

Mass spectrometry analysis and *in silico* prediction of allergenicity of peptides in tryptic hydrolysates of the proteins from *Ruditapes philippinarum*

Yue Yu^a, Hongwei Liu^a, Maolin Tu^b, Meiling Qiao^a, Zhenyu Wang^a, Ming Du^{a,*}

^a School of Food Science and Technology, National Engineering Research Center of Seafood, Dalian Polytechnic University, Dalian 116034, China.

^b Department of Food Science and Engineering, Harbin Institute of Technology, Harbin 150090, China;

*Corresponding author: Dr. Ming Du; E-mail: duming@dlpu.edu.cn

Postal address: School of Food Science and Technology, National Engineering Research Center of Seafood, Dalian Polytechnic University, Dalian 116034, China; Tel: +86-411-86332275; Fax: +86-411-86323262;

ABSTRACT

BACKGROUND: The *Ruditapes philippinarum* is nutrient-rich and widely-distributed, but little attentions had been paid on the identification and characterization of bioactive peptides in the bivalve. In this study, we evaluated the peptides of the *Ruditapes philippinarum* enzymolysed by trypsin using a combination of UPLC separation and a ESI-Q-TOF-MS/MS spectrometer, followed by data processing, sequence-similarity database searching. The potential allergenicity of the peptides was assessed *in silico*.

RESULT: The enzymolysis was performed under conditions as following, E:S 3:100

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/jsfa.8389

(w/w), pH 9.0, 45 °C for 4 h. After separation and detection, a Swiss-Prot database and a *Ruditapes philippinarum* sequence database were used, totally 966 unique peptides were identified by non-error tolerant database searching, 173 peptides matched 55 precursor proteins were highly conserved cytoskeleton protein. The rest 793 peptides were identified from the *Ruditapes philippinarum* sequence database. Results showed that 510 peptides were labeled as allergens and 31 peptides were potential allergens, 425 peptides were predicted to be non-allergens.

CONCLUSION: The abundant peptide information would contribute to further studying the structure and potential function of the *Ruditapes philippinarum*. And further vitro studies would be performed in order to demonstrate and ensure correct production of the hydrolysates if used in food industry of the *Ruditapes philippinarum*.

Keywords: *Ruditapes philippinarum*; UPLC-ESI-Q-TOF-MS/MS; Sequence-similarity; Peptides; *In silico* analysis

Introduction

The short-necked clam, *Ruditapes philippinarum*, is a member of the *Veneridae* family. It is an intertidal bivalve, which preferentially inhabits the inner bay.¹ The *R. philippinarum* (*Ruditapes philippinarum*) is also a nutrient-rich food which rich in minerals, vitamins, various of amino acids, as well as functional proteins and lipid.² At present, the *R. philippinarum* has been usually economically exploited as a food source, but little attentions had been paid on the identification and characterization of bioactive peptides in this bivalve. Some studies have investigated the structure and bioactivity of compounds from other shellfishes.³ Mytilan, isolated from the mantle of the mussel *Crenomytilus grayanus*, proved to show high immunomodulating activity.⁴ A polysaccharide (glycogen) generated from *Perna canaliculus*, has been found to exhibit antiinflammatory activity.⁵ A heparin-like substance separated from the marine clam *Anomalocardia brasiliana*, was able to bind antithrombin III, and was proved to have effective anticoagulant activity.⁶ However, at present, few studies about the analysis and identification of bioactive peptides from *R. philippinarum* proteins were reported, which are obstructive for the applications in many kinds of products of the *R. philippinarum*.

Peptidomics is a comprehensive qualitative and quantitative analysis technology of peptides within a tissue or a biological sample. Peptidomics is being applied in characterizing food-processing related proteolysis, identifying peptides in connection with disease and comprehensive mapping of food protein digestion and so on.⁷ With the development of mass spectrometry and related proteomics analyses, sequencing large numbers of peptides in complex mixtures became possible and efficient, and the

abundant peptide information would contribute to further studying of structure and potential function of proteins.

In the food industry, although the preparations of protein hydrolysates are common and usually use food-grade materials, proteins and peptides may produce allergic reactions after ingest in the human body due to their special amino acid sequence.⁸ Peptides can also cause anaphylactic reaction as peptides can be used as allergens in specific allergen immunotherapy.⁹ The terminology, food allergy, refers to an immunoreaction aimed at food and can be defined as an “adverse health effect caused by a specific immune reaction that repeated occur by reason of a given food”.¹⁰ Generally speaking, the allergenicity of protein consist in food must be evaluated *in vivo* and *in vitro*, the European Food Safety Authority (EFSA) prefers the use of prediction methods which achieved by *in silico* tools to evaluate potential allergens from food proteins.¹¹

In the present study, we performed an approach using UPLC combined with ESI-Q-TOF-MS/MS to separate and character the mixture of trypsin hydrolysate of the *R. philippinarum* based on peptidomics, the use of high resolution liquid phase separation and in the data independent acquisition mode, instead of selecting one precursor, an on overlapped m/z window is selected and all the precursor ions included in the window are co-fragmented outputting more fragment ions per peptide and more peptides per protein, which increases the protein identification score,¹² and a rigorous and hierarchical database searching was performed against two diverse database to high through screening the peptides, after acquiring hundreds of unique peptides matching

dozens of proteins with high confidence, some bioinformatics summarize were performed about the peptides and proteins of the *R. philippinarum*, combined with the enzymolysis condition, and the potential allergenicity was assessed *in silico*, which could provide useful information for developing biological activity or related products of the *R. philippinarum*.

Materials and methods

Materials and reagents

Fresh clams *R. philippinarum* (shell length of 3–4 cm, dry weight of 0.3–0.4 g) were collected from a commercial market (Dalian, China). Formic acid (FA) from Fluka (Switzerland), trypsin (EC 3.4.21.5) was purchased from Sigma (St. Louis, MO, USA).

Enzymatic hydrolysis

Prior to enzymatic hydrolysis, *R. philippinarum* was lyophilized and pulverized into a powder using a grinder, 2.5 g of the lyophilized *R. philippinarum* powder was added to 50 mL milli-Q water at 45 °C, after blending for 3 min, pH condition was adjusted to 9.0, then trypsin was added in the mixture with an enzyme to substrate ratio of 3:100 (w/w). The enzymatic hydrolysis reactions were performed for 4 h at a constant pH=9.0 adjusting by 0.1 M NaOH. Agitation was maintained at a constant of 60 rpm. At the end of hydrolysis, the mixture was immediate heated at 100 °C for 10 min to inactivate the enzyme. Finally, the enzymatic hydrolysates were rapidly cooled to 25 °C in an ice bath and centrifuged at 10000 g for 10 min. The supernatant was collected, lyophilized and stored at -30 °C.

DH is defined as the percent ratio of the number of peptide bonds ruptured during

the hydrolysis to the total number of peptide bonds in the substrate^{13,14} and can be calculated by following equations (1) and (2):

$$DH(\%) = \frac{BN_b}{M_p \alpha h_{tot}} \times 100 \quad (1)$$

$$\alpha = \frac{10pH - pK}{1 + 10pH - pK}, \quad pK = \frac{7.8 + 2400 * (298 - t)}{298(273.15 + t)} \quad (2)$$

In this formula, B is the amount of NaOH added to the substrate to keep the pH constant during the hydrolysis; N_b is the normality of NaOH, in this experiment was 0.1 M; α is the degree of dissociation of the α-NH₂ groups liberating during hydrolysis; M_p is the mass of protein (in grams), which was 1.25 g in this experiment (determined by kjeldahl method); h_{tot} is the total number of the peptide bonds which is assumed to be 7.5 meq.g; and pK is the dissociation value for the α-amino groups releasing during hydrolysis, valued 7.5.

Desalting by SPE column

Hydrolysate samples were redissolved in 0.1% FA-H₂O. The SPE column was activated by adding 3 mL of methyl alcohol following the equilibration with 3mL of 0.1% FA-H₂O. Then samples were loaded to the activated column followed by desalination with 3 mL of 0.1% FA-H₂O. Finally, samples were eluted with the mobile phase (80% methyl alcohol H₂O) of 1.5 mL, lyophilized, and stored at -30 °C.

Peptides separation by UPLC

Samples were redissolved in 0.1% FA-H₂O and filtered through a 0.22 μm syringe filter and injection volume was 15 μL. Separation of peptide was performed on a Dionex UltiMate3000 RSLC System (ThermoScientific, Waltham, MA, USA),

consisting of a vacuum degasser, an auto-sampler and a quaternary Surveyor pump at 20 °C. This instrument was equipped with a Phenomenex Luna C₁₈ column (3×150 mm, 3 µm).

Acidified water (0.1% formic acid, v/v) and ACN (with 0.1% formic acid, v/v) were used as mobile phases A and B, respectively. Gradient was programmed as following: 0 min, 10% B; 2 min, 10% B; 30 min, 35% B; 32 min, 35% B; 34 min, 80% B; 36 min, 80% B; 36.1 min 10% B, and finally, the initial condition was held for 4 min as a re-equilibration step. The flow rate was set at 0.40 mL/min throughout the gradient.

ESI-Q-TOF-MS/MS analysis

The UPLC system was coupled to Quadrupole-Time-of-Flight (microTOF-Q™, Bruker Daltonik GmbH, Bremen, Germany), an orthogonal accelerated Q-TOF mass spectrometer, equipped with an electrospray ionization source (ESI). Parameters for analysis were set using positive ion mode with spectra acquired over a mass range from 50 to 2200 m/z with the scan rate of 5 spectra/s. The optimum values of the ESI-MS parameters as following: capillary voltage, +4.5 kV; drying gas temperature, 200 °C; drying gas flow, 6.0 L/min; nebulizer gas pressure, 2.0 bar; Funnel 1RF, 400 Vpp, Funnel 2RF, 600 Vpp, Hexapole RF, 350 Vpp; transfer time, from 50 to 110 µs, and pre-pulse storage, 10 µs. Fragmentation of protonated molecular ions was conducted by auto MS/MS mode with the top 20 most intense ions were isolated. Each sample was analyzed in triplicate to eliminate technical errors.

Database searching by MASCOT

The raw data for peptide sequence was processed through Data Analysis 4.0

software (Bruker Daltonics, Bremen, Germany) and searched by using MASCOT (version 2.5.1.0, Matrix Science, London, UK) server (<http://www.expasy.ch/sprot/>). Since the *R. philippinarum* is an organism with unsequenced genomes, the protein database utilized in this study consisted in the combination of the Swiss-Prot protein database (546000 sequences released at July in 2014), and a *Ruditapes philippinarum* database (361 sequences, unpublished data) downloaded from National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>). The parameters used in this search were as follows: no restrictions on protein molecular weight or species of origin of analyzed proteins, trypsin was selected for protein cleavage allowing for two missed cleavage, and oxidation(M) was selected as modification, peptide mass tolerance and fragment mass tolerance were 10 ppm and 0.03 Da, respectively. Peptide charge including +1, +2, and +3, both instrument type CID and instrument type ETD choose ESI-QUAD-TOF. Identifications were validated by performing a decoy database search for the estimation of False Discovery Rate (FDR). Ions score cut-off and peptide rank cut-off was adjusted to 20 and 10, respectively. Ions score threshold for significant peptide IDs defined by user as 35, and a significance threshold was set as $p < 0.05$ and FDR of 1%. Protein assessment was determined using MASCOT protein scores, with identification confidence estimated by the coverage of protein sequence and the number of matched peptides. For each peptide, accept top hit compound only. In case of MS/MS spectra matching peptides from more than one identified protein, only the highest scoring peptide was accepted.

Tricine-SDS-PAGE

Electrophoresis was performed by using a BIO-RAD Mini-PROTEAN® 3 cell System (Bio-Rad Laboratories Ltd., Hertfordshire, UK) and compounds was separated by 16% separation gel and 4% spacer gel. Commercial molecular weight (MW) markers (3.4-100 kDa) were used as the standards.

Prediction of allergenicity *in silico*

Peptides identified were evaluated for potential allergenicity using an online tool, AlgPred, available at [Http://www.imtech.res.in/raghava/algpred/](http://www.imtech.res.in/raghava/algpred/)¹⁵ with two method mode were used, the SVM module based on amino acid composition prediction and the MEME/MAST motif prediction approach.¹⁶

Results and discussions

To identified proteins and peptides, in a typical identified experiment always perform as follow: sample protein are digested by a proteolytic enzyme (usually trypsin), then use liquid chromatography method to separate peptides in the mixture and emitted directly into a tandem mass spectrometer, the ESI-Q-TOF-MS/MS is more and more commonly used for its distinguished mass accuracy and resolution. Peptides are isolated and fragmented, and the generated spectra are performed a series of pre-processing then searched against a sequence database to identify the respective peptide sequences.¹⁷ This workflow is now widely applicable in most proteomics research, due to a combination of sturdy commercially available tandem mass spectrometers and well-established database searching software,¹⁸ such as Mascot, SEQUEST, and X!Tandem.

Database searching a collection of sequences with tandem mass spectrometry

information is performed as follow: protein in the collection is digested theoretically according to the specificity cleavage of the enzyme used in the experiment. The masses and MS/MS fragment of resulting peptides are calculated, then the theoretical tandem mass spectrum is established. The practical MS/MS spectrum is then compared with the theoretical MS/MS spectrum and given a score to represent the degree of correlation. This process is repeated for each protein in the sequence collection. Finally, the proteins and peptides are ranked according to the calculated scores.¹⁹ Severe matching immensely increases the specificity of database searches, contributes to overcome the innate lack of tandem mass spectra and enables convinced identification of proteins using spectra of a few fragmented peptide precursors.²⁰ However, there is a significant obstacle in the characterization of proteomes of species which genomes still unsequenced.^{21,22} In this case, we can combine mass spectrometry with sequence-similarity database searches to avert these limitations and pave the way to high throughput and efficient exploration of peptides, in spite of lack of available sequence resources.

Trypsin hydrolysis of *R. philippinarum* proteins

The whole hydrolysis process of digestion continued for 4 h, the degree of hydrolysis (DH) was investigated. As shown in Fig.1 of the hydrolysis curve, DH of trypsin attained 17.54%, the DH changed drastically in the first 50 min, then verge to smooth and steady gradually. This might attribute to the decrease of solubility of proteins and could interpreted by observation by Lv et al.(2017),²³ who demonstrated that the trypsin treatment to myosin could lead to increase of surface hydrophobicity by

exposing the concealed hydrophobic groups and unfolding natural α -helix, and the solubility has been indicated to be most importantly dependent on the balance of hydrophilic and hydrophobic forces of proteins, so the process of enzymolysis could lead to the increase of hydrophobicity and the decrease of solubility, then the DH turned steady relaxedly.

In addition, an electrophoresis separation using Tris-Tricine peptide gel was performed to have a rough idea of molecular weight distribution profile of peptides (proteins). Results were presented in Fig.2, the MW markers are in the left side which MW ranging between 3.4-100 kDa, the hydrolysate proved four obvious stripes which represent molecular weight about 103.72 kDa, 72.89 kDa, 41.66 kDa and 26.57 kDa, respectively. Since the trypsin hydrolysate was a mixture, the gel failed to separate absolutely and peptides or proteins which molecular weight less than 26 kDa.

Peptide identification

Fig.3 showed the total intensity chromatogram for *R. philippinarum* trypsin hydrolysate. Survey full-scale of the triplicate analyses showed excellent repeatability in elution times (data not shown). The chromatographic separation was a 40 min run. The 10–80% 90 min ACN (with 0.1%formic acid, v/v) gradient was used, but the elution effective in separation of peptides was 10-35%, a high concentration of ACN can forcefully elute the column residual. The better separation of peptides enabled the identification of the lower abundant peptides, which otherwise might be masked by the abundant peptides in the mixture.²⁴ We also attempted several elution processing which remain the ACN (with 0.1%formic acid, v/v) concentration but lengthen each gradient

(data not shown). It turn out that the optimized elution of peptides remain 10-35%, which could attribute to hydrophobicities and polarities of peptides, so that the UPLC condition is regard as effective and efficient.

After separation by the UPLC, the peptides were instilled to the mass spectrometry source through a pico tip needle for the analysis. In the mass spectrographic analysis, an alternant scanning approach was applied, where the peptide information was collected at low collision energy and peptide sequence information was gathered at higher collision energy. In this way, more fragment ions per peptide and more peptides per protein became detectable, which provides higher confidence protein identification. The spectra mostly included common peptide signals, but this did not impact the resolution capability of the setup since the peptides were separated over 30 min and then, analyzed in the mass spectrometer. Since the procedure was timed for throughput, sequence coverage and robustness rather than maximum sensitivity. In addition, Zhang et al. (2009)²⁵ has paid attention to the fact that the searching results should be ulteriorly confirmed, and the number and consecutiveness of the -y and -b ion series in the MS/MS spectra were taken into consideration.

Total 966 unique peptides were identified in this study, the similar peptide sequences were deleted from the result list only retain the highest peptide score entry, these peptides were summarized along with their amino acid residues in the Fig.4, the shortest peptide composed of seven amino acid residues was identified, in total 74. It is well known that bioactive peptides generated from food materials mostly consisted of 2–20 amino acids.²⁶ In the present study, the peptides with 2-20 amino acids form the

majority (82%). On the other hand, all identified unique peptides were classified by molecular mass and were studied to evaluate the hydrolysis process and the degree of fragmentation by the ion source (Fig.5), the MW of all peptides ranged from 700 to 7290 Da, the majority of peptides distributed in the MW range of 1000 to 1500 Da, only 5 corresponded to large proteins.

Protein Identification

Total 966 unique peptides were identified from 82 proteins by searching against two disparate databases, 173 peptides were identified with the Swiss-Prot database and 793 were the *R. philippinarum* database. A sequence similarity search strategy was applied for the protein identified because the genome of the *R. philippinarum* has not been sequenced. Out of total 82 unique proteins, 55 proteins were identified by matching the Swiss-Prot database, confident cross-species identification was performed. As shown in Table 1, the Swiss-Prot accession numbers along with the full names of the proteins, the source of the species, protein molecular mass, isoelectric points (pI), protein score, and peptides matched of each protein. These proteins were found in a wide range of pI and MW, the pI values ranged from 4.43 to 11.14. The MW of all proteins identified against Swiss-Prot database were ranged from 10 kDa to 223 kDa, the MW of the majority proteins was in the range 30-60 kDa (63.6%).

Although Mascot can only acquire peptides identical to sequences consist in database, its specificity in identification of cross-species matching is higher than other approach, such as using the peptide mass fingerprinting, which attributed to the fact that tandem mass spectra were rich in sequence specific fragment ions.²⁰ After the error-tolerant and

stringent database searching, the identified proteins were taken into consideration of conservatism and phylogenetic distance. Among 55 identified proteins against Swiss-Prot database, 41(74%) were animalia, 8 (15%) were plantae and 6 (11%) were protozoa. Only the protein of *Spisula sachalinensis* identified with protein entry at the order level (Veneroida), proteins of *Placopecten magellanicus*, *Mytilus californianus*, *Mytilus galloprovincialis* and *Argopecten irradians* at the class level (Bivalvia). Proteins of *Biomphalaria glabrata*, *Turbo cornutus* and *Haliotis diversicolor* at Phylum level (Mollusca). Although most of the proteins were cross-species and from distantly related species, these hits with high protein score that matched upon a MASCOT database were homologous proteins, and the proteins that did match across species in our research were highly conserved cytoskeleton proteins like actin, tropomyosin, histone, tubulin, myosin, paramyosin and coronin, which are known to be widely exist in organism. It is well known that actin is a major structural protein in all eucaryote cells. Histone, as basic components of nucleosome, is essential to chromatin structure and function. Myosin is of great importance in living muscle and constitutes >50% of myofibrillar proteins, Its molecule consists of two heavy chains, each of ~200 kDa, and four light chains, each of ~20 kDa.²⁷ Paramyosin, is present in thick myofilaments of invertebrate muscle, has been proved molecular weight of 98-101 kDa.²⁸ Tropomyosin is widely distributed in all cell types associated with actin as a fibrous molecule and about 40kDa.²⁹ Based on these findings, as shown in Fig.6, we could roughly judged the 200kDa band, 99kDa band and 44kDa band were highly possible to be myosin heavy chain, paramyosin and tropomyosin, these identified proteins are also high-abundant

proteins of the *R. philippinarum*, since the identification were to be expected and reliable.

Peptides identified by using the *R. philippinarum* database, as shown in Table 2 Provided the protein accession number of each protein in the database, along with its full name, and protein score which given by the Mascot, molecular mass of proteins and the number of matched peptides of each ptotein. The peptide sequence source were varied, peptides were identified from a type of troponin (Troponin C), glycoprotein (Complement C1q-like protein 1, Complement C1q-like protein 4), heat shock protein (Heat shock protein 22 isoform 1), a small cysteine-rich cationic proteins (Defensin), a family of proteins defined by their binding specificity for β -galactoside sugars (Tandem repeat galectin), multiple NADH dehydrogenese subunits and cytochrome oxidase subunits, a kind of lysozyme (C-type lysozyme 2).

Assessment of allergenicity of identified peptides *in vitro*

The potential allergenicity of peptides identified was assessed *in silico* using AlgPred with two typical modules, SVM and MEME/MAST. AlgPred utilizes an SVM which uses a radial basic function kernel, amino acid composition and dipeptide composition of each protein sequence were served as input vectors for this study.¹⁵ SVM module relay on amino acid composition prediction. SVM are supervised learning models applied for regression, classification and outlier detection. MEME is a tool for finding motifs in a set of related protein sequences. MAST is a tool for discovering biological sequence database for sequences which include at least one of a group of known motifs.¹⁵ AlgPred has a limitation that only those peptides longer than ten amino

acids in length can be analyzed, so that the data on potential allergenicity of shorter peptides were omitted. As a result, by using the SVM module which based on amino composition suggested that 510 peptides were labeled as allergens and 31 peptides were potential allergens, 425 peptides were predicted to be non-allergens. Numbers of peptides which predicted to be allergens and potential allergens just a few amino acid residues in length. These peptides can hardly maintain stable structures in solution,³⁰ decreasing the presence of potential conformational epitopes. Furthermore, using hydrolysates as food ingredients usually experience thermal treatment which can destroy conformational epitopes and influence the interaction with other food constituent.³¹ In addition, the presence of linear epitopes depends on the type of enzyme used and the degree of hydrolysis.³² For this purpose, peptides generated from the *R. philippinarum* using trypsin might potentially contain linear epitopes and maintain the allergenicity of their parent protein.

In addition, AlgPred based approach using MEME/MAST module turned out that none of the identified peptides predicted to be allergens, but it does not express that the identified peptides cannot cause anaphylases to humans after digestion, as the developers of AlgPred elucidated that the MEME/MAST module had a low specificity and could show a high percentage (nearly 66%) of wrong mission of allergenicity.¹⁶

Conclusion

In this study, total 966 unique peptides were identified from trypsin hydrolysate of the *R. philippinarum*, these peptides were matched with 82 unique proteins of various organisms, some even phylogenetically far from the *R. philippinarum*. The combined

using of UPLC separation assisted by mass spectrometric detection with ESI-Q-TOF-MS/MS spectrometer, and custom peptidomic database searching proved to be a useful tool to increase the representation of protein sequences in database and will contribute to characterize more proteins and peptides from more phylogenetically distant species. But peptides composed under seven amino acids could not be detected because of the limitations of the mass spectrometry approaches, and small peptides containing some hydrophobic amino acids can retained on classical C18 columns, small peptides composed of only hydrophilic amino acids are not. Further studies are needed to detected more peptides. *In silico* tools were used for predicting potential allergenicity of the identified peptides, the hydrolysate in this study was generated using the food-grade enzyme trypsin, however, a large number of peptides were predicted to be allergenic. Further studies *in vitro* and *in vivo* should be performed in order to demonstrate and ensure correct productions of this hydrolysate if used in food industry.

Acknowledgements

This study was financially supported by the National Natural Science Foundation of China (31371805), the National Science & Technology Pillar Program (2013BAD18B06-03).

References

1. Yang Z, Zhao Y, Yan H, Xu L, Ding G, Yu D, Sun Y. Isolation and purification of oligopeptides from *Ruditapes philippinarum* and its inhibition on the growth of DU-145 cells in vitro[J]. MOL MED REP, **11**: 1063-68(2015).
2. Lee SJ, Kim EK, Kim YS, Hwang JW, Lee KH, Choi DK, Kang H, Moon SH, Jeon

- BT, Park PJ. Purification and characterization of a nitric oxide inhibitory peptide from *Ruditapes philippinarum*[J]. FOOD CHEM TOXICOL, **50**: 1660-66(2012).
3. Kim EK, Hwang JW, Kim YS, Ahn CB, Jeon YJ, Kweon HJ, Bahk YY, Moon SH, Jeon BT, Park PJ. A novel bioactive peptide derived from enzymatic hydrolysis of *Ruditapes philippinarum*: Purification and investigation of its free-radical quenching potential[J]. PROCESS BIOCHEM, **48**: 325-30(2013).
4. Ovodova RG, Glazkova VE, Mikheyskaya LV, Molchanova VI, Isakov VV, Ovodov YS, Fernandez Molina LE. The structure of mytilan, a bioglycan-immunomodulator isolated from the mussel *Crenomytilus grayanus*[J]. Carbohydr Research, **223**: 221-6(1992).
5. Miller TE, Dodd J, Ormrod DJ, Geddes R. Anti-inflammatory activity of glycogen extracted from *Perna canaliculus* (NZ green-lipped mussel)[J]. INFLAMM RES, **38**: C139-42(1993).
6. Cesaretti M, Luppi E, Maccari F, Volpi N. Isolation and characterization of a heparin with high anticoagulant activity from the clam *Tapes philippinarum*: evidence for the presence of a high content of antithrombin III binding site[J]. Glycobiology, **14**: 1275-84(2004).
7. Dallas DC, Guerrero A, Parker EA, Robinson RC, Gan J, German JB, Barile D, Lebrilla CB. Current peptidomics: applications, purification, identification, quantification, and functional analysis[J]. Proteomics, **15**: 1026-38(2015).
8. Wang W, Mejia EGD. A new frontier in soy bioactive peptides that may prevent age-related chronic diseases[J]. COMPR REV FOOD SCI F, **4**: 63-78(2005).

9. Larché M. Immunotherapy with Allergen Peptides[J]. ALLERGY ASTHMA CL IM, **3**: 53-9(2007).
10. A Wesley B, Mimi T, Scott S, Antonella M, Eigenmann PA, Motohiro E, Alessandro F, Wen C, Kirsten B, Robert W, Jonathan H, Stacie MJ, Gideon L, Hugh AS. ICON: food allergy[J]. J ALLERGY CLIN IMMUN, **129**: 906-20(2012).
11. Christer H, Andersson S, Arpaia D, Casacuberta J, Davies H, Jardin P, Flachowsky G. Scientific Opinion on the assessment of allergenicity of GM plants and microorganisms and derived food and feed[J]. EFSA Journal, **8**: 1700(2010)
12. Pak H, Pasquarello C, Scherl A. Label-free protein quantification on tandem mass spectra in an ion trapping device[J]. Journal of Integrated OMICS, **1**: 7-11(2011).
13. Adlernissen J. Enzymic hydrolysis of food proteins[J]. CAN MED ASSOC J, **172**: 1783-85(1986).
14. Costa ELD, Gontijo JADR, Netto FM. Effect of heat and enzymatic treatment on the antihypertensive activity of whey protein hydrolysates[J]. INT DAIRY J, **17**: 632-40(2007).
15. Saha S, Raghava G. AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes[J]. NUCLEIC ACIDS RES, **34**: W202-9(2006).
16. Lafarga T, Wilm M, Wynne K, Hayes M. Bioactive hydrolysates from bovine blood globulins: Generation, characterisation, and in silico, prediction of toxicity and allergenicity[J]. J FUNCT FOODS, **24**: 142-55(2016).
17. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass

- Spectrom[J]. J AM SOC MASS SPECTR, **5**: 976-89(1994).
18. Bern M, Finney G, Hoopmann MR, Merrihew G, Toth MJ, MacCoss MJ. Deconvolution of Mixture Spectra from Ion-Trap Data-Independent-Acquisition Tandem Mass Spectrometry[J]. ANAL CHEM, **82**: 833-41(2010).
19. Chen Y, Kwon SW, Kim SC, Zhao Y. Integrated approach for manual evaluation of peptides identified by searching protein sequence databases with tandem mass spectra[J]. J PROTEOME RES, **4**: 998-1005(2005).
20. Shevchenko A, de Sousa MM, Waridel P, Bittencourt ST, de Sousa MV. Sequence similarity-based proteomics in insects: characterization of the larvae venom of the Brazilian moth *Cerodirphia speciosa*[J]. J PROTEOME RES, **4**: 862-9(2005).
21. Liska AJ, Shevchenko A. Expanding the organismal scope of proteomics: Cross-species protein identification by mass spectrometry and its implications[J]. Proteomics, **3**: 19-28(2003).
22. Wijk KJV. Challenges and Prospects of Plant Proteomics[J]. PLANT PHYSIOL, **126**: 501-8(2001).
23. Lv T, Wang Y, Pan D, Cao J, Zhang X, Sun Y, Chen Y, Liu Y. Effect of trypsin treatments on the structure and binding capacity of volatile compounds of myosin[J]. FOOD CHEM, **214**: 710-6(2017).
24. Yilmaz MT, Kesmen Z, Baykal B, Sagdic O, Kulen O, Kacar O, Yetim H, Baykal AT. A novel method to differentiate bovine and porcine gelatins in food products: Nano UPLC-ESI-Q-TOF-MS E, based data independent acquisition technique to detect marker peptides in gelatin[J]. FOOD CHEM, **141**: 2450-8(2013).

25. Zhang G, Liu T, Wang Q, Chen L, Lei J, Luo J, Ma GH, Su ZG. Mass spectrometric detection of marker peptides in tryptic digests of gelatin: A new method to differentiate between bovine and porcine gelatin[J]. FOOD HYDROCOLLOID, **23**: 2001-7(2009).
26. Meisel H, FitzGerald RJ. Biofunctional peptides from milk proteins: mineral binding and 2531 cytomodulatory effects[J]. CURR PHARM DESIGN, **9**: 1289-95(2003).
27. Marie Togashi, Makoto Kakinuma, Misako Nakaya, Tatsuo Ooi, Shugo Watabe. Differential Scanning Calorimetry and Circular Dichroism Spectrometry of Walleye Pollack Myosin and Light Meromyosin[J]. J AGR FOOD CHEM, **50**: 4803-11(2002).
28. Gobert GN, Mcmanus DP. Update on paramyosin in parasitic worms[J]. PARASITOL INT, **54**: 101-7(2005).
29. Leung PS, Chu KH. Molecular and immunological characterization of shellfish allergens[J]. Frontiers in Bioscience A Journal & Virtual Library, **3**: d306-310(1998).
30. Irvine GB. High-performance size-exclusion chromatography of peptides[J]. J BIOCHEM BIOPH METH, **56**: 233-42(2003).
31. Rahaman T, Vasiljevic T, Ramchandran L. Effect of processing on conformational changes of food proteins related to allergenicity[J]. TRENDS FOOD SCI TECH, **49**: 24-34(2016).
32. Sabadin IS, Villas-Boas MB, Zollner RDL, Netto FM. Effect of combined treatment of hydrolysis and polymerization with transglutaminase on β -lactoglobulin

antigenicity[J]. EUR FOOD RES TECHNOL, **235**: 801-9(2012).

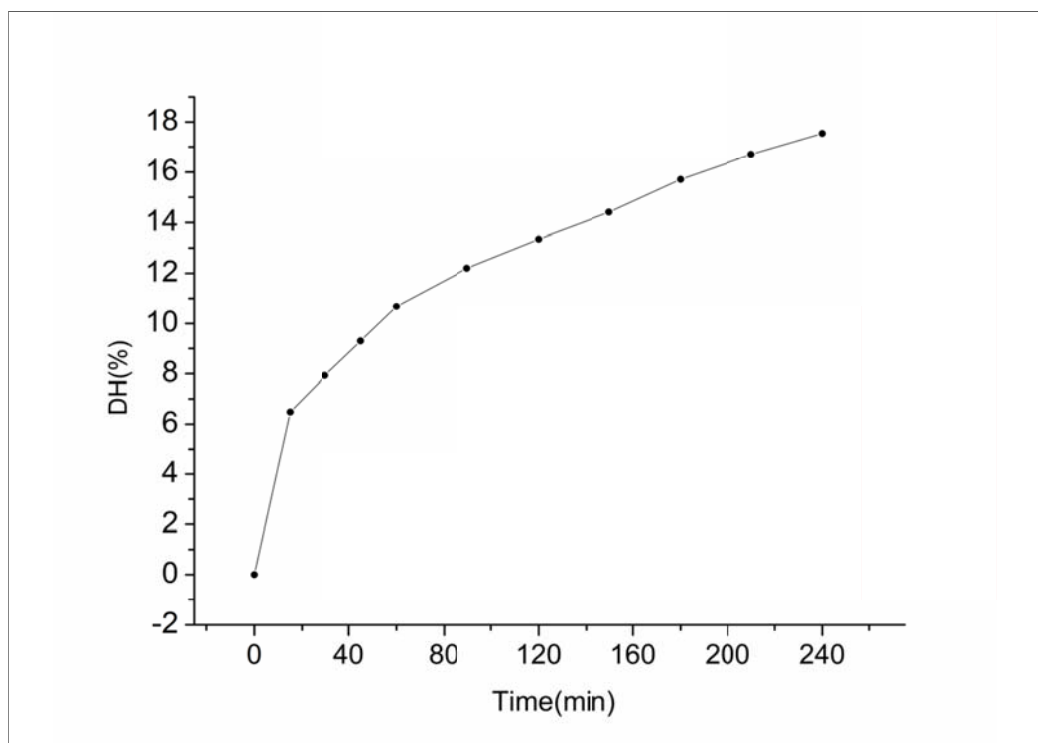


Fig.1. Hydrolysis curve of *R. philippinarum* protein with trypsin. The hydrolysis was performed under the conditions: E:S 3:100 (w/w), pH 9.0, 45 °C, 4 h.

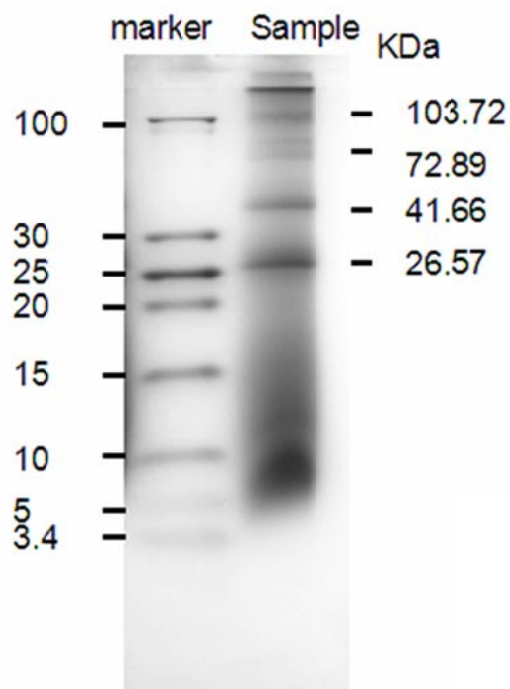


Fig.2. Tris-Tricine SDS-PAGE of trypsin hydrolysate of *R. philippinarum* proteins.

Sample of final concentration of 4 mg/mL was separated by 16% separation gel and 4% spacer gel. Molecular weight standards are shown in the left lane.

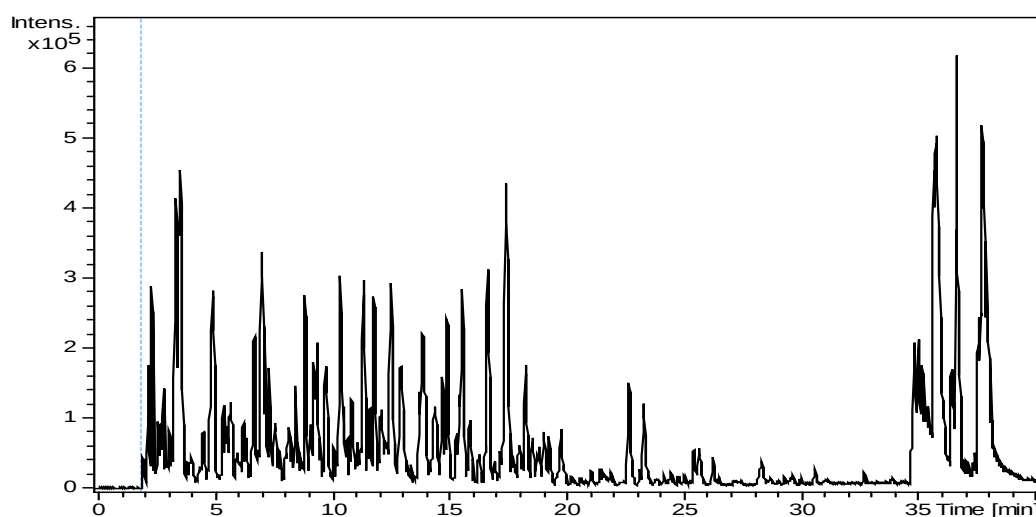


Fig.3.Total intensity chromatogram of *R. philippinarum* trypsin hydrolysate. Peptides were separated with a 40 min 10–80% ACN (with 0.1% formic acid, v/v) gradient elution on a 3 ×150 mm length LC column. The vast majority of the peptides eluted within 25 min, only a few peptides eluted after 35 min, the intense chromatographic peaks arising after 35 min were inevitable loss of column, which was shown in the spectra

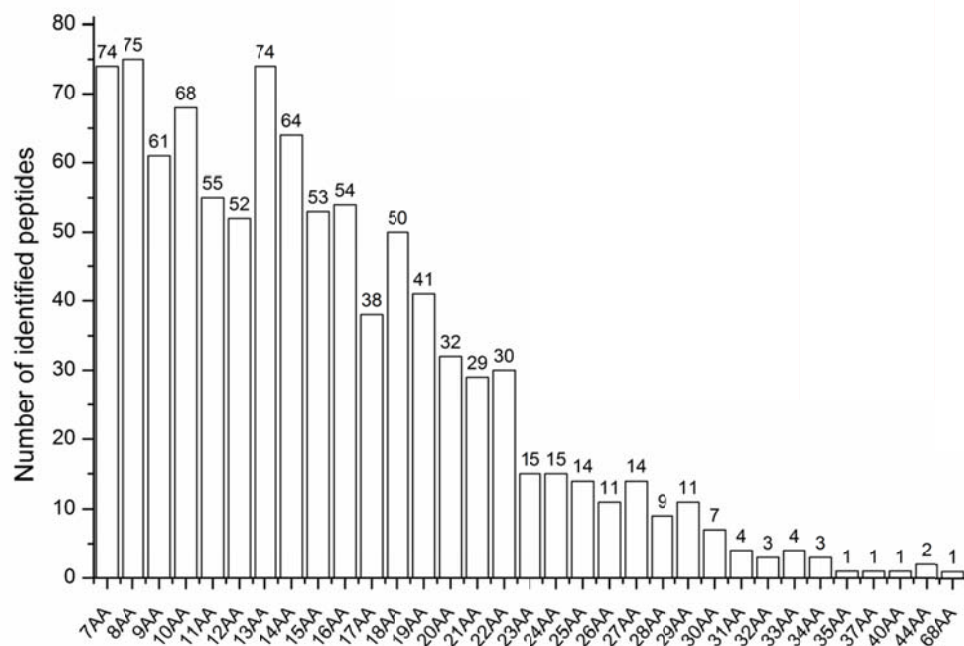


Fig.4. Chain length distribution of all identified unique peptides. The amino acid residue numbers were range from 7 AA to 68 AA.

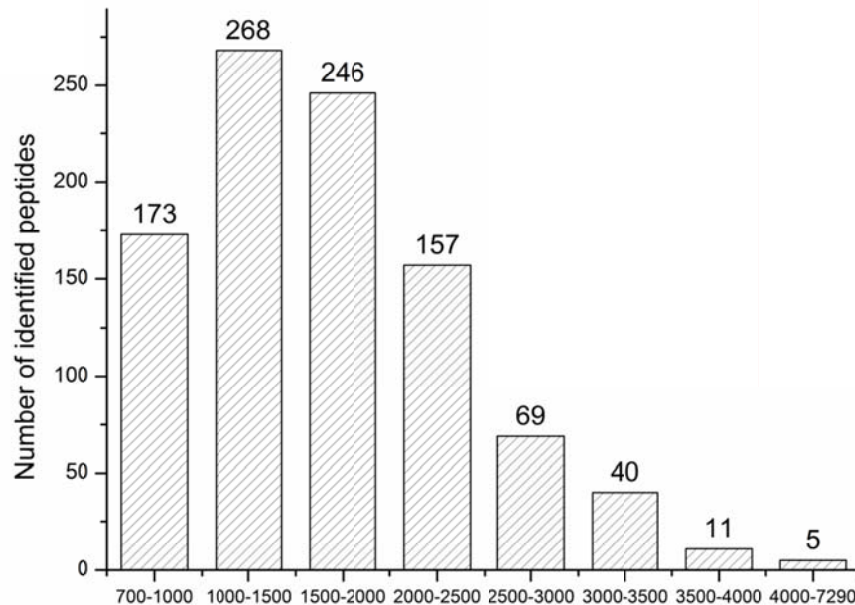


Fig.5. Molecular mass of all identified unique peptides. The MW of all peptides ranged from 700 to 7290 Da, the majority of peptides distributed in the MW range of 1000 to 1500 Da, only 5 corresponded to large proteins.

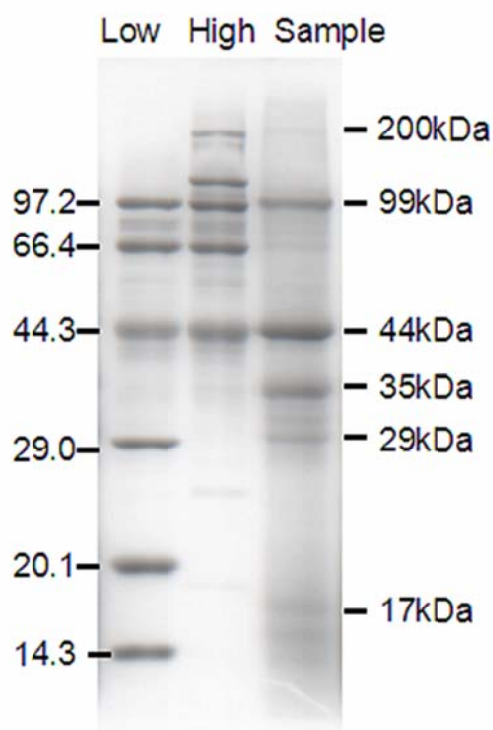


Fig.6. SDS-PAGE of *R. philippinarum* proteins. Sample of final concentration of 4 mg/mL was separated by 12% separation gel and 5% spacer gel. Premixed protein marker (high) and premixed protein marker (low) were used, and the molecular weight standards were shown in the left side.

Table.1.Complete description of proteins searched by Swiss-Prot database.

Prot accession number	Full name of protein	Organism	Prot score	Mass/pI	Peptides matched by MASCOT
ACT_PLAMG	Actin, adductor muscle	Placopecten magellanicus	1199	41735/5.3	45
ACTC_BRABE	Actin, cytoplasmic	Branchiostoma belcheri	1177	41678/8.11	1
ACT5_CHICK	Actin	Gallus gallus	1128	41809/8.11	3
ACT2_LUMTE	Actin-2	Lumbricus terrestris	1095	41813/5.6	1
ACTC_BRAFL	Actin, cytoplasmic	Branchiostoma floridae	1078	41694/8.11	1
ACTB3_TAKRU	Actin, cytoplasmic 3	Takifugu rubripes	1006	41756/8.11	1
ACTF_STRPU	Actin, cytoskeletal 3B	Strongylocentrotus purpuratus	1003	41805/5.3	3
ACT2_MOLOC	Actin	Molgula oculata	989	42236/5.12	1
ACT1_ACACA	Actin-1	Acanthamoeba castellanii	985	41649/4.58	2
ACTM_HELER	Actin	Heliocidaris erythrogramma	950	41721/5.04	1
ACT_BRUMA	Actin	Brugia malayi	864	41683/9.43	1
ACT2_LYTPI	Actin, cytoskeletal 2	Lytechinus pictus	831	41847/5.17	1
ACT2_HALRO	Actin, muscle 2/4/4A	Halocynthia roretzi	824	41987/5.3	2
ACT3_DICDI	Actin-3	Dictyostelium discoideum	809	41814/6.19	1
ACT1_LYTPI	Actin, cytoskeletal 1	Lytechinus pictus	798	41815/9.43	2
ACT3_DIPDE	Actin-3	Diphyllbothrium dendriticum	735	41885/8.11	2
ACTM_STRPU	Actin, muscle	Strongylocentrotus purpuratus	724	41530/8.43	1
ACT22_DICDI	Putative actin-22	Dictyostelium discoideum	718	41708/4.99	1

POTEI_HUMAN	POTE ankyrin domain family member I	Homo sapiens	542	121205/5.3	1
ACT1_ECHGR	Actin-1	Echinococcus granulosus	486	41801/4.97	6
ACT_FUCVE	Actin	Fucus vesiculosus	471	41863/8.11	1
ACT18_DICDI	Actin-18	Dictyostelium discoideum	414	42504/6.19	2
ACT_FUCDI	Actin	Fucus distichus	380	41767/6.13	1
ACT1_ARTSX	Actin, clone 205	Artemia sp	379	41759/8.43	1
ACT2_PEA	Actin-2	Pisum sativum	356	41741/9.43	2
TPM2_BIOGL	Tropomyosin-2	Biomphalaria glabrata	331	32663/4.78	1
TPM_TURCO	Tropomyosin (Fragments)	Turbo cornutus	312	16829/4.58	1
H2A_MYTCA	Histone H2A	Mytilus californianus	298	13352/10.9	9
H2B_CHITH	Histone H2B	Chironomus thummi thummi	277	13851/9.23	7
H2A7_ARATH	Probable histone H2A.7	Arabidopsis thaliana	270	15956/10.55	1
MYS_ARGIR	Myosin heavy chain, striated muscle	Argopecten irradians	270	222684/5.3	9
H4_ACRAS	Histone H4	Acrolepiopsis assectella	270	11374/5.3	7
H4_CAEEL	Histone H4	Caenorhabditis elegans	265	11362/6.13	1
H4_ASCSU	Histone H4	Ascaris suum	259	11373/6.13	2
ACT1_SOYBN	Actin-1	Glycine max	259	41330/8.43	1
H2A1_PEA	Histone H2A.1	Pisum sativum	253	15870/10.99	3
ACT1_NAEGR	Actin-1	Naegleria gruberi	248	41541/9.43	2
TBAT_ONCMY	Tubulin alpha chain,	Oncorhynchus mykiss	233	49962/8.43	5
H2A2_ARATH	histone H2A.2	Arabidopsis thaliana	233	13841/6.07	1

TPM_HALDV	Tropomyosin	Haliotis diversicolor	210	32803/5.68	1
ACT_GIAIN	Actin	Giardia intestinalis	191	41560/9.43	1
TBA3_HOMAM	Tubulin alpha-3 chain	Homarus americanus	175	49975/4.43	1
H3_ACRFO	Histone H3	Acropora formosa	173	15275/5.83	5
MYH6_HUMAN	Myosin-6	Homo sapiens	122	223596/8.36	2
MLR_SPISA	Myosin regulatory light chain, smooth muscle	Spisula sachalinensis	108	17983/5.38	3
MYSP_MYTGA	Paramyosin	Mytilus galloprovincialis	97	99513/7.64	2
TBA5_CHICK	Tubulin alpha-5 chain	Gallus gallus	96	49916/11.14	1
ACT_EUPCR	Actin, cytoplasmic	Euplotes crassus	92	42596/8.43	1
MYH1_BOVIN	Myosin-1	Bos taurus	89	222851/8.43	1
H2B1_CAEEL	Histone H2B 1	Caenorhabditis elegans	87	13492/5.7	1
TBA2_HOMAM	Tubulin alpha-2 chain	Homarus americanus	46	50219/8.43	1
COR2A_BOVIN	Coronin-2A	Bos taurus	46	59807/6.69	1
gi 353558656	Insoluble matrix shell protein 5		40	15596/8.23	9
CDPKE_ARATH	Calcium-dependent protein kinase 14	Arabidopsis thaliana	48	60016/6.75	1
ANLNALDR	POTE ankyrin domain family member I	Homo sapiens	542	121205/5.83	1

Table.2. Description of proteins searched by *Ruditapes philippinarum* database.

Prot accession number	Full name	Prot score	Prot mass	Prot matches	Prot sequences
gi 378408555	Troponin C	45	17361	4	34
gi 167888935	Tandem repeat galectin	66	34487	9	5
gi 363990577	Sigma glutathione s-transferase 3	29	24067	7	19
gi 363990573	Sigma glutathione s-transferase 1	54	23268	13	10
gi 378408551	Sigma class glutathione s-transferase	50	23202	5	44
gi 256549332	Selenium-dependent glutathione peroxidase	117	26377	20	7
gi 363990579	Omega glutathione s-transferase	64	27470	10	10
gi 18157446	NADH dehydrogenase subunit 4 (mitochondrion)	174	47885	44	8
gi 32140731	NADH dehydrogenase subunit 6 (mitochondrion)	52	18846	10	4
gi 32140754	NADH dehydrogenase subunit 5 (mitochondrion)	129	62676	34	5
gi 32140695	NADH dehydrogenase subunit 4 (mitochondrion)	97	50078	24	17
gi 18157462	NADH dehydrogenase subunit 3 (mitochondrion)	40	15597	11	27
gi 18157454	NADH dehydrogenase subunit 2 (mitochondrion)	171	37343	28	23
gi 18157453	NADH dehydrogenase subunit 1 (mitochondrion)	110	34659	26	26
gi 256549354	Heat shock protein 22 isoform 1	116	20284	11	29
gi 256549324	Glutathione s-transferase A	34	23769	7	9
gi 399762273	Defensin	32	8025	3	25
gi 666364473	Cytochrome oxidase subunit I, partial (mitochondrion)	48	20501	6	9
gi 18157458	Cytochrome c subunit II (mitochondrion)	157	58772	25	24

gi 18157466	Cytochrome c subunit I (mitochondrion)	162	59369	29	5
gi 18157445	Cytochrome b (mitochondrion)	106	49571	28	10
gi 512991689	C-type lysozyme 2	36	17354	5	8
gi 966344233	Complement C1q-like protein 4	37	24271	5	5
gi 966344227	Complement C1q-like protein 1	38	25728	8	3
gi 32140696	ATP synthase subunit 6 (mitochondrion)	101	27201	17	7
gi 325977003	AP-1 protein	37	30094	8	7