal (AATAAA) is found at position 951, which is 16 nucleotides upstream from the poly(A) tail. The N-terminal sequence of the first 20 amino acids determined by this method is fairly consistent to the previous result by Edman degradation (*29*) (Figure 2), except that the cDNA-derived EP protein sequence shows both Gly

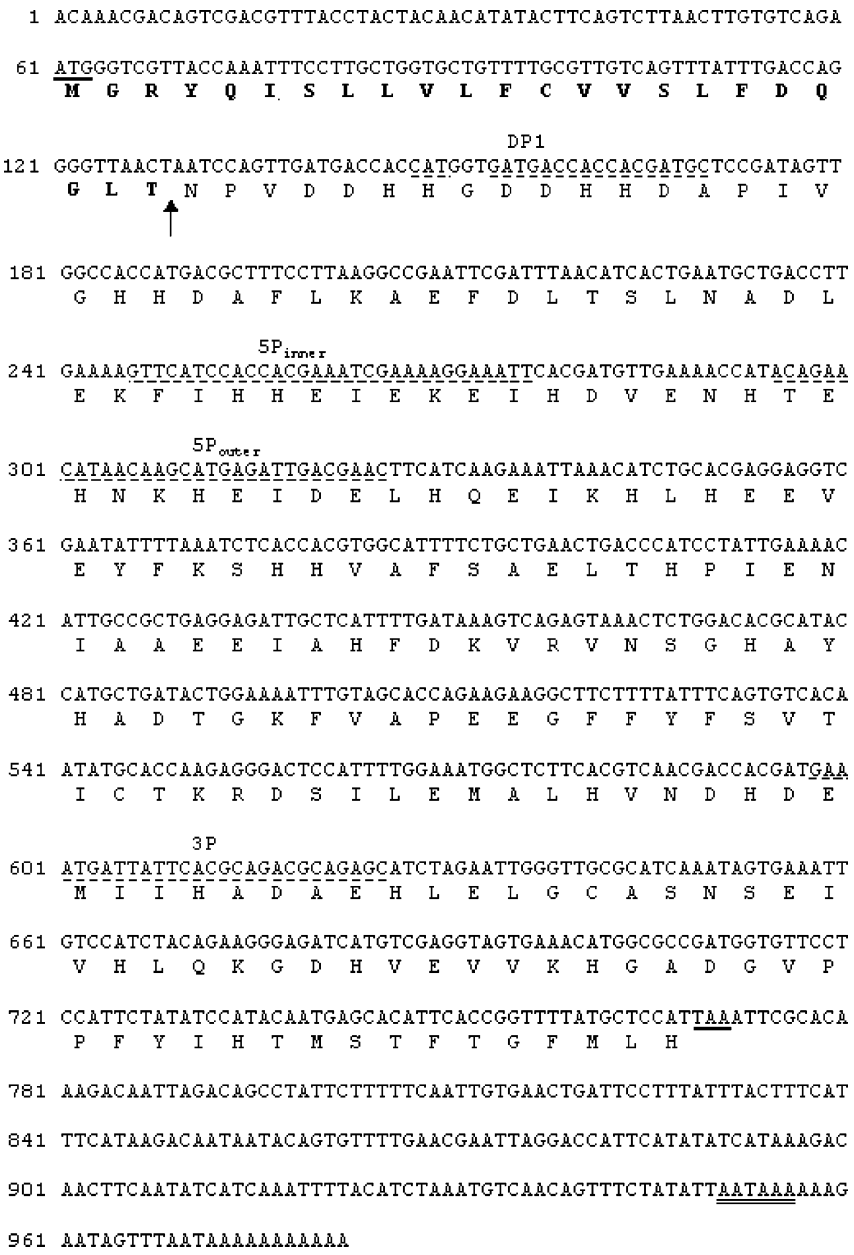


FIGURE 1: cDNA-derived protein sequence of the EP protein. The putative signal peptide is in boldface. The vertical arrow designates the cleavage site. The initiation codon (ATG) and the stop codon (TTA) are underlined. The polyadenylation signal (AATAAA) is doubly underlined. Dashed underlined cDNA sequences denote primers used to determine the complete sequence: DP1 for RT-PCR; 3P, sense primer for 3′ RACE; $5P_{inner}$, antisense primer for inner 5′ RACE; $5P_{outer}$, antisense primer for outer 5′ RACE.

```
EP  NPVDDH_DDHHDAPIVEHHD(Edman degradation)                    20
EP  NPVDDHHGDDHHDAPIVGHHDAFLKAEFDLTSLNADLEKFIHHEIEKEIHDVENHTEHNK 60
HIP NPVDDHQNDDHHDAPIVGHHDAFLKAEFDLTSLNADLEKFIHHEIEKEIHDVENHTEHNK 60

EP  HEIDELHQEIKHLHEEVEYFKSHHVAFSAELTHPIENIAAEEIAHFDKVRVNSGHAYHAD 120
HIP HEIDALHLEIKQLHEEVEYFKSHHVAFSAELTHPIENLGAEEIAHFDKVRVNSGDAYHVD 120

EP  TGKFVAPEEGFFYFSVTICTKRDSILEMALHVNDHDEMIIHADAEHLELGCASNSEIVHL 180
HIP TGKFVAPEEGFFYFSVTICTKRDSILEMALHVNDHDEMIIHADAEHLELGCASNSEIVHL 180

EP  QKGDHVEVVKHGADGVPPFYIHTMSTFTGFMLH 213
HIP QKGDHVEVVKHGADGVPPFYIHTMSTFTGFMLH 213
```

FIGURE 2: Amino acid sequence alignment of the EP protein, the HIP protein, and the N-terminal sequence of the EP protein acquired by Edman degradation. Dissimilar residues are shaded. The skipped residue is underlined. The N-glycosylation site (NHTE) is inversely shaded. Two Cys residues that form an intramolecular disulfide bond are in boldface and doubly underlined.

8 and Gly 18, which has also been confirmed by mass spectrometry data, whereas by Edman degradation Glu 18 was found but there was no residue at the eighth position. Together with the molecular weight and amino acid composition analysis (Table 1), the data further confirm that this full cDNA sequence encodes the entire EP protein. The

Table 1: Amino Acid Composition of the EP Protein without Signal Peptide

| aa residue | results based on EP sequence derived from cDNA sequence | | previous result[b] (mol %) |
|---|---|---|---|
| | mol/mol[a] | mol % | |
| Ala (A) | 17 | 8.0 | 6.9 ± 1.1 |
| Arg (R) | 2 | 0.9 | 2.2 ± 1.1 |
| Asn (N) | 8 | 3.8 | |
| Asp (D) | 18 | 8.5 | |
| Cys (C) | 2 | 0.9 | 1.2 |
| Gln(Q) | 2 | 0.9 | |
| Glu (E) | 25 | 11.7 | |
| Gly (G) | 10 | 4.7 | 6.2 ± 1.2 |
| **His (H)[c]** | **30** | **14.1** | **11.1 ± 1.9** |
| Ile (I) | 15 | 7.0 | 5.1 ± 1.2 |
| Leu (L) | 13 | 6.1 | 6.0 ± 0.7 |
| Lys (K) | 11 | 5.2 | 5.4 ± 0.6 |
| Met (M) | 4 | 1.9 | 1.3 ± 0.7 |
| Phe (F) | 13 | 6.1 | 4.6 ± 0.2 |
| Pro (P) | 6 | 2.8 | 3.0 ± 0.1 |
| Ser (S) | 9 | 4.2 | 5.8 ± 1.2 |
| Thr (T) | 9 | 4.2 | 5.5 ± 0.4 |
| Trp (W) | 0 | 0 | |
| Tyr (Y) | 4 | 1.9 | 2.2 ± 0.3 |
| Val (V) | 15 | 7.0 | 8.2 ± 0.8 |
| **Asx (B)[c]** | **26** | **12.2** | **11.5 ± 1.0** |
| **Glx (Z)[c]** | **27** | **12.7** | **13.7 ± 1.8** |
| total | 213 | 100 | 100 |

[a] Mole of amino acid per mole of protein. [b] Reference *29*. [c] Especially abundant residues.

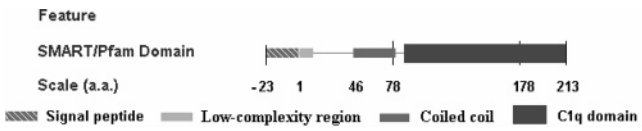


FIGURE 3: Domain profile of the EP protein predicted by SMART.

complete cDNA sequence and the derived amino acid sequence of the EP protein have been deposited in GenBank (Accession Number AY364453).

According to SignalP v2.0 (www.us.expasy.org), a signal peptide cleavage site between Thr 23 and Asn 24 is predicted by both neural networks and hidden Markov models with the maximum cleavage probability of 0.833 (*32*), indicating that the first 23 amino acid residues form a signal peptide (probability of 0.995) (Figure 1). The secreted EP protein is composed of 213 amino acids with a mass of 24.3 kDa and a theoretical p*I* of 5.24; both are predicted from the derived amino acid sequence (ProtParam tool). On the basis of the deduced amino acid sequence, the EP protein is rich in His (14.1%) as well as Asx (12.3%) and Glx (12.6%) residues (Table 1), which is in good agreement with previous results from amino acid analysis (*29*). In addition, the EP protein domain profile, including a low-complexity region of Asp 4−Asp 13, a coiled region (intimately associated bundles of α-helices) of Glu 46−Glu 78, and a globular C1q (complement 1q) domain from Ala 86 to Lys 212, was predicted by the simple modular architecture research tool (SMART) (Figure 3) (*33*).

*Molecular Mass Determination.* The molecular mass of the EP protein was determined by MALDI-TOF. The MS profile of intact EP protein showed three peaks (Figure 4A). The peak centered at *m*/*z* 28.2 kDa indicates a singly charged protein monomer $[M + H]^+$, and the *m*/*z* of the 55.8 kDa ion represents the singly charged protein dimer $[M_2 + H]^+$ from the known property of dimerization (*29*). The lower mass ion at *m*/*z* 14.1 kDa represents the doubly charged protein monomer $[M + 2H]^{2+}$.

To examine the existence of N-linked glycosylation, the protein was treated with PNGase F, an endoglycosidase that specifically cleaves N-linked glycans from proteins, and the treated sample was subsequently analyzed by MALDI-TOF-MS under the same conditions as the native EP protein. The MS profile of the PNGase F treated EP protein showed a related mass distribution peak pattern (Figure 4B). Peaks at *m*/*z* 24.2 and *m*/*z* 48.2 kDa are consistent with a singly charged protein monomer and dimer, respectively, and the peak at *m*/*z* 12.1 kDa indicates a doubly charged monomer. Comparison between the spectra indicates that the molecular masses of the monomer and the dimer are shifted down by ∼4.0 and ∼8.0 kDa, respectively, after deglycosylation, which is consistent with the MW of the N-linked glycans (*34*, *35*). Polyacrylamide−SDS gel electrophoresis of the EP protein and deglycosylated EP protein also confirms the presence of a glycan of mass ∼4.0 kDa in the native protein (gel not shown).

*2D Gel Electrophoresis.* The 2D gel of the EP protein showed two rows of spots approximating the monomer MW of the intact glycoprotein (Figure 5A). The first row showed eight spots of MW ∼30−33 kDa, with incrementing differences in p*I* (ranging from 4.8 to 5.3), which is somewhat higher than previous observations (*29*). The second row of spots has the same p*I* but lower MW. The peptide mass fingerprint (PMF) of four protein spots, two spots from each row, showed that they all contain the same peptide ions, only differing in relative ion intensity (data not shown). This observation suggests that these spots have identical protein backbones and supports the hypothesis that the lower row is a clipped version of the protein. There is another row of lighter spots at about 55 kDa corresponding to the EP protein dimer as observed previously by size exclusion chromatography and analytical ultracentrifugation (*29*).

The gel of the PNGase F treated sample showed a similar pattern with all spots shifted to lower MW dimension and higher pH dimension, ranging from 5.1 to 5.5 (Figure 5B). In the MW dimension, the downward shift accounted for ∼4 kDa, due to the loss of glycan. The upward shift in pH dimension indicates that the glycan is acidic.

Often such heterogeneity in the charge dimension is attributed to protein phosphorylation, indicated by a horizontal trail of spots on 2D gels as seen in Figure 5A (*36*). However, treatment of the protein with λ-protein phosphatase did not change the pattern of spots significantly (data not shown), arguing against this interpretation. The existence of residual charge heterogeneity after deglycosylation suggests that the majority of the charge difference resides in the protein backbone and not in the glycan, a finding contrary to the previous suggestion (*29*). The charge heterogeneity possibly arises from the incorporation of posttranslationally modified amino acids, such as carboxylation of glutamate or aspartate, or sulfation of tyrosine.

