CHARACTERIZATION OF AIRBORNE ALLERGENS OF SELECTED CRUSTACEANS AND MONITORING OF THEIR OCCUPATIONAL EXPOSURE LEVELS USING MASS SPECTROMETRY

ANAS M ABDEL RAHMAN


CENTRE FOR NFLD. STUDIES
MEMORIAL UNIVERSITY
OF NEWFOUNDLAND

Characterization of airborne allergens of selected crustaceans and monitoring of their occupational exposure levels using mass spectrometry

by

© Anas M Abdel Rahman

A Thesis submitted to the

School of Graduate Studies in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

Faculty of Science

Memorial University of Newfoundland

April, 2011

St. John's Newfoundland

Abstract

The number of workers in the processing and harvesting sector of the fishery has increased substantially in recent decades, seeing more than 41 million workers worldwide engaged in various activities related to seafood production. Serious allergies and asthma have developed in the seafood occupational workplace as the result of airborne allergenic proteins being aerosolized on fishing vessels and in processing plants. Traditionally, the identification and quantification of these allergens are performed by immunoactive-based techniques such as radioallergosorbent test and enzyme-linked immunosorbent assay.

In this study, a novel strategy of monitoring airborne seafood allergens was developed using tandem mass spectrometry. First, the major allergenic proteins of snow crab and black tiger prawn were identified and profiled using an allergeomics approach. The immunoblotting was performed against sensitized patients' sera and the active protein bands were characterized using peptide mass fingerprinting mass spectrometry. The purified forms of the target allergenic proteins, tropomyosin and arginine kinase, were used as reference materials for primary structural identification and for quantification method development.

For protein sequencing, the samples were exposed to different enzymatic digestions (Trypsin, Glu-C V8, and Asp-N), the peptides analyzed by different mass spectrometry ion sources, and various derivatization protocols were used to maximize the amino acid sequence coverage. The product mass spectra of the enzymatically produced peptides were uploaded to a Mascot search engine and the national center for

biotechnology information non-redundant database searched. Signature peptides that surrogate each individual allergenic protein were determined and chemically synthesized (purchased from third party) for use as analytical standards in both light and heavy forms. Subsequently, a quantification approach using liquid chromatography-multiple reaction monitoring isotopic dilution tandem mass spectrometry was developed to measure snow crab tropomyosin and arginine kinase in air samples. A comprehensive validation study was performed for the quantification method with respect to the international conference on harmonization guidelines. The air sampling protocol, allergen extraction recovery, and the completeness of the protein digestion were optimized for real air sample collection. Actual air samples were collected from a simulated crab processing plant and analyzed by mass spectrometry.

Acknowledgements

This work could only have been accomplished through the mutual collaboration with different people, who I want to acknowledge. I want to express my gratitude to Dr. Robert Helleur, my supervisor, for his faith in me, constructive advice, and patient guidance. In addition, I would like to acknowledge Dr. John Robinson (co-supervisor), Dr. Andreas Lopata, and Mr. Sandip Kamath, for their contributions in performing some of the project's experiments in their labs and the constructive advice they brought from their knowledge and experience. Also, sincere thanks to Dr. John Molgard for the loan his crab butchering table and to Wade Murphy (Marine Institute) for setting up the simulated crab plant. I would also like to express my thanks to Dr. Travis Fridgen (supervisory committee) for reviewing a draft of this thesis and for his helpful advice.

I am grateful to Dr. Barbara Neis and Memorial University SafetyNet Centre for Occupational Health & Safety Research for their financial assistance moral support and advice. In addition, I would like to thank Dr. Edward Randell, Dr. Christina Bottaro, Mrs. Donna Jackman, and Dr. Joseph Banoub for assistance in allowing me to use their laboratory for sample analysis.

I would like to thank Stephana Egli, Suliman Barhoum, Ali Atoom, Meagan White, Farid Jahouh, and Natalie Adams for being friends during these joyous years.

I would like also to express my appreciation to my parents; Mr. Mahmoud Abdel Rahman and Mrs. I'tidal Khasawneh, my brothers, Muhannad, Muayyad, Mohmmad, Moad and my sisters, Da'ad, Ghada, and Enas and my grandfather Radi Khasawneh.

Finally, I would like to thank the Department of Chemistry and the School of Graduate Studies at Memorial University, and the National Sciences and Engineering Research Council (NSERC) of Canada for their funding support.

Dedication

I dedicate this work to my parents who spent their lives waiting for this tremendous moment and who illuminated my way in the darkest days of my life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AA:	Amino acid
ACN:	Acetonitrile
AK:	Arginine kinase
AQUA:	Absolute quantification
BIRD:	Blackbody infrared dissociation
BLAST:	Basic local alignment search tool
BSA:	Bovine serum albumin
CID:	Collision induced dissociation
DNA:	Deoxyribonucleic acid
ECD:	Electron capture dissociation
ELISA:	Enzyme-linked immunosorbent assay
ESI:	Electrospray ionization
ETD:	Electron transfer dissociation
FAB:	Fast atom bombardment
FDA:	Food and drug administration
FTICR:	Fourier transforms ion cyclotron resonance
FTMS:	Fourier transforms mass spectrometry
HPLC:	High performance liquid chromatography
HRP:	Horse radish peroxidase
ICAT:	Isotope-coded affinity tag
ICH:	International conference on harmonization
IgE:	Immunoglobulin E
IRMPD:	Infrared multiphoton dissociation

iTRAQ:	Isobaric tag for relative and absolute quantitation
KE:	Kinetic energy
LC MS/MS:	Liquid chromatography tandem mass spectrometry
LOD:	Lower limit of detection
LOQ:	Lower limit of quantification
mAb:	Monoclonal antibody
MALDI:	Matrix-assisted laser desorption/ ionization
MHT:	Mascot homology threshold
MIS:	Mascot ions score
MIT:	Mascot identity threshold
MRM:	Multiple Reaction Monitoring
MS:	Mass Spectrometry
NCBI:	National Center for Biotechnology Information
NMR:	Nuclear magnetic resonance
OA:	Occupational asthma
OAl:	Occupational allergies
OHS:	Occupational health and safety
PBS:	Phosphate buffer saline
PBZ:	Personal breathing zones
PCR:	Polymerase chain reaction
PFF:	Peptide fragment fingerprinting
pI:	Isoelectric point
PITC:	Phenylisothiocyanate
PMF:	Peptide mass fingerprinting

PTC:	Phenylthiocarbamyl
PTFE:	Polytetrafluoroethylene
PTH:	Phenylthiohydantoin
PTM:	Post-translational modification
PVDF:	Polyvinylidene fluoride
QC:	Quality control
QqToF MS:	Quadruple-quadruple-time of flight mass spectrometer
RAST:	Radioallergosorbent test
RIA:	Radioimmuno assay
RNA:	Ribonucleic acid
SC:	Snow crab
SDS-PAGE:	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC:	Stable-isotope labeling by amino acid
SP:	Signature peptide
SRM:	Selected reaction monitoring
T:	Tons
TBS:	Tris-buffered saline
TM:	Tropomyosin
TMT:	Tandem mass tag
TFA:	Trifluoroacetic acid
TOF:	Time-of-flight
TSP	Tropomyosin signature peptide

Chapter 1: Introduction

1.1. Seafood allergens and occupational asthma

Crustaceans are the most consumed seafood worldwide, and therefore play a major role in human nutrition and in the world economy. The international trade of seafood has been growing rapidly in the last decades, which reflects on the popularity and frequency of consumption worldwide. Seafood was incorporated significantly into the North American diet as an effort to attain healthier living (1). The United States has become the third largest consumer of seafood in the world, having consumed 1.86 billion kg of crustaceans in 2007(6.04 kg/capita/year) (2). Since seafood ingestion can cause severe acute hypersensitivity reactions and is recognized as one of the most common food allergies, the increased production and consumption of seafood has resulted in more frequent adverse health problems (4). Exposure to seafood can cause a variety of health problems, one problem reported to be significant an increase in occupational asthma (5,6). Dermatitis and asthma are the major hazards, caused by both irritative and immunological reaction to occupational exposure in seafood production.

Occupational asthma is a disease characterized by variable airflow obstruction and/or airway hypersensitivity due to active agents in the work environment (7). A true allergy is known as type-one hypersensitivity that activates the mast cells, a certain type of white blood cells producing an immunoglobulin E-mediated (IgE) response. This activation releases histamine and other inflammatory mediators such as cytokines. These immunological activities result in different allergenic symptoms such as itchiness, dyspnea, and anaphylaxis. In the seafood processing industry, the high-molecular mass

molecules such as proteins and polysaccharides mainly cause chronic dermatitis (8). Indeed, the severity of the atopic individuals' symptoms was reported to increase as the time between fishing and processing increased (9). As well, a non-atopic worker was reported to have developed chronic dermatitis and recurrent episodes of edema of the lips and face after chronic exposure to crustaceans (9). Atopy is a genetic disposition to develop an allergenic reaction and elevates the levels of immunoglobulin E-mediated (IgE) upon exposure to environmental allergens by inhalation or ingestion. In another case a worker suffered repeated urticaria after feeding fish and crabs (10).

Previous studies have evaluated some of these seafood allergens and reported that the prevalence for occupational protein contact dermatitis ranges from 3% to 11% and from 7% to 36% for occupational asthma (11). The diagnostic assay is usually performed on affected workers who show a high rate of reactivity in skin testing with crab extracts. Inhaling the allergen species from workplace environments develops a food-like allergen with a classical IgE reaction (12, 13). Other occupational asthma incidents have occurred among Polish fishermen working in processing plants on factory ships (14). In addition, other reports have shown sensitization to red soft corals among spiny lobster fishermen and occupational asthma in workers employed in fish-freezing and fish smoking factories (15, 8). Moreover, using live fish bait such as bee moth, mealworm, and other worms and larvae may also lead to the development an IgE-mediated occupational asthma (16).

High molecular-mass agents such as proteins, polysaccharides, and large peptides induce allergenic responses by producing specific IgE and sometimes IgG antibodies. These high molecular-mass agents would also give a positive skin test reaction (7). The overall prevalence of occupational asthma is still unknown, although it is significantly higher when processing and consumption of seafood plays a crucial part in the community. Seafood consumption has increased worldwide; consequently, the number of workers who are involved in seafood production has also increased (4). Among US workers, a recent study shows the prevalence of shellfish allergy was 1.9% and 1.3% among respondents with undiagnosed self-reported food allergy and doctor diagnosed food allergy, respectively (17). This study was undertaken by a Food Safety Survey of over 4400 participants (17).

Snow crab processing became a major fishery in Newfoundland and Labrador, after the ground fish collapse in the early 1990s. Some of the crab processing plants were adapted from ground fish facilities with limited ventilation (18). In 2001, a multidisciplinary study commenced on Canada's east coast, which evaluated the prevalence of occupational asthma (OA) and occupational allergies (OAl) in workers from four snow crab processing plants in Newfoundland and Labrador (18). In this study the relationship between the OA and OAl, and exposure to snow crab allergens was determined along with the host risk factors such as smoking. The comprehensive study reported that the OA and OAl prevalence were 15.8% and 14.9%, respectively; which

agrees with the first epidemiological study of snow crab processing workers in Quebec, in 1984, which obtained a value of 15.6% for OA (13).

1.1.1. Crustaceans' allergens

Tropomyosin (TM), one of the muscle proteins, is the major allergen responsible for ingestion-related allergic reactions. Tropomyosins belong to the family of actin filament-binding proteins with different isoforms found in muscles, and non-muscle cells. A complex of TM and troponin regulates the calcium sensitive interaction of actin and myosin. Crustaceans' TM was first identified in shrimp in 1981 by Hoffman *et al.* (19). Shanti *et al.* (20) reported an 86% amino acid sequence homology between the *Penaeus indicus* shrimp allergen Pen i 1 and fruit fly Drosophila melanogaster TM. The open reading frame of the cloned tropomyosin in invertebrates was reported to be 281-amino acids with a monomeric molecular weight range of 38-41 kDa. The highly conserved amino acid sequence of TM is responsible for its identification as a panallergen for cross-reactivity between crustaceans, insects, arachnids, and different classes of mollusks (1, 15). In addition, the alleregenicity of TM was confirmed in six species of crustaceans; black tiger prawn, kuruma prawn, pink shrimp, king crab, snow crab, and horsehair crab by immunoblotting and the overall sequence identity showed more than 90% homology (21).

Many other allergens have been identified in crustaceans. Yu's group identified arginine kinase (40 kDa) as a novel shrimp allergen. The amino acid sequence of this protein showed 60% similarity to arginine kinase of the crustacean, Kuruma prawn (*Penaeus japonicas*) (22). Arginine kinase (AK) was recently reported as an allergen in different crustacean and invertebrate species. AK was identified in white shrimp (*Litopenaeus vannamei*) (23), gulf shrimp (*Penaeus aztecus*) (24), chinese shrimp (*Fenneropenaeus chinensis*) (25), and other shrimp species using a proteomics approach (26). Moreover, AK has been identified in other invertebrates such as the house dust mite (*Dermatophagoides farinae*) (27), Indian-meal moth (*Plodia interpunctella*) (28), and silkworm larvae (*Bombyx mori*) (29).

Recent studies have reported other novel crustacean allergens. The sarcoplasmic reticulum Ca-binding protein (20 kDa) was identified and characterized in black tiger shrimp (*Penaeus monodon*). In addition, a myosin light chain (20 kDa) was also identified in white shrimp (*Litopenaeus vannamei*), and in black tiger prawn (*Penaeus monodon*) (30). Since TM is a common allergen in both crustaceans and mollusks, other potent allergens such as myosin heavy chain, hemocyanin and amylase could exist in crustaceans. These allergens were identified and characterized in different mollusks such as mussel, oyster, and could exist alone in crustaceans (31).

Parvalbumin proteins, which are one of the Ca-transporting muscle proteins and reported as the major codfish allergen Gad c-1, were studied comprehensively to exploit

the closeness between animal allergens and their human homologs (32). Parvalbumin's molecular mass ranges from 10-13 kDa, and contains heat-resistant linear epitopes that are stabilized by the interaction of metal-binding domains (4). The alleregencity of the parvalbumin was studied in purified forms from different type of fish (33). In addition, other fish allergens were also characterized such as collagen isolated from skin (34) and muscle tissues, and fish hormones like vitellogenin (35). The literature reports that there is no cross reactivity between fish allergens and shellfish (4).

1.1.2. Immunological analysis

Aeroallergen analysis requires highly selective and sensitive techniques, which is an analytical challenge when dealing with high molecular mass compounds and very low air concentrations. The allergenic protein causes asthma-like symptoms, and is usually identified by immunoblotting. In this technique, the crude airborne protein isolate is first profiled on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and then the protein bands transferred to nitrocellulose membrane. This protein blot is incubated with patients' sera (those with clinical histories of crustaceans' allergy and giving positive skin test) (36, 37). The identity of the immuno-reactive proteins is confirmed by their cloned genes, which are introduced to the immunoblotting. Specifically, the target gene is isolated and cloned in bacteria, after culturing this bacteria

the gene will be translated to its equivalent protein where is subsequently isolated and purified to be used as a reference and a standard protein.

Motoyama *et al.* (21), used the gene cloning technique to identify the reactive TM protein in six species of crustaceans using immunoblotting against sensitized patient sera and also a specific monoclonal antibody (mAb). By the same approach, air samples were collected from active crab-processing plants in Newfoundland and Labrador and the aeroallergens were also identified using immunoblotting against a pool of patients' sera (3). This study did not use molecular cloning or a high level of structural analysis but just determined the molecular masses of these proteins using molecular mass markers in SDS-PAGE.

There are several quantification approaches utilized to measure the levels of aeroallergens in workplace environments. These techniques are based mainly on the immunoreactivity of allergenic proteins with the specific IgE of sensitized patients' sera. Inhibition radioimmuno assay (RIA), radioallergosorbent test (RAST), and enzyme-linked immunosorbent assay (ELISA) are the most common techniques implemented for aeroallergenic proteins from an occupational health and safety (OHS) perspective. Taylor *et al.* (38) measured the level of raw fish aeroallergens in open-air fish markets using the inhibition RIA. Measuring the levels of specific IgE in the patients' sera is usually performed using the RAST test, where an antigen water extract is used as a standard to develop the quantification approach (39, 40). Basically, the target allergen in the RAST

test is bonded to a solid polymer. The specific IgE in the patient sera reacts with the bound allergen. After washing away non-specific IgE, radiolabeled anti-human IgE antibody is added to form a complex, and then the unbound secondary antibody is washed away after incubation. The amount of radioactivity is proportional to the amount of specific IgE (41). The RAST test was introduced to measure the airborne allergens of different species of mice (42), pigs (43), house flies (44), and cats (45). In 1997 Malo *et al.* (41) applied the RAST test for the first time to measure seafood aeroallergens, where the air samples were collected from snow crab processing plants. In this study, total allergen extracts were prepared as a standard and the activity of these extracts was evaluated by skin test with diagnosed patients. The amount of the allergenic protein was estimated by percent inhibition of the RAST profile. A comprehensive study with four snow crab plants was accomplished in Newfoundland and Labrador, where air samples were analyzed using a developed approached of the RAST test (46).

Inhibition ELISA techniques were performed to study seafood aeroallergens. The water or meat protein extract is incubated in the wells of a microtiter plate surface; a pool of patients' sera is added to allow the specific IgE to bind to the target allergen. The non-bound IgE is subsequently washed away, and then the labeled anti-human IgE antibody binds to the bound IgE. The antihuman IgE is labeled with florescent dye or radio isotope (e.g. ^{125} I), where the response is detected on an ELISA reader and scintillation counter, respectively. Several studies have utilized these ELISA techniques. Weytjens *et al.* (47) measured the levels of snow crab allergen in workplaces and studied the gradient profile

of the aeroallergens inside a processing plant using the ELISA technique with an ^{125}I-radiolabel. Using the same approach, Beaudet's group studied aerosolized crab levels aboard crab processing vessels in Alaska, where the prevalence of crab allergen was reported to be similar across all job categories (48). ELISA techniques with a colorimetric detection system were involved for measuring the fish aeroallergen in processing plants (49). In 2005, Jeebhay *et al.* (50) quantified the level of aeroallergens to study the correlation between the levels of the aerosolized allergens and other major determinants of variability, and then established dose-response relationships for fish antigen exposure. Another comprehensive study was performed for examining the levels of bioaerosols in the seafood industry using the ELISA reader for quantification purposes (6).

1.2. Protein analysis

1.2.1. Protein biosynthesis

Proteins are macromolecules made up of heterogeneous chains of amino acids which are linked to each other by covalent peptide bonds. In peptide bonds, the carboxylic site of one amino acid reacts with the amine site of the adjacent amino acid, thereby releasing a water molecule. Twenty different amino acids 'letters' are used to translate a distinctive gene to obtain the sequence of its related protein, which can result in a huge number of possibilities for deducing protein sequences (51). The terminal

residues of the protein chain that contain free α-NH_3^+ and COO^- groups are referred to as N-terminal and C-terminal, respectively. Naturally, during protein biosynthesis, the cell gets a message to make a certain quantity of a specific protein. The responsible gene for this protein is exposed by unwinding a portion of the deoxyribonucleic acid (DNA). Nucleotides move along one strand of the exposed gene and form a messenger ribonucleic acid (mRNA) molecule with a help of enzymes. According to the received message, multiple copies of the mRNA are made and they leave the nucleus to the cytoplasm. A ribosome binds with an mRNA molecule and starts reading the message (decoded), where the message is read three nucleotide bases (codon) at time. Since each codon stands for a specific amino acid, enzymes activate the corresponding amino acid once its codon is read. The transfer RNA molecule has two ends. One end has a specific binding site for a particular amino acid; the other end has a particular sequence of three nucleotides (anticodon) that can base pair with the codon. The appropriate molecule of tRNA carries the activated amino acid to the ribosome. The anticodon end base pair with the codon in order to bring the specific amino acid to the correct place. A second tRNA molecule picks up another activated amino acid to bind with the second codon, and then the first tRNA releases its amino acid to the second tRNA and leaves the site. The two amino acids form a peptide bond using ATP as its energy source. The amino acids "load" of the previous tRNA molecule is released to be linked up with the amino acid of the consecutive tRNA and forming a polypeptide. The polypeptide chain folds into its final conformation and released. Some proteins have different isoforms, whereby these

isoforms would be produced from different genes or from the same gene by alternative splicing.

1.2.2. Protein structure

The structure of a protein inside the cells consists of a combination of four different levels in order to have its proper function. In a hierarchical fashion, each level of the protein structure is constructed upon the one below (52). Protein structures are recognized by four levels; primary, secondary, tertiary, and quaternary (Figure 1.1) (52).

The primary structure refers to the sequence of the amino acids that translated from its corresponding gene. The chemical and biophysical properties of a protein come from the contribution of the same properties of the involved amino acids. The primary structure of proteins is usually deduced by the direct DNA sequencing of its corresponding gene. Also the primary structure of a protein can be identified by Edman degradation or mass spectrometry techniques.

The polypeptide chain of a protein is in an ordered array to confer regular conformational forms. This confirmation constitutes the secondary structure, which is formed by intramolecular hydrogen bonding of amide groups between amino acid residues in close proximity. In general, two broad classes of secondary structure are termed; globular proteins and fibrous proteins (53).

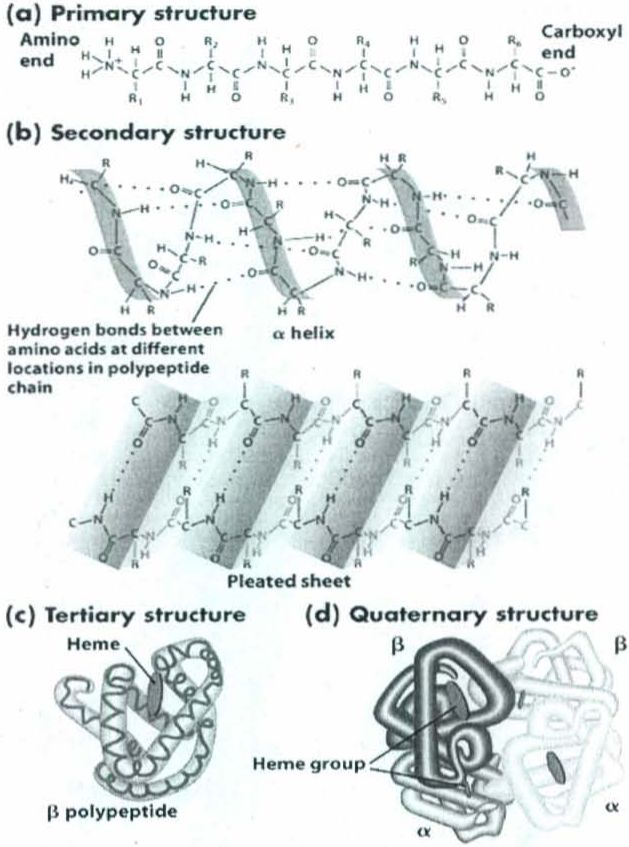


Figure 1. 1: Levels of protein structures; (a) primary (b) secondary (α helix and pleated sheet) (c) tertiary (β polypeptide) and (d) quaternary structure (54).

The α-helix structure, which is a common secondary structure in globular proteins, of polypeptides is coiled, where the backbone resides in the inner part and the side chain of the amino acids extend out from the coil. This kind of structure is identified by its pitch (rise per residue), period (number of residues per turn), handedness (right or left), and diameter (52). The individual protein chains in the β-pleated sheet structure are arranged side-by-side in two forms; parallel and anti-parallel. While the H-amide of the amino acid is bonded to the O-amide of the neighboring polypeptide chain, the side chains of the amino acids are positioned alternately up and down along a single β strand. The 3-D structure of a protein can be elucidated from different experimental techniques such as X-ray crystallography (55), solid-state nuclear magnetic resonance (NMR) (56, 57), and dual polarization interferometery (58). Once the protein primary structure is identified, the secondary structure can be predicted using several computational methods such as the *Ab initio* prediction method (59).

The spatial relationship between a protein's secondary structures gives full three-dimentional structure known as tertiary structure. This structure is known as the overall, unique, three dimensional folding of a protein (60). Several forces govern the interaction between different secondary structure domains within a protein. These include ionic bonding, hydrogen bonding, hydrophobic interactions, and disulfide bonds. The tertiary structure of proteins can be revealed by X-ray crystallography and NMR (52). The non-covalent interactions between helices and β-structures together with the side chain and

backbone interactions are a unique feature of a given protein. The disulfide bonds are the major chemical forces that stabilize the conformation of a protein after native folding.

The polypeptides' interactions within the oligomeric proteins give a quaternary structure. The oligomeric protein can be composed of one type of polypeptides, which are termed homo-oligomers or composed of several distinct polypeptides, which known as hetero-oligomers. This structure plays a significant role in the function of the oligomer proteins (e.g. hemoglobin $\alpha 2\beta 2$). The amino acid sequences for the subunits play a crucial role in the quaternary structures, which is determined by the stoichiometry and spatial arrangement of those subunits. Once the interface region between the adjacent subunits is highly complementary, the hydrophobic and polar interactions are the dominant forces between these subunits. In conclusion, the non-specific interaction of the close proximity amino acids along the primary structure of the polypeptide not only confers the secondary and tertiary structure, but also dictates the quaternary structure of the protein (52).

1.2.3. Protein purification

Characterization of a target protein requires a certain degree of purity, which depends mainly on what information are needed and the technique used to purify. The main factors that control the direction of purification steps are the actual target protein, the complexity of the crude starting material, the required degree of target purity, and the

chemistry of coexisting protein species. In addition, the scale of the purified material and the cost of purification strategy also should be considered.

The target proteins, in this study, are purified from the muscle tissue of the seafood, which makes the purification strategy more straightforward. Since there are many purification techniques that can be followed to purify a protein, this section will focus on those used in the present study. Acetone, isoelectric point, and ammonium sulfate precipitation techniques are used to purify the target proteins from seafood tissue.

As discussed before, the three main forces contributing to protein interactions are electrostatic, hydrophobic, and van der Waals forces. When electrostatic forces are dominant between protein molecules, the like charges will repel each other because they have the same net charge. On the other hand, "unlike proteins," having different charge states will attract each other. In isoelectric point precipitation, proteins are most likely to aggregate and precipitate when the pH of the solution is at or near to their isoelectric point (pI). There is no net charge and no repelling of molecules of the target protein (61).

The salting out precipitation procedure is the most common process used for protein purification. A high concentration of salt, usually ammonium sulfate $(NH_4)_2SO_4$, competes with the protein for water molecules, leading to charged protein group interaction. Water molecules are attracted preferentially to salt molecules during this method since the surfaces of the protein is more hydrophobic. Thus, the proteins become

associated by hydrophobic interactions and precipitate out at a high concentration of salt (62).

The dielectric constants of the proteins make the interactions between the charged sites on the surfaces of the protein stronger. Adding a water-miscible solvent, such as acetone, decreases the dielectric constant of the protein. The interactions between proteins become weaker and they begin to aggregate and precipitate. These solvents associate with water molecules in a stronger manner than with the protein, which leads to dehydrated protein surfaces and the van der Waal forces become dominant (63).

A combination of several precipitation techniques is highly feasible to achieve target purity. But since solvent and isoelectric precipitation techniques are highly sensitive to the salt concentration, it is sometime more effective to perform the ammonium sulfate precipitation (63). These techniques usually give a purified enough protein to be analyzed by MS. Further purification can be achieved with chromatography such as ionic exchange and affinity chromatography as well as using filters or membranes to remove coexistence molecules (52).

1.2.4. Protein separation using gel electrophoresis

Gel electrophoresis plays a central role in proteomics research wherein it provides a powerful separation and quantification method for proteins in complex mixtures. The

high resolution of separation and the sensitivity of detection both contributed to the presence of this technique in most protein labs worldwide. The principle of separation in the gel electrophoresis technique is mainly based on the ability of molecules to move under the influence of an electric field. In these experiments, proteins carry charges from a uniform coating of sodium dodecyl sulphate (SDS). The rate of protein mobility is based mainly on the size of the protein, where larger differences in mobility give better resolutions. The gel materials also participate in protein movement under the influence of an electric field. There are mainly two systems of gel electrophoresis for protein separation; gel that separates proteins according to their molecular weight or by their isoelectric point (64).

Polyacrylamide gel electrophoresis (PAGE) is a central technique for protein separation. The permeation of the gel is the most important property, where the pore size of the gel should be small enough to restrict the movement of proteins. Either the percentage of acrylamide (%T) or the cross-linker (%C) can control this factor. In brief, the highest percentage of acrylamide gives very small pore sizes, making the gel highly restrictive and favoring the movement of small proteins (64).

Combining both the isoelectric focusing gel and the molecular sizing gel (SDS-PAGE) gives a better resolution in two dimensions. In the first dimension, the proteins are separated on the basis of their pI, where the protein charges are determined by the pH of the local environment. Consequently, the strip of isoelectric focusing gel is saturated in

SDS solution and then transferred to the SDS-PAGE, the second dimension. The resulting gel shows pI values in the abscissa and molecular weight markers along the ordinate. This combination allows high resolution of proteins separation from a complex mixture. In addition, this techniques also introduces some analytical parameters for the separated proteins such as molecular weight, isoelectric point (pI), and the amount of the protein in each single spot (65).

Visualization is a very important step after protein separation. Moreover, this step improves the sensitivity and visibility of the gel electrophoresis technique. The validity of the method is evaluated with respect to its compatibility with mass spectrometry. In addition to sensitivity, linearity, and reproducibility, the following visualization methods should be fully capable to interface with proteomics MS technology (61, 66).

Coomassie blue-staining is developed by an ionic interaction between the dye and the basic amino acid moieties of the protein in an acidic environment. The separated proteins will appear as dark blue dye with almost background-free detection (67). This dye detects as little as 0.5-1.0 pmol in a 1D-gel band and 0.2-0.5 pmol in 2D-gels. The ease of using Coomassie blue-staining, its good quantitative linearity, and its compatibility with mass spectrometry are the main advantages accounting for its use in protein labs.

1.2.5. Immunoblotting

Immunoblotting is a powerful selective and sensitive technique, used to detect a small amount of protein or to study antibody-antigen reactivity. After the proteins are separated in gel electrophoresis, they are electrophoretically transferred to a membrane which is usually nitrocellulose or polyvinylidene fluoride (PVDF) for visualization. The membranes are soaked with a blocking reagent, either 3-5% bovine serum albumin (BSA) or skim milk in tris-buffered saline (TBS), to prevent any non-specific binding. The blot is then incubated with the primary antibody, after that, the peroxidase-conjugated secondary antibody is added against the target antibody for the purpose of detection (68). Then the colored solution is added to visualize the binding bands. The allergencity of a protein can be also studied using the immunoblotting technique. Specifically, the blot of proteins incubated with sensitized patients' sera that include an IgE has been provoked by the exposed allergen.

1.3. Protein sequencing using mass spectrometry

1.3.1. Protein sequencing

For any species, each protein has a unique amino acid sequence, which can therefore be used to identify it specifically. This sequence information helps to understand the other proteins' structural features and functional activity. In addition, the amino acid sequence directly corresponds to the DNA sequence of the corresponding

gene (s), which helps to study the activity of proteins by detecting the common mutations at the molecular level (64).

In protein identification by the amino acid sequencing, the identity of each amino acid residue and its position within the protein must be determined. The selected method should be sensitive enough to use a small amount of material and rapid enough to determine the amino acid sequence in a timely manner. In addition, it should provide reliable information with sufficient mass accuracy, and high throughput (64). Proteins are large molecules which can be digested with proteolytic enzymes to produce a large number of peptides. The peptide sequence information can be combined to obtain the whole protein sequence. However, the chemical diversity of these peptides makes the analysis challenging (69).

1.3.2. Edman degradation

Edman degradation is the oldest method used for protein sequencing. The Edman reaction is applied to the protein N-terminus, where the amine side of the N-terminal amino acid reacts with phenylisothiocyanate (PITC) to form a phenylthiocarbamyl (PTC) protein, after which trifluoroacetic acid is used to cleave the PTC, which then gives phenylthiohydantoin (PTH) amino acid derivatives as final products. These final products have different chemical structures, depending on the N-terminal amino acid, and can subsequently be separated by HPLC and compared to a standard to identify the N-

terminal amino acid. By this reaction, the N-terminal amino acid is removed from the protein, and produces a cleaved derivative of that amino acid along with a protein that is shortened by one amino acid and has a new N-terminus. This shortened protein is re-exposed to the same procedure for removing the next amino acid (70).

Edman degradation is a part of modern protein analysis because of its ability to sequence a protein in its intact form. Furthermore, Edman sequencing has many advantages, making it the main technique used worldwide in last few decades before the use of mass spectrometry took over this kind of analyses. Advantages include the sensitivity, which is sufficient for analyzing proteins detected by Coomassie staining, the ease of operation, and the clarity of the data. However, this technique has a number of serious drawbacks compared to tandem mass spectrometry that make it a poor technique for proteomics work. Additionally, the time required to accomplish each cycle is very high (~45 min/amino acid) and it cannot be used with proteins having blocked N-termini. Tandem mass spectrometry is the technique of choice in this study and it will be discussed later (64).

1.3.3. Peptide mass fingerprinting (PMF)

Mass spectrometry (MS) has become a major analytical tool for protein structural analyses. Early on, the major problem in MS analyses was the need for generation of gas phase ions from peptides. The development of ion sources such as matrix-assisted laser

desorption ionization (MALDI) and electrospray ionization (ESI), along with accurate mass analyzers, offered the biochemist a valuable tool for protein analyses. Mass spectrometry not only offered information about the peptide and/or protein molecular weight, but it also gave details about the primary structures of these species in more accurate detail. Following the completion of the early genome sequencing projects, several search engines (i.e. Mascot and Sequest) were developed for the genome databases with mass spectrometry interfaces. The enzymatic digestion of a protein generates a highly unique peptide profile in mass spectrometry. Once these peptides are introduced to a mass spectrometer a fingerprint profile will be generated, which is known as a peptide mass fingerprint (PMF). This technique is highly important for protein identification. There are several factors that control the precision of the database search for instance the resolving power of the MS, type of protease, availability of the gene data in data bank, and the purity of the target proteins (71).

A single mass analyzer was able to provide '*de novo*' sequence information for the purified peptide by combining Edman degradation with mass spectrometry. This technique is known as peptide ladder sequencing, which works by chemically degrading the N-terminal amino acid from the targeted peptide using Edman reagent. Equal amounts of the starting peptide material should be collected before each cycle. Consequently, these collected materials are combined to each other and introduced to MS. The mass difference between the consecutive ion represents loss of an amino acid (64).

1.3.4. Tandem Mass spectrometry

Using tandem mass spectrometry, the fragmentation of gaseous peptide and protein ions gives a series of ions, which can be manipulated to give the amino acid sequence. This development in protein mass spectrometry was extended to give high throughput protein identification, '*de novo*' sequencing, and identification of post translational modifications (71).

Protein tandem mass spectrometry-based techniques are divided into two major types; '*bottom up*' or '*top down*'. These terminologies are assigned for the point of sample handling before the mass spectrometry analysis and from the data processing and sequencing manipulation (64).

1.3.4.1. '*Bottom up*' approach

In the '*bottom up*' approach, the enzymatically-digested proteins are subjected to MS analyses. The precursor ions of the peptides are separated in the first mass event, and then fragmented to produce a series of informative product ions. These product ions for each single peptide are then analyzed in the second mass event (64).

A distinctive series of ions of the product ion spectra that relate directly to the amino acid sequence are extracted from the product ion spectra and then compared with '*in-silico*' ions that are generated theoretically from the DNA databases by computers.

The sequence will be assigned when the both series of ions, experimental and theoretical, are matched perfectly as shown in peptide number 3 (Figure 1.2).

This technique usually provides useful information about the post-translational modifications (i.e. phosphorylated, acetylated). Some of these post translationally modified residues are difficult to be identified due to poor chromatographic resolution or detector sensitivity (72), as seen for the two peptides number 2 and 5 (Figure 1.2). On the other hand, introducing the whole protein to the mass spectrometer without any digestion, *'top down'*, is a very useful technique to identify most of the labile post translation modification groups with high amino acid sequencing coverage (73).

Enzymatic digestion of the intact protein following purification and electrophoretic separation prior to MS analyses is routine but tedious work for *'bottom up'* approach. The enzymatic peptides are normally separated by reversed phase chromatography and subjected to electrospray ionization (ESI) or matrix-assisted laser desorption/ ionization (MALDI) with analysis a tandem mass spectrometry. The sequence of each peptide is deduced from the product ion spectra after gas phase collision induced dissociation (CID) (64). The *'bottom up'* approach can take place at two levels of mass analyses. The first is known as peptide mass fingerprinting (PMF), which determines the masses of the intact peptides. These masses result in a fingerprint spectrum related to a specific protein. The other level is known as peptide sequencing or fragment fingerprinting (PFF), which is based on isolation of each peptide's molecular

ion in the first mass filter followed by the collection of the product spectrum after CID fragmentation (64) as described above. The amino acid

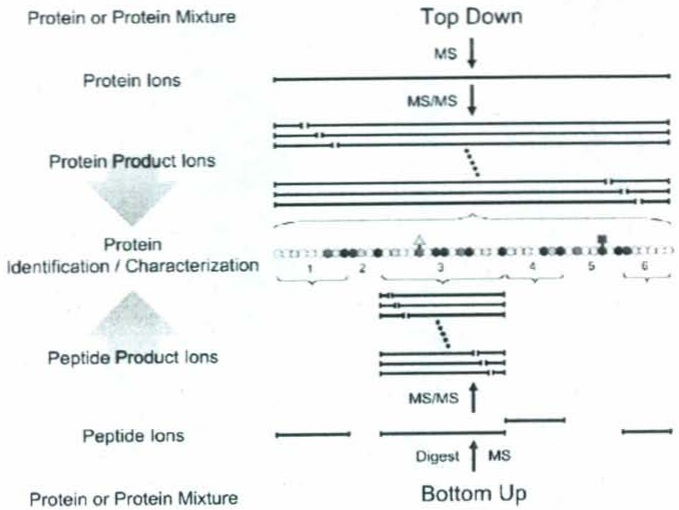


Figure 1. 2: Schematic overview of *'bottom-up'* and *'top-down'* approaches employed for tandem mass spectrometry-based protein identification and characterization. The shaded circles represent the amino acids of a protein, while the shaded triangle and rectangle represents post-translational modifications. (Reproduced with permission; Cherperel, G and Reid, G. E. *The Analyst*. **2007**, 132, 500-506).

sequence of the isolated peptide can be manipulated from the product spectrum by forming a useful fragmentation ladder and comparing it with DNA-based databases (e. g. NCBInr).

Generating a good sequencing coverage for the intact protein requires producing peptides that are readily soluble for chromatography. The chromatographic separation of peptides prior to MS ionization is very important to generate high energy gas-phase ions inside the field free region required for facile fragmentation or dissociation. Sequencing the intact protein using a regular tandem mass spectrometer is highly limited to the energy of the dissociation and the resolving power of the mass analyzer, which lead to erroneously manipulation from the algorithms of the database. The labile post-translational and environmental modifications of proteins are difficult to be identified using energetic dissociation *'bottom up'* method. As well as, the modified peptides are more sensitive to be detected by MS than the large hydrophobic proteins (51).

The '*bottom up*' approach can usually deal with most types of proteins regardless of their size or function. This approach is still the most common, but the sequence coverage is limited (51). The approach also has difficulty dealing with a protein mixture and identifying the post translation modification groups (74).

1.3.4.2. *'Top down'* approach

Manipulating the *'bottom up'* approach data, to give the sequence for the target protein, requires data availability in the gene bank. Thus, the *'top down'* approach was found to improve the protein identification even in case of the lack of genetic information in the databases. Understanding the chemical composition, molecular weight, and stability/solubility of the target protein is paramount for achieving proper ionization and gas phase dissociation in the mass spectrometer. Kellher *et al.* (75) introduced the *'top down'* approach by analyzing an intact protein using an ESI ion source. This type of analysis requires a high resolution mass analyzers as well as a high energy collision cell. The amino acid sequence, and post translation modification locations and their structures were determined in this approach using *'in-time'* tandem mass spectrometry experiments. The mass difference between the generated precursor ions and their equivalent predicted ions in the cDNA-based databases were considered for amino acid sequencing of the intact protein, as detailed in Figure 1.3 (76).

The *'top down'* approach was further developed for the analysis of a mixture of proteins without classical separation (chromatography or electrophoresis). Whereby the precursor ion of the target protein is selected and isolated in the first mass analysis event for amino acid sequencing (76). This approach still offers the comprehensive mapping features for post translation modification identification. Several bioinformatics tools can be implemented to map the post translation modification information in protein, which are mainly defined by the mass differences between the experimental and theoretical

spectral data. For proteins <40 kDa the molecular weight error measurement is usually no more than 1 Da using Fourier-transform mass spectrometry (FTMS) instruments (75). This level of resolution is high enough to identify a disulfide bond, which increases the molecular weight of the protein by 2 Da after reduction. The *'top down'* method has clearly been shown to identity even the labile post-translation modifications (72). Two-dimensional correlation between MS/MS and MS/MS/MS was also developed to confirm the amino acid sequence in FTMS instruments (77).

The *'top down'* approach affords another valuable and powerful tool for proteomic analyses. Post translation modification mapping can become a reliable tool for determining chemistry, location, and quantity of these modifications along with the primary structure of a protein. Using this method as the sequencing approach for an intact protein can provide close to 100% coverage. Moreover proteins can be part of a mixture, or have unusually large molecular weight. Analysis of these types of samples will still give the same accuracy and with less complexity compared to the *'bottom up'* approach. Simple algorithm databases with more sophisticated data processing are readily available for *'top down'* approaches. Including powerful and advanced MS instruments such as FT-ICR allowing for maximum mass accuracy and resolution reinforces the strength behind this approach. But the high price of the FT-ICR and corresponding maintenance /operating and infrastructure costs are the only limitations for the *'top down'* approach to be used.

Larger sample size to maintain the generation of ions in the ICR cell, and longer analysis times are the major limitations but this is encouraging researchers to build up a more high throughput *'top down'* approach with sophisticated data analysis. This is one of the few MS tools which offer both a comprehensive study for protein structure and PTM mapping, an important tool to help us to understand more complex protein processes such as cell signaling virus-cell interaction, or cancer cell development.

1.3.5. Peptide fragmentation

Tandem mass spectrometry provides useful structural information of a series product ions generated from the fragmentation of selected molecular ions in the collision cell. Subsequently, these fragment ions provide structural information which are combined together to give a complete structure of the target molecule. Accordingly, tandem mass spectrometry is the only technique that can provide this kind of information, whereby the first mass event separates the precursor molecular ion to be fragmented, and then the resultant fragments are analyzed in the second mass event giving informative spectrum.

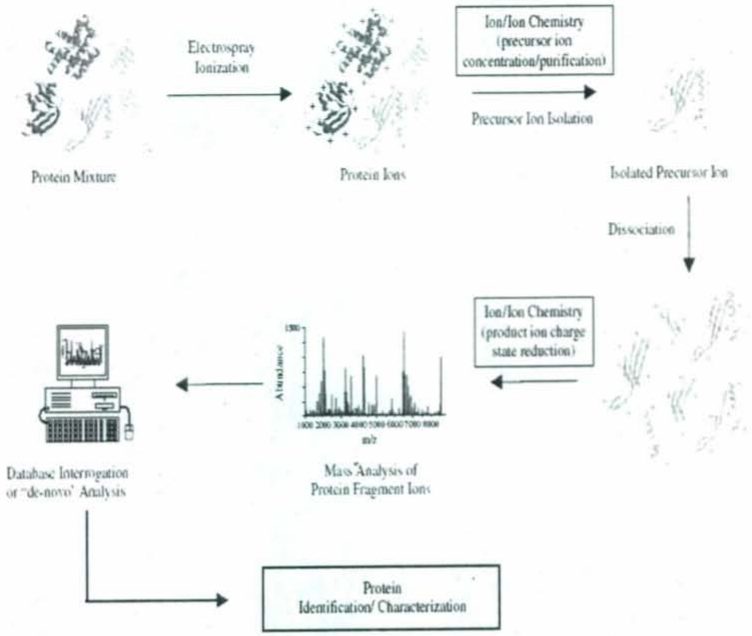


Figure 1. 3: A schematic diagram of typical *'top down'* protein characterization experiment. (Reproduced, with permission; Reid, G. E.and McLuckey S. A. *J Mass spectrom.* **2002.** 37; 663–675).

The multiply charged (ESI) or singly charged ions (MALDI) are analyzed by tandem mass experiment. The accelerated ions exit the first mass event with a specific kinetic energy (KE). The first mass event selects the chosen precursor ion which collides with neutral gas molecules, where the KE is transformed to the vibrational energy, which

then equally distributes among the molecule bonds. Once the internal energy is significantly higher than the barrier of bonding energy; the bond will break to give fragments. Since the weakest bond in the peptide structure is the peptide linkage, this bond will preferentially break to produce y and b ions, with respect to Beiman nomenclatures (78), as the two major ion series. However, in high collision energy cells, there are three other pathways of fragment ion formation (Fig 1.4).

Tandem mass spectrometry is performed by either a combination between any two mass analyzers (by space) or having an ion trap-based mass analyzer (by time). At high-energy CID, a huge amount of kinetic energy will be used to generate the full informative ions from the protein molecules. This kind of collision cell is only available in magnetic sectors, and TOF instruments, where the kinetic energy reaches around several keV (79). Similarly, peptide fragmentation can also be observed by the post-source decay technique in MALDI-TOF. Indeed after MALDI ionization, the produced ions have enough energy (metastable ions) to be fragmented during flight time and produce informative ions. These product ions are separated after reaching the reflectron with respect to their m/z (80).

Protein ion dissociation techniques

The fragmentation of peptides in mass spectrometry collision event occurs at three types of chemical bonds in the protein structure. These bonds, Cα-C, C-N, or N-Cα, yield six types of fragments that are respectively labeled a_n, b_n, c_n when the positive charge is on the N-terminus and x_n, y_n, z_n, when the positive charge is on the C-terminus as shown in Figure 1.4. The subscript n is related to the number of amino acids in the peptide sequence. The difference between the mass of consecutive ions within the same series is related to a specific amino acid, which is extended to deduce the amino acid sequence of the targeted peptide or protein (79).

The bonding energy of the chemical bonds varies with respect to the type of linkage. The minimum energies required to break the chemical bonds are 335 kJ/mol, 8-17 kJ/mol, and 272 kJ/mol for Cα-C, C-N, and N-Cα bonds, respectively (81). In CID, the peptide linkage being the weakest is broken first as long as the acquired internal energy is high enough. But other peptide fragmentations can be achieved selectively using special activation techniques associated with FTICR; infrared multiphoton dissociation (IRMPD) and electron capture dissociation (ECD). These dissociation techniques can be used to fragment peptide ions in the ICR cell by using laser (in IRMPD) or electron beam (in ECD). Targeted proteins can be ionized by either MALDI or ESI sources, but ESI is the only ion source that gives useful multi-charged ions. The activation and dissociation of multi-charged ions in the field-free regions require less KE

to generate useful fragmentation pattern for protein sequencing. Different activation techniques for peptide ions have recently been developed (76). The most popular one is collision induced dissociation (CID), and there are others such as surface-induced dissociation, infrared multiphoton dissociation (IRMPD), or blackbody infrared dissociation (BIRD). These techniques are all efficient methods to activate the protein ions to above the dissociation threshold (76, 82). This is accomplished by increasing the internal energy slowly over milliseconds to above the dissociation threshold by applying energy to a multi-charged precursor protein ion. Non-covalent bonds which maintain the tertiary structure of the proteins (energy around 4-20 kJ/mol) are the first to be dissociated, followed by the weakest of the covalent bonds (82).

Electron-based activation methods have also been invented for the ion trap mass analyzers. Electron capture dissociation (ECD) was developed to fit the FTICR, which cleaves the inter-residue bonds at different positions along the protein backbone. The fragmentation mechanism of the electron-based techniques is quite different than the other activation techniques. The multi-protonated protein ions inside the trapping cell are activated by a beam of low energy electrons (<1 eV) at the sites with highest electron affinity, whereby the vibrational energy is increased enough to produce a radical ion. The protein's excited radical ion is rapidly cleaved at specific bonds (83). By this approach, the hypervalent species (RNH_3•) generates two series of complimentary fragment ions, c_n and z_n• (Fig. 1.4).

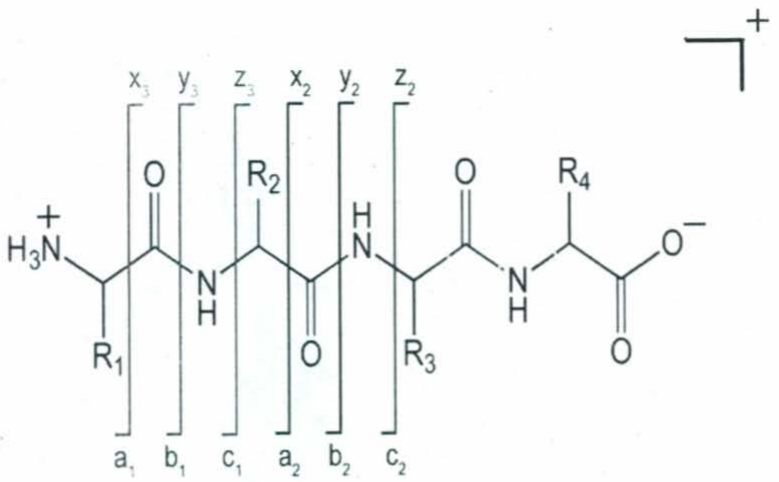


Figure 1. 4: Types of fragmentation ions of peptides in tandem mass spectrometry. When the charge is retained on the C-terminal the ions' types are a, b, c, and on the N-terminal x, y, z. (Reproduced with permission; Roepstorff, P. and Fohlman, J. *Biomed Mass Spectrom.* **1984.** 11; 601-601).

In the ergonic processes i.e., CID or IRMPD, the peptide linkage is the first target for dissociation because it has the lowest energy barrier and thus the resulting spectra predominantly produce b_n and y_n ions. In contrast, in ECD, the N-αC linkages on the peptide backbone are the ones which rapidly dissociate through radical ions to yield c and $z^{\bullet}$ (~90%) or $a^{\bullet}$ and y (~10%) ions as complementary pairs of fragment ions. Side chain

dissociations are negligible and the ions produced from the polypeptide backbone are most informative in protein sequencing.

In ECD, an emitter cathode produces the beam of low energy electrons. However it is not compatible with instruments such as the QToF, conventional ion traps, or the Orbitrap due to the affect of the strong electric field that is produced from these analyzers on the movement of the electron (from the ECD) inside the trap. However, an ECD-like activation technique has developed to be used in ion traps (e.g. Orbitrap). Electron transfer dissociation (ETD) uses gas phase ion/ion chemistry to transfer the low energy electron from singly charged anthracene anions to multiply charged protein ions. The ECD and ETD are readily used to identify the labile PTM groups that have the bond dissociation barrier less than the peptide linkage. A combination of CID and ECD can increase the overall efficiency for sequence coverage, partially as this can specify post translation modification locations, or preserve the PTM groups during tandem MS analysis (76).

1.4. Bioinformatics tools

The proteomics shotgun-approach basically commences with enzymatic digestion of the target proteins into their peptides. To simplify the complexity of the peptide mix, this mixture can be fractionated using multidimensional liquid chromatography (LC) followed by tandem MS. The first mass analyzer selects, in sequence, the molecular ion

of the individual peptides. The peptide molecular ions are subjected to CID MS analysis for peptide sequencing. After obtaining product ion mass spectra, computational approaches are utilized to ascertain the peptide sequence and later the protein identity. Most of the available computational programs begin by manipulating the amino acid sequence of each peptide, and then these data (experimental spectra) are compared with theoretical spectra, generated for each protein sequence available in the database (84). Several bioinformatics tools are used to transform the mass spectrometry output to give meaningful biological knowledge.

1.4.1. Mascot

The MS/MS database search tools work in a similar manner, where the MS/MS experimental spectra are used as input and compared against the theoretical fragmentation patterns of the peptides that are available in the cDNA databases (e.g. National Center for Biotechnology Information (NCBI)). Mascot is one of the most popular database search engines. The experimental data are evaluated in three different scores; Mascot ions score (MIS), Mascot identity threshold (MIT), and Mascot homology threshold (MHT) (85). The MIS represents the similarity between experimental spectra and theoretical spectra and it is identified as $-10\log_{10}(P)$, where P is the probability that a peptide spectrum match is a random event. The algorithm employed to calculate the P value is unknown to the user. The identity threshold factor is also defined as $-10\log_{10}$

(20p/qmatch), where qmatch is the number of candidate peptides with masses close to the precursor ion in a database, and p is the significant threshold. For controlling the significant outlier peptides, a lower empirical threshold, (Mascot homology threshold), was invented by Mascot, but this algorithm remains as a blind stage for the users (86).

1.4.2. Basic Local Alignment Search Tool (BLAST)

Basic local alignment search tool (BLAST) is a computational tool used to find the local alignment between a query sequence and its target. This method begins by "sowing" the search with a query word, which is a small subset of letters from the query sequence. The BLAST finds not only this certain query word but also related words where conservative substitutions have been introduced. The scoring matrices develop what is called "the neighborhood", which is used to determine the relationship between the examined word and the original query word. Some cutoff is used to control the words that are closely related to the original query word. The neighborhood score threshold (T) controls this cutoff, which is determined automatically by the BLAST program. Increasing this value by the user can push the search toward exact matches and speed up the search process (87).

The BLAST method for local alignment provides several outputs to express the degree of matching, mismatching, and gaping. The visual section of the BLAST report shows a color key for alignment scores and a list of colored bars with their distinctive

color related to their alignment scores against the query sequence. Another list of the alignments appears immediately after the visual section, which gives the accession number, the score, and the E value for each hit. The E value represents the number of high-scoring segment pairs that would be expected purely by chance (87). While browsing each specific hit, a more detailed report is shown with the E-value, the number of identity (exact matches), the number of " positives"(exact matches and conservative substitutions), and the number of residues that fell into a gapped region (number of amino acid residues that remained unaligned because of the introduction gap) (87).

1.4.3. NetPhos 2.0 server

NetPhos is a tool used for evaluating potential phosphorylation sites at serine, threonine, or tyrosine in a query protein sequence using the mass spectrometric data (88). This information is highly significant while searching for a phosphorylation-free signature peptide for quantitation purposes. This evaluation is very important prior to selecting, then synthesizing the signature peptide as an analytical standard for developing an isotopic dilution MS method (89).

1.4.4. NetAcet 1.0 server

Predicting the acetylation in the N-terminal residue is also important, particularly in eukaryotes, where blocked N-terminals are very hard to be ionized, subsequently affecting the '*de novo*' sequence coverage of the N-terminal peptide (37). The NetAcet tool produces a very useful algorithm to evaluate these sites which was initially developed using yeast proteins (90).

1.5. Protein quantification by mass spectrometry

Studying the large scale of proteins belonging to a certain biological system is known as proteomics. In recent decades, elegant MS-based proteomics techniques have developed for protein analysis and identification with high throughput. Identification tools using tandem mass spectrometry have been very influential in field of protein science as the genome sequence projects were being completed. The discovery and development of ESI and MALDI sources has given protein mass spectrometry a huge push to become one of the dominating fields in proteomics. At its inception, MS proteomics research was mainly a qualitative discipline, where the outputs were limited to a list of identified proteins without any further information about abundance, distributions, or stoichiometry (91). In contrast, quantitative strategies were involved in analyzing the gene expression by microarray technology, real time polymerase chain reaction (PCR), or evaluating the enzymes' activity that directly represent their quantity.

Several MS-based tools are being developed for proteomics quantification, which help in characterizing the proteome complexities. For example, quantitative data can assist in the study the true protein interactions (interactoms) of a given "bait" protein over a "background" one. These proteomics strategies are used to differentiate between stressed and normal samples or between 'knock out' and wild types for certain biological systems.

1.5.1. Proteomics quantification

Quantitative proteomics approaches are used as part of the larger framework of the available techniques for studying regulatory processes in the living cell. Choosing the quantification method is less important than the methods practical aspects which include good technical reproducibility. By the same token, the biological system variations must be considered in the experimental design, along with financial costs. Since there is no amplification step at the protein level, the sample size is the most important limiting factor affecting the strategy of quantitative differential proteomics. Accordingly, enrichments and purification of protein are carried out to meet the sensitivity of the quantitation technique. In this way, the sample size will influence the choice of analytical method (91).

The differences in the physical and chemical properties of different peptides, directly affect the proton affinity and the MS signal intensity. These phenomena makes

the LC-MS/MS method not as absolute quantitative as such. These peptide variables are their charge state, peptide length, amino acid composition, and any post translational modification. These structural variations contribute to the peptides' ion intensities even if they belong to the same protein. Because of this, many MS-based quantitative methods produce only relatively quantitative data, which is based mainly on the MS response in contest with its sample matrix. Accordingly, careful experimental design and suitable data analysis are needed to obtain the desired analytical information.

The relative quantitation MS-based techniques are categorized into two major groups; stable-isotope-labeling and the label-free approach. The relative standard deviation for most stable-isotope-labeling techniques is below 10% and the accuracy is approximately 30%, based on peak intensities or extracted ion chromatograms. However, the precision obtainable from label-free approaches, based on spectral counting or derived indices, is as high as 50% RSD (91).

The label-free quantification methods are based mainly on precursor signal intensity, which requires a high precision mass spectrometer. The label-free approaches are inexpensive (no labeled material to purchase) and can be applied to any biological material. The proteome coverage for quantified proteins can be very high, because any protein with one or more identified peptides can be quantified. Thus the sample complexity does not normally increase by mixing different proteomes together. Hence the protein expression can be quantified between different biological samples from different

treatments. These samples can be analyzed by the free-label MS approache due to its analytical power and flexible dynamic range (92). There are two main label-free approaches; protein-based methods (spectral count or derived indices) and peptide-based methods (ion intensities and protein correlation profile). The protein-based method look to the protein expression levels among proteins of the same sample and compares them with another sample. In contrast, the peptide-based method utilizes the averaged and normalized ion intensities of the identified peptides (of specific proteins) for quantification (91).

The isotopic-labeling methods are classified with respect to the labeling strategy; metabolic labeling and isobaric mass tags. The metabolic labeling approach introduces the whole cell or organism to a labeled culture media (e.g. stable-isotope labeling by amino acid (SILAC) in cultured cells). On the other hand, chemical labeling (isobaric mass tags) can be attached to the protein or proteolytic peptides by a chemical reaction. Equivalent quantities of the labeled and unlabeled samples are mixed and treated in the same manner prior the MS analysis. The differences between the labeled peptides' intensities and those of unlabelled represent the levels of that specific protein in the original sample. The drawback in this method is interference by co-eluting isobaric compounds when tandem MS is not used.

The isobaric mass tags method can overcome any co-eluting interferences by developing reagents that co-elute with the target peptide ion, then they will have a

significant resolution after fragmentation in the product ion spectra. Each product ion spectrum is the result of a different chemical tag with its intensity equivalent to the level of protein expression in the original sample. These fragments (tags) are in the low mass range, which make interferences from other peptides' fragments difficult. There are some commercial isobaric reagents that offer this type of quantification strategy such as the isobaric tag for relative and absolute quantitation (iTRAQ) and the tandem mass tag (TMT).

1.5.2. Absolute quantitation and isotopic dilution mass spectrometric techniques

A technique called absolute quantification (AQUA) using the stable-isotopic-labeled technique was reported for the first time by Desiderio and Kai (1983) (93). By using selected reaction monitoring (SRM) in triple quadruple tandem mass spectrometer, the AQUA became a standard in protein quantification (91). The combination of information, specifically the retention time, peptide precursor ion, and the fragmentation profile gives this method high specificity toward the particular target peptide. The signal-to-noise ratio in SRM is very high, which gives an extended linear dynamic range up to 5 orders of magnitude. By spiking a protein sample with a known amount of the stable-isotopic-labeled peptide, the concentration of the native protein can be calculated. However, this may not cover all levels of the expressed protein in the tissue, because sample preparation steps may lead to loss or enrich of the target peptide that are not

addressed by AQUA method. Improving the reliability of the AQUA approach will be discussed in Chapter 4.

The absolute quantification of large proteins (MW>15kDa) by mass spectrometry can be performed using the *'bottom up'* approach. The intact protein is digested and the signature peptide is selected to represent the target protein. The stable-isotopic-labeled and natural (light) forms of the signature peptide are synthesized, where the heavy form is used as internal standard. Before MS analysis, the heavy form of the signature peptide is added to the sample; hence the method is called "isotope dilution" MS. Small proteins (<15kDa) can also be quantified using the *'top down'* approach, whereby the internal standard is prepared as a metabolic labeling strategy using labeled culture media (94).

1.5.3. Validation of bioanalytical methods

Once the signature peptide of the target protein is determined and evaluated, standard peptides are chemically synthesized in both light and heavy forms. These standard peptides are used to develop a quantification method using liquid chromatography tandem mass spectrometry (LC-MSMS). The developed method is validated to successfully meet the minimum standards of the Food and Drug Administration (FDA) guidelines (95) for accuracy, precision, selectivity, sensitivity, reproducibility, and stability. This guideline was specifically developed for studying the bioanalytical method validation of small molecules or drugs. In this study, a modified

protocol is developed to study the AQUA (or stable isotope dilution) method for allergen quantification using multiple reaction monitoring (MRM) (89).

1.6. Thesis objectives and future work

In 2003, a comprehensive study for evaluating the factors that play a crucial role in crab occupational asthma in Newfoundland was initiated by Memorial the SafteyNet Centre for occupational health and safety research. The outcome of their study showed the importance of the crab asthma in the crab industry, which motivated our group to start a study to identify the most common crab allergens and evaluate their levels in the workplace environment. Mass spectrometry was the technique of choice in this project, whereas the SafteyNet project was using immunological-based techniques for allergen determination. This thesis will identify the main protein allergens that are primarily responsible for crab asthma and propose a novel sensitive and specific quantification method for airborne allergen using mass spectrometry. The MS approach will offer occupational health and safety (OHS) researchers the opportunity to manage and evaluate the threshold of crab allergen exposure and to upgrade the indoor air quality in processing plants.

Thesis objectives

1) Characterization of the major allergens in both snow crab (*Chionoecetes opilio*) and black tiger prawn (*Penaeus monodon*). The crude protein extract will be profiled on gel electrophoresis. The immunological activity of each gel-separated protein will be studied using immunoblotting with the sensitized patinets' sera.

2) The active proteins will be characterized by determining the primary structure (amino acid sequence) using mass spectrometry.

3) A signature peptide that represents its protein stoichiometrically will be determined for each target protein. High performance analytical techniques such as liquid chromatography and tandem mass spectrometry will be used in this study.

4) The synthetic signature peptides will be used to develop a quantification method using LC multi-reaction monitoring tandem mass spectrometry. A full validation and optimization study will be accomplished for measuring the levels of both TM and AK in aerosolized forms.

5) Real air samples will be collected from a simulated crab processing plant. These samples will be analyzed using the developed mass spectrometry method.

1.7. Co-authorship statement

All experimental work was performed by the principle author (AAR) except the immunoblotting against antibodies and patients' sera as found the figures 2.1, 3.1, and 5.1. These were performed in Dr. Lopata's and Mr. Kamath's laboratory (RMIT University, Australia).The purification of TM from black tiger prawn using ion-exchange chromatography (Figure 5.3) was also done by Mr. Kamath. The phylogenetic tree in figure 3.3 was also generated by Dr. Lopata group.

The principle author prepared the draft of the manuscripts based on chapters 2, 3, 4, and 5 after intellectual discussion with the other coauthors, the he toke the role to reply on reviewers/editors concerns by Dr. Helleur, Dr. Lopata and other co-authors.

1.8. References

1. Laurianne, W. G., Lehrer, S. B. Fish and shellfish allergy. *Curr Allergy Asthma Rep*. **2005.** 5; 74-79.

2. Food and Agriculture Organization of the United Nations. 2007 Core consumption data.Available at: http://faostat.fao.org/site/610/DesktopDefault.aspx?PageID=610#ancor. Accessed March 28, 2011

3. Gill, B. V., Rice, T. R., Cartier, A., Gautrin, D., Neis, B., Horth-Susin, L., Jong, M., Swanson, M., Lehrer, S.B. Identification of crab proteins that elicit IgE reactivity in snow crab–processing workers. *J. Allergy Clin. Immunol*. 2009. 124;1055-1061.

4. Lopata, A. L, Lehrer, S. B. New insights into seafood allergy. *Curr Opin Allergy Clin Immunol*. 2009. 9; 270-277.

5. Malo, J-L, Cartier, A. Occupational reaction in the seafood industry. *Clin Rev Allergy*. **1993**. 11; 223-240.

6. Bang, B., Aasmoe, L., Aamodt, B. H., Aardal, L., Andorsen, G. S., Bolle, R. Bøe, R., Van, T. D., Evans, R. Florvåg, E., Gram, I. T., Huser, P. O., Kramvik, E. Løchen, M-L., Pedersen, B., Rasmussen, T. Exposure and Airway Effects of Seafood industry workers in Northern Norway. *J Occup Environ Med*. **2005**.47; 482-492.

7. Rosenstock, L., Cullen, M.. Textbook of clinical occupational and environemntal medicine. *W. B. Saunders Company*, **1994**. 1st edition.

8. Greenberg, M. I., Hamilton, R. J., Phillips, S. D., McCluskey, G. J. Occupational, industrial, and environmental toxocology. *Philadelphia, Pennsylcania : Mosby*, **2003**. 2nd edition.

9. Beck, H., Kundsen, B. Contact urticaria to commercial fish in atopic persons. *Acta Derm (Stockh)*. **1983**. 63; 257-260.

10. Medicine dictionary. http://www.medterms.com/script/main/art.asp?articlekey=31081. Accessed March 28, 2011

11. Jeebhay, M. F., Cartier, A. Occupational seafood allergy: a review. *Occup Environ Med.* **2008**. 58; 553-562.

12. Cartier, A., Malo, J. L., Ghezzo, H., McCants, M., Lehrer, S. B. IgE sensitization in snow crab-processing workers. *J. Allergy Clin. Immunol.***1986**. 78; 344-348.

13. Cartier, A., Malo, J. L., Forest, F., Lafrance, M., Pineau, L., St-Aubin, J. J., Dubois, J.Y. Occupational asthma in snow crab-processing workers. *J. Allergy Clin. Immunol.* **1984**.74; 261-269.

14. Tomaszunas, S., Weclawik, Z., Lewinski, M. Allergic reactions to cuttlefish in deep-sea fishermen. *The Lancet.* **1988.** 14; 1116-1117.

15. Rodríguez, J., Reaño, M., Vives, R., Canto, G., Daroca, P., Crespo, J. F., Vila, C., Villarreal, O., Bensabat, Z. Occupational asthma caused by fish inhalation. *Allergy*. **1997**. 52; 866-869.

16. Siracusa, A., Bettini, P., Bacoccoli, R., Severini, C., Verga, A., Abbritti, G. Asthma caused by live fish bait. *J. Allergy Clin. Immunol*. **1994**. 93; 424-430.

17. Vierk, K. A., Koehler, K. M., Fein, S. B, Street, D. A. Prevalence of self-reported food allergy in American adults and use of food labels. *J. Allergy Clin. Immunol.* **2007**. 119; 1504-1510.

18. Gautrin, D., Cartier, A., Howse, D., Horth-Susin, L., Jong, M., Swanson, M., Lehrer, S., Fox, G., Neis, B. Occupational asthma and allergy in snow crab processing in Newfoundland and Labrador. *Occup Environ Med.* **2010**. 67; 17-23.

19. Hoffman, D. R., Day, E. D., Miller, J. S. The major heat stable allergen of shrimp. *Ann. Allergy*. **1981.** 47; 17.

20. Shanti, K. N., Martin, B. M., Nagpal, S., Metcalfe, D. D., Rao, P. V. Identification of tropomyosin as the major shrimp allergen and characterization of its IgE-binding epitopes. *J. Immunol.* **1993**. 151; 5354-5363.

21. Motoyama, K., Suma, Y., Ishizaki, S., Nagashima, Y., Shiomi, K. Molecular cloning of tropomyosins identified as alleregns in six species of crustaceans. *J. Agric. Food. Chem.* **2007**. 55; 985-991.

22. Yu, C. J., Lin, Y. F., Chiang, B. L., Chow, L. P. Proteomics and immunological analysis of a novel shrimp allergen, Pen m 2. *J. Immunol.* **2003.** 170; 445-453.

23. García-Orozco, K. D., Aispuro-Hernández, E., Yepiz-Plascencia, G., Calderón-de-la-Barca, A. M., Sotelo-Mundo, R. R. Molecular characterization of arginine kinase, an allergen from the shrimp Litopenaeus Vannamei. *Int Arch Allergy Immunol.* **2007.** 144; 23-28.

24. France, R. M., Sellers, D. S., Grossman, S. H. Characterization, and hydrodynamic properties of arginine kinase from gulf shrimp (Penaeus aztecus). *Arch. Biochem. Biophys.* **1997**. 345; 73-78.

25. Yao, C. L., Wu, C. G., Xiang, J. H., Dong, B. Molecular cloning and response to laminarin stimulation of arginine kinase in haemolymph in Chinese shrimp, Fenneropenaeus chinensis. *Fish Shellfish Immun.* **2005**. 19; 317-329.

26. Ortea, I., Cañas, B., Gallardo, J. M. Mass spectrometry characterization of Species-Specific peptides from arginine kinase for the identification of commercially relevant shrimp species. *J. Proteome Res.* **2009**. 8; 5356-5362.

27. Bi, X. Z., Chew, F. T. Molecular, proteomic and immunological Characterization of isoforms of arginine kinase, a cross-reactive invertebrate pan-allergen, from the house Dust Mite, dermatophagoides farinae. *J Allergy Clin Immunol.* **2004**. S226.

28. Binder, M., Mahler, V., Hayek, B., Sperr, W. R., Schöller, M., Prozell, S., Wiedermann, G., Valent, P., Valenta, R., Duchêne, M. Molecular and immunological characterization of arginine kinase from the indianmeal moth, Plodia interpunctella, a novel cross-reactive invert Pan-allergen. *J Immunol.* **2001.** 167; 5470-5477.

29. Liu, Z., Xia, L., Wu, Y., Xia, Q., Chen, J., Roux, K. H. Identification and characterization of an arginine kinase as a major allergen from Silkworm (Bombyx mori) larvae. *Int Arch Allergy Immunol.* **2009**. 150; 8-14.

30. Ayuso, R., Grishina, G., Bardina, L., Carrillo, T., Blanco, C., Ibáñez, M. D., Sampson, H. A., Beyer, K. Myosin light chain is a novel shrimp allergen, Lit v 3. *J. Allergy Clin. Immunol.* **2008**. 122; 795-802.

31. Taylor, S. L. Molluscan shellfish allergy. *Adv Food Nutr Res.* **2008**. 54; 139-177.

32. Jenkins, J. A., Breiteneder, H., Mills, E. N. Evolutionary distance from human homologs reflects allergenicity of animal food proteins. *J Allergy Clin Immunol.* **2007**. 120; 1399-1405.

33. Beale, J. E., Jeebhay, M. F., Lopata, A. L. Characterisation of purified parvalbumin from five fish species and nucleotide sequencing of this major allergen from Pacific pilchard, Sardinops sagax. *Mol Immunol.* **2009**. 46; 2985-2993.

34. Sakaguchi, M., Toda, M., Ebihara, T., Irie, S., Hori, H., Imai, A., Yanagida, M., Miyazawa, H., Ohsuna, H., Ikezawa, Z., Inouye, S. IgE antibody to fish gelatin (type I collagen) in patients with fish allergy. *J Allergy Clin Immunol.* **2000**. 106; 579-584.

35. Escudero, R., Gamboa, P. M., Antón, J., Sanz, M. L. Food allergy due to trout roe. *J Investig Allergol Clin Immunol.* **2007**. 5; 346-347.

36. Abdel Rahman, A. M., Kamath, S., Lopata, A. L., Helleur, R. J. Analysis of the allergenic proteins in black tiger prawn (Penaeus monodon) and characterization of the major allergen tropomyosin using mass spectrometry. *Rapid Commun Mass Spectrom.* **2010**. 24; 2462-2470.

37. Abdel Rahman, A. M., Lopata, A. L., O'Hehir, R. E., Robinson, J. J., Banoub, J. H., Helleur, R. J. Characterization and '*de novo*' sequencing of snow crab tropomyosin enzymatic peptides by both electrospary ionization and matrix-assisted laser desorption ionization QqToF tandemmass spectrometry. *J Mass Spectrom.* **2010**. 45; 372-381.

38. Taylor, A. V., Swanson, M. C., Jones, R. T., Vives, R., Rodriguez, J., Yunginger, J. W., Crespo, J. F. Detection and quantitation of raw fish aeroallergens from an open-air fish market. *J Allergy Clin Immunol.* **2000.** 105; 166-169.

39. Cartier, A., Malo, J. L., Ghezzo, H., McCants, M., Lehrer, S. B. IgE sensitization in snow crab-processing workers. *J Allergy Clin Immunol*. **1986.** 78; 344-348.

40. Barraclough, R. M., Walker, J., Hamilton, N., Fishwick, D., Curran, A. D. Sensitization to king scallop (Pectin maximus) and queen scallop (Chlamys opercularis) proteins. *Occup Med (Lond)*. **2006**. 56; 63-66.

41. Malo, J. L., Chrétien, P., McCants, M., Lehrer, S. Detection of snow-crab antigens by air sampling of a snow-crab production plant. *Clin Exp Allergy*. **1997.** 27; 75-78.

42. Twiggs, J. T., Agarwal, M. K., Dahlberg, M. J., Yunginger, J. W. Immunochemical measurement of airborne mouse allergens in a laboratory animal facility. *J Allergy Clin Immunol*. **1982**. 69; 522-526.

43. Swanson, M. C., Agarwal, M. K., Yunginger, J. W., Reed, C. E. Guinea-pig-derived allergen Clinicoimmunologic studies, characterization, airborne quantitation, and size distribution. *Am Rev Respir Dis.* **1984**; 1295; 844-849.

44. Tee, R. D., Gordon, D. J., Lacey, J., Nunn, A. J., Brown, M., Taylor, A. J. Occupational allergy to the common house fly. (Musca domestica): use of immunologic response to identify atmospheric allergen. *J Allergy Clin Immunol.* **1985**. 76; 826-631.

45. Wentz, P.E., Swanson, M. C., Reed, C. E. Variability of cat-allergen shedding. *J Allergy Clin Immunol*. **1990**. 85; 94-98.

46. Howse, D., Gautrin, D., Neis, B., Cartier, A,. Horth-Susin, L., Jong, M., Swanson, M. C. Gender and snow crab occupational asthma in Newfoundland and Labrador Canada. *Environ Res*. **2006**. 101; 163–174.

47. Weytjens, K., Cartier, A., Malo, J. L., Chrétien, P., Essiembre, F., Lehrer, S., Swanson, M. Aerosolized snow-crab allergens in a processing facility. *Allergy.* **1999.** 54; 892-902.

48. Beaudet, N., Brodkin, C. A., Stover, B., Daroowalla, F., Flack, J., Doherty, D. Crab allergen exposures aboard five crab-processing vessels. *AIHA Journal.* **2002**. 63; 605-609.

49. Lopata, A. L., Jeebhay, M. F., Reese, G., Fernandes, J., Swoboda, I., Robins, T. G., Lehrer,S. B. Detection of fish antigens aerosolized during fish processing using newly developed immunoassays. *Int Arch Allergy Immunol.* **2005**. 138; 21-28.

50. Jeebhay, M. F., Robins, T. G., Seixas, N., Baatjies, R., George, D. A., Rusford, E., Lehrer ,S. B., Lopata, A. L. Environmental exposure characterization of fish processing workers. *Ann Occup Hyg*. **2005**. 49; 423-437.

51. Eidhammer, I., Flikka, K., Martens, L., Mikalsen, S-O. Computational methods for mass spectrometry proteomics: *Wiley*. **2007**. 1st edition.

52. Rosenberg, I. M., Protein analysis and purification benchtop techniques. *Boston Birkhauser*. **1996**.

53. Pauling, L., Corey, R. B., Branson, H. R. The Structure of Proteins Two Hydrogen-Bonded Helical Configurations of the polypeptide chain. *Proc Natl Acad Sci USA*. **1951**. 37; 205–211.

54. Bilingual biology11. http://2.bp.blogspot.com/_207DNIaL-gc/TBTbM65cR1I/AAAAAAAAAII/PIsQ_ycOciw/s640/protein+structure.JPG. Accessed March 28, 2011.

55. Drenth, J. Principles of protein X-ray crystallography. *Springer*, **2007**. 3rd edition.

56. Wüthrich, K. NMR studies of structure and function of biological macromolecules. *Nobel Lecture*. **2002**.

57. Cavanagh, J., Fairbrothe, W. J., PalmerIII, A. G., Skilton, N. J. Protein NMR spectroscopy: principles and practice. *Elsevier Academic press*. **2007**. 1st edition.

58. Swann, M. J., Peel, L. L., Carrington, S., Freeman, N. J. Dual-polarization interferometry: an analytical technique to measure changes in protein structure in real time, to determine the stoichiometry of binding events, and to differentiate between specific and non-specific interactions. *Anal Biochem*. **2004**. 29; 190-198.

59. Zhang, Y. Progress and challenges in protein structure prediction. *Curr Opin Struct Bio*. **2008**. 18; 342-348.

60. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science*. **1973**. 181; 223-230.

61. Janson, J. C., Ryden, L. Protein purification principles, high resolution methods, and applications. *John Wiley and sons*. **1998**. 1st edition.

62. Jakoby, W. B. Crystallization as a purification technique. *Methods Enzymol*. **1971**. 22; 248-252.

63. Scopes, R. K. Protein purification: principles and practice. *Springer*. **1993**, 3rd Edition.

64. Kinter, M., Sherman, N. E. Protein sequencing and identification using tandem mass spectrometry. *John Wiley and sons*. **2000**,1st Edition.

65. Rabilloud, T. Proteome research: two-dimensional gel electrophoresis and identification methods (principles and practice). *Springer*. **2000**, 1st Edition.66.

Lopez, M. F., Berggren, K., Chernokalskaya, E., Lazarev, A., Robinson, M., Patton, W. F. A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis*.**2000**. 21; 3673–3683.

67. Chrambach A, Reisfeld R. A, Wyckoff M., Zaccari J. A procedure for rapid and sensitive staining of protein fractionated by polyacrylamide gel electrophoresis. *Anal Biochem*. **1967**. 20:150-4.

68. Towbin, H., Staehelin, T., Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci US.*, **1979**. 76; 4350–4354.

69. Findlay, J. B., Geisow, M. J. Protein sequencing, practical approach series. *Oxford University Press*. **1989**.

70. Edman, P. Method for Determination of the amino acid sequence in peptides. *Acta Chemica Scandinavica*. **1950**. 4; 283-293.

71. Aebersold, R., Mann, M. Mass spectrometry-based proteomics. *Nature*. **2003**. 422; 198-207.

72. Scherperel, G., Reid, G. E. Emerging methods in proteomics: multistage tandem mass spectrometry for top-down protein characterization. *The Analyst.* **2007**; 132; 500-506.

73. Siuti, N., Kelleher, N. L. Decoding protein modifications using top-down mass spectrometry. *Nat. Methods*. **2007**. 4; 817-821.

74. Chait, B. T. Mass Spectrometry: Bottom-up or top-down?. *Science*. **2006**. 314; 65 - 66.

75. Kelleher, N. L., Lin, H. Y., Valaskovic, G. A., Aaserud, D. J., Fridriksson, E. K., McLafferty, F. W. *'Top down'* versus *'bottom up'* protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc*. **1999**. 121; 806–812.

76. Reid, G. E., McLuckey S. A. *'Top down'* protein characterization via tandem mass spectrometry. *J Mass spectrom.* **2002.** 37; 663–675.

77. Zhang, Z., McElvain, J. S. *'De novo'* peptide sequencing by two-dimensional fragment correlation mass spectrometry. *Anal Chem.* **2000**. 72; 2337-2350.

78. Roepstorff, P., Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom.* **1984.** 11; 601-601.

79. De Hoffmann, E., and Stroobant, V. Mass spectrometry principles and applications. *John Wiley and sons.* **2007.** 3rd edition.

80. Spengler, B. Post-source decay analysis in matrix-assisted laser desorption/ionization mass spectrometry of biomolecules. *J Mass Spectrom.* **1997**. 32; 1019-1036.

81. Martin, R. B. Free energies and equilibria of peptide bond hydrolysis and formation. *Biopolymers.* **1998.** 45; 351–353.

82. Breuker, K., Jin, M., Han, X., Jiang, H., McLafferty, F. W. Top-down identification and characterization of biomolecules by mass spectrometry. *J Am Soc Mass Spectrom*. **2008**. 19; 1045-1053.

83. Zubarev, R. A. Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol.* **2004.** 15; 12-16.

84. Nesvizhskii, A. I. Protein identification by tandem mass-spectrometry and sequence database searching. (book auth.) Rune, M. Mass spectrometry data analysis in proteomics. *Humana Press.* **2007**. 1st edition.

85. Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* **1999**. 20; 3551-3567.

86. Shi, J., Wu, F-X. Assigning probabilities to Mascot peptide identification using logistic regression. (book auth.) Hamid, R. A. Advances in computational biology. *Springer.* **2010**. 1st edition.

87. Baxevanis, A. D. Assessing pairwise sequence similarity: BLAST and FASTA. (book auth.) Andreas, D B., Ouellette, B. F. Bioinformatics apractical guid to the analysis of genes and proteins. *Wiley Interscienc.* **2005.** 3rd edition.

88. Blom, N., Gammeltoft, S., Brunak, S. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol.* **1999.** 294; 1351-1362.

89. Abdel Rahman, A. M., Lopata, A. L., Randell, E. W., Helleur, R. J. Absolute quantification method and validation of airborne snow crab allergen tropomyosin using tandem mass spectrometry. *Analytica Chimica Acta.* **2010.** 681; 49–55.

90. Kiemer, L., Bendtsen, J. D., Blom, N. NetAcet: Prediction of N-terminal acetylation sites. *Bioinformatics.* **2005.** 21; 1269-1270.

91. Schulze, W. X., Usadel, B.Quantitation in Mass-spectrometry-based proteomics. Annu. *Rev. Plant Biol.* **2010.** 61; 491–516.

92. Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A., Ahn, N. G. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics.* **2005.** 4; 1487-1502.

93. Desiderio, D. M., Kai, M. Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biol Mass Spectrom.* **1983**. 10; 471-479.

94. Brun, V., Masselon, C., Garin, J., Dupuis, A. Isotope dilution Stratgies for absolute quantitative proteomics. *J proteomics.* **2009.** 72; 740-749.

95. Administration, Food and Drug. FDA Guidance for Industry:Bioanalytical Method Validation. US Department of Health and Human Services, *Food and Drug Administration, Center for Drug Evaluation and Research.* **2001.**

96. Abdel Rahman, A. M., Kamath, S. D., Lopata, A. L. Robinson, J. J., Helleur, R.J. Biomolecular characterization of allergenic proteins in snow crab (*Chionoecetes opilio*) and '*de novo*' sequencing of the second allergen arginine kinase using tandem mass spectrometry. *J proteomics.* **2010.** 74; 231-241.

97. Shiomi, K., Sato, Y., Hamamoto, S., Mita, H., Shimakura, K. Sarcoplasmic Calcium-Binding Protein: Identification as a new allergen of the black tiger shrimp penaeus monodon. *Int Arch Allergy Immunol.* **2008.** 146; 91-98.

Chapter 2: Characterization and *'de novo'* sequencing of snow crab tropomyosin enzymatic peptides by both electrospray ionization and matrix-assisted laser desorption ionization QqToF tandem mass spectrometry [1]

[1] The Chapter has been published; Abdel Rahman, A. M, Lopata, A. L., O'Hehir, R. E., Robinson, J. J., Banoub, J. H., Helleur, R. J. Characterization and *de novo* sequencing of snow crab tropomyosin enzymatic peptides by both electrospray ionization and matrix-assisted laser desorption ionization QqToF tandem mass spectrometry. *J Mass Spectrom.* **2010**. 45; 372-381.

Abstract

The protein tropomyosin (TM) is a known major allergen present in shellfish causing frequent food allergies. TM is also an occupational allergen generated in the working environment of snow crab (*Chionoecetes opilio)* processing plants. The tropomyosin protein was purified from both claw and leg meats of Snow crab and analyzed by electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) using a hybrid quadruple time of flight tandem mass spectrometry (QqToF-MS). The monomeric molecular weight of TM was determined to be 32733 Da. The protein was further characterized using the '*bottom-up*' mass spectrometry approach. A peptide mass fingerprinting (PMF) was obtained by two different enzymatic digestions and '*de novo*' sequencing of the most abundant peptides were performed. Any post translational modifications (PTM) were identified by searching their calculated and predicted molecular weights in precursor ion spectra. The immunological reactivity of snow crab extract was evaluated using specific antibodies and allergenic reactivity assessed with serum of allergic patients. Subsequently a signature peptide for TM was identified and evaluated in terms of identity and homology using the basic local alignment search tool (BLAST). The identification of a signature peptide for the allergen tropomyosin using MALDI-QqToF-MS will be critical for the sensitive and specific quantification of this highly allergenic protein in the workplace.

2.1. Introduction

Seafood plays an important role in human nutrition and health. The growing international trade in seafood reflects the popularity and frequency of consumption of a variety of seafood products in many countries. Unfortunately, increased production and consumption of seafood has resulted in more frequent adverse health problems (i. e. food allergies) among consumers of seafood (1). The fishing and fish processing industry has also experienced tremendous growth in recent years with over 41 million workers worldwide engaged in various activities and exposed to seafood (2). Increased levels of production and processing of seafood have led and continue to lead to more frequent reporting of occupational health problems such as asthma and other allergic reactions particularly in the crustacean processing sector (3, 4).

Tropomyosin (TM) is one of the common muscle proteins; it mediates the interaction between the troponin complex and actin which regulates muscle contraction (5). Crustaceans TM was first identified in shrimp by Hoffman and Miller in 1981 (6). It is a water soluble and heat stable protein with molecular weights (MW) ranging from 34 to 39 kDa (7). It has a highly conserved amino acid sequence among different invertebrate organisms and is present in muscle as well as in non-muscle cells (8). Differential splicing of the pre-messenger ribonucleic acid (pre-mRNA) produces

isoforms of TM (9). Generally, the TM muscle isoform contains 284 amino acids as a highly conserved N-terminal region (10-12).

Research has shown that TM isolated from shrimp and crab is the major food allergen (13). Snow crab is among the seafood that are most frequently associated with an IgE-mediated type I hypersensitivity in food-allergic patients. Urticaria, asthma and diarrhea are the major clinical symptoms that are caused by type I hypersensitivity (14).

Monitoring the level of TM in the snow crab processing workplaces is essential in reducing the worker's risk of developing allergic airway disease. The total amount of snow crab allergens have previously been characterized and measured through immunological reactivity by enzyme-linked immunosorbent assay (ELISA) (15), radioallergosorbent rest (RAST) (13), and immunoblotting (15, 16). The main disadvantage in analyzing the total crab protein by conventional methods is the inclusion of additional proteins such as troponin, which are not necessarily involved in the allergic reactions of the exposed worker (17). However, by targeting and quantifying the major crustacean allergen proteins once can correlate the amount of allergen with the severity of the allergen and determine the threshold values (3).

In the present study, the snow crab TM was isolated and purified. For studying different isoforms of TM, tissues were obtained from claw and leg and the extracted proteins were separated by SDS-PAGE to yield homogeneous bands. Purified TM was

used in Western blot analysis to study its reactivity with specific monoclonal and polyclonal antibodies as well as sera of crustacean allergic patients.

The molecular weight of the extracted TM was determined by ESI-QqToF. Following fractionation by SDS-PAGE, gel bands were excised and subjected to enzymatic digestion. The peptides were analyzed using peptide mass fingerprinting (PMF) by MALDI-QqToF-MS. Peptides were also '*de novo*' sequenced using the '*bottom up*' mass spectrometry approach (17-20).

It has been well-established that PMF and the various outcomes of the MS approach can give information about the identity of proteins which includes the amino acid sequence and post translational modification (PTM) mapping (20). This data, along with immunological characterization data of the extracted protein, will lead to identification of a signature peptide. In the future, this signature peptide will be chemically synthesized and utilized as a peptide standard to quantify Snow crab TM.

A good score in the Basic Local Alignment Search Tool (BLAST) analysis (20) for the target peptide is very important result used to select the unique peptide for representing the target protein. In addition, this signature peptide should be separated by HPLC with no potential post-translational modification (i.e. glycosylation and phosphorylation) sites (18,20). The sensitivity and specificity that mass spectrometry can provide with a high throughput sample analysis in protein characterization, sequencing and PTM mapping is ideal for this study.

This is the first study to analyze the full amino acid sequence of a tropomyosin protein while searching for any protein mutations, mapping of the sites of PTM, and to identify the signature peptide of Snow crab tropomyosin using MALDI and ESI-QqToF-MS techniques.

2.2. Experimental

2.2.1. Chemicals and materials

All chemicals were used as purchased without further purification. Ethanol, acetone, potassium chloride, ammonium sulfate, acetonitrile, hydrochloric acid and methanol were supplied from ACP (Montreal, Canada). Trypsin and endoproteinase (Glu-C V8) sequencing grade enzymes were purchased from Promega (WI, USA). Tris(hydroxymethyl)aminomethane (Tris), ditheotheritol (DTT), ethylenediaminetetraacetic acid (EDTA), formic acid (FA), ammonium bicarbonate, o-methylisourea hemisulfate, ammonium hydroxide, horseradish peroxidase (HRP), 3, 3′, 5, 5′-tetramethylbenzidine (TMB), and α-cyano-4-hydroxycinamic acid (HCCA) matrix were purchased from Sigma-Aldrich (St.Louis, MO, USA). The Bradford assay kit was from BioRad (Hercules, CA, USA). The dialysis bags were from Fisher Scientific (Roncho Dominguez, CA, USA). For desalting, ZipTip C_{18} (10μL bed) filters were purchased from Millipore Corporation (Bedford, MA, USA).

2.2.2. Tropomyosin purification

2.2.2.1. Acetone powder extract

The tissue of the claw and legs (~500 g) was removed from fresh-frozen crab sections, homogenized in a pre-cooled stainless steel homogenizer (Polytron, Brinkmann instruments), and then mixed with an equal volume of de-ionized water for 2 min. After standing for 20 min, the mixture was strained through two layers of cheesecloth. The crude protein extract was washed with 500 ml 95% ethanol for 3 min, washed again with 500 ml 50% ethanol, followed by 2 times with 500 ml 95% ethanol and finally 2 times with 500 ml acetone.

2.2.2.2. Purified extract

The acetone powder (10 g) was dissolved in 70 mL of buffer A (1M KCl, 25 mM Tris-HCl, pH 8.0, 0.25 M DTT and 0.5 mM EDTA), left overnight at 4°C, followed by centrifugation at 8000 RPM for 20 min. Total protein was determined in the supernatant using the Bradford assay and the supernatant diluted to a concentration of 1-2 mg/mL with buffer A. With isoelectric precipitation (pI 4.6) the proteins were precipitated with 1M HCl. The precipitant proteins were re-dissolved in buffer B (0.20 M KCl, 25mM Tris–HCl, pH 8.0, 0.25M DTT and 0.5 mM EDTA). TM precipitated from 40-70% (w/v) ammonium sulfate after the other proteins have been precipitated in 0-40% (w/v). Finally, the pellets were collected after centrifugation at 8000 RPM for 20 min, and then

reconstituted by 200 ml of deionised water and dialyzed against 200mM ammonium bicarbonate in dialysis tubes (MWCO =12-14000) overnight at room temperature then lyophilized to a powder.

2.2.3. Sodium dodecylsulfate polyacrylamide gel electrophoresis

A pair of 12% SDS-PAGE slab gel was used for profiling the purified TM and the acetone powder extract of Snow crab meat from legs and claw. Protein solution (4-10 μg) was added to each of the wells, and electrophoresis was run at a voltage of 170 V until the tracker dye was seen at the base of the gel. One gel was treated with Coomassie brilliant blue using standard protocol. The second gel proteins were transferred to a nitrocellulose membrane at 100 V for 1 hr. After the transfer was completed the membrane was placed in a blocking solution (5% skim milk in Tris-buffered saline) for immunoblotting.

2.2.4. Immunoblotting

The rabbit antibody used in this study was generated in Dr. Lopata lab (Millburn, Australia) by exposing the animal to heat treated protein extracts from crab, prawn and lobster using 500 μg of a mixture of these three protein sources for each injection. Each protein extract was prepared from raw crustaceans, the generated protein extract was

heated to 100°C and the supernatant evaluated by SDS gel-electrophoresis for the presence of TM. Blood samples from the rabbits were taken at week 0 and 6 to analyse for antibody production and the final bleed conducted at week 9 (21). The immunoblot was blocked with 5% skim milk solution for 1 hr at room temp following an overnight incubation (4°C) with the rabbit polyclonal antibody (primary antibody) using a working dilution of 1:40000. The blot was washed three times with TBS-Tween. The secondary antibody used was Polyclonal Goat anti-rabbit antibody conjugated to horse radish peroxidase (HRP), with a working dilution of 1:20000. The blot was incubated for 30 min in the secondary antibody solution, washed again three times with TBS-Tween and incubated with the substrate 3, 3′, 5, 5′-tetramethylbenzidine (TMB). The bands were then visualized using the enhanced chemiluminescence (ECL) technique (21).

To demonstrate the allergenicity of the isolated crab proteins, different extracts were analyzed for IgE antibody binding from allergic patients. The human sera were collected from patients with strong allergic reactivity to shellfish. Ethics approval for this study was acquired by Monash University (Victoria, Australia) as part of an ongoing survey. For immunoblotting, protein extracts were electrophoretic separated (see section 2.2.3), proteins transferred and incubated with human serum (diluted 1:10 in 1% skim milk) overnight at 4°C. Subsequently, blots were washed three times with PBS-Tween and the membrane incubated for 1 hr in 5 mL of rabbit anti-human IgE antibody (diluted 1:1000) in PBS-T containing 1% skim milk. After washing the membrane with PBS-T three times, it was incubated for 30 min in 5 mL of HRP tagged goat anti-rabbit

polyclonal antibody (DAKO, Carpentaria, CA, USA) (diluted 1:1000) in PBS-T containing 1% skim milk. Finally the membrane was washed with PBS three times incubated with the substrate TMB and the immunoblot membranes analyzed for IgE reactivity using the ECL technique (21).

2.2.5. In-gel digestion and guanidation

The 33 kDa protein bands were excised from the SDS-PAGE gel. The *in-gel* guanidation procedure was performed on all protein samples that were analyzed by MALDI-QqToF-MS using the protocol developed by Sergeant *et al.* (22). This procedure increases the sensitivity of the lysine-containing peptides by changing the lysine residue to the homoarginine. The gel pieces were destained by washing three times with 200 mM of NH_4HCO_3 in solution of 50% acetonitrile in dH_2O, at 30°C for 30 min. The destained piece was dried under a stream of N_2, and then covered by a solution of 50mM NH_4HCO_3, pH 7.8 containing 5 ng/μL trypsin or endoproteinase Glu-C V8 in ice for 30 min for rehydration. After rehydration, the excess solution was removed out by suction. The gel was covered by a solution of 50 mM of NH_4HCO_3 and incubated at 37°C overnight to enhance protein digestion. The water soluble peptides were extracted twice with the incubation solution and other remaining peptides extracted twice with 0.15%TFA in 50% ACN after a 2 min vortex mixing. The samples were lyophilized, and reconstituted prior to analysis with 10 μL of 0.1% TFA and desalted with C_{18} ZipTip.

2.2.6. MALDI-QqToF-MS

2.2.6.1. Matrix/Sample Preparation

The two-layer sample/matrix preparation for plate spotting was employed (20). The first layer solution consisted 20 mg of HCCA in 1 mL (1/9) methanol/acetone. The second layer solution consisted 40% ACN in H_2O saturated by HCCA. A 0.5 μL of the first layer matrix solution was applied to a MALDI target. A 1-μL aliquot of the second layer matrix solution was mixed with 1 μL of sample. One microliter of the sample/matrix mixture was deposited onto the first layer and allowed to dry, followed by an on-target wash step. This was performed by adding of 1 μL of water on top of the dry spot and after 10 seconds blowing the water off using a pulse of air.

2.2.6.2. MALDI-QqToF-MS and CID-MS/MS

MALDI-MS and low-energy (CID) analyses were carried on a QSTAR XL hybrid quadrupole-quadrupole/time-of-flight tandem mass spectrometer (QqToF-MS/MS) (Applied Biosystems/MDS Sciex, Foster City, CA, USA) equipped with an o-MALDI ion source (Applied Biosystems, Foster City, CA, USA)

2.2.6.3. ESI- QqToF-MS

Peptide separation was conducted using a DIONEX UltiMate3000 Nano LC System (Germering, Germany). A 250 *f*mol sample of protein digest dissolved in 0.1%

TFA was loaded onto a precolumn (300 μm ID X 5 mm, C_{18} PepMap 100, 5 μm (LC Packing, Sunnyvale, CA)) for desalting and concentrating. Peptides were then eluted from the pre-column and separated on a nanoflow analytical column (75 μm ID x 15 cm, C18 PepMap 100, 3 μm, 100 °A, (LC Packing, Sunnyvale, CA) at 180 nL/min using the following gradient. The aqueous mobile phases consisted of (A) 0.1% formic acid/ 0.01% trifluoroacetic acid/ 2% ACN and (B) 0.08% formic acid/ 0.008% trifluoroacetic acid / 98% ACN. A gradient of 0% B for 10 min, 0-60% B in 55 min, 60-90% in 3 min, 90% B for 5 min was applied. Including a column regeneration step, one run was 106 min long.

The ESI-MS of the LC-eluting peptides were measured with the same hybrid QqToF-MS/MS system equipped with a nanoelectrospray source (Protana XYZ manipulator). The nanoelectrospray was generated from a PicoTip needle (10 μm i.d., New Objectives, Wobum, USA) at a voltage of 2400 V.

In further experiments the ESI-MS of the desalted 33 kDa protein was directly injected into the ESI source and the resulting multiple-charged spectrum of the protein was deconvoluted by *Analyst QS 1.1* software.

Snow crab tropomyosin was further analyzed by CID-MS/MS and the resulting peptides spectra were searched by using the National Center for Biotechnology Information non-redundant database (NCBInr) with the Matrix Science (Mascot) search engine (precursor and product ion mass tolerance set at 0.2 Da). Methionine oxidation was allowed as a variable modification and guanidinyl (K) as a fixed modification since

the guanidation derivatisation has been performed. Peptides were considered identified if the Mascot score was over 95% confidence limit.

2.3. Result and Discussion

The proteins of snow crab obtained as an acetone powder extract and as the purified TM extract were profiled by SDS-PAGE, as seen in Fig. 1a. The strong band at 33 kDa is presumed to be TM. The much weaker band ~ 65kDa, which has a high reactivity against anti-crustacean polyclonal antibody (Fig. 2.1 b) along with patients' sera (Fig. 2.1 c and d) appeared more clearly, was presumed to be a dimer form of TM.

The purified protein for both the claw and leg extracts was further investigated using direct flow injection into ESI source of QqToF instrument. The multiply-charged ion spectrum of the purified tropomyosin gave a series of multicharged ions, which were rationalized as $[M+45H^{45+}$ m/z 729.6096; $[M+44H]^{44+}$ m/z 746.1454; $[M+43H]^{43+}$ m/z 763.2390; $[M+42H]^{42+}$ m/z 781.5457; $[M+41H]^{41+}$ m/z 800.7028; $[M+40H]^{40+}$ m/z 820.3127; $[M+39H]^{39+}$ m/z 841.6777; $[M+38H]^{38+}$ m/z 863.9742; $[M+37H]^{37+}$ m/z 886.9603; $[M+36H]^{36+}$ m/z 912.1664; and $[M+35H]^{35+}$ m/z 938.3027. This series of molecular ions was deconvoluted with the Bayesian protein reconstruct tool which is available in *Analyst QS 1.1* software to give a molecular weight of claw TM extract at 32783 ± 0.02% Da. Another series of multicharged ions was also deconvoluted using the same tools to give 32776 ± 0.02% Da as a molecular weight of the leg TM extract. The

differences in mass between the calculated and the experimental are related to the post translational modification or to intramolecular modifications (i. e. N-terminal acetylation (42 Da)). (23)

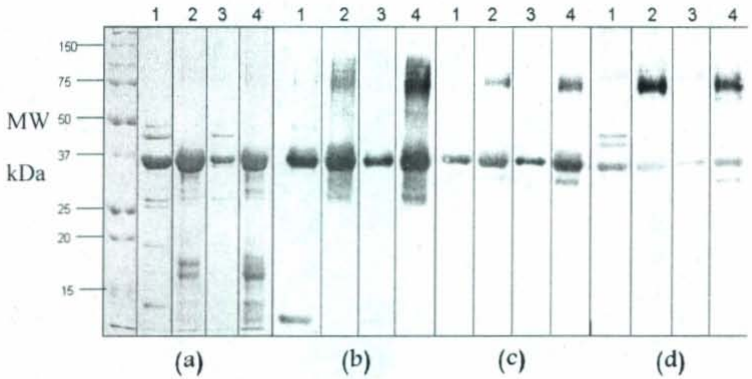


Figure 2. 1: Snow crab extract profiling by (a) SDS-PAGE, (b) immunoblotting anti-crustacean polyclonal antibody, (c) immunoblotting patient serum A, and (d) Immunoblotting patient serum B, where (1) claw acetone powder extract, (2) claw purified TM extract, (3) leg acetone powder extract, and (4) leg purified TM extract.

The separated proteins in SDS-PAGE were transferred to the nitrocellulose membrane and incubated with anti-tropomyosin mAb, anti-crustacean TM polyclonal antibody, and sera of two sensitized patients with different symptoms, as seen in Fig. 2.1 b, c, and d. This permitted the examination of the bio-reactivity of the extracted protein with sera of allergic patients to Snow crab TM as well as the cross linking reactivity against specific mAb.

The 33 kDa band was excised and subjected to *in-gel* trypsin and another sample to endoproteinase (V8 Glu-C) digestions. For both enzymatic digestions, the resultant peptides were extracted and analyzed by MALDI-QqToF mass spectrometry (Fig. 2.2). The combination of these two enzymes generates two sets of peptides with different termini and mass spectrometric sensitivity, which covers more amino acid motifs during sequencing experiments. (24) The wide sequencing coverage plays a significant role in the signature peptide selectivity approach and in the PTM site determination.

The precursor ions of enzymatic digested TM peptide from both leg and claw extracts were uploaded to the Mascot PMF search engine. The results were in average score (95) for tryptic digestion, which is the top probability based on Mowse scores of matching with the snow crab TM protein (cDNA based Library (NCBInr)). The above scores are matched with Mascot algorithms' criteria of which "Individual ions scores > 82 indicate identity or extensive homology ($p<0.05$)". (25) To improve the sensitivity of lysine containing peptides in the MALDI source and increase the sequencing coverage, the *in-gel* guanidation reaction was performed for the excised bands prior to the digestion steps (results are not shown). (21) Interestingly, both TM samples that were extracted from claws and legs gave the same amino acid sequence without any indication of any isoforms as could be deduced from the SDS profile (Fig. 2.1a).

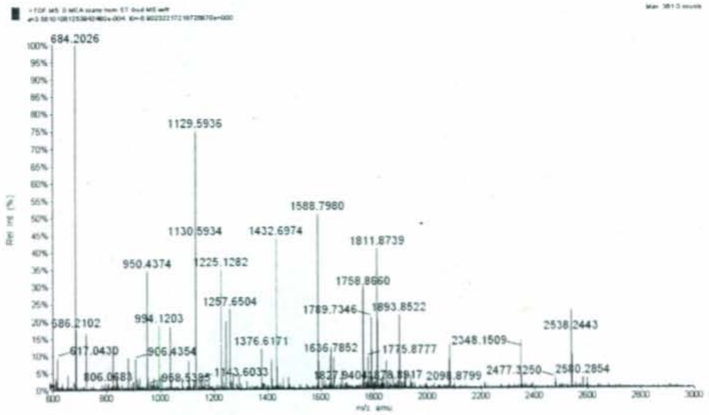


Figure 2. 2: MALDI-QqToF mass spectrum of 'in-gel' trypsin digestion of snow crab tropomyosin. Each peak in the spectrum represents a tryptic peptide. Their masses in this spectrum are used for PMF characterization.

Both MALDI and ESI techniques were used for the '*de novo*' sequencing of the resulting peptides from both types of protein digestion. In MALDI-CID-MS/MS, the high abundant peptides' precursor ions (i.e. 2348.1438, 1432.6904, 1588.8092, 1257.6347, 1758.8116, 1376.6111, 1893.8459, 1433.7121, 950.4362, 880.4612, 1129.5890, 722.3220, 1789.7741, etc.), as some shown in Fig. 2.2, were selected in the first quadruple, fragmented in the low energy CID, and the product ions scanned by ToF analyzer. The full sequenced peptides are summarized in the Table 2.1.

In the ESI-CID-MS/MS analyses the peptides were first separated by nano-HPLC and ionized by a nano-electrospray source. The data for the precursor and product ions of peptides were uploaded to the Mascot MS/MS ion search engine against NCBInr database. The Mowse scores were 746 and 505 as Snow crab TM for the tryptic peptides of TM that has been extracted from claws and legs respectively. Table 2.1 shows the amino acid sequencing and its distribution between the two enzymatic types. As an example, two of peptides and their product ion spectra are shown in Fig. 2.3.

The sequence of some peptides produced from the same enzyme were confirmed by both MALDI- and ESI-CID-MS/MS. Henceforth, many of the amino acid motifs were confirmed by either of the ionization sources or by two different digestion enzymes.

Analyzing the results in Table 2.1 was interesting to note that the N-terminus peptide (residues 1-7, with sequence MDAIKKK) as a major ion was difficult to be observed (ionized) and isolate for product ion spectrum. This is an indication that it was blocked with an acetyl group at the N-terminus which is commonly found in the *Eukaryotic* TM (detailed below).

The size difference between trypsin and V8 enzymes limits the V8 enzyme to penetrate far inside the gel pores. This can lead to reduction in *in-gel* digestion efficiency. Many of the V8 resultant peptides are identified with a lot of missed cleavages up to 5 whereas the tryptic peptides had a maximum of 1 missed cleavage (Table 2.1).

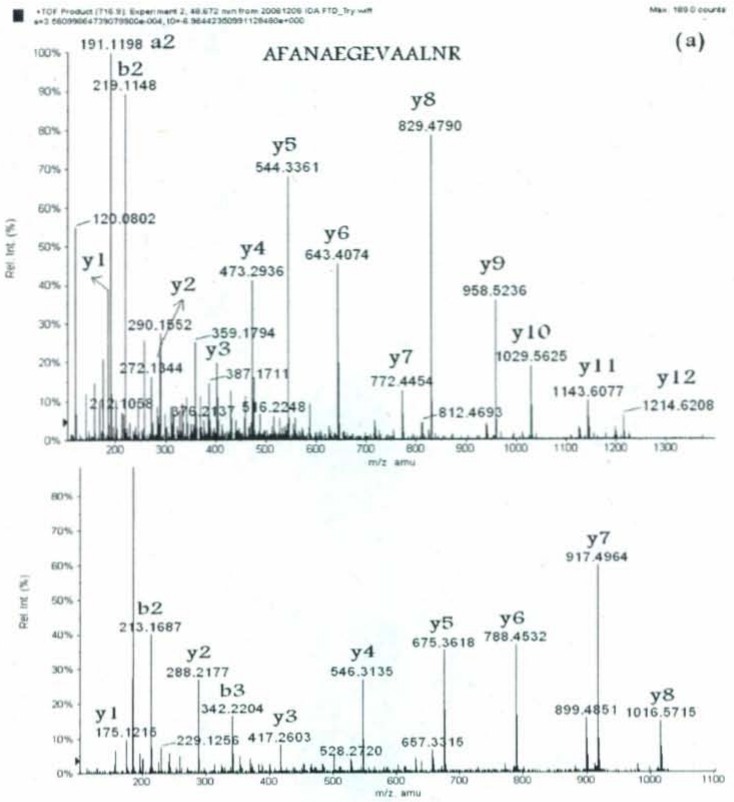


Figure 2. 3: Representative ESI product ion spectra of selected precursor ions of (a) $[M+2H]^{2+}$ m/z 716.9 and (b) $[M+2H]^{2+}$ m/z 565.3 of two selected peptides obtained from the tryptic digestion of SC tropomyosin having sequences AFANAEGEVAALNR, and IVELEEELR respectively.

In addition, a study to identify the chemical nature of the higher molecular bands, which have strong antibody reactivity both to specific rabbit and patients IgE antibodies was undertaken. The 65 kDa band of the SDS-PAGE were excised, destained, and digested using the same protocols described above. Expectedly, the same amino acid sequence was obtained by the amino acid sequencing, but with higher relative intensities of the peptide signals.

To the best of our knowledge, this study conclusively demonstrates that the higher molecular weight IgE antibody binding proteins is a dimer of the allergen tropomyosin. This is of significant importance as several aerosolized allergenic proteins have been demonstrated in the crustacean processing industry since the first discovery of occupational asthma in the snow crab processing industry (26) and confirmed in a recent study on IgE binding proteins (27). However the molecular nature of these allergens was not known. In addition to determining the full sequence of SC tropomyosin protein, other important factor in selecting a signature peptide is for the absence of PTM groups (e.g. phosphorylation and glycosylation). Consequently, all the precursor mass spectra were evaluated for the presence of any potential PTM groups. The ESI-MS results show molecular weights 32783 Da and 32776 Da for both claw and leg extracts, respectively.

Table 2. 1: Amino acid sequencing of the product ion spectra obtained from the MALDI-QqToF and ESI-QqToF experiments of the major peptides of snow crab TM by trypsin and V8 digestion.

Protease	Peptide sequence	Residues #	Missed cleavage	ESI	MALDI
Trypsin	MDAIKKK	1-7	-	-	-
	MQAMKLEK	8-15	1	+	-
	DNAMDKADTLEQQNK	16-30	1	+	-
	AEKTEEEIR	36-44	1	+	+
	SQLVENELDHAQEQLSAATHK	50-70	0	+	+
	AFANAEGEVAALNRR	77-91	1	+	+
	IQLLEEDLER	92-101	0	+	-
	IQLLEEDLERSEER	92-105	1	+	+
	LAEASQAADESER	113-125	0	+	+
	SLSDEERMDALENQLK	134-149	1	-	+
	MDALENQLK	141-149	0	+	-
	FLAEEADR	153-160	0	+	+
	KYDEVAR	161-167	0	+	-
	KLAMVEADLERAEER?	168-172	1	+	+
	IVELEEELR	190-198	0	+	+
	SLEVSEEK	206-213	0	+	-
	EETYKEQIK	218-226	1	+	-
	AEFAER	239-244	0	-	+
	LEDELVNEKEK	256-266	1	+	-
	NIADEMDQAFSELSGF	269-284	0	-	+
Endoproteinase Glu-C	KDNAMDKADTLEQQNKE	15-31	1	+	-
	ANLRAEKTEEE	32-42	3	+	+
	IRANQKKSQLVENE	43-56	1	+	+
	LDHAQEQLSAATHKLVE	57-73	1	+	+
	QLSAATHKLVEKE	63-75	1	+	+
	KEKAFANAEGE	74-84	2	+	-
	VAALNRRIQLLEEDLERSEE	85-104	4	-	+
	RLNTATTKLAEASQAADE	105-122	1	+	-
	SERMRKVLE	123-131	1	+	-
	NRSLSDEERMDALE	132-145	2	+	+
	ARFLAEEADRKYDE	151-164	2	+	+
	VARKLAMVE	165-173	0	+	-
	VARKLAMVEADLE	165-177	1	+	-
	RAESGESKIVELEEE	182-196	5	+	-
	SKIVELEEE	188-196	3	+	-

(Cont'd) **Table 2. 2:** Amino acid sequencing of the product ion spectra obtained from the MALDI-QqToF and ESI-QqToF experiments of the major peptides of snow crab TM by trypsin and V8 digestion(

Protease	Peptide sequence	Residues #	Missed cleavage	ESI	MALDI
Endoproteinase Glu-C	LRVVGNNLKSLE	197-208	0	+	-
	KANQREETYKE	213-223	2	-	+
	QIKTLANKLKAAEARAE	224-246	1	+	-
	FAERSVQKLQKE	241-252	1	+	+
	RSVQKLQKE	244-252	0	+	-
	VDRLEDELVNE	253-263	2	+	+
	KEKYKNIADE	264-273	1	+	-

The experimental MW is compared with the calculated molecular weight of TM (32655 Da) which based on the cDNA sequence (29), the difference in mass indicates the presence of acetylated N-terminus and some variable modifications (i. e. oxidation of methionine). This was confirmed by the theoretical algorithm of the *NetAcet 1.0* server tool (29), which indicated that the only site applicable for acetylation was the N-terminal methionine. This type of modification is quite common for the Eukaryotic TM, as reported for the bovine, chicken (P04268) and human (P09493) in the UniProtKB/Swiss-Prot data bases.

The "non-matched peptides" precursor ion and their intensity data generated from PMF experiments were uploaded on the ExPASy FindMod tool, to check if there is any potential peptide having any PTM motif(s). The obtained report indicated an absence of any type of modifications. Further confirmation of the absence of modified peptide ions

was undertaken by usually searching for the calculated molecular ions of PTM motifs in precursor spectra.

The possible sites of phosphorylation of SC TM were identified using the *NetPhos 2.0* server (30), which can reveal the T, S and Y motifs of major phosphorylated amino acids in Eukaryotic TM protein (Table 2.2). The scores of predicted sites, positions, and molecular ions of the peptides that possess these phosphorylated sites are estimated, and then compared with the present (ESI or MALDI) precursor spectra.

Therefore, this evaluation suggests that there was no matching between the predicted phosphorylated molecular ions and the precursor experimental molecular ion in PMF analyses. This result was further confirmed by the absence of any phosphorylated immonium ions on the product ion spectra.

All the resultant peptides produced by both trypsin and V8 enzymes, where introduced to the BLAST test to show which of peptides is the best candidate as a signature peptide to be used as a TM chemical surrogate in future quantitative work in the environmental researches. The NCBI BLAST test, which used to find regions of local similarity between sequences of the NCBI database and calculates the statistical significance of matches, reported the peptide located at 50-70 (SQLVENELDHAQEQLSAATHK) is an ideal signature peptide (Fig. 2.4 and 2.5) for tropomyosin *(Chionoecetes opilio)* with 100% identity, score = 68.5 bits (154), and expect = $6e^{-11}$. This peptide also got the best scores as homology and identity by the

Mascot search engine (25). Therefore the chosen signature peptide for future quantitation development to shown below:

SQLVENELDHAQEQLSAATHK

Table 2. 3: List of putative peptides for phosphorylation as calculated from NetPhos 2.0 with actual mass spectra data reported.

Peptide (charge states)	*NetPhos 2.0* Phosphorylation Prediction				This study Mass Spectrometric data			
	position	*Score	**AA	m/z	Observed	Mr (exp)	Mr (calc)	Delta
DNAMDKADTLEQQNK (+3)	24	0.449	-	600.9205	574.2832	1719.8278	1719.7733	0.0544
AEKTEEEIR (+2)	39	0.891	T	592.7634	552.7990	1103.5835	1103.5458	0.0377
SQLVENELDHAQEQLSAATHK (+3)	50	0.046	-	810.0429	783.4223	2347.2450	2347.1404	0.1046
SQLVENELDHAQEQLSAATHK (+3)	68	0.598	T	810.0429	783.4223	2347.2450	2347.1404	0.1046
SQLVENELDHAQEQLSAATHK (+3)	65	0.434	-	810.0429	783.4223	2347.2450	2347.1404	0.1046
IQLLEEDLERSEER (+3)	102	0.936	S	613.6226	586.9802	1757.9187	1757.8795	0.0392
RLNTATTKLAE (+2)	108	0.149	-	649.3292	609.3691	1216.7236	1216.6775	0.0461
RLNTATTKLAE (+2)	110	0.364	-	649.3292	609.3691	1216.7236	1216.6775	0.0461
RLNTATTKLAE (+2)	111	0.569	T	649.3292	609.3691	1216.7236	1216.6775	0.0461
LAEASQAADESER (+2)	117	0.203	-	728.8012	688.8408	1375.6670	1375.6215	0.0454
LAEASQAADESER (+2)	123	0.616	S	728.8012	688.8408	1375.6670	1375.6215	0.0454
SLSDEERMDALENQLK (+3)	134	0.988	S	653.2906	626.6616	1876.9628	1876.8836	0.0792
SLSDEERMDALENQLK (+3)	136	0.997	S	653.2906	626.6616	1876.9628	1876.8836	0.0792
KYDEVAR (+2)	162	0.925	Y	480.7130	440.7463	879.4781	879.4450	0.0331

(Cont'd.) **Table 2. 4:** List of putative peptides for phosphorylation as calculated from NetPhos 2.0 with actual mass spectra data reported

Peptide (charge states)	*NetPhos 2.0* Phosphorylation Prediction				This study Mass Spectrometric data			
	position	*Score	AA	m/z	Observed	Mr (exp)	Mr (calc)	Delta
RAESGESKIVELEEE (+2)	185	0.996	S	892.9011	852.9547	1703.8948	1703.8213	0.0735
RAESGESKIVELEEE (+2)	188	0.669	S	892.9011	852.9547	1703.8948	1703.8213	0.0735
SLEVSEEK (+2)	206	0.958	S	500.7154	460.7420	919.4694	919.4498	0.0196
SLEVSEEK (+2)	210	0.943	S	500.7154	460.7420	919.4694	919.4498	0.0196
EETYKEQIK (+2)	220	0.985	T	624.2814	584.3200	1166.6255	1166.5819	0.0436
EETYKEQIK (+2)	221	0.936	Y	624.2814	584.3200	1166.6255	1166.5819	0.0436
QIKTLANKLKAAEARAE (+3)	227	0.056	-	645.6856	619.0606	1854.1601	1854.0686	0.0915
FAERSVQKLQKE (+2)	245	0.009	-	771.8874	731.9433	1461.8720	1461.7939	0.0781
KEKYKNIADE (+2)	267	0.993	Y	659.3080	619.3475	1236.6804	1236.6350	0.0454
YKNIADEMDQAFSELSGF (+2)	279	0.646	S	1072.948	1033.005	2063.9961	2063.9146	0.0816
YKNIADEMDQAFSELSGF (+2)	282	0.017	-	1072.948	1033.005	2063.9961	2063.9146	0.0816

* The NetPhos 2.0 threshold is >0.500. **Amino acid

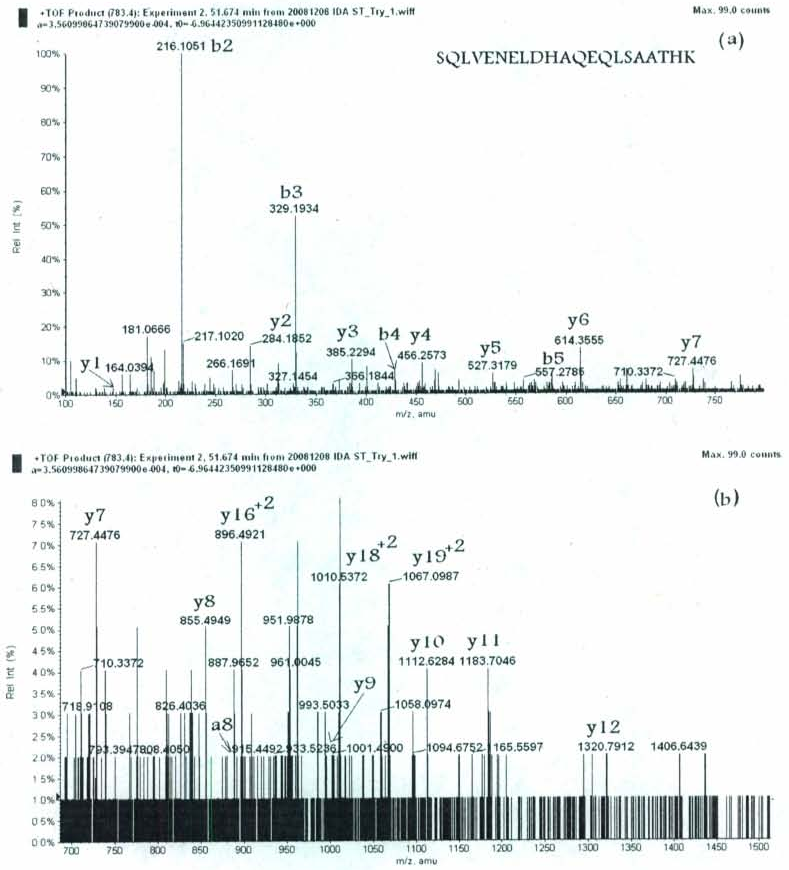


Figure 2. 4: ESI product ion spectrum of Snow crab TM signature peptide $[M+3H]^{3+}$, m/z 783.4 Da. The axes have been cut to improve peak visualization. The top (a) and bottom (b) correspond to m/z ranges of 100-800 and 700-1500, respectively.

```
1          11         21         31         41         51
MDAIKKMQMQAMKLEKDNAMDKADTLEQQNKEANLRAEKTEEEIRALNQKKSQLVENELDHA
61         71         81         91         101        111
QEQLSAATHKLVEKEKAFANAEGEVAALNRRIQLLEEDLERSEERLNTATTKLAEASQAA
121        131        141        151        161        171
DESERMRKVLENRSLSDEERMDALENQLKEARFLAEEADRKYDEVARKLAMVEADLERAE
181        191        201        211        221        231
ERAESGESKIVELEEELRVVGNNLKSLEVSEEKANQREETYKEQIKTLANKLKAAEARAE
241        251        261        271        281
FAERSVQKLQKEVDRLEDELVNEKEKYKNIADEMDQAFSELSGF
```

Figure 2. 5: Full amino acid sequence of Snow crab tropomyosin with the calculated molecular weight 32655 Da.

2.4. Conclusion

It was demonstrated for the first time that the full length of a major crustacean allergen, tropomyosin, can be sequenced utilizing a combination of electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) using a hybrid quadruple time of flight tandem mass spectrometry (QqToF-MS).

Using a '*bottom up*' mass spectrometry approach (31) the sequence of tropomyosin from legs and claws of the snow crab (*Chionoecetes opilio)* demonstrated to have the same amino acid sequence and an average molecular weight of 32733±0.02% Da, which is slightly higher than the previously reported. Using precursor ion spectra it was determined that N-terminal acetylation is the site for post translational modification, this accounting for the small molecular weight difference (stated above). In addition a higher molecular weight band, which is immunologically recognized by patient sera, was

identified as the tropomyosin dimer. This finding is of significant importance as previous studies on environmental allergen exposure among Snow crab processors where not able to identify the nature of all the generated protein allergens during processing. Based on molecular data obtained, a signature peptide of 21 amino acids was chosen as surrogate that could be correlated directly to the presence of the full length SC tropomyosin. The identification of a signature peptide for the allergen tropomyosin using MALDI-QqToF-MS is fundamental for the development of sensitive and specific quantification methods of this highly allergenic protein in the work place. Furthermore, the results of this study clearly demonstrate the strength of mass spectrometry in characterizing minute amounts of unknown allergens, which will be vital in the identification of yet unknown snow crab allergens, responsible for occupational sensitization.

2.5. References

1. Lopata, A.L., Lehrer, S. B. New insights into seafood allergy. *Current Opinion in Allergy and Clinical Immunology.* **2009.** 9; 270-277

2. Food and Agriculture Organization of the United Nations (FAO), The state of world fisheries and aquaculture,, *Fisheries and Aquaculture Department,* Rome, **2007**.

3. Jeebhay, M. F., Robins, T. G., Miller, M. E., Bateman, E., Smuts, M., Baatjies, R., Lopata, A. L. Occupational allergy and asthma among salt water fish processing workers. *Am J Ind Med.* **2008.** 51; 899-910.

4. Jeebhay, M. F., Robins, T. G., Lehrer, S. B., Lopata, A. L. Occupational seafood Allergy- A review. *Occup Env Med.* **2001.** 58; 553-562.

5. Lu, Y., Ohshima, T., Ushio, H., Hamada, Y., Shiomi, K. Immuological characteristics of monoclonal antibodies against shellfish major allergen tropomyosin. *Food Chemistry.* **2007**. 100; 1093-1099.

6. Hoffman, D. R., Day, E. D., Miller, J. S. The major heat stable allergen of shrimp. *Ann Allergy.* **1981.** 47; 17-22.

7. Liang, Y. L., Cao, M. J., Su, W. J., Zhang, L. J., Huang, Y. Y., Liu, G. M. Identification and characterization of the major allergen of Chinese mitten crab (*Eriocheir sinensis*). *Food Chemistry.* **2008**. 111; 998-1003.

8. Jeebahay, M. F., Robins, T. G., Seixas, N., Baatjies, R., George, D. A., Rusford, E., Lehrer. S. B., Lopata, A. L. Environmental Exposure Characterization of Fish Processing Workers. *Ann Occup Hyg.* **2005**. 49; 423-437.

9. Malo, J. L., Chretien, P., Mccants, M., Lehrer, S. B. Detection of Snow crab antigens by air sampling of a snow crab production plant. *Clin Exp Allergy.* **1997**. 27; 75-78.

10. Wolska, B. M., Wieczorek, D. F. The role of tropomyosin in the regulation of myocardial contraction and relaxation. *Pflugers Arch-Eur J Physiol.* **2003**. 446; 1-8.

11. Lewis, W. G., Smillie, L. B. The amino acid sequence of rabbit cardiac tropomyosin. *J Biol Che.* **1980.** 255; 6854-6895.

12. Heeley, D. H., Bieger, T., Waddleton, D. M., Hong, C., Jackman, D. M., McGowan, C., Davidson, W. S., Beavis, R. C. Characterisation of fast, slow and cardiac muscle tropomyosins from salmonid fish. *Eur J Biochem.* **1995**. 232; 226-234.

13. Reese, G., Schicktanz, S., Lauer, I., Randow, S., Lüttkopf, D., Vogel, L., Lehrer, S. B., Vieths, S. Structural, immunological and functional properties of natural recombinant Pen a 1, the major allergen of Brown Shrimp, Penaeus aztecus. *Clin Exp Allergy.* **2006.** 36; 517-524.

14. Geng, M., Ji, J., Regnier, F. E. Signature-peptide approach to detecting protein in complex mixtures. *J Chromatogr A.* **2000.** 870; 295-313.

15. Kluwe, L., Maeda, K., Miegel, A., Fujita-Becker, S., Maeda, Y., Talbo, G., Houthaeve, T., Kellner, R. Rabbit skeletal muscle αα-tropomyosin expressed in baculovirus-infected insect cells possessed the authentic N-terminus structure and function. *Journal of Muscle Research and Cell motility.* **1995.** 16; 103-110.

16. Rosmilah, M., Shahnaz, M., Masita, A., Noormalin, A., Jamaludin, M. Identification of major allergens of two species of local snappers: Lutjanus argentimaculatus (merah/ red snapper) and Lutjanus johnii (jenahak/ golden snapper). *Trop Biomed.* **2005**. 22; 171-177.

17. Lorenz, A. R., Scheurer, S., Haustein, D., Vieths, S. Recombinant food allergens. *J Chromatogr B.* **2001**. 756; 255-279.

18. Cohen, A. M., Mansour, A. A., Banoub, J. H. *'De novo'* sequencing of Atlantic cod vitellogenin tryptic peptides by matrix-assisted laser desorption/ionization quadrupole time-of-flight tandem mass spectrometry: similarities with haddock vitellognin. *Rapid Commun Mass Spectrom.* **2005.** 19; 2454-2460.

19. Banoub, J. H., Cohen, A. M., Mansour, A. A., Thibault, P. Characterisation and '*de novo*' sequencing of Atlantic salmon Vitellogenin protein by electrospray tandem and matrix-assisted laser desorption/ionization mass spectrometry. *Eur J Mass Spectrom.* **2004.** 10; 121-34.

20. Kinter, M., Sherman, N. E. Protein sequencing and identification using tandem mass spectrometry. *John Wiley*; New York, **2000**.

21. Lopata, A. L., Jeebhay, M. F., Reese, G., Fernandes, J., Fenemore, B., Elliott, A., Robins, T. G., Lehrer S. B. Detection of fish antigens aerosolized during fish processing using newly developed immunoassays. *Int Arch Allergy Immunol.* **2005**. 138; 21-28.

22. Sergeant, K., Samyn, B., Debyser, G., Beeumen, J. V. *'De novo'* sequence analysis of N-terminal sulfonated peptides after in gel quanidination. *Proteomics.* **2005**. 5; 2369-2380.

23. Huang, M. C., Ochiai, Y. Fish fast skeletal muscle tropomyosins show species-specific thermal stability. *Comp Biochem Physiol B Biochem Mol Biol.* **2005**. 141; 461-471.

24. Hoffmann, E. Stroobant, V. Mass Spectrometry: Principles and Applications. *John Wiley;* New York, **2007**.

25. Pappin, D. J., Hojrup, P., Bleasby, A. J. Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol.* **1993.** 3; 327-332.

26. Cartier, A., Malo, J. L., Forest, F., Lafrance, M., Pineau, L., St-Aubin, J. J., Dubois, J. Y. Occupational asthma in snow crab-processing workers. *J Allergy Clin Immunol* **1984**. 74; 261-269.

27. Gill, B. V., Rice, T. R., Cartier, A., Gaùtrin, D., Neis, B., Horth-Susin, L., Jong, M., Swanson, M., Lehrer, S. B. Identification of crab proteins that elicit IgE

reactivity in snow crab–processing workers. *J Allergy Clin Immunol.* **2009.** 124; 1055-1061.

28. Motoyama, K., Suma, Y., Ishizaki, S., Nagashima, Y., Shiomi, K. Molecular cloning of tropomyosins identified as allergens in six species of crustaceans. *J Agric Food Chem.* **2007**. 55; 985-991.

29. Kiemer, L., Bendtsen, J. D., Blom, N. NetAcet: Prediction of N-terminal acetylation sites. *Bioinformatics.* **2005.** 21; 1269-1270.

30. Blom, N., Gammeltoft, S., Brunak, S.Sequence and structure-based prediction of Eukaryotic protein phosphorylation sites. *J Mol Biol.* **1999.** 294; 1351-1362.

31. Cohen, A. M., Jahouh, F., Sioud, S., Rideout, R. M., Morgan M. J., Banoub, J. H. Quantification of Greenland halibut serum vitellogenin: a trip from the deep sea to the mass spectrometer. *Rapid Commun. Mass Spectrom.* **2009.** 23; 1049-1060.

Chapter 3: Biomolecular characterization of allergenic proteins in snow crab (*Chionoecetes opilio*) and '*de novo*' sequencing of the second allergen, arginine kinase, using tandem mass spectrometry [2]

[2] This chapter has been published; Abdel Rahman, A. M., Kamath, S. D., Lopata, A. L., Robinson, J. J., Helleur, R. J. Biomolecular characterization of allergenic proteins in snow crab (*Chionoecetes opilio*) and *de novo* sequencing of the second allergen arginine kinase using tandem mass spectrometry. *J. Proteomics*. **2011**. 74; 231-241.

Abstract

Snow crab (*Chionoecetes opilio*) proteins have been recognized as an important source of both food and occupational allergens. While snow crab causes a significant occupational allergy, only recently has one novel allergen (tropomyosin) been fully characterized. The muscle proteins from snow crab legs were profiled by SDS-PAGE. Several of these proteins were characterized using tandem mass spectrometry. Five proteins were identified; sarcoplasmic Ca-binding (20 kDa), arginine kinase (40), troponin (23 kDa) α –actine (42 kDa) and smooth endoplasmic reticulum Ca^{2+}ATPase (113 kDa).

Immunoblotting using serum of sixteen allergic patients resulted in strong reactivity with the 40-kDa protein in seven patients (43%). This protein was purified by chromatography and subsequently '*de novo*' sequenced using matrix assisted laser desorption ionization and electrospray tandem mass spectrometry. We identified a second important allergen, arginine kinase, in snow crab, designated Chi o 3. Based on identity and homology analysis, and using bioinformatics tools, a signature peptide was identified as a chemical surrogate for arginine kinase. The suitability of this signature peptide was tested to determine analytically if it could be used to chemically represent arginine kinase. Multiple-reaction monitoring tandem mass spectrometry was used for the analysis of actual air samples collected on filters from a simulated crab processing plant.

3.1. Introduction

The snow crab (SC) fishery in Atlantic Canada represents the world's largest SC fishery; it accounts for almost 90% of world landings (by weight) in 2004, with over half of these landings coming from Newfoundland and Labrador (1). In 2007, the NL fishery alone landed 50,000 t, valued at $177 million (1). Unfortunately, SC meat can be one of the most important causes of severe acute hypersensitivity reactions, including fatal anaphylaxis and severe asthma among the fishermen and processing plant workers (2). Fish and shellfish are a leading cause of IgE-mediated food hypersensitivity (2-4). IgE-mediated reactions that cause nausea, vomiting, abdominal pain, and diarrhea may be triggered within minutes of ingestion (5).

The molecular structure of tropomyosin, the major allergen in crustaceans (5-7), was recently characterized in snow crab and black tiger prawn using mass spectrometry (8, 9). Besides tropomyosin, other allergenic proteins from crustaceans have been reported. In 2003, Yu *et al.* (10) identified a novel shrimp allergen designated as Pen m 2 from (*Penaeus monodon)* by two-dimensional immunoblotting using sera from shrimp allergenic patients. The allergen was identified using cDNA cloning. The open reading frame encoded 356 amino acids with a theoretical molecular weight of ~40 kDa. The amino acid sequence of this protein showed 60% similarity to arginine kinase of the crustacean, Kuruma prawn *(Penaeus japonicus)* (10).

Arginine kinase (AK) was recently reported as an allergen in different crustacean and invertebrate species. AK was identified in white shrimp *(Litopenaeus vannamei)* (11), gulf shrimp *(Penaeus aztecus)* (12), chinese shrimp *(Fenneropenaeus chinensis)* (13), and other shrimp species using a proteomics approach (14). Moreover, AK has been identified in other invertebrates such as the house dust mite *(Dermatophagoides farinae)* (15), Indian-meal moth *(Plodia interpunctella)* (16), silkworm larvae *(Bombyx mori)* (17), and american cockroach *(Periplaneta americana)* (18-20).

Monitoring airborne allergens in SC harvesting and processing workplaces is essential to reducing the worker's risk of developing allergenic airway diseases (21). Normally, allergens are characterized and measured through immunological reactivity by enzyme-linked immunosorbent assay (ELISA) (22-24), radioallergosorbent rest (RAST) (25), and immunoblotting (22-26). These techniques evaluate the total protein concentration which includes non-allergenic proteins in addition to non-SC allergens (27).

Quantifying the major SC allergens as a way of correlating their amounts with the severity of the allergen exposure and in determining the threshold values requires a highly sensitive, specific, and reproducible technique. Isotopic dilution mass spectrometry has played a crucial role in protein quantification in the last two decades, provoking the use of this technique for the present study (27).

Snow crab tropomyosin was previously characterized and '*de novo*' sequenced (8). Arginine kinase is also a protein of interest since it has been identified in other crustaceans but not snow crab (11-14). In this study, snow crab crude extract was profiled on SDS-PAGE and screened against patients' sera. Different allergenic proteins from this profile were characterized using tandem mass spectrometry. Previously, we identified and characterized the major allergen tropomyosin (8, 9). Arginine kinase, which is one of the most immunoactive proteins was isolated and purified from the crude extract. The allergencity of purified AK was examined by immunoblotting with allergenic patients' sera. The tryptic peptides were generated along with different types of derivatization reactions used in amino acid sequencing. The most abundant peptides were characterized using both ESI and MALDI ion sources, for protein identification and '*de novo*' sequencing using peptide mass fingerprinting (PMF) and peptide fragment fingerprinting (PFF), respectively. The mass spectra data were uploaded to the Mascot database search engine. The AK homology between snow crab and orange mud crab, which is the closest species, was studied. The active site motifs were evaluated using selected bioinformatics algorithms to confirm a signature pattern for AK. A unique and abundant tryptic peptide was selected and evaluated as a chemical surrogate for AK. Finally, this signature peptide and its deuterated isotopic homolog using d_3-*L*-alanine-were chemically synthesized and used in a preliminary study to develop a sensitive and specific quantification method for AK using multi reaction monitoring (MRM) LC-MS/MS. The levels of the snow crab AK, for the first time, were monitored in actual air samples collected from a simulated

processing plant. A tryptic-digested snow crab crude extract sample was used as a positive control.

3.2. Experimental

3.2.1. Chemicals and materials

All chemicals were used without further purification. Ammonium sulfate, acetonitrile, hydrochloric acid, and methanol were supplied by ACP (Montreal, Canada). Trypsin sequencing grade enzymes were purchased from Promega (WI, USA). Tris(hydroxymethyl)aminomethane (Tris), dithiotheritol (DTT), ethylenediaminetetraacetic acid (EDTA), formic acid (FA), ammonium bicarbonate, o-methylisourea hemisulfate, ammonium hydroxide, horseradish peroxidase (HRP), Chemiluminescent substrate, Sodium dodecylsulphate (SDS), ammonium formate and α-cyano-4-hydroxycinamic acid (HCCA) matrix were purchased from Sigma-Aldrich (St. Louis, MO, USA). The Bradford assay kit and PVDF immunoblot membrane were from BioRad (Hercules, CA, USA). The dialysis bags were purchased from Fischer Scientific (Roncho Dominguez, CA, USA). For desalting, the ZipTip C_{18} filters were purchased from Millipore Corporation (Bedford, MA, USA). The skimmed milk powder was purchased from the local supermarket. Tris Buffered Saline (TBS) and Phosphate Buffered Saline (PBS) tablets were purchased from Amresco, USA. The photo-sensitive

films were purchased from Amersham Bioscience, Germany. The developer and fixer were purchased from Kodak, USA. RapiGest SF surfactant from Waters Corporation (Milford, MA, USA), TopTip filters with Poly-Hydroxyethyl (HILIC) resin from PolyLC Inc. (Columbia, MD, USA), and 33-mm polytetrafluoroethylene (PTFE) filters for air sampling was bought from SKC Inc. (Eighty Four, PA, USA). The signature peptide, **LVSAVNEIEK** (purity > 98.33%; molar mass = 1101.27 Da) and its deuterated isotopic homolog using d_3-*L*-alanine- (purity > 96.80%; molar mass 1104.27 Da) were synthesized by GeneMed Synthesis (San Francisco, CA, USA).

3.2.2. Arginine kinase purification and SDS profiling

Arginine kinase was purified from crab leg muscle extract by ammonium sulfate precipitation following a protocol developed by Garcia-Orozco *et al.* (11). The total protein concentration of the crude extract and the AK isolate was determined using the Bradford assay. Since this crude extract does not contain many different proteins we utilized a 12% SDS-PAGE gel electrophoresis for protein profiling of the crude extract, the ammonium sulfate precipitated fractions, and the purified AK. A protein solution (10μg) was added to each of the wells, and electrophoresis was run at 170 V until the tracker dye was seen at the base of the gel. One gel was treated with Coomassie Brilliant Blue R-250, using standard protocol. For the second gel, proteins were transferred to a PVDF membrane at 15V for 20 min. After the transfer was completed, the membrane

was placed in a blocking solution (5% skimmed milk in Tris-buffered saline (TBS)) for immunoblotting.

3.2.3. Immunoblotting

To demonstrate the allergenicity of the isolated crab proteins, different bands were analyzed for IgE antibody binding from allergic patients with marked allergic symptoms to shellfish. Patients for this study were selected based on clinical reactivity to shellfish. Ethics approval for this study was acquired by Monash University (Victoria, Australia) as part of an ongoing survey. For immunoblotting, protein extracts were electrophoretically separated (see section 3.2.2), proteins transferred and incubated with human serum (diluted 1:10 in 1% skimmed milk) overnight at 4°C using a slot-blot apparatus, (Idea Scientific, USA). Subsequently, blots were washed three times with PBS-Tween and the membrane incubated for 1 hr, in 5 mL of HRP labeled rabbit polyclonal antihuman IgE antibody (DAKO, USA) (diluted 1:2000) in PBS-T containing 1% skimmed milk. After washing the membrane with PBS-T three times the blots were incubated with the chemiluminescent substrate and analyzed for IgE reactivity using ECL technique (8, 22).

3.2.4. In-gel digestion and guanidation

Six immunoreactive proteins, as identified by SDS-gel electrophoresis and subsequent immunoblotting with allergic patient serum, were excised from SDS-PAGE gels. An in-gel guanidation procedure was performed on these selected proteins for subsequent MALDI-QqToF-MS using the protocol developed by Sergeant *et al.* (28). The gel pieces were destained by washing three times with 200 mM of $NH_4H\,CO_3$ in solution of 50% acetonitrile in dH_2O, at 30°C for 30 min. The destained piece was dried under a stream of N_2, and then covered by a solution of 50mM NH_4HCO_3 pH 7.8, containing 5 ng/μL trypsin in ice for 30 min for rehydration. After rehydration, the excess solution was sucked out. The gel was covered by a solution of 50 mM of NH_4HCO_3 and incubated at 37°C overnight to enhance protein digestion. The water soluble peptides were extracted twice with the incubation solution and other remaining peptides extracted twice with 0.15%TFA in 50% ACN after a 2 min vortex mixing. The samples were lyophilized, and reconstituted, prior to analysis with 10μL of 0.1% TFA and desalted with C_{18} ZipTip.

3.2.5. MALDI-QqToF-MS

3.2.5.1. Matrix/Sample Preparation

The two-layer sample/matrix preparation for plate spotting was employed as described in Chapter 2, section 2.2.6.1. Whereas the first layer solution consisted 20 mg of HCCA in 1 mL (1/9) methanol/acetone the second layer solution consisted 40% ACN

of H_2O saturated by HCCA. A 0.5 µL of the first layer matrix solution was applied to a MALDI target. 1 µL of the second layer matrix solution was mixed with 1 µL of sample. 1 µL of the sample/matrix mixture was deposited onto the first layer and allowed to dry, followed by an on-target wash step. By adding 1 µL of water on top of the dry spot and blowing the water off using a pulse of air (after 10 seconds), a subsequent amount of salt was removed.

3.2.5.2. MALDI-QqToF-MS and CID-MS/MS

MALDI-MS and low-energy (CID) analyses were carried out on a QSTAR XL hybrid quadrupole-quadrupole/time-of-flight tandem mass spectrometer (QqToF-MS/MS) (Applied Biosystems/MDS Sciex, Foster City, USA) equipped with an o-MALDI ion source (Applied Biosystems, Foster City, CA).

3.2.6. LC- ESI- QqToF-MS

Peptide separation was conducted using a DIONEX UltiMate3000 Nano LC System (Germering, Germany). A 250 *f*mol sample of protein digest dissolved in 0.1% TFA was loaded onto a precolumn (300 µm ID x 5 mm, C_{18} PepMap 100, 5 µm (LC Packing, Sunnyvale, CA) for desalting and concentrating. Peptides were then eluted from the pre-column and separated on a nanoflow analytical column (75 µm ID x 15 cm, C18

PepMap 100, 3 μm, 100 A, (LC Packing, Sunnyvale, CA) at 180 nL/min using the following gradient. The aqueous mobile phases consisted of (A) 0.1% formic acid/ 0.01% trifluoroacetic acid/ 2% ACN and (B) 0.08% formic acid/ 0.008% trifluoroacetic acid / 98% ACN. A gradient of 0% B for 10min, 0-60% B in 55 min, 60-90% in 3 min, and 90% B in 5 min was applied. Including a regeneration step one run was 106 min long.

The ESI-MS of the LC-eluting peptides were measured with the same hybrid QqToF-MS/MS system equipped with a nanoelectrospray source (Protana XYZ manipulator). The nanoelectrospray was generated from a PicoTip needle (10 μm i.d., New Objectives, Wobum, USA) at a voltage of 2400 V.

This protein was further analyzed by CID-MS/MS and the resulting peptides spectra were searched by using the National Center for Biotechnology Information non-redundant database (NCBInr) with the Matrix Science (Mascot) search engine (precursor and product ion mass tolerances set at 0.2 Da. Methionine oxidation was allowed as a variable modification and guanidinyl (K) as a fixed modification since the guanidation derivatisation has been performed. Peptides were considered identified if the Mascot score was over 95% confidence limit.

3.2.7. Air sample collection, protein extraction and tryptic digestion

Airborne snow crab proteins were generated by processing crab in a simulated crab processing station and air samples were collected as described in Chapter 4, section 4.2.3, where airborne crab tropomysin was measured . The procedures for air sample collection, protein extraction from filters, tryptic digestion, and detergent removal prior to LC-MS/MS analysis were performed as described elsewhere (27).

3.2.8. Multiple reaction monitoring tandem mass spectrometry quantification method

Signature peptide analysis was performed following the MRM LC-MS/MS protocol that developed and described in Chapter 4, section 4.2.6, however, some mass spectrometry parameters were changed. The eluted peptides' droplets were desolvated at the ESI with gas flow rate of 400 l/hr and a temperature of 250 °C. The ions were accelerated through the capillary and orifice cone at 3.02 kV and 40 V, respectively. The precursor ions were fragmented at low energy CID using argon gas and the collision energy of 13 eV. The $[M+2H]^{2+}$ precursor ions of the unlabeled and labeled forms of the signature peptide were 551.52 and 552.64 m/z, respectively. Data processing was performed with Mass Lynx 4.1 software.

3.3. Result and discussion

The crude extract of the SC proteins (from leg sections) was profiled by SDS-PAGE as shown in Figure 5.1A. Proteins from numerous bands were excised and treated

with tryptic digestion for further characterization by '*bottom up*' tandem mass spectrometry approach. Several protein bands within the range 20-113 kDa were analyzed and their relevant peptides '*de novo*' sequenced using PFF by tandem mass spectrometry. The precursor and product spectra of the peptides were uploaded and searched in Mascot search engine against the NCBInr database. The proteins' identity and their relevant information are summarized in Table 3.1.

The number of peptides that has been used in protein identification varied with the natural abundance of each protein in the crude extract. On the other hand, the lack of specific genomic information in the databases for snow crab made the matching of the MS results very difficult since most of the identified proteins were obtained by comparison with the closest species available in the databases.

All the SC proteins identified in Table 3.1 are also immunoreactive as these proteins demonstrate serum IgE antibody reactivity by immunoblotting in our study. This current study further identifies the proteins which were previously noted as being involved in allergic reactions to snow crab, ranging from 20 to 113 kDa (7). One of these allergens was recently sequenced by Dr. Helleur's group as the 33 kDa tropomyosin novel snow crab allergen (see Chapter 2).

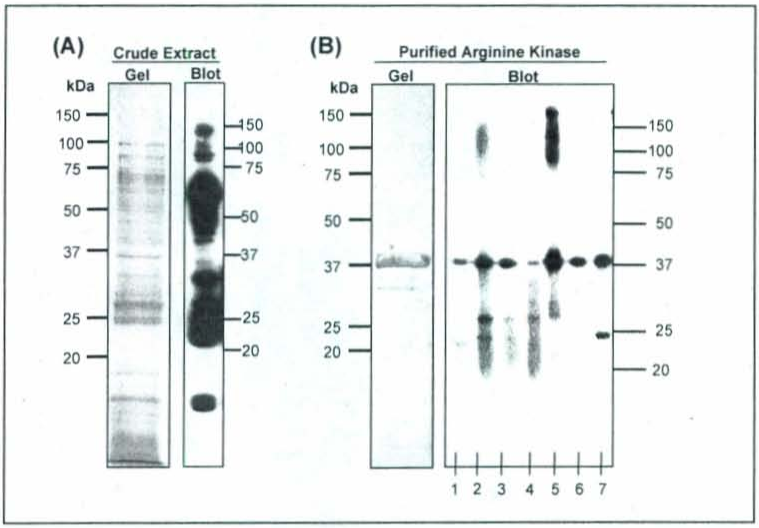


Figure 3. 1: (A) Crude extract of snow crab muscle protein analyzed by SDS-PAGE Coomassie staining and characterized by immunoblotting using a pool of patients' IgE antibody. (B) Purified arginine kinase analyzed by SDS-PAGE Coomassie staining, and by immunoblot IgE antibody binding using serum from seven patients.

Subsequently, arginine kinase (AK) (40-kDa) was isolated in the present study from frozen non-processed SC meat and purified using ammonium sulphate precipitation. The

purification steps were monitored by SDS-PAGE (results are not shown) and the final purified AK shown by its SDS band in Figure 3.1B.

Table 3. 1: List of the allergenic muscle proteins from snow crab and the '*de novo*' sequences of their relevant peptides

Protein Name	Molecular mass (kDa)	Mascot score	Selected peptides
Sarcoplasmic Ca-binding	20	44	[108] VATVSLPR [115]
Troponin	23	118	[38] KGFMTPER [45] [162] AAEFNFR [168]
Tropomyosin	33	494	ref (8)
Arginine kinase	40	885	table 2
α-Actine	42	891	[21] AGFAGDDAPR [30] [31] AVFPSIVGRPR [41] [53] DAYVGDEAQSKR [64] [64] RGILTLK [70] [98] IAPEESPVLLTEAPLNPK [115] [150] TTGIVLDTGDGVTHTVPIYEGYCLPHAILR [179] [180] LDLAGRDLTAYLTK [193] [199] GYSFTTTAEREIVR [212] [241] SYELPDGQVITIGNER [256] [287] CDIDIRK [293] [293] KDLFANNVLSGGTTMYPGIADR [314] [318] EITALAPPTIK [328] [329] IKIIAPPER [337] [331] IIAPPERK [338] [362] EEYDESGPGIVHR [374] [36] YGPNELPAEEGK [47] [111] NAESAIEALKEYEPEMGK [128] [144] EIVPGDLVEISVGDKIPADLR [164]
Smooth Endoplasmic Reticulum Ca^{2+} ATPase (SERCA)	113	585	[175] IDQSILTGESVSVIK [189] [206] NILFSGTNVAAGK [18] [237] TQMAETEEIKTPLQQK [252] [437] VGEATETALIVLGEK [451] [482] EFTLEFSR [489] [621] VIVITGDNK [629] [713] KAEIGIAMGSGTAVAK [728]

Further, the allergencity of the AK extract was evaluated by sixteen different patients' sera with allergic symptoms. Seven of these patients' IgE have strong reactivity with AK (~43%). The IgE binding profile for these seven patients is demonstrated by Figure 3.1B. Serum of a non-allergic patient did not show any IgE reactivity (data not shown). Interestingly, the 43% of the patients' sera reacted with snow crab AK, confirming that AK is one of the important crab allergens along with tropomyosin. The AK band was excised and its identity confirmed by PMF mass spectrometry. The precursor ion spectra (Figure 3.2) of the generated peptides were uploaded to the Mascot database search engine. The probability based Mowse score was 80 for arginine kinase, where the data of any protein scored >75 is highly significant ($P<0.05$). This search was matched with AK from orange mud crab (*Scylla olivacea*), marbled crab (*Pachygrapsus marmoratus*), and green crab (*Carcinus maenas*). A comparison was performed among the different crustaceans using multisequence alignment freeware. A phylogenetic tree based on AK's amino acid sequence was constructed by the neighbour joining method. A cladogram was described based on the maximum probability of distances. Orange mud crab was found to have the closest similarity in evolution to snow crab as shown in Figure 3.3.

Several of the snow crab AK tryptic peptides produced by both in-solution and in-gel digestions were *'de novo'* sequenced using 'bottom up' tandem MS approach. MALDI and ESI ion sources were used in order to obtain a maximum number of product ions. The peptide's amino acid sequence was interpreted manually and later confirmed

using bioinformatics tools. The MALDI ion source is known to ionize preferentially the arginine-containing peptides (28).

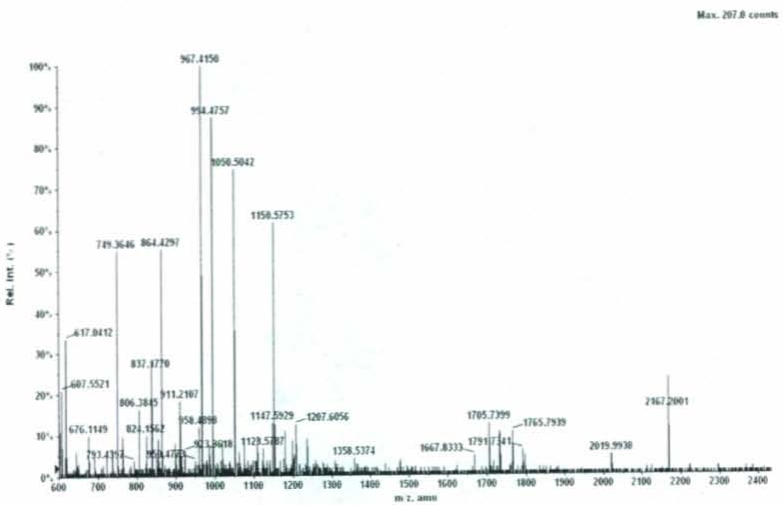


Figure 3. 2: An ESI MS spectrum from snow crab AK in-solution tryptic digestion. Spectrum was uploaded to PMF Mascot mode for protein characterization.

To increase the sensitivity of the lysine-containing peptides, guanidation of lysine was performed whereby lysine residues were converted to homoarginine. This amino acid has the same chemical behaviour and proton affinity as arginine in a MALDI source (results are not shown). For minimizing ion suppression effect and increasing the sensitivity of

mass spectrometry, the tryptic peptides were also separated based on their hydrophobicity using nano-HPLC prior to ionization by an ESI source.

The peptide ions from both the ESI (Figure 3.2) and MALDI precursor ion spectra, (i.e $[M+H]^{1+}$ m/z 923.3618, 850.4348, 1020.5053, 864.4297, etc) were exposed to low energy collision induced dissociation (CID) for peptide '*de novo*' sequencing using a quadruple time-of-flight (QqToF) tandem mass spectrometer. Data interpretation was performed manually and the peptide sequence confirmed by uploading them to the Mascot MS/MS mode. Two representative product ion spectra are shown in Figure 3.4. The Mascot software was tuned to search against non-redundant proteins NCBInr database.

Since the orange mud crab is the closest species to the SC, most of the SC AK's tryptic peptides matched with the orange mud crab's AK. The Mascot outcome for the snow crab AK meets the criteria of identity and extensive homology, wherein the Mows score average ~ 1185 as orange mud crab AK (*S. olivacea*).

The amino acid sequence of snow crab AK (this study) that matched totally or partially with orange mud crab AK is summarized in Table 3.2. The protein sequence data reported in this chapter will appear in the UniProt knowledgebase under the accession number **P86699** for AK (*Chionoecetes opilio*).

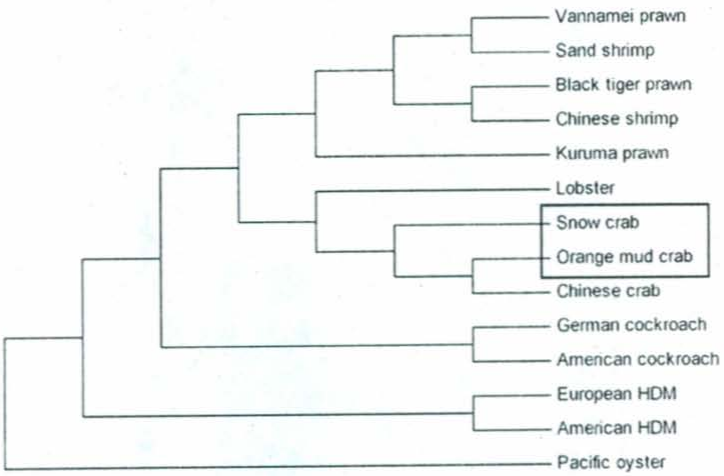


Figure 3. 3: Phylogenetic tree based on the amino acid sequences of arginine kinase from snow crab (this study) in comparison with other invertebrate and vertebrates

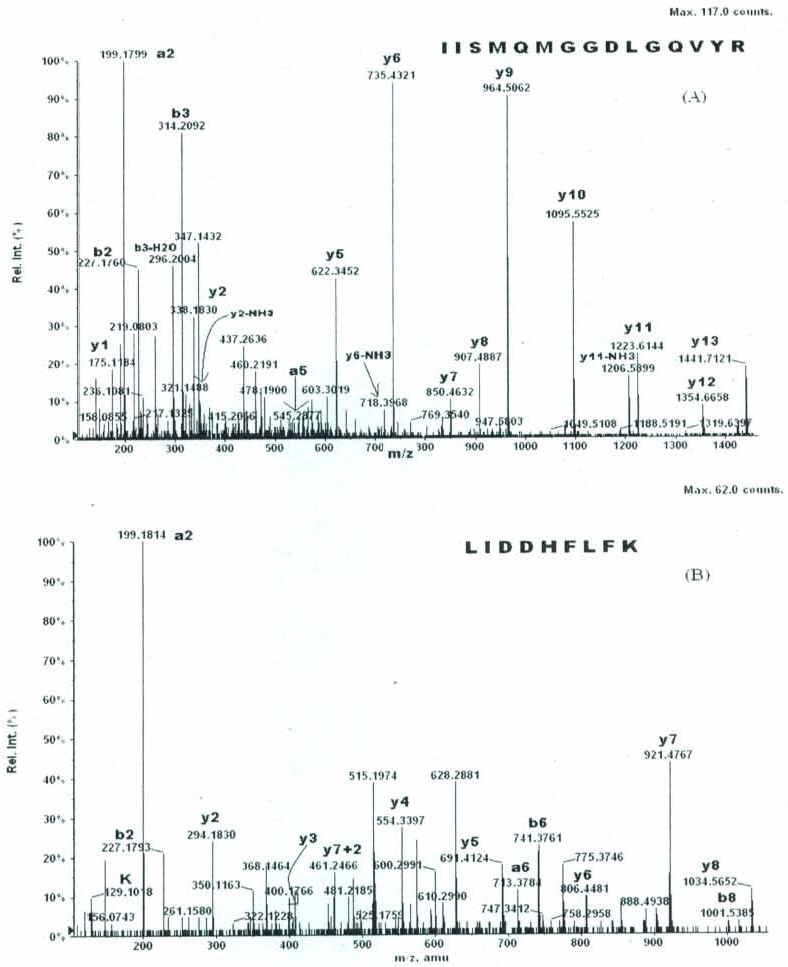


Figure 3. 4: Representative ESI product ion spectra of arginine kinase tryptic peptides. The selected precursor ions are (A) $[M+2H]^{2+}$, m/z 834.4 and (B) $[M+2H]^{2+}$,m/z 574.3. The amino acid sequence is shown for each spectrum.

Using BLAST 2.0 (http://www.ch.embnet.org/cgi-bin/blast20_parser.pl) the alignment scores for the two AK sequenced proteins was 701 bits (1810) with 96% identity (Figure 3.5).

```
SCAK:   1   MADAATISKLEEGFKKLQGATDCKSLLKKYLTKDVFDQLKAKKTSLGATLLDVIQSGVEN  60
            MADAATI+KLEEGFKKL+ ATDCKSLLKKYLTK VFDQLK KKTSLGATLