Proteomic characterization of oyster shell organic matrix proteins (OMP)

Abhishek Upadhyay[1,2,*], Vengatesen Thiyagarajan[1,2], Ying Tong[2]

[1]The Swire Institute of Marine Science (SWIMS), The University of Hong Kong (HKU), Hong Kong; [2]School of Biological Sciences (SBS), The University of Hong Kong (HKU), Hong Kong; Abhishek Upadhyay - E-mail: abhiup01@gmail.com; *Corresponding author

Abstract:
Oysters are economically and ecologically important bivalves, with its calcareous shell and delicious meat. The shell composition is a blend of inorganic crystals and shell proteins that form an organic matrix which protects the soft inner tissue of the oyster. The objective of the study was to compare the composition of organic matrix proteins (OMP) of two phylogenetically related species: the Hong Kong oyster (*Crassostrea hongkongensis*) and the Portuguese oyster (*Crassostrea angulata*) which differ in their shell hardness and mechanical properties. *C. hongkongensis* shells are comparatively stronger than *C. angulata*. Modern shotgun proteomics has been used to understand the nature of the OMP and the variations observed in the mechanical properties of these two species of oyster shells. After visualizing proteins on the one (1DE) and two-dimensional electrophoresis (2DE) gels, the protein spots and their intensities were compared using PDQuest software and 14 proteins of *C. hongkongensis* were found to be significantly different (student's t-test; $p<0.05$) when compared to the *C. angulata*. Furthermore, shell OMP separated on 1DE gels were processed using Triple TOF5600 mass spectrometry and 42 proteins of *C. hongkongensis* and 37 of *C. angulata* identified. A Circos based comparative analysis of the shell proteins of both oyster species were prepared against the shell proteome of other shell forming gastropods and molluscs to study the evolutionary conservation of OMP and their function. This comparative proteomics expanded our understating of the molecular mechanism behind the shells having different hardness and mechanical properties.

Keywords: Oyster, Shell, Organic matrix proteins, Proteomics

Background:
Oyster meat is considered a delicacy throughout the world and is a part of our culture and tradition. Three species of oysters e.g. Hong Kong oysters (*Crassostrea hongkongensis*), Portuguese oysters (*Crassostrea angulata*) and Pacific oysters (*Crassostrea gigas*) have been extensively cultivated and used as sea food around Hong Kong and southern China. Among the three species, *C. gigas* and *C. angulata* are evolutionarily closely related species and divergent from *C. hongkongensis* **[1].** Oysters are well known for their protein rich meat **[2].**

An oyster shell is a composite comprising of organic matrix and minerals. An oyster shell contains 95% or more of calcium carbonate ($CaCO_3$) and 0.1-5% of organic matrix proteins (OMP) which are called skeleton/shell proteins. This OMP comprises of proteins, glycoproteins, chitin and polysaccharides **[3].** The presence of OMPs in a naturally occurring biomaterial shell is the major difference between naturally assembled and manufactured minerals. These unique proteins having shell forming properties,

whilst designing and regulating functions are synthesized by mantle tissues in an oyster to make tightly packed mineral structures. OMPs play an essential role in a shell's morphology, structure, crystal size and thus its overall mechanical properties. The shell matrix presumably attracts the precursor ions present in seawater to form a three-dimensional framework for crystal formation. OMPs nucleate the crystal arrangement, decide the $CaCO_3$ polymorph and control the crystal elongation. However, the mechanism and pathways involved in biomineralization are still ambiguous and not well studied **[4].** Based on their acidic nature, the oyster shell proteins can be classified into three categories – extremely acidic shell proteins (iso-electric point; pI below 4.5), moderately acidic shell proteins (pI between 4.5 to 7) and basic shell proteins (pI above 7). A challenging step is to isolate and purify the most acidic shell proteins, due to their detection on SDS-PAGE, because of their negative charge. It is expected that they might easily diffuse out of the electrophoresis gel **[5].**

Shell mechanical strength is dependent on mineral (i.e. calcite crystal) orientation (or organization) and the quality and quantity of occluded OMP. The $CaCO_3$ crystal polymorph form and shell structure of Hong Kong oysters are similar to Portuguese oysters. The question is being addressed as to, what makes Hong Kong oyster shells so strong? It is hypothesized that the OMPs of Hong Kong oysters are significantly different in terms of both quality and quantity from Portuguese oyster shells. This hypothesis was tested and addressed using an interdisciplinary approach. Firstly, we have examined the OMPs of the two oyster species (both soluble and insoluble fractions) using 1-DE and 2-DE gel electrophoresis. This conventional quantitative proteomics approach helped us to understand why Hong Kong oyster shells are stronger. The aim in the second part of this research is to investigate the qualitative differences between the OMP samples of the two oyster species shells. This was accomplished using a 1-DE LC–MS/MS proteomics approach. Using these two proteomic approaches, we have addressed whether there is any significant difference in OMPs between the two species. Using the observed OMPs difference and the comparative OMP proteome structure analysis, this study demonstrated why Hong Kong oyster shells are relatively stronger than Portuguese oyster shells. This study, however, did not examine the mechanical properties of these two species shells and also did not attempt to correlate the observed expression pattern of the OMPs with their shell mechanical properties. The primary objective of this research was to comparatively characterize the organic matrix proteins of mechanically stronger as well as weaker oyster shells. This OMP quantitative and comparative analysis, using two different proteomics approaches, could help us gain new insights into the mechanisms that determine shell mechanical properties in oyster shells.

Methodology:
Study materials
The Hong Kong adult oysters (*C. hongkongensis*) and the Portuguese adult oysters *C. angulata* were respectively collected from Lau Fu Shan (LFS) located on the north-west coast of Hong Kong and 26° 05' 53.36'' N 119° 47' 45.81'' W in Fujian, China. Similarly sized (about 15 cm in length) and aged (about a year) shells of these two species were thoroughly washed and adhered epifauna was removed for the following analysis.

Organic matrix protein (OMP) extraction
About 10 oyster shells from each species of *C. hongkongensis* and *C. angulata* were rinsed with distilled water and dried at room temperature for 48 hours. The shells were then broken into small pieces, prior to which, any encrusting biofouling organisms were carefully removed from the shell surface. The broken pieces of shell were freeze dried using a freeze dryer machine (Martin Christ; Alpha 1-4 LSC Plus) then crushed into a powder of less than 200µm using a grinder (Mill, Disc Type/ N.V. Tema/ SiebTechnik/ TS100) to obtain a uniform powder to facilitate easy extraction by solvents. The extraction of protein was done in duplicate (2 x 30 gm of powder) in order to check the reproducibility of the results. The crushed shell powder was bleached using commercial bleach (20%) NaOCl, Colorox™), bleach was removed by repeatedly rinsing with distilled water

and the powder was freeze dried before adding pre-chilled 50% acetic acid (20 mL/gm of shell powder) for decalcification at 4°C for 16 hours. The resulting dark-brown suspension was dialyzed against 10% acetic acid (3 volume x 3 changes), 5% acetic acid (3 volume x 3 changes) and distilled water (3 volume x 3 changes) at 4°C for 3 days using SE membrane (Spectra/Por 6 dialysis membrane, mw cut-off 3000; Spectrum Europe, Breda, The Netherlands). The solution was centrifuged (HisacCR 22G; High-Speed Refrigerated Centrifuge) at 14,000 x g for 1 hour at 4°C to separate the ASM fraction as a supernatant and the AIM fraction as a pellet in plastic bottles (Amicon, 200 ml) (Mann and Jackson, 2014). Later, a freeze dryer was used to reduce the volume of the ASM fraction of the OMP and dry off the AIM pellet. The laemmli sample buffer [1M Tris pH 6.8, glycerol (87%), SDS (10%), DTT, milli Q water] was added to the ASM and the AIM pellet to dissolve the proteins in the powder formed from the freeze drying process and it was then sonicated (Banson Sonifier 150) using 3 cycles of 10 sec to properly mix and dissolve the proteins.

Protein precipitation and quantification
Trichloroacetic acid (TCA) and acetone precipitation was performed to clean and precipitate proteins. One volume of 60% (wt/vol) TCA was added to five volumes of acid soluble matrix (ASM) samples and 1 ml 60% (wt/vol) TCA was added to AIM pellets. Both the fractions were incubated at 4°C overnight, centrifuged at 10,000 x g at 4°C for 30 min, washed twice with ice-cold acetone at 4°C for 15 min. and centrifuged at 10,000 x g at 4°C for 30 min [6]. The protein pellets were again suspended in a laemmli sample buffer [1M Tris pH 6.8, glycerol (87%), SDS (10%), DTT, milli Q water] for shell proteins quantification using the Bradford-based method by Bio-Rad RC-DC Protein Assay Kit.

Protein separation by one dimensional gel electrophoresis (1-DE)
40 µg of shell protein samples were deglycosylated using a set of enzymes kit (New England BioLabs) **[6].** The deglycosylated mixture was further complemented with 1% bromophenol blue, β-mercaptoethanol, heated to 95°C for 5 min., 40 µg of protein from each was loaded in 10% SDS-PAGE Criterion Tris-HCl gels (Bio-rad), run at room temperature at 80 V for the first 15 min., then at 150 V for the next 40 min., stained with the coomassie brilliant blue G-250 (CBB) and gels were scanned at an optical resolution of 400 dpi using the GS-800 densitometer (Bio-Rad, Hercules, CA, USA).

Protein separation by two dimensional gel electrophoresis (2-DE)
Protein fractions were cleaned using a ReadyPrep Protein 2-DE purification kit (BioRad), 2-DE rehydration/sample buffer was added to the protein the quantification using The RC DC Protein Assay. Precast 11 cm. linear pH 3-10 immobilized pH gradient (IPG) strips (BIO-RAD Laboratories Inc) were re-hydrated for 16 hours at 50 V (20°C) with 250 microL buffer containing 60 ug of ASM and AIM of both *C. hongkongensis* and *C. angulata* in 7 M urea, 2 M thiourea, 2% (w/w) Chaps, 20 mM dithiothreitol, 0.2% ampholytes and 1% bromophenol blue. After this step, IEF was carried out at 250 V for 15 min, followed by 8000 V for 2 and half

hours and 8000 V until 40,000 Vh was reached. The IPG strips were then transferred for 20 min into 2 mL of equilibration buffer (6 M urea, 2% SDS, 1.5 M Tris/HCl pH 8.8, 50% glycerol) containing 130 mM dithiothreitol and 20 min into the same buffer containing 135 mM iodoacetamide. Strips were then rinsed in 25 mM Tris, 192 mM glycine and 0.1% SDS, placed on top of precast 12.5 % Bio Rad gels and sealed in place with an overlay solution of 0.5% agarose (w/v). Immediately after, electrophoresis was performed at 80 V for the first 15 min and then at 150 V for 40 min. [7]. Gels were run in triplicates for reproducibility, stored in a fixing solution overnight and stained using the vorum silver method [8] and scanned at an optical resolution of 400 dpi using the GS-800 densitometer (Bio-Rad, Hercules, CA, USA) for representative images.

Proteome analysis
PDQuest software (ver. 8.0; Bio-Rad) was used to analyze and compare the gels, which models protein spots mathematically as a three-dimensional Gaussian distribution and determines the maximum absorption after correcting the raw image and subtracting the background. To ensure proper detection of spots, automatic spot detection in every gel was visually inspected. For this compatible gel picture files were added to the PDQuest platform. Replicate gel images were grouped and named accordingly. By choosing spot centres, the gaussian model was selected with filter sensitivity. Now a master gel was chosen from all the gels based on the best manual visualization. Gel spot intensities were normalized using total-density values.

Thereafter, gels were grouped by consensus for analysis where spots were cross checked manually in PDQuest generated results. Consistently detected spots across all the gels were used for subsequent analysis. For a spots oriented comparison between *C. hongkongensis* and *C. angulata* the spot intensities dataset was analyzed using a student's t-test. Spots that displayed statistical difference (p<0.05) and with 1.25 or greater fold changes (*C. hongkongensis / C. angulata*) were considered differentially expressed (PDQuest™ 8.0 2D analysis software quick guide BioRad Laboratories 2005).

LC-MS/MS analysis of OMPs
The eight groups of protein bands from the gel were manually excised and followed by in-gel trypsin digestion. The gel pieces were washed twice for 15 min each with milliQ water, twice with H_2O/ACN (1:1 v/v) and were then placed in 100% ACN. The gel pieces were dried in a centrifugal SpeedVac before adding 1:50 sequencing grade trypsin (Promega) in 20 mM NH_4HCO_3 buffer. The gel pieces were covered with the buffer solution and digestion was allowed for 16 hours at 37° C. The peptides were extracted using several volumes of an H_2O/ACN/trifluoroacetic acid mixture (80:20:1). These fractions were dried in a vacuum centrifuge and subjected to analysis on a TripleTOF 5600 mass spectrometer (AB SCIEX) in order to generate raw wiff files to further identify proteins.

Trypsin digested fractions was dissolved in water and subjected to a reversed phase nano-LC-MS/MS consisting of a nano pump equipped with a 10-well-plate auto-sampler (Agilent

Technologies, Wilmington, DE, USA) coupled with a TripleTOF 5600 system (AB SCIEX, Concord, ON, Canada) fitted with a Nanospray III source (AB SCIEX, Concord, ON, Canada). The isolated peptide elution was applied with a 5-35% ACN gradient in 0.5% formic acid during 60 min gradient after loading on a 5 cm. reverse-phase C18 trap column. The following MS settings were used for MS/MS analysis of eluted peptides: ion spray voltage, 2.8 kV; curtain gas, 20 psi; nebulizer gas, 6 psi; interface heater temperature, 125°C. For IDA, full scans were acquired within 250 ms over the range m/z 400-1250. The eluted peptides were scanned by MS/MS, for which, the 20 most abundant peaks which exceeded 125 counts per second and carried a charge between +2 to +5 in the range m/z 100-1500, were manually selected. The exclusion time for MS/MS analysis of the acquired ions was set at 20 s. The raw data acquired from MS/MS analysis was examined using the Paragon algorithm in Protein Pilot 4.5 software (Applied Biosystems, Framingham, MA, USA).

Table 1: Comparison of protein expression in the acid soluble matrix (ASM) for both *C. hongkongensis* and *C. angulata* is given. Fold changes were obtained by dividing the average mean of normalized spot intensity of triplicate gels of *C. hongkongensis* to *C. angulata*. A student's t-test (p-value <0.05) was applied to determine whether the differentially expressed proteins had a fold change equal or greater than 1.25.

	ASM Spot number		Fold changes (HK/ANG)
PDQuest	**manually assigned to HK**	**manually assigned to ANG**	
SSP301	HS1	AS1	**1.69**
SSP1306	HS2	AS2	**1.33**
SSP2505	HS3	AS3	**1.29**
SSP2506	HS4	AS4	<1.25
SSP4407	HS5	AS5	**1.34**
SSP5404	HS6	AS6	**1.53**
SSP5502	HS7	AS7	<1.25
SSP6502	HS8	AS8	<1.25
SSP6503	HS9	AS9	<1.25
SSP6602	HS10	AS10	**2.59**
SSP6604	HS11	AS11	<1.25

Shell proteome databases for protein identification
In the process of workflow of shell proteomics, the raw wiff files obtained from mass spectrometer were run using Protein Pilot™ software to identify proteins against all the previously profiled shell proteome of the gastropods and molluscs which includes 42 shell proteins of *Pinctada maxima*, 78 shell proteins of *P.*

ISSN 0973-2063 (online) 0973-8894 (print)

268

Bioinformation 12(5): 266-278 (2016)

©2016

BIOMEDICAL
INFORMATICS

margaritifera, 253 shell proteins of *C. gigas* **[9]**, 94 shell proteins of *Haliotis asinine*, 63 shell proteins of *L. giganteum*, 553 shell proteins of *C. nemoralis* [10] and 36 shell proteins of *Stylophora pistillata* [6]. BLASTp based comparisons of the *C. hongkongensis* and *C. angulata* were made against the above mentioned eight previously published calcifying marine animals (gastropods and molluscs) shell proteomes **[11].**

Circos ideograms
The NCBI BLAST comparison result based text files and .conf files were provided to Circos using custom Perl scripts in order to generate two ideograms for both species analysis. Circos is offline software which generates publishable ideograms that allow for ease of comprehension and helps visualize the extensive analysis. In order to find out the similarity between our species and other species for protein sequences, sequence alignment was applied on our sequences and the 8 other species dataset, using NCBI BLASTp software **[12]**. An e-value 1e-6 was used as a cutoff to reserve those most likely results.

Table 2: Comparison of protein expression in the acid insoluble matrix (AIM) of both C. *hongkongensis* and *C. angulata* is given. Fold changes were obtained by dividing the average mean of normalized spot intensity of triplicate gels of *C. hongkongensis* to *C. angulata*. A student's t-test (p-value <0.05) was applied to determine whether the differentially expressed proteins had a fold change equal or greater than 1.25.

AIM Spot number			Fold changes (HK/ANG)
PDQuest	manually assigned to HK	manually assigned to ANG	
SSP2202	HI1	AI1	**1.38**
SSP3404	HI2	AI2	<1.25
SSP3405	HI3	AI3	**1.84**
SSP4305	HI4	AI4	**1.84**
SSP4403	HI5	AI5	**1.35**
SSP5303	HI6	AI6	**1.26**
SSP5304	HI7	AI7	**1.47**
SSP6605	HI8	AI8	<1.25
SSP6606	HI9	AI9	**1.29**
SSP7604	HI10	AI10	**2.24**

Based on this, a Circos figure was drawn to show the similarities between our sample proteins and the proteins in other species, using the software Circos-0.67-7. A full length of karyotype file was created to set the position around the Circos for each sequence, based on the total sequence number. In this way, each

outer line on the circle represents one protein sequence, and the width of the line represents the length of the sequence. The similarity between sequences was shown by inner lines. Each inner line connected two sequence lines. Colors of inner lines were used to indicate the level of global similarity, with classification based on the four quartiles. The green lines are in the top quartile (4th quartile), which means they have highest global similarity. And the blue, yellow and red lines are in 3rd, 2nd and lowest quartile, separately **[13].**
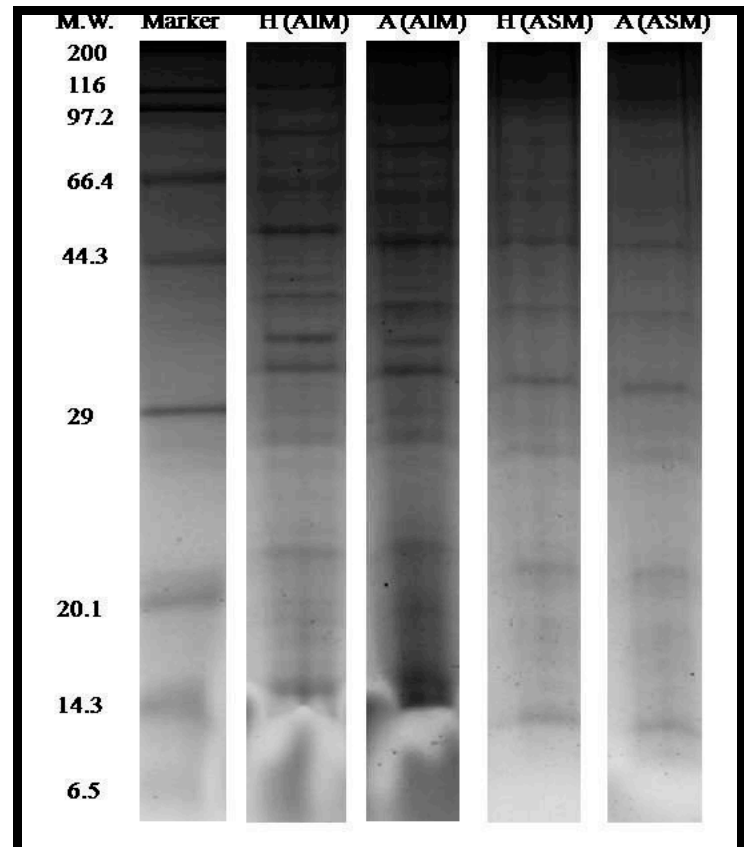
Figure 1: One-dimensional gel electrophoresis (1-DE SDS PAGE) results of the shell OMPs analysis. The first lane represents the marker bands while the second, third, fourth and fifth lanes of H (AIM), A (AIM), H (ASM) and A (ASM) represent the shell proteins bands where H: C. *hongkongensis* and A: C. *angulata*.

Results:
One dimensional gel (1-DE/SDS-PAGE)
10% SDS-PAGE (1-DE gel) of shell proteins of Hong Kong oysters (*C. hongkongensis*) and Portuguese oysters (*C. angulata*) with the AIM and ASM fractions **(Figure 1)** depict the observed proteins band pattern including the duplicate run (**Figure 9**). Protein bands were seen to be properly resolved to a sufficient enough extent especially considering that shell proteins are much harder to be visualized than tissue proteins. Use of the deglycosylation enzyme mixture, in particular to get rid of additional sugar chains, was positively found to be helpful in properly visualizing protein bands on the gel. Analysis of the two fractions of OMPs from the two species, using the 1-DE revealed both species and

fraction differences in the detection of OMPs. The Hong Kong species appears to have a slightly higher number of OMPs in both ASM and AIM fractions than those found in Portuguese oyster shells. It is also evident that, ASM fractions of both species have a higher number of OMPS than those found in the AIM. Nevertheless, the separation patterns using 1-DE gels generally failed to distinctively separate both the species and fractions due inherent limitations of 1-DE. Therefore, it was decided to further investigate these results using 2-DE.

Two dimensional gels
2-DE gel electrophoresis of shell proteins of Hong Kong oysters (*C. hongkongensis*) and Portuguese oysters (*C. angulata*) with AIM and ASM fractions depict the observed protein spots **(Figures 2 and 3)** in the triplicate runs to see the reproducibility. Other repeats of the triplicate runs **(Figures 10, 11, 12 & 13)** are mentioned.

**OMPs analysis using a proteomics approach**
Since equal amounts of protein (60 micro-grams) were loaded onto all the gels irrespective of the replicates and treatments, further qualitative comparison could be performed. Protein spots of Hong Kong *C. hongkongensis* and Portuguese *C. angulata* oysters were analyzed using PDQuest software to study the proteins expression **(Figures 2 and 3)**.
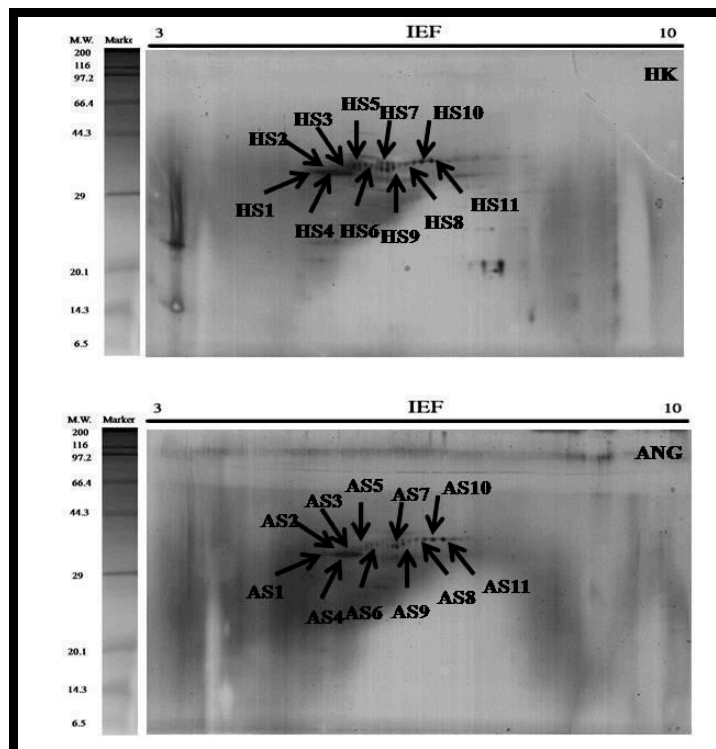
Figure 2: 2-DE Gel (A) The total protein extracts (60 µg) were separated on 11 cm. linear IPG strips (pH/pI 3-10) followed by 12.5% polyacrylamide gel electrophoresis. Gels were stained with silver staining to visualize protein spots & PDQuest spot analysis of the acid soluble matrix (ASM) of *C. hongkongensis* (HS) and *C. angulata* (AS).

Exactly 10-11 numbers of spots were found to be consistently expressed in all the gels which are why they are comparable for further conclusions. Protein spots of different characteristics were seen as they were diversely brighter, fainter, smaller and larger.

Each bar represents the normalized spot intensity obtained from biological triplicate gels (± S.D.) of *C. hongkongensis* and *C. angulata* **(Figures 4 and 5)**. Briefly, each gel was compared against one another to generate a normalization factor, by which the spot volumes in a gel were then normalized according to the respective factor. The normalized volume of each spot was exported to the SPSS statistical software (ver. 16.0; IBM) and compared between the *C. hongkongensis* and the *C. angulata* groups of ASM and AIM fractions. The spot analyses in this study assumed normal distribution of spot volumes in replicate gels within each group (*C. hongkongensis* or *C. angulata*) **(Tables 1 and 2)**. This is a common approach adopted in many 2-DE based proteomics studies **[14].**
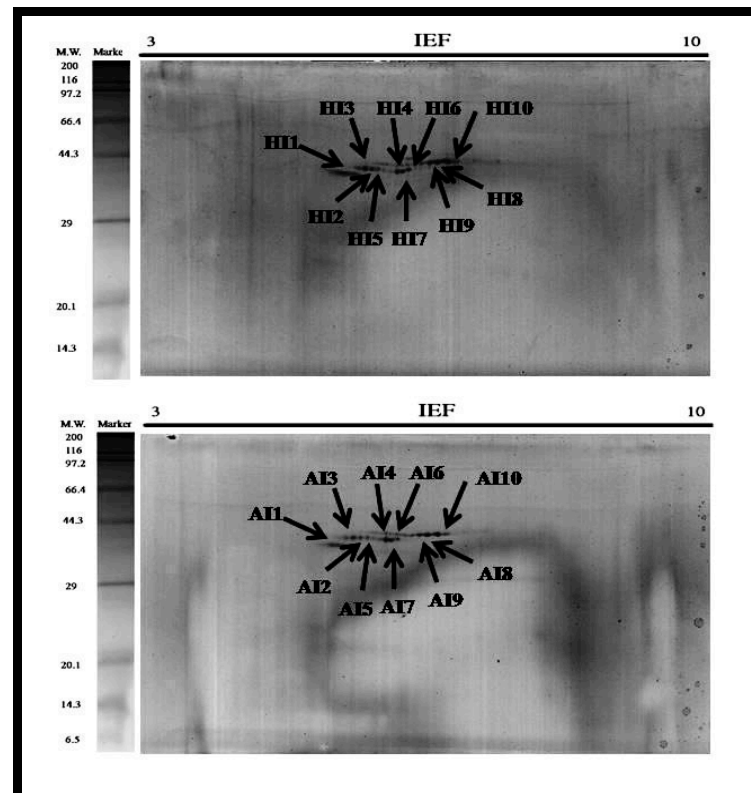
Figure 3: 2-DE Gel (B) The total protein extracts (60 µg) were separated on 11 cm. linear IPG strips (pH/pI 3-10) followed by 12.5% polyacrylamide gel electrophoresis. Gels were stained with silver staining to visualize protein spots & PDQuest spot analysis of the acid insoluble matrix (AIM) of the *C. hongkongensis* (HI) and *C. angulata* (AI).
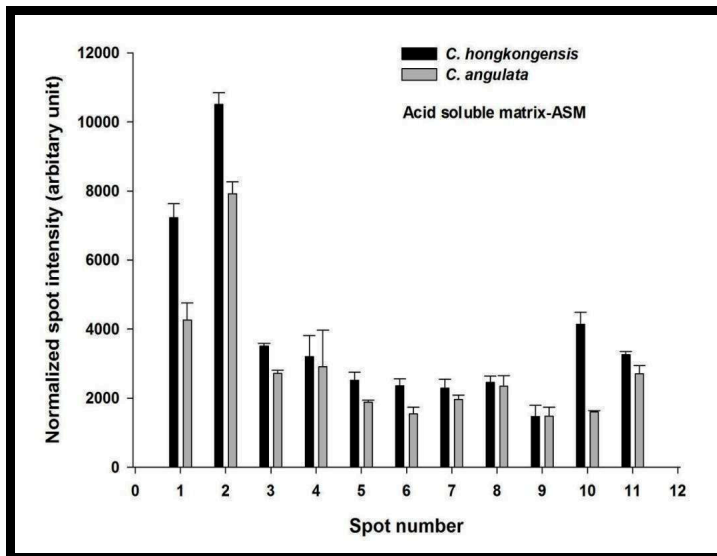
ISSN 0973-2063 (online) 0973-8894 (print)

270

Bioinformation 12(5): 266-278 (2016)

©2016

BIOMEDICAL
INFORMATICS

Figure 4: PDQuest spot intensity vs spot number comparison between the ASM of *C. hongkongensis* and *C. angulata*
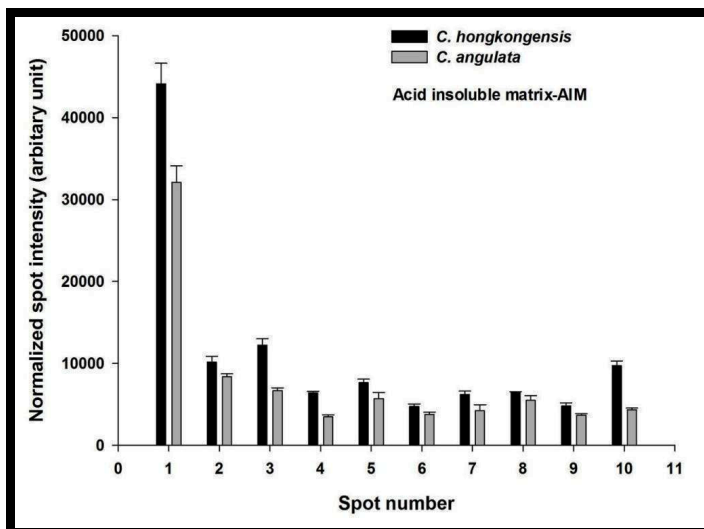

Figure 5: PDQuest spot intensity vs spot number comparison between the AIM of *C. hongkongensis* and *C. angulata*

Organic matrix proteins (OMP) identification

In total 42 of *C. hongkongensis* and 37 of *C. angulata* shell proteins were identified after in-gel digestion method from SDS-PAGE with Triple TOF mass spectrometry using Protein Pilot™ software. Out of the 42 OMP of *C. hongkongensis*, the AIM consists of 25, the ASM consists of 10, and 7 are common in both fractions whereas out of 37 organic matrix proteins of *C. angulata*, the AIM consists of 26, the ASM consists of 10, and 4 are common to both fractions **(Figure 6)**. The proteins role in biological processes and their molecular functions were retrieved using UniProt online software. **Tables 3, 4, 5 & 6** have the list of accession codes, protein names, number of peptides matching the derived sequence against the theoretical sequence for the confirmation of identification of proteins with or more than 95% threshold,

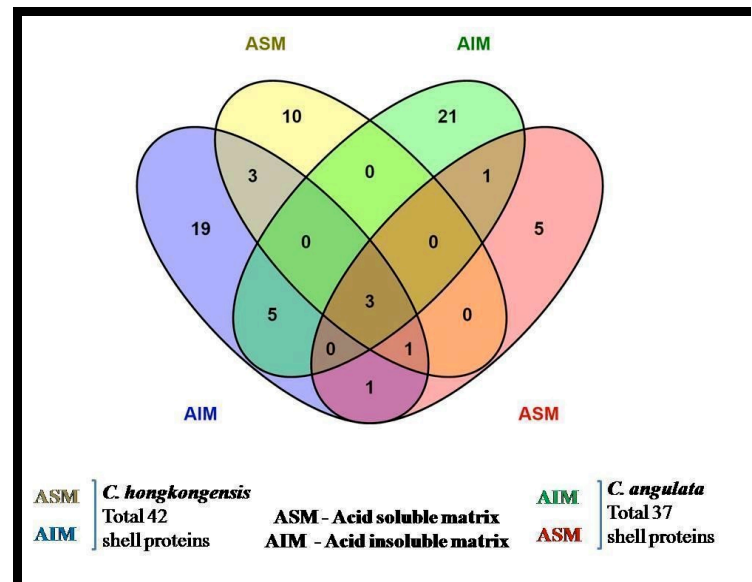biological processes and molecular functions of a few of the important proteins.

Figure 6: Venn diagram representation of the first identified shell proteins of *C. hongkongensis* and *C. angulata*.

Circos BLASTp analysis

The protein wise BLASTp comparison of shell proteome was recently done using Circos ideograms for the first time **[10]**. A clearly different pattern of significant similarities was found **(Figures 7 and 8)** in the first identified shell proteome of both the species *C. hongkongensis* and *C. angulata* when the BLASTp results were compared against the set of gastropods and molluscs shell proteomes. For 39 HK sequences, 16 of them were found to have 56 similar sequences in 8 other species. And for 37 ANG sequences, 20 of them find 84 similar sequences. The global similarity for each similar sequence pair was calculated, and these pairs were classified into 4 quartiles. Proteins, for example, Collagen alpha-1 (IV) a chain of *H. asinina* and Wnt inhibitory factor 1 of *P. margaritifera* and *P. maxima*, Myosin essential light chain, striated adductor muscle, Heat shock 70 kDa protein 12A and Enolase of *S. purpuratus*, Kielin/chordin-like protein of *L. gigantea*, Tripartite motif-containing protein 3, Paramyosin and Spore cortex-lytic enzyme of *C. gigas* are uniquely matched against shell proteins of *C. angulata* within the threshold of global similarity of sequences ranging from the top quartile to the lowest quartile of global similarity. This is not the case in *C. hongkongensis*.

These findings suggest that the evolutionary conservations of the shell proteins of *C. hongkongensis* have been quite limited in diversity whereas in *C. angulata* they are more diversely related and conserved. The functions of distinctive proteins hint towards different shell mechanical properties and protein pathways involved in biomineralization mechanisms in both the species of oyster *C. hongkongensis* and *C. angulata*.

BIOMEDICAL INFORMATICS

©2016

Table 3: Total identified acid insoluble matrix (AIM) fraction of organic matrix proteins (OMP) of *Crassostrea hongkongensis* oyster shell

Accession Code	Protein	Peptides (95%)	Biological Processes	Molecular Functions
CGI_10024572	Actin	7	movement of cell or subcellular component, structural constituent of cytoskeleton	ATP binding
CGI_10013249	Poly [ADP-ribose] polymerase 4	4	DNA repair, cellular protein modification process	DNA binding, enzyme binding
Lotgi1_222542	Hypothetical protein LOTGIDRAFT_222542	3		
Lotgi1_175997	Hypothetical protein LOTGIDRAFT_175997	3		
Lotgi1_239271	Hypothetical protein LOTGIDRAFT_239271	3		
CGI_10024501	ATP synthase subunit alpha, mitochondrial	2	ATP hydrolysis coupled proton transport	transmembrane transporter activity
CGI_10012474	Elongation factor 1-alpha	2	biosynthesis of proteins	translation elongation factor activity
CGI_10010974	Glyceraldehyde-3-phosphate dehydrogenase	2	Glycolysis	Oxidoreductose
CGI_10008055	Histone H3	2		
CGI_10008087	Histone H4	2		
CGI_10015004	Malate dehydrogenase, mitochondrial	2		
CGI_10001959	Putative polysaccharide export protein wza	2		
CGI_10026091	Insulin receptor substrate 1	2		
Lotgi1_206617	Hypothetical protein LOTGIDRAFT_206617	2		
CGI_10022995	Cyclin-Y-like protein 1	1	regulation of cyclin-dependent protein serine/threonine kinase activity	cyclin
CGI_10027743	Dynein heavy chain 8, axonemal	1	ATP binding	Motor protein
CGI_10005287	GTPase IMAP family member 1	1		
CGI_10008084	Histone H2A	1		
CGI_10008057	Histone H2B.3	1		
CGI_10011765	Probable phosphoglycerate mutase	1	gluconeogenesis	phosphoglycerate mutase activity
CGI_10016028	Katanin p60 ATPase-containing subunit	1	protein localization	ATP binding
CGI_10010911	CAD protein	1	calcium-dependent cell-cell adhesion, organ regeneration	Ca ion, Zn ion, metal, calmodulin binding, Transferase
CGI_10014688	Vezatin	1	single organismal cell-cell adhesion	Developmental protein
CGI_10020348	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	1		
Lotgi1_209765	Hypothetical protein LOTGIDRAFT_209765	1		
Lotgi1_104289	Hypothetical protein LOTGIDRAFT_104289	1		
CGI_10015060	Structural maintenance of chromosomes protein 2	1	cell division	ATP, nucleotide binding
CGI_10010727	Estrogen sulfotransferase	1		
Lotgi1_231652	Hypothetical protein LOTGIDRAFT_231652	1		
Lotgi1_225558	Hypothetical protein LOTGIDRAFT_225558	1		
CGI_10021973	Mediator of RNA polymerase II transcription subunit 23	1		
Lotgi1_181153	Hypothetical protein LOTGIDRAFT_181153	1		

Table 4: Total identified acid soluble matrix (ASM) fraction of organic matrix proteins (OMP) of *Crassostrea hongkongensis* oyster shell

Accession Code	Protein	Peptides (95%)	Biological Processes	Molecular Functions
CGI_10010974	Glyceraldehyde-3-phosphate dehydrogenase	5		
CGI_10015060	Structural maintenance of chromosomes protein 2	4		
CGI_10008725	CGI_10008725 NA	3		
CGI_10022260	Linear gramicidin synthetase subunit D	3		
CGI_10015492	78 kDa glucose-regulated protein	1	maintenance of protein localization in endoplasmic reticulum	Ca ion binding
CGI_10010911	CAD protein	1	calcium-dependent cell-cell adhesion, organ regeneration	Ca ion, Zn ion, metal, calmodulin binding, Transferase
CGI_10018899	CGI_10018899 NA	1		
CGI_10022856	Eukaryotic translation initiation factor 4 gamma 3	1		
CGI_10010828	Kinesin-like protein KIF20A	1	transport of an ion, a molecule (metabolite, protein, etc) or an electron	transporter activity
CGI_10027677	Negative elongation factor E	1		
CGI_10011765	Probable phosphoglycerate mutase	1		
CGI_10015004	Malate dehydrogenase, mitochondrial	1		
CGI_10010021	Adenylate kinase 2, mitochondrial	1		
Lotgi1_222542	hypothetical protein LOTGIDRAFT_222542	1		
Lotgi1_225558	hypothetical protein LOTGIDRAFT_225558	1		
Lotgi1_173072	hypothetical protein LOTGIDRAFT_173072	1		
Lotgi1_207121	hypothetical protein LOTGIDRAFT_207121	1		

Table 5: Total identified acid insoluble matrix (AIM) fraction of organic matrix proteins (OMP) of *Crassostrea angulata* oyster shell

Accession Code	Protein	Peptides	Biological Processes	Molecular Functions

		(95%)		
CGI_10001653	Paramyosin	22	metabolic process	motor activity
CGI_10010974	Glyceraldehyde-3-phosphate dehydrogenase	4		
CGI_10005425	CGI_10005425 NA	3		
CGI_10018876	Actin, adductor muscle	2		
CGI_10024777	Hepatocyte growth factor receptor	2	organ regeneration	differentiation, proliferation
CGI_10008087	Histone H4	2		
CGI_10022997	CGI_10022997 NA	2		
CGI_10013249	Poly [ADP-ribose] polymerase 4	2		
CGI_10020413	Actin	2		
CGI_10003417	Heat shock protein 70 B2	1		
CGI_10025376	Histone H2A	1		
CGI_10028414	Kielin/chordin-like protein	1	Enhances bone morphogenetic protein (BMP) signaling	metal ion binding
CGI_10003589	Probable E3 ubiquitin-protein ligase HERC1	1		
CGI_10011765	Probable phosphoglycerate mutase	1		
CGI_10020665	Sonic hedgehog protein	1	anatomical structure development, anterior/posterior/dorsal/ventral pattern formation, organ formation, signal transduction	Ca ion, Zn ion, metal, Glycoprotein binding
CGI_10015060	Structural maintenance of chromosomes protein 2	1		
CGI_10017032	Transcriptional adapter 2-alpha	1		
CGI_10001464	Tripartite motif-containing protein 3	1	transport of an ion, a molecule (metabolite, protein, etc) or an electron	metal ion binding
CGI_10017544	Wnt inhibitory factor 1	1	signal transduction	developmental protein
CGI_10014529	CGI_10014529 NA	1		
CGI_10020348	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	1		
CGI_10018042	Collagen alpha-1(IV) chain	1	Angiogenesis	extracellular matrix structural constituent
CGI_10009635	CGI_10009635 NA	1		
Lotgi1_239242	hypothetical protein LOTGIDRAFT_239242	1		
Lotgi1_137344	hypothetical protein LOTGIDRAFT_137344, partial	1		
CGI_10025038	Spore cortex-lytic enzyme	1		
Lotgi1_194265	hypothetical protein LOTGIDRAFT_194265	1		
CGI_10018235	Myosin essential light chain, striated adductor muscle	1	calcium regulation	Ca ion binding
CGI_10002446	CGI_10002446 NA	1		
Lotgi1_227913	hypothetical protein LOTGIDRAFT_227913	1		

Table 6: Total identified acid soluble matrix (ASM) fraction of organic matrix proteins (OMP) of *Crassostrea angulata* oyster shell

Accession Code	Protein	Peptides (95%)	Biological Processes	Molecular Functions
CGI_10015060	Structural maintenance of chromosomes protein 2	5		
CGI_10028838	Coiled-coil domain-containing protein 123, mitochondrial	2	shell coil patterning	shell coiling
CGI_10010974	Glyceraldehyde-3-phosphate dehydrogenase	2		
CGI_10010415	Cysteine desulfurase, mitochondrial	1		
CGI_10022154	Enolase	1	glycolytic process	Mg ion and metal ion binding
CGI_10013525	Heat shock 70 kDa protein 12A	1		
CGI_10016028	Katanin p60 ATPase-containing subunit	1	protein localization	microtubule-severing ATPase activity
CGI_10015004	Malate dehydrogenase, mitochondrial	1		
CGI_10003608	Mitochondrial fission factor homolog B	1	mitochondrial fission	protein homodimerization activity
CGI_10011765	Probable phosphoglycerate mutase	1		
CGI_10020665	Sonic hedgehog	1		

Discussion
Comparative analysis of oyster shell proteins

Shell proteomes of various shell forming animals have been revealed in the past few years. Out of these shell proteomes, some are of *C. gigas* (pacific oyster) **[14],** *C. nemoralis* (snail) **[10],** *S. purpuratus* (sea urchin), *H. asinina* (abalone), *L. gigantea* (limpet), *A. millepora* (coral) **[15],** *P. maxima* and *P. margaritifera* (pearl oysters). Not only identification of the shell proteins of the above mentioned shell-forming animals has been done, various evolutionary and molecular studies have also been performed. Whereas shell proteins of *C. hongkongensis* (Hong Kong oyster) and *C. angulata* (Portuguese oyster) were unknown before

conducting this study. So, a gap has been filled by identifying the shell proteins of these two species.

Not only were the shell proteins of these two species *C. hongkongensis* and *C. angulata* identified, but these shell proteins were the first to be visualized using 1-DE and 2-DE gel electrophoresis gels. These two shells have different material properties and as it is known that shell proteins play a vital role in interacting with inorganic crystals during shell formation that give rise to the mechanical properties of the shell **[16]**. In order to study the difference in shell mechanical properties of both these species of *C. hongkongensis* and *C. angulata*, a comparative analysis was carried out on protein spots observed from 2-DE

BIOMEDICAL
INFORMATICS

©2016

gels. In the study, it was demonstrated that shell protein expression was significantly (student's t-test, p-value<0.05) higher in *C. hongkongensis* than in *C. angulata* and in total seventeen (17) spots were differentially expressed whereas fourteen (14) spots were seen to have 1.25 or greater fold changes using mean normalized spot intensity.

Furthermore, in order to understand the evolutionary conservation of shell proteins, involvement of shell proteins in calcification and bio-mineralization related biological processes, as well as molecular pathways, we conducted Circos based BLASTp analysis comparisons **[13]** with other shell forming proteomes for the two species of *C. hongkongensis* and *C. angulata* . This particular study gives us a partial indication of how the

shells of these two oysters' species have evolved with different patterns of shell proteins conservation. This also tells that the shell proteins of these two oysters' species are conserved in a different manner deriving from the other molluscs and the gastropods over the course of years in the past and therefore possibly explicit different shell properties.

This study attempted to extract, identify and compare the shell proteins of the two phylogenetically related Hong Kong and Portuguese oyster species *C. hongkongensis* and *C. angulata* and possibly understand the shell proteins role in the mechanical properties of both the oyster shells. The results obtained seem to fulfill the objectives of the study with a range of experiments conducted.
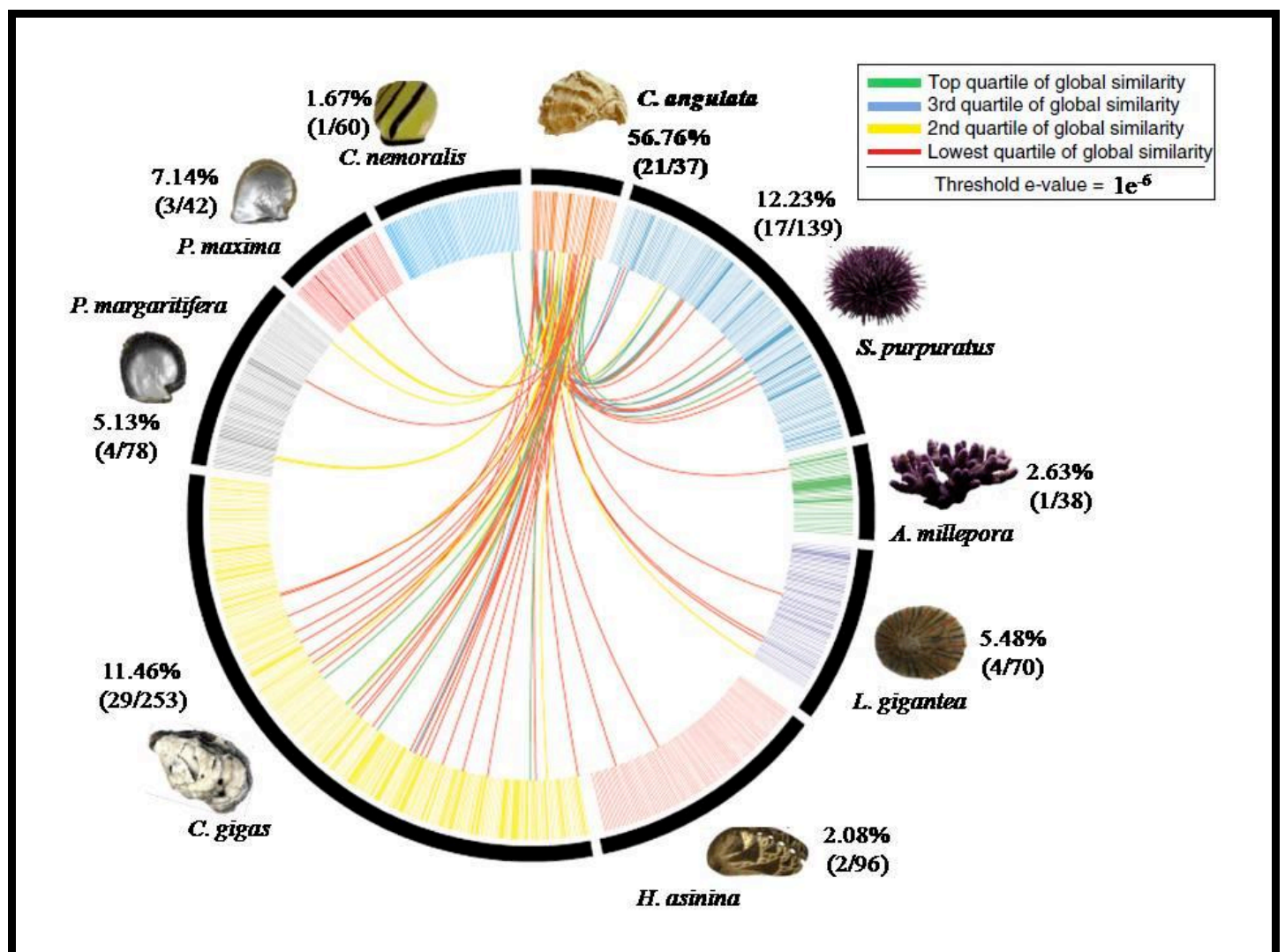

Figure 7: BLASTp comparisons of the *C. angulata* shell proteome against the shell proteomes derived from other shell forming molluscs and gastropods. Individual line spanning the ideogram connect proteins that share a significant similarity (e values< 1e-6). Red lines connect proteins with the lowest quartile of similarity (with a threshold of 1e-6) and green lines with the highest quartile of similarity. The percentage of each shell proteome that shared similarity with the *C. angulata* proteome is provided.
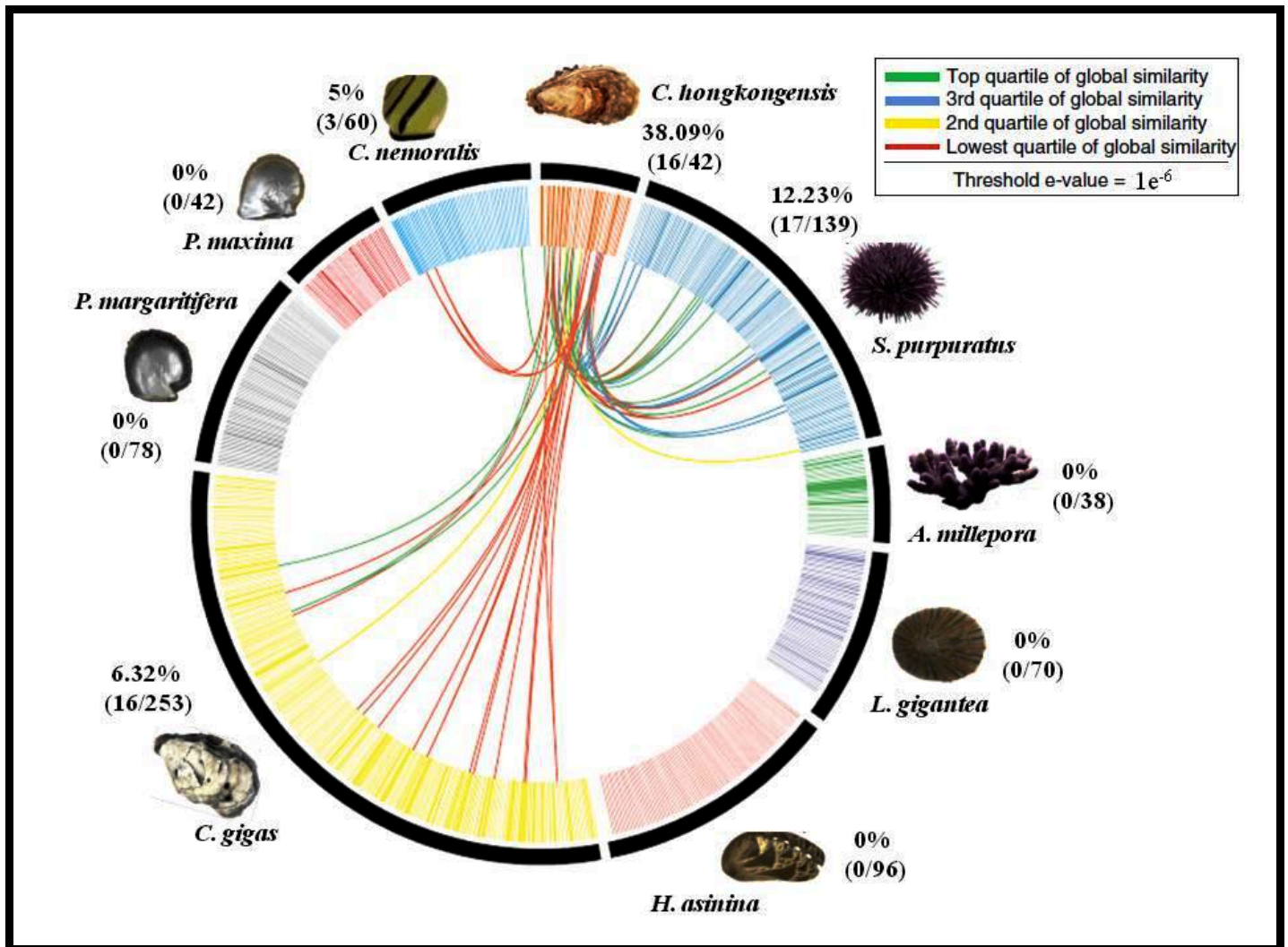
ISSN 0973-2063 (online) 0973-8894 (print)

274

Bioinformation 12(5): 266-278 (2016)

©2016

BIOMEDICAL
INFORMATICS

Figure 8: BLASTp comparisons of the *C. hongkongensis* **shell proteome against the shell proteomes derived from other shell forming molluscs and gastropods.** Individual lines spanning the ideogram connect proteins that share a significant similarity (e values< 1e-6). Red lines connect proteins with the lowest quartile of similarity (with a threshold of 1e-6) and green lines with the highest quartile of similarity. The percentage of each shell proteome that shared similarity with the *C. hongkongensis* proteome is provided.

Limitations of the present study and proposals for future research

Due to the pros and cons of the resolution of the technique in separating a highly complex blend of proteins, some proteins that are present in a low abundance might have been missed in the present study. By improving the resolution and separation of the complex proteome, a future study can further inculcate those overlooked proteins as well. A future study could split the pH range of IPG strips that used pH 3-10 into a set of two coupled range of pH 3.0-6.0 and pH 5.0-8.0 strips for improved resolution in 2-DE gels. Also the size of the gel used in this study can be increased to 17 cm. from 11 cm. for clearer spot/bands resolution in 1-DE and 2-DE gels **[9, 17].**

While 2DE-based is widely used for comparative proteome analysis, its application in non-model species was limited because it was not feasible to identify proteins using mass spectrometer and then validate with western blot analysis **[18, 19, 20, 21, 22, 23].** PDQuest software analyzed differentially expressed spots could further be submitted to the mass spectrometer for proteins identification. But due to time constraints of the study this was not performed, though it can possibly be done in the future by other researchers in order to get more information. A New Hybrid quadrupole Time-of-Flight Tandem Mass Spectrometer AB Sciex TripleTOF 5600 was used for identification of proteins from 1-DE gel bands. Using this, sufficiently good numbers were identified (approximately 50) but in the future it would be better to use the advanced Thermo Scientific Orbitrap (Elite/LTQ - Orbitrap Velos/ Q Exactive) to identify a higher number of

BIOMEDICAL
INFORMATICS

proteins because of the higher sensitivity and productivity of this machine **[24, 25].** An extensive shell proteome profiling of *C. hongkongensis* and *C. angulata* is needed in order to have a better comparison of shell proteins of the two species and BLASTp analysis with other shell forming molluscs and gastropods shell proteomes. It is believed that this study fulfilled the research objectives set down.
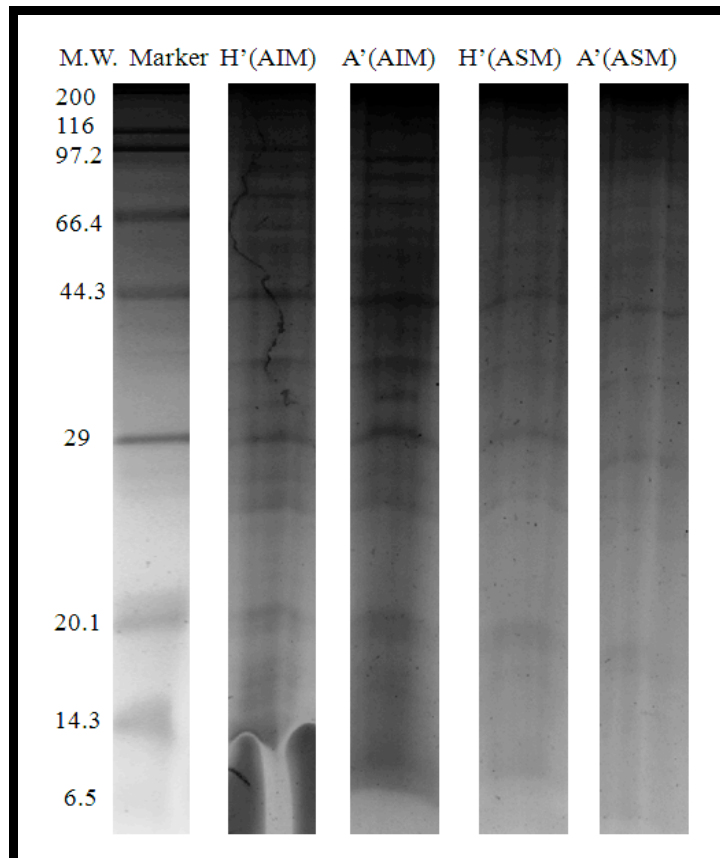

Figure 9: Duplicate gel of the one dimensional gel electrophoresis (1DE SDS PAGE) results of the shell OMPs analysis
The first lane represents the marker bands while the second, third, fourth and fifth lanes of H (AIM), A (AIM), H (ASM) and A (ASM) represent the shell proteins bands where H: C. hongkongensis and A: C. angulata.

Conclusion:
In total eleven spots in the ASM and ten spots in the AIM proteins of both species *C. hongkongensis* and *C. angulata* were consistently observed using PDQuest software analysis in biological triplicate gels. Out of the eleven ASM, seven protein spots and ten AIM, all ten protein spots significantly (student's t-test, p-value<0.05) show differentially expression. Fold changes between *C. hongkongensis* to *C. angulata* were calculated using differentially expressed spots normalized intensity. Out of these, six spots of the ASM i.e. PDQuest numbered SSP 301, 1306, 2505, 4407, 5404 & 6602 and eight spots of the AIM i.e. PDQuest numbered SSP 2202, 3405, 4305, 4403, 5303, 5304, 6606 & 7604

from both species were significantly (student's t-test, p-value < 0.05) seen expressed higher in *C. hongkongensis* than *C. angulata* with a fold change equal or greater than the 1.25 threshold. This result likely strengthens the hypothesis of one of the objectives of the study that: shell protein expression in *C. hongkongensis* might be higher than *C. angulata* which in turn could possibly give rise to harder and bigger (much more calcified) shells of *C. hongkongensis* Hong Kong oysters. Nonetheless, the current shell protein spots analysis demonstrated the significant difference among observed spots which stated that Hong Kong oysters shell proteins were differentially expressed higher than Portuguese oysters. Furthermore, these spots can be identified as protein names using mass spectrometry and their functions will likely help more in understanding the mechanism of biomineralization involved in these species.
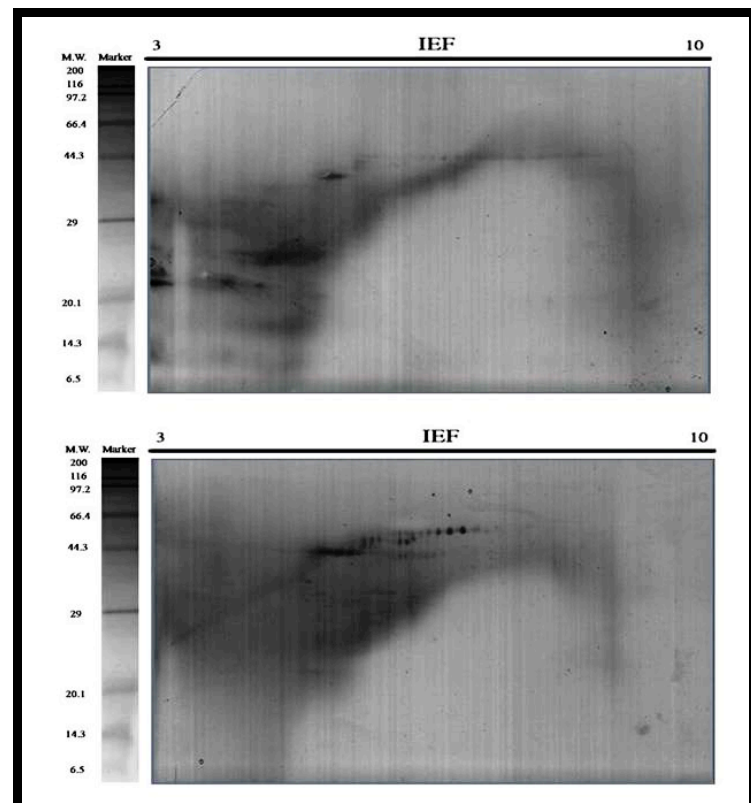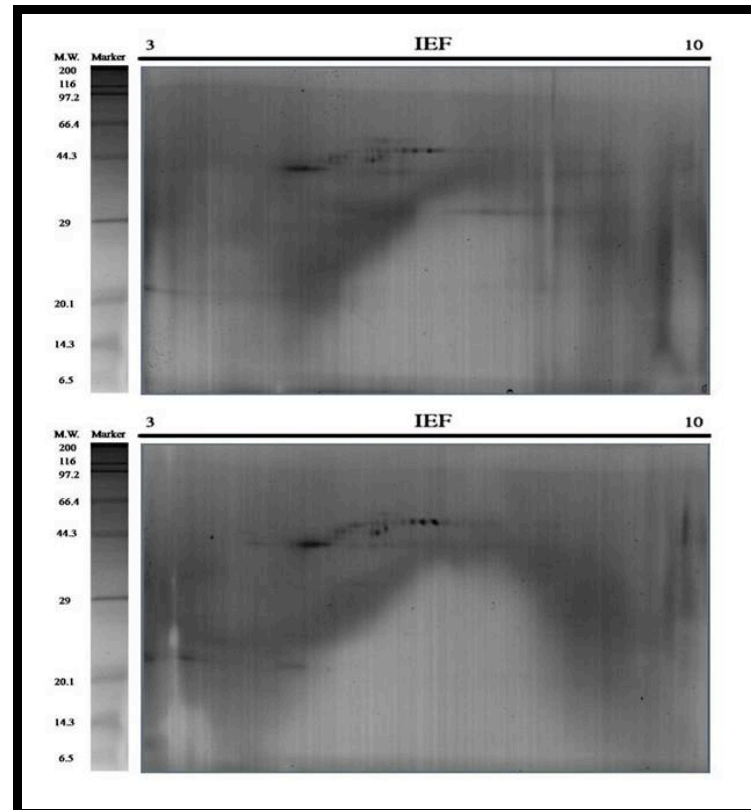

Figure 10: Repeat two of the triplicates 2-DE Gels. Representative 2D gels of the acid soluble matrix (ASM) and the acid insoluble matrix (AIM) of *C . hongkongensis* oyster shell proteins together with the protein marker. The total protein extracts (60 μg) were separated on 11 cm. linear IPG strips (pH/pI 310) followed by 12.5% polyacrylamide gel electrophoresis. Gels were stained with silver staining to visualize protein spots.

Diversity of shell mechanical properties is determined by interaction between inorganic crystals and the organic matrix, the oyster species, *C. hongkongensis* and *C. angulata* have served as a model for marine biologists and ecologists for years. Albeit, precise molecular mechanisms by which these species develop

ISSN 0973-2063 (online) 0973-8894 (print)

276

Bioinformation 12(5): 266-278 (2016)

©2016

BIOMEDICAL
INFORMATICS

diverse shells remains largely unknown. One way of addressing this would possibly be to compare the evolutionary conservation of shell proteins with other shell producing marine animals shell proteomes e.g. molluscs and gastropods. Since the shell proteins of *C. hongkongensis* and *C. angulata* were not known before, this study briefly profiled the proteomes for further identification using 1-DE gel bands with the help of Triple TOF 5600. Shell proteins of both species were visualized on 1-DE and 2-DE gels using conventional electrophoresis methods. To realize the objective broad level BLASTp comparisons of the *C. hongkongensis* and *C. angulata* shell proteome against the shell proteomes derived from mollusc, gastropods were made. Since all the shell proteomes databases primarily do contain proteins that have been isolated from the shells of their respective species (by mapping back to either RNASeq mantle transcriptomes or mainly genomes), it is therefore assumed to be somehow directly involved in shell formation.
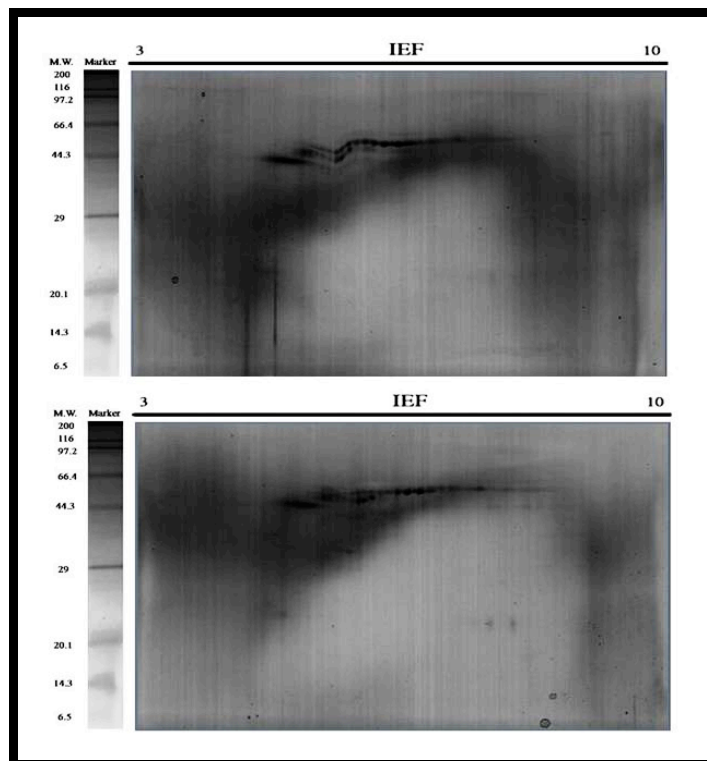
Figure 11:   Repeat three of the triplicates 2-DE Gels. Representative 2D gels of the acid soluble matrix (ASM) and the acid insoluble matrix (AIM) of *C . hongkongensis* oyster shell proteins together with the protein marker. The total protein extracts (60 μg) were separated on 11 cm. linear IPG strips (pH/pI 310) followed by 12.5% polyacrylamide gel electrophoresis. Gels were stained with silver staining to visualize protein spots.

Author contributions
AU carried out the whole experiment, obtained the results, performed the data analysis of the results and wrote the manuscript. VT helped to draft the manuscript and received the

research grant for the experiment. YT helped in the Circos analysis. All authors read and approved the manuscript.
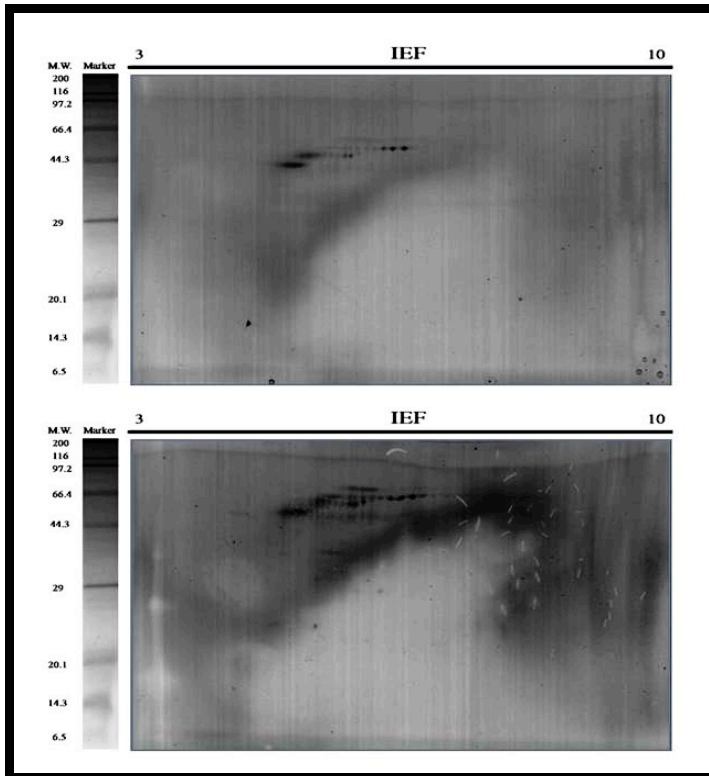

Figure 12:   Repeat two of the triplicates 2-DE Gels. Representative 2D gels of the acid soluble matrix (ASM) and the acid insoluble matrix (AIM) of *C . angulata* oyster shell proteins together with the protein marker. The total protein extracts (60 μg) were separated on 11 cm. linear IPG strips (pH/pI 310) followed by 12.5% polyacrylamide gel electrophoresis. Gels were stained with silver staining to visualize protein spots.

ISSN 0973-2063 (online) 0973-8894 (print)

277

Bioinformation 12(5): 266-278 (2016)

©2016

BIOMEDICAL
INFORMATICS

Figure 13: Repeat three of the triplicates 2-DE Gels. Representative 2D gels of the acid soluble matrix (ASM) and the acid insoluble matrix (AIM) of *C . angulata* oyster shell proteins together with the protein marker. The total protein extracts (60 µg) were separated on 11 cm. linear IPG strips (pH/pI 310) followed by 12.5% polyacrylamide gel electrophoresis. Gels were stained with silver staining to visualize protein spots.

References:
[1] Lam K & Morton B (2004) *Raffles Bulletin of Zoology* 52: 11.
[2] Yoon G-L *et al.* (2003) *Waste Management* 23: 825. [PMID: 14583245]
[3] Marin F *et al.* (2008) *Current Topics in Developmental Biology* 80: 209. [PMID: 17950376]
[4] Pérez-Huerta A & Cusack M (2008) *Zoology* 111: 9. [PMID: 17976966]
[5] Marin F *et al.* (2001) *Protein Expression and Purification* 23: 175. [PMID: 11570860]
[6] Drake JL *et al.* (2013) *Proceedings of the National Academy of Sciences* 110: 3788. [PMID: 23431140]
[7] Marie B *et al.* (2007) *FEBS Journal* 274: 2933. [PMID: 17488282]
[8] Mortz E *et al.* (2001) Proteomics 1: 1359. [PMID: 11922595]
[9] Zhang G *et al.* (2012) *Nature* 490: 49. [PMID: 22992520]
[10] Mann K & Jackson DJ (2014) *BMC Genomics* 15: 249. [PMID: 24684722]
[11] Jackson DJ *et al.* (2015) *Genome Biology and Evolution* 7:1349. [PMID: 25912046]
[12] Mcginnis S & Madden TL (2004) *Nucleic Acids Research* 32: 20. [PMID: 15215342]
[13] Krzywinski M *et al.* (2009) *Genome Research* **19:** 1639. [PMID: 19541911]
[14] Zhang Y *et al.* (2010) *Journal of Proteome Research* 9: 4851. [PMID: 20666481]
[15] Ramos-Silva P *et al.* (2013) *Molecular Biology and Evolution* 109. [PMID: 23765379]
[16] Weiner S (1979) Calcified Tissue International, 29: 163. [PMID: 116756]
[17] Zhang Y *et al.* (2010) *Journal of Proteome Research* 9: 3146. [PMID: 20397722]
[18] Diz AP *et al* (2012) *Proteomics* 12, 1949. [PMID: 22623270]
[19] Johnson S *et al.* (2007) *Mar Ecol Prog Ser*, 332, 247.
[20] Rabilloud T & Lelong C (2011) *Journal of Proteomics*, 74: 1829. [PMID: 21669304]
[21] Rodrigues PM *et al.* (2012) *Journal of Proteomics*, 75, 4325. [PMID: 22498885]
[22] Sanchez BC *et al.* (2011) *Environmental Toxicology and Chemistry* 30: 274. [PMID: 21072841]
[23] Slattery M *et al.* (2012) *Journal of Natural Products*, 75: 1833-1877. [PMID: 23009278]
[24] Jones KA *et al.* (2013) *Journal of Proteome Research* 12: 4351. [PMID: 24004147]
[25] Frese CK *et al.* (2011) *Journal of Proteome Research* 10: 2377. [PMID: 21413819]

ISSN 0973-2063 (online) 0973-8894 (print)

278

Bioinformation 12(5): 266-278 (2016)

©2016

BIOMEDICAL
INFORMATICS