*De Novo Peptide Sequencing by Trypsin and Glu-C Mapping.* De novo sequencing of tryptic and Glu-C digested EP peptides was performed by tandem mass spectrometry (CID of MALDI-QIT and ESI-QTOF, PSD of MALDI-TOF) with isotope labeling and chemical-assisted fragmentation. The PMF of the trypsin-digested EP protein acquired by

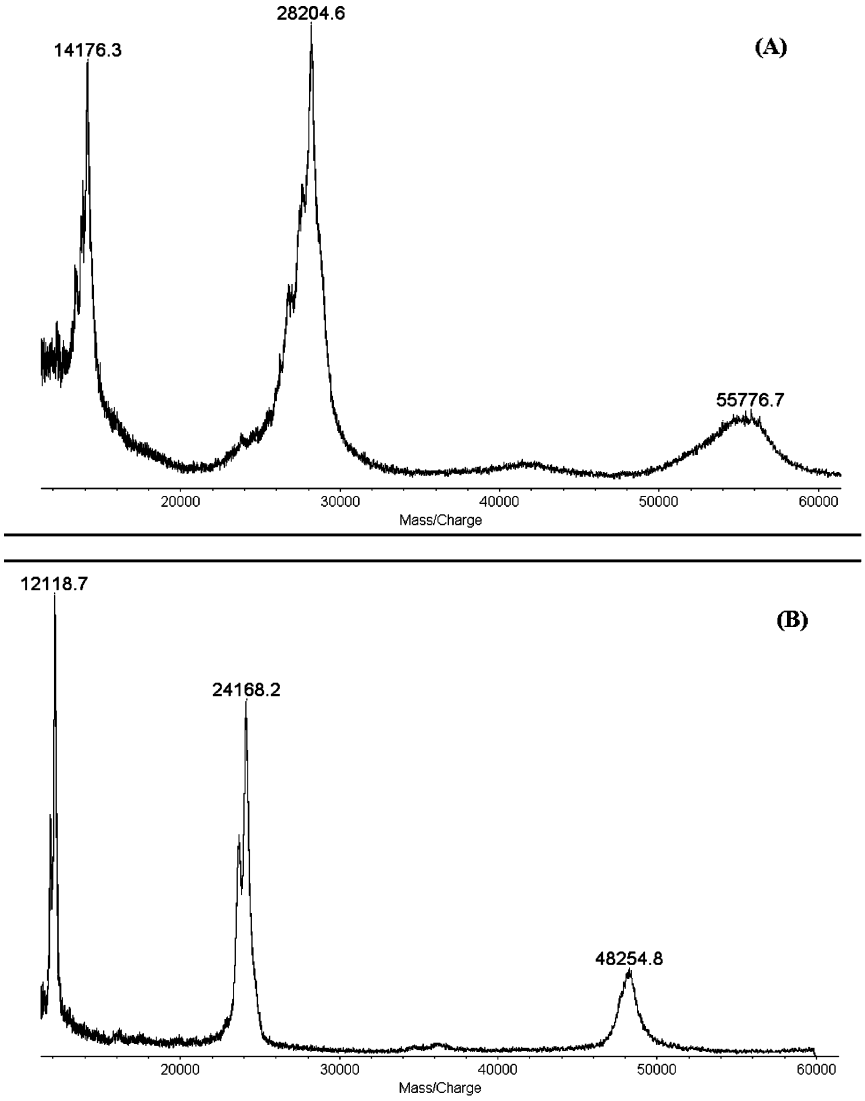


FIGURE 4: MS profiles of the EP protein (A) and deglycosylated EP protein (B) acquired by MALDI-TOF (linear mode). (A) The peak centered at *m*/*z* 28.2 kDa indicates a singly charged protein monomer, and the additional peaks at *m*/*z* 14.1 and 55.8 kDa represent the doubly charged monomer and singly charged dimer, respectively. (B) The peak centered at *m*/*z* 24.2 kDa represents a singly charged protein monomer; peaks at *m*/*z* 12.1 and 48.2 kDa represent a doubly charged monomer and a singly charged dimer, respectively.

MALDI-QIT-TOF are shown in Figure 6. The de novo sequencing results of the trypsin- and Glu-C-digested peptides are summarized in Table 2.

*Identification of the N-Glycosylation Site.* The EP protein is modified by N-glycosylation, and the attached oligosaccharide is an acidic glycan with a molecular mass of ~4 kDa. However, the expected glycopeptides were not detected in peptide mass fingerprints possibly because of the acidity and total size of the glycan. To characterize the glycosylation site, the intact protein was enzymatically deglycosylated with PNGase F, and $^{18}O$ was introduced at the Asn residue to label the glycosylation site.

Protease Glu-C was used to acquire smaller peptides that would provide glycosylation sequence information. A new ion, *m*/*z* 1094, was detected after the digestion of deglycosylated EP protein by Glu-C. The expanded mass range of this ion showed a 2 Da split isotope pattern in $H_2{}^{16}O$/ $H_2{}^{18}O$ digestion buffer, indicating the presence of an N-linked consensus site. The new peptide ion, *m*/*z* 1094, was fragmented by postsource decay, and the spectrum showed the full sequence as I/LHDVENHTE (Figure 7). As shown in the inset of Figure 7, those fragments that carried $^{18}O$-labeled aspartate keep their 2 Da doublet isotope pattern such as the "$b_6$" and "$y_5$" ions, and those ions without aspartate still retain the normal isotope distribution such as "$b_5$" at an *m*/*z* of 594.2. The result further confirms the successful $^{18}O$ labeling of the glycosylation site "NHTE" at Asn-54, in the predicted coiled region of the protein (Figure 3).

*Characterization of the Intramolecular Disulfide Bond.* In the MALDI PMF spectrum (Figure 8), two peptide peaks at *m*/*z* 2904 and 3033 did not correspond to any known peptides that would be generated by the Glu-C digest. Their masses match the Glu-C-digested sequence segment 168−176 cysteine-linked with segments 130−147 (*m*/*z* 2904) and 129−147 (*m*/*z* 3033), respectively, and therefore correspond to peptides containing a disulfide bridge.

Two new peptide peaks at *m*/*z* 2184 and 2313 appeared in the PMF spectrum when the EP protein was reduced by DTT and the resulting free sulfhydryl groups were alkylated with iodoacetamide (data not shown). Their masses matched the cysteine carboxymethylated peptide fragments from the peptide ions *m*/*z* 2904 and 3033, respectively, indicating that

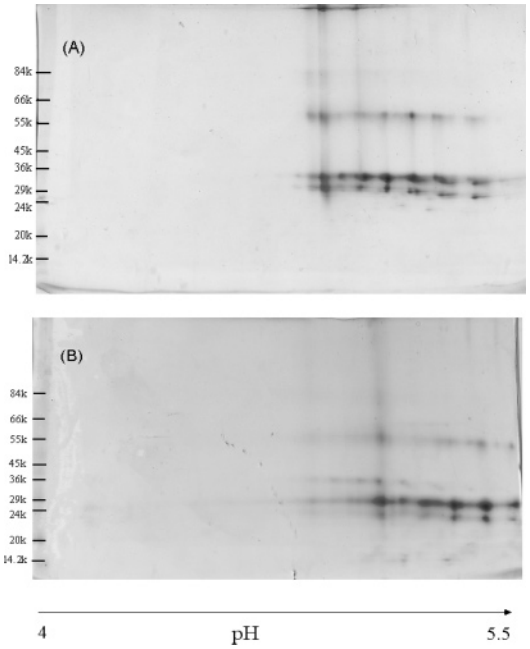


FIGURE 5: 2D gel electrophoresis of the intact EP protein (A) and N-glycan released EP protein (B). Both gels are a 24 cm, pH 4−7, IPG strip in the first dimension (gels of pH 4−5.5 are shown here) and 6−15% polyacrylamide in the second dimension.

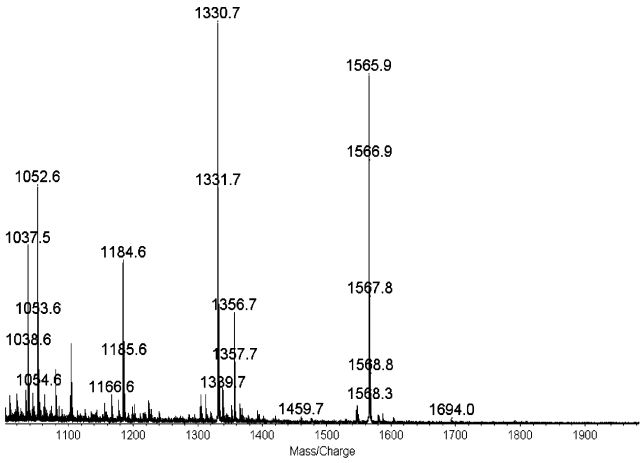


FIGURE 6: Tryptic PMF of the EP protein acquired by MALDI-QIT-TOF.

they were fragments from cleavage of the intramolecular disulfide bond of the protein. The other cysteine carboxymethylated peptide fragment (*m*/*z* 838) corresponding to sequence segment 168−176 was not detected, probably due to the low elution efficiency of hydrophilic peptides when the mixture is passed through the Ziptip C-18 column.

In conclusion, these experiments demonstrate the presence of a disulfide bond between Cys 139 and Cys 171, the only two Cys residues in the EP protein (Figure 2). As expected, the cysteine bridge occurs in the predicted C1q globular domain of the protein (Figure 3).

*Preliminary Search of the EP Protein in the Shell Matrix.* The mollusc shell organic matrix was extracted by demineralizing the shell with EDTA or acetic acid. The protein components of the organic matrix were separated by 2D GE and analyzed by MALDI-TOF/PSD mass spectrometry. To date, tropomyosin and paramyosin, from EDTA-soluble and acetic acid-insoluble shell matrices, respectively, have been identified by peptide mass fingerprinting in conjunction with a MASCOT database search (data not shown). These two proteins are consistently observed in significant amounts in all 2D gels of the organic matrix of meticulously cleaned shells; thus, their presence seems unlikely to be due to contamination. So far, we have not found evidence for the EP protein in the shell matrix using proteomic methods.

## DISCUSSION

This study extends our previous work on the characterization of the EP protein, the major protein content of the extrapallial fluid from *M. edulis* (*29*). The amino acid sequence of the EP protein was determined by RT-PCR, RACE, cloning, and DNA sequencing methods and confirmed by various mass spectrometry techniques. This is the first report on the structural characterization of a protein component isolated from a molluscan extrapallial fluid. Figure 1 shows the complete cDNA (982 bp) and amino acid (236 aa) sequences encoding the EP protein. A signal peptide consisting of 23 amino acid residues was discovered for the first time. It has a hydrophobic core of 8 residues and 6 other hydrophobic residues flanked by hydrophilic residues. A basic arginine residue found near the N-terminus is similar to known signal peptides found in proteins that are processed in the endoplasmic reticulum and subsequently secreted from the cell (*22*).

On the basis of the derived amino acid sequence, the calculated molecular mass of the secreted EP protein before posttranslational modification is 24.3 kDa, and its theoretical p*I* is 5.24 for the protein component alone. They are in good agreement with the previous measurements of MW (28.3 kDa) and median p*I* (4.43) of the mature EP protein (*29*), when taking into account the size (∼4 kDa) and acidic property of its N-linked glycan. The derived amino acid sequence of the protein exhibits a high proportion of Glu (11.7%) and Asp (8.5%) residues. Both kinds of residues are present in acidic form at the p*I* of 4.43. These results are in agreement with the observation that the soluble matrix is composed of proteins rich in acidic amino acid residues (*37*−*40*). Some of these Glu and/or Asp residues may provide carboxylic groups as ligands for chelating $Ca^{2+}$, accounting for the intrinsic $Ca^{2+}$ binding and induced assembly properties of the EP protein demonstrated earlier (*29*).

No sequence similarity was found between the deduced primary structure of the EP protein and any known molluscan shell matrix protein using BLAST search tools. Repetitive sequence motifs, such as the $(Asp\text{-}Y)_n$ type (Y = glycine or serine) (*8*, *41*−*43*) or $(Asp\text{-}Gly\text{-}X\text{-}Gly\text{-}X\text{-}Gly)_n$ type (*44*), have been predicted as calcium binding sites in molluscan shell matrix proteins. In contrast to these prevalent hypotheses, neither of these motifs exists in the EP protein. It is notable that these predictions were based on indirect observations and there have been several exceptions so far. A phosphoprotein of oyster shell is absent from the $(Asp\text{-}Y)_n$ motifs (*45*). The major soluble protein of pearl nacre, nacrein, contains an acidic domain, but it is not of the $(Asp\text{-}Y)_n$ type (*46*). These observations imply that those theories of protein−mineral interaction probably need further refinement.

Interestingly, the N-terminal sequence (Asp 4−Asp 13) of the mature EP protein is predicted by SMART to contain

Table 2: De Novo Sequencing Results of Trypsin and Glu-C Digested Peptides of EP Protein and Corresponding PMF Database Sequence of HIP[a]

| peptide | *m/z* (Da) | de novo sequencing of EP by both MS and cDNA | database sequence of HIP |
|---|---|---|---|
| trypsin | | | |
| 1 | 1052 | FI/LHHEI/LEK | FIHHEIEK (40−47) |
| 2 | 1330 | **H**I/LHEEVEYFK | **Q**LHEEVEYFK (72−81) |
| 3 | 1356 | V̄NSG**H**AYH**A**DTGK | V̄NSG**D**AYH**V**DTGK (111−123) |
| 4 | 1565 | AEFDL̄TSI/LN̄ADI/LEK | AEFDL̄TSLN̄ADLEK (26−39) |
| 5 | 1541 | HGADGVPPFYI/LHTM | HGADGVPPFYIHTM (201−214) |
| Glu-C | | | |
| 1 | 664 | HNKHE | HNKHE (57−62) |
| 2 | 810 | KFI/LHHE | KFIHHE (39−44) |
| 3 | 905 | I/LK/Q**H**LHEE | IK**Q**LHEE (70−76) |
| 4 | 1133 | I/LK/Q**H̄**LHEEVE | IK**Q̄**LHEEVE (70−78) |
| 5 | 1180 | MAI/LH̄VNDHDE | MAL̄HVNDHDE (148−157) |
| 6 | 1274 | I/LVHI/LK/QK/QGDHVE | IVHLQKGDHVE (177−187) |
| 7 | 1336 | I/LTHPI/LENI/L**A̲**AEE | LTHPIENL**G̲**AEE (91−102) |

[a] The differing amino acid residues are indicated in underlined boldface type.

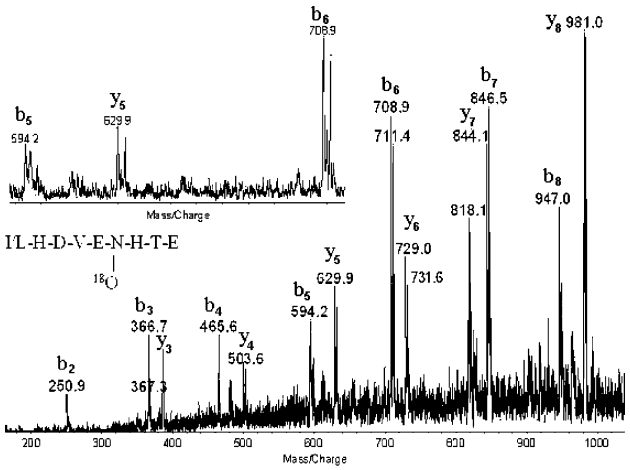


FIGURE 7: PSD spectrum of the peptide ion *m/z* 1094 acquired by MALDI-TOF. The complete sequence was derived, and the consensus site was identified as NHTE. The inset shows that those fragments containing the labeled aspartate retain the characteristic of a 2 mass split pattern whereas those without retain the normal isotope distribution.

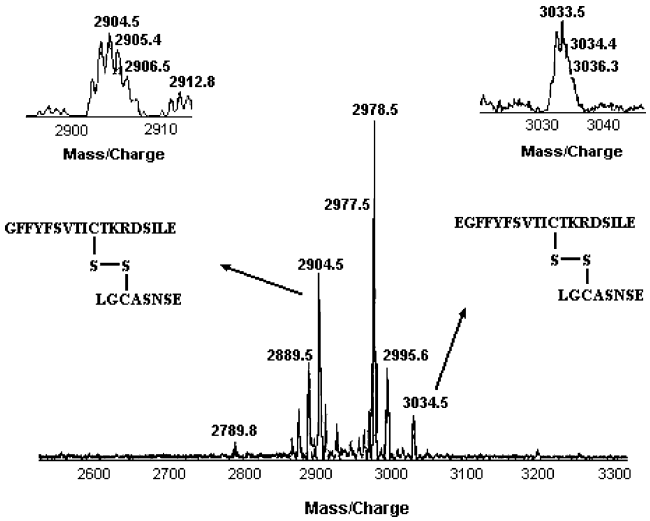


FIGURE 8: Glu-C PMF of the EP protein acquired by MALDI-TOF. The peptide ions *m/z* 2904 and 3033 matched the fragments with a disulfide bond.

an unfolded, conformationally labile polypeptide structure (Figure 3). The sequence also features the presence of several carboxylate residues (i.e., Asp) that may serve as $Ca^{2+}$ binding sites and hydrogen-bonding proton receptors, as well as hydrogen-bonding proton donor residues, His. It resembles the key attributes found in other polypeptide−$CaCO_3$ interaction domains (*47*, *48*). A flexible peptide backbone may allow strong electrostatic interactions with the surface of $CaCO_3$ through Asp diad residues, which are stabilized by hydrogen bonds between Asp and His and/or His and mineral surface water interactions. Close to the C-terminus, over half of the sequence of this protein forms a conserved globular complement C1q domain that has been found in the C-terminus of vertebrate-secreted or membrane-bound proteins, mostly short-chain collagens and collagen-like molecules (*49*−*51*). While it is conceivable that the N-terminus of the EP protein may play a crucial role in the protein−$Ca^{2+}$ interaction, the biological significance of the predicted C-terminal C1q domain needs further investigation.

The N-glycosylation site NHTE of the EP protein is located at amino acid positions 54−57. A dimer peak of the EP protein was detected by MALDI-TOF for both intact and deglycosylated protein, suggesting that the presence of glycan is not a prerequisite for the protein to assemble into its known dimeric structure in solution (*29*). However, since the MALDI measurement is done in a vacuum without solvent, dimer formation may be favored in the experiment. Glycosylation is a common modification of soluble shell proteins (*11*) and may provide sites for high anionic charge density through the addition of phosphorylated or sulfated oligosaccharides or sialic acid. However, the composition, structure, and functional significance of the N-linked glycan in the EP protein remain to be fully determined (*34*, *35*). The only two Cys residues present in the EP protein form an intramolecular disulfide bond in the predicted globular domain, indicating that the quaternary structure of the homodimer of the native EP protein is stabilized not by chemical cross-links but by hydrophobic and electrostatic interactions.

The amino acid sequence of the EP protein resembles that of a recently reported heavy metal binding protein (HIP) isolated from the hemolymph of *M. edulis* whose sequence has been determined with cDNA methods by a German group (Accession Number P83425). The HIP protein binds heavy

metal ions, such as $Zn^{2+}$, $Cd^{2+}$, and $Cu^{2+}$, and is proposed to serve as a carrier of divalent cations in the plasma for detoxification purposes. The sequences of the EP protein and HIP are compared in Figure 2. Only 9 out of 213 amino acid residues of the EP protein are different from those of HIP. EP protein has three additional His residues. The identities of these nine residues in the EP protein sequence have been confirmed in the present work either by de novo peptide sequencing or by the molecular weights of the corresponding peptide fragments obtained by mass spectrometry. Our previous study showed that the EP protein also binds $Cu^{2+}$, $Cd^{2+}$, $Mg^{2+}$, and $Mn^{2+}$, besides $Ca^{2+}$ (*29*). It binds 3−4 $Cu^{2+}$ per subunit, as investigated by EPR spectrometric titration and contains $Cu^{2+}$as isolated from the extrapallial fluid (*52*). The abundant His residues (14.1%) present in the EP protein could contribute to the binding sites of these metal ions. In the case of $Cu^{2+}$ binding, nitrogen coordination has been demonstrated by EPR (*52*). The EP protein also resembles a histidine-rich glycoprotein (HRG) from the plasma of *Mytilus edulis* whose $Cd^{2+}$ binding and transport properties have been studied by Nair and Robinson (*53*−*56*). The amino acid sequence translated from the preliminary partial cDNA sequence (327 bp) of the HRG shows at least 79% identity with the EP protein (*57*). It seems likely that the EP protein, HIP, and HRG are the same protein in *M. edulis*. Whether they all bear the same posttranslational modifications still remains to be established. The limited number of differences in protein sequence between the EP protein and HIP is possibly due to the fact that, in the two studies, the proteins were isolated from mussels on different sides of the Atlantic Ocean or that the EP protein and HIP are products of different but evolutionarily related genes.

Neither the EP protein nor the HRG protein from *M. edulis* shows significant sequence similarities to the multifunctional metal-binding HRG protein of human plasma (Accession Number P04196) (*58*) or to the sequences of the many other known HRG proteins of different phyla. However, pernin, a histidine-rich glycoprotein from the hemolymph of the green-lipped mussel *Perna canaliculus*, has been recently isolated and characterized (*59*). While the precise function(s) of pernin is (are) unclear, the protein bears some similarities to the EP protein. Pernin is also rich in His (13.7%) and Asp (12.3%). Although the overall sequences of the EP protein and pernin do not correspond, the N-terminal sequence "DDHHGDDHHD" of the EP protein is 90% homologous to the N-terminal segment of "DDHHDDHHD" in pernin. This unique repetitive sequence in pernin is proposed to be a spacer separating the partial binding motif of a thrombin-inhibitory protein from the three domains related to Cu−Zn superoxide dismutases (SODs), although no demonstrable SOD activity was detected for pernin. The sequence comparisons between the EP protein and the active sites of Cu−Zn SODs in yeast (Accession Number P00445), mouse (Accession Number P08228), and human (Accession Number P00441) reveal in total 19 identical residues across the aligned regions. Five out of seven His or Asp residues involved in Cu−Zn coordinating of typical SODs are conserved in the EP protein. However, like pernin, we found that the EP protein has no detectable SOD activity using the assay based on inhibition of cytochrome *c* reduction (*60*).

The EP fluid and the blood of mollusc have been traditionally thought to be separate fluids, based in part on differences in their inorganic composition (*16*). However, a recent in vivo study demonstrated that $Ca^{2+}$ and small organic molecules, in the range of a few hundred daltons (e.g., tyrosine), are readily exchanged between the EP fluid and plasma of the quahog *Mercenaria mercenaria*, while the passage of large molecules (e.g., bovine serum albumin, MW 66 kDa) across the outer mantle epithelium is restricted (*61*). Eighty-five percent of the calcium in plasma and EP fluid was found bound to macromolecules greater than ∼1000 Da and proposed to be transported by some relatively weakly binding protein and small organic molecules. Considering the calcium-binding properties of the EP protein, the size of the EP protein, and the similar ionic radii of $Ca^{2+}$ and $Cd^{2+}$, it is conceivable that this protein is also involved in the transport of $Ca^{2+}$ between the EP fluid and the plasma. Moreover, $Ca^{2+}$ and $Cd^{2+}$ appear to compete for the same binding sites and transport mechanisms in marine organisms (e.g., ref *62*). Thus, it is possible that the EP protein plays different roles in *M. edulis*, as a heavy metal detoxification protein, a $Ca^{2+}$-transport and a shell matrix protein. These putative roles remain to be further established.

We have recently initiated studies of the organic matrix of the shell using 2D GE coupled with MALDI-TOF mass spectrometry to identify the protein components present. To date, none of the identified shell matrix proteins resembles the EP protein. It is plausible that the EP protein becomes highly modified or cross-linked when incorporated into the shell and does not lend itself readily to identification by this approach or is not present in the matrix. Interestingly, tropomyosin from the EDTA-soluble matrix and paramyosin from the acetic acid-insoluble shell matrix have been found in the shell matrix. Tropomyosin has been found to act as a regulatory component in the muscle contraction and relaxation cycle in a $Ca^{2+}$-dependent manner when bound to troponin and actin thin filaments in skeletal muscle (*63*). For tropomyosin in invertebrate and smooth muscles, its myosin-containing thick filaments are primarily responsible for $Ca^{2+}$ regulation (*64*, *65*). Paramyosin forms the core of the invertebrate thick filaments and is proposed to have a structural role instead of as an obligatory component of the contractile apparatus (*66*). Previous studies have also revealed the presence of muscle proteins in seashells. For instance, tendon cells were found to insert into the gastropod shell by an extensive network of extracellular organic fibers when the muscle−shell attachment of the shell was investigated at the ultrastructure level (*67*). Therefore, the presence of tropomyosin and paramyosin in mollusc shells, if not involved in calcium mineralization, is possibly due to their penetration into the shell at the muscle scar sites.

## ACKNOWLEDGMENT

We thank Dr. Fadi Bou Abdallah for performing the SOD activity assay.

## REFERENCES

1. Schuler, D. (1999) Formation of magnetosomes in magnetotactic bacteria, *J. Mol. Microbiol. Biotechnol. 1*, 79−86.
2. Bazylinski, D. A., and Frankel, R. B. (2004) Magnetosome formation in prokaryotes, *Nat. Rev. Microbiol. 2*, 217−230.
3. Mann, S. (2001) *Biomineralization: principles and concepts in bioinorganic materials chemistry*, pp 1−5, Oxford University Press, New York.

4. Belcher, A. M., Wu, X. H., Christensen, R. J., Hansma, P. K., Stucky, G. D., and Morse, D. E. (1996) Control of crystal phase switching and orientation by soluble mollusk-shell proteins, *Nature 381*, 56−58.
5. Wierzbicki, A., Sikes, C. S., Madura, J. D., and Drake, B. (1994) Atomic force microscopy and molecular modeling of protein and peptide binding to calcite, *Calcif. Tissue Int. 54*, 133−141.
6. Addadi, L., and Weiner, S. (1992) Control and design principles in biological mineralization, *Angew. Chem., Int. Ed. Engl. 31*, 153−169.
7. Mann, S. (1988) Molecular recognition in biomineralization, *Nature 332*, 119−124.
8. Addadi, L., and Weiner, S. (1985) Interactions between acidic proteins and crystals: Stereochemical requirements in biomineralization, *Proc. Natl. Acad. Sci. U.S.A. 82*, 4110−4114.
9. Watabe, N., and Wilbur, K. M. (1960) Influence of the organic matrix on crystal type in molluscs, *Nature 188*, 334.
10. Weiner, S. (1979) Aspartic-acid rich proteins: major components of the soluble organic matrix of mollusk shells, *Calcif. Tissue Int. 29*, 163−167.
11. Lowenstam, H. A., and Weiner, S. (1989) *On Biomineralization*, pp 7−24, Oxford University Press, New York.
12. Crenshaw, M. A. (1980) Mechanisms of shell formation and dissolution, in *Skeletal growth of aquatic organisms* (Rhoads, D. C., and Lutz, R. A., Eds.) pp 115−132, Plenum Press, New York.
13. Wilbur, K. M., and Bernhardt, A. M. (1984) Effects of amino acids, magnesium, and molluscan extrapallial fluid on crystallization of calcium carbonate: in vitro experiments, *Biol. Bull. 166*, 251−259.
14. Eyster, L. S., and Morse, M. P. (1984) Early shell formation during molluscan embryogenesis, with new studies on the surf clam, *Spisula solidissima*, *Am. Zool. 24*, 871−882.
15. Young, S. D., Crenshaw, M. A., and King, D. B. (1977) Mantle protein excretion and calcification in the hardshell clam *Mercenaria mercenaria*. I. Protein excretion in the intact clam, *Mar. Biol. 41*, 253−257.
16. Crenshaw, M. A. (1972) Inorganic composition of molluscan extrapallial fluid, *Biol. Bull. 143*, 506−512.
17. Peitrzak, J. E., Bates, J. M., and Scott, R. M. (1976) Constituents of unionid extrapallial fluid. II. The pH and metal ion composition, *Hydrobiologia 50*, 89−93.
18. Wada, K., and Fujijuki, T. (1980) in *Mechanisms of mineralization in the invertebrates and plants* (Wataba, N., and Wilbur, K. M., Eds.) pp 175−190, University of South Carolina Press, Columbia, SC.
19. Weiner, S., Lowenstam, H. A., and Hood, L. (1977) Discrete molecular weight components of the organic matrices of mollusc shells, *J. Exp. Mar. Biol. Ecol. 30*, 45−51.
20. Misogianes, M. J., and Chasteen, N. D. (1979) A chemical and spectral characterization of the extrapallial fluid of *Mytilus edulis*, *Anal. Biochem. 100*, 324−334.
21. Halloran, B. A., and Donachy, J. E. (1995) Characterization of organic matrix macromolecules from the shells of the Antarctic scallop, *Adamussium colbecki*, *Comp. Biochem. Physiol. 111B*, 221−231.
22. Shen, X., Belcher, A. M., Hansma, P. K., Stucky, G. D., and Morse, D. E. (1997) Molecular cloning and characterization of lustrin A, a matrix protein from shell and pearl nacre of *Haliotis rufescens*, *J. Biol. Chem. 272*, 32472−32481.
23. Kono, M., Hayashi, N., and Samata, T. (2000) Molecular mechanism of the nacreous layer formation in *Pinctada maxima*, *Biochem. Biophys. Res. Commun. 269*, 213−218.
24. Marin, F., Corstjens, P., de Gaulejac, B., de Vrind-De Jong, E., and Westbroek, P. (2000) Mucins and molluscan calcification. Molecular characterization of mucoperlin, a novel mucin-like protein from the nacreous shell layer of the fan mussel *Pinna nobilis* (*Bivalvia, pteriomorphia*), *J. Biol. Chem. 275*, 20667−20675.
25. Weiss, I. M., Kaufmann, S., Mann, K., and Fritz, M. (2000) Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the mollusc *Haliotis laevigata*, *Biochem. Biophys. Res. Commun. 267*, 17−21.
26. Sarashina, I., and Endo, K. (2001) The complete primary structure of molluscan shell protein 1 (MSP-1), an acidic glycoprotein in the shell matrix of the scallop *Patinopecten yessoensis*, *Mar. Biotechnol. 3*, 362−369.
27. Marxen, J. C., Nimtz, M., Becker, W., and Mann, K. (2003) The major soluble 19.6 kDa protein of the organic shell matrix of the freshwater snail *Biomphalaria glabrata* is an N-glycosylated dermatopontin, *Biochim. Biophys. Acta 1650*, 92−98.
28. Zhang, Y., Xie, L., Meng, Q., Jiang, T., Pu. R., Chen. L., and Zhang, R. (2003) A novel matrix protein participating in the nacre framework formation of pearl oyster, *Pinctada fucata*, *Comp. Biochem. Physiol. 135B*, 565−573.
29. Hattan, S. J., Laue, T. M., and Chasteen, N. D. (2001) Purification and characterization of a novel calcium-binding protein from the extrapallial fluid of the mollusc, *Mytilus edulis*, *J. Biol. Chem. 276*, 4461−4468.
30. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
31. Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes, *Cell 44*, 283−292.
32. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, *Protein Eng. 10*, 1−6.
33. Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998) SMART, a simple modular architecture research tool: Identification of signaling domains, *Proc. Natl. Acad. Sci. U.S.A. 95*, 5857−5864.
34. Hattan, S. J., and Trimble, R. B. (1996) The purification and characterization of an extrapallial fluid glycoprotein from the mollusc, *Mytilus edulis*, *Glycobiology 6*, 755 (abstract).
35. Naggar, E., Ye, S., Chasteen, N. D., Reinhold, V. N., and Reinhold, B. (2000) Structural characterization of N-linked glycan from biomineralization matrix protein of *Mytilus edulis*, *Glycobiology 10*, 1119 (abstract 157).
36. Mann, M., and Jensen, O. (2003) Proteomic analysis of post-translational modifications *Nat. Biotechnol. 21*, 255−261.
37. Weiner, S. (1983) Mollusk shell formation: isolation of two organic matrix proteins associated with calcite deposition in the bivalve *Mytilus californianus*, *Biochemistry 22*, 4139−4145.
38. Crenshaw, M. A., and Ristedt, H. (1980) in *Mechanisms of mineralization in the invertebrates and plants* (Watabe, N., and Wilbur, K. M., Eds.) pp 355−367, University of South Carolina Press, Columbia, SC.
39. Greenfield, E. M., Wilson, D. C., and Crenshaw, M. A. (1984) Ionotropic nucleation of calcium carbonate by molluskan matrix, *Am. Zool. 24*, 925−932.
40. Halloran, B. A., and Donachy, J. E. (1995) Characterization of organic matrix macromolecules from the shells of the Antarctic scallop, *Adamussium colbecki*, *Comp. Biochem. Physiol. 111B*, 221−231.
41. Weiner, S., and Traub, W. (1984) Macromolecules in mollusc shells and their functions in biomineralization, *Philos. Trans. R. Soc. London, Ser. B 304*, 425−434.
42. Weiner, S., and Addadi, L. (1991) Acidic macromolecules of mineralized tissues: the controllers of crystal formation, *Trends Biochem. Sci. 16*, 252−256.
43. Weiner, S., and Hood, L. (1975) Soluble protein of the organic matrix of mollusk shells: a potential template for shell formation, *Science 190*, 987−989.
44. Runnegar, B. (1984) Crystallography of the foliated calcite shell layers of bivalve molluscs, *Alcheringa 8*, 273−290.
45. Wheeler, A. P. (1992) in *Hard tissue mineralization and demineralization* (Suga, S., and Watabe, N., Eds.) pp 171−187, Springer-Verlag, Tokyo.
46. Miyamoto, H., Miyashita, T., Okushima, M., Nakano, S., Morita, T., and Matsushiro, A. (1996) A carbonic anhydrase from the nacreous layer in oyster pearls, *Proc. Natl. Acad. Sci. U.S.A. 93*, 9657−9660.
47. Wustman, B. A., Morse, D. E., and Evans, J. S. (2004) Structural characterization of the N-terminal mineral modification domains from the molluscan crystal-modulating biomineralization proteins, AP7 and AP24, *Biopolymers 74*, 363−376.
48. Michenfelder, M., Fu, G., Lawrence, C., Wustman, B. A., Taranto, L., and Evans, J. S. (2003) Characterization of two molluscan crystal-modulating biomineralization proteins and identification of putative mineral binding domains, *Biopolymers 70*, 522−533.
49. Smith, K. F., Haris, P. I., Chapman, D., Reid, K. B. M., and Perkins, S. J. (1994) $\beta$-Sheet secondary structure of the trimeric globular domain of C1q of complement and collagen types VIII and X by Fourier-transform infrared spectroscopy and averaged structure predictions, *Biochem. J. 301*, 249−256.
50. Brass, A., Kadler, K. E., Thomas, J. T., Grant, M. E., and Boot-Handford, R. P. (1992) The fibrillar collagens, collagen VIII,

collagen X and the C1q complement proteins share a similar domain in their C-terminal non-collagenous regions, *FEBS Lett. 303*, 126−128.
51. Petry, F., Reid, K. B., and Loos, M. (1992) Isolation, sequence analysis and characterization of cDNA clones coding for the C chain of mouse C1q. Sequence similarity of complement subcomponent C1q, collagen type VIII and type X and precerebellin, *Eur. J. Biochem. 209*, 129−134.
52. Liang, C. (1998) A copper study of the extrapallial fluid protein of M*ytilus edulis*, M.S. Thesis, University of New Hampshire, Durham, NH.
53. Nair, P. S., and Robinson, W. E. (1999) Purification and characterization of a histidine-rich glycoprotein that binds cadmium from the blood plasma of the bivalve *Mytilus edulis*, *Arch. Biochem. Biophys. 36*, 8−14.
54. Nair, P. S., and Robinson, W. E. (2001) Cadmium binding to a histidine-rich glycoprotein from marine mussel blood plasma: Potentiometric titration and equilibrium speciation modeling, *Environ. Toxicol. Chem. 20*, 1596−1604.
55. Nair, P. S., and Robinson, W. E. (2000) Cadmium speciation and transport in the blood of the bivalve *Mytilus edulis*, *Mar. Environ. Res. 50*, 99−102.
56. Nair, P. S., and Robinson, W. E. (2001) Histidine-rich glycoprotein in the blood of the bivalve *Mytilus edulis*: Role in cadmium speciation and cadmium transfer to the kidney, *Aquat. Toxicol. 52*, 133−142.
57. Robinson, W. E., Sugumaran, M., Wallace, G., Abebe, A., Gaudette, M., and Catanzano, S. (2002) Ongoing molecular characterization of a metal-binding, histidine-rich glycoprotein (HRG) in marine mussel, *Mytilus edulis*, blood plasma, Proceedings of 29th Annual Aquatic Toxicology Workshop, Oct 21−23, 2002, Whistler, British Columbia, *Can. Technol. Rep. Fish. Aquat. Sci. 2438*, 44−47.
58. Jones, A. L., Hulett, M. D., and Parish, C. R. (2005) Histidine-rich glycoprotein: a novel adapter protein in plasma that modulates the immune and vascular and coagulation systems, *Immunol. Cell Biol. 83*, 106−118.
59. Scotti, P. D., Dearing, S. C., Greenwood, D. R., and Newcomb, R. D. (2001) Pernin: a novel, self-aggregating haemolymph protein from the New Zealand green-lipped mussel, *Perna canaliculus* (Bivalvia: Mytilidae), *Comp. Biochem. Physiol. 128B*, 767−779.
60. McCord, J. M., and Fridovich, I. (1969) Superoxide dismutase: an enzymatic function for erythrocuprein (hemocuprein), *J. Biol. Chem. 244*, 6049−6055.
61. Nair, P. S., and Robinson, W. E. (1998) Calcium speciation and exchange between blood and extrapallial fluid of the Quahog *Mercenaria mercenaria* (L.), *Biol. Bull. 195*, 43−51.
62. Wright, D. A. (1977) The effect of calcium on cadmium uptake by the shore crab *Carcinus meanas*, *J. Exp. Biol. 67*, 163−173.
63. Konno, K., Arai, K., and Watanabe, S. (1979) Myosin-linked calcium regulation in squid mantle muscle: Light-chain components of squid myosin, *J. Biochem. 86*, 1639−1650.
64. Ebashi, S., and Endo, M. (1968) Calcium ion and muscle contraction, *Prog. Biophys. Mol. Biol. 18*, 123−183.
65. Ohtsuk, I., Maruyama, K., and Ebashi, S. (1986) Regulatory and cytoskeletal proteins of vertebrate skeletal muscle, *Adv. Protein Chem. 38*, 2579−2583.
66. Watabe, S., and Hartshorne, D. J. (1990) Paramyosin and the catch mechanism, *Comp. Biochem. Physiol. 96B*, 639−646.
67. Tompa, A. S., and Watabe, N. (1976) Ultrastructural investigation of the mechanism of muscle attachment to the gastropod shell, *J. Morph. 149*, 339−352.

BI0505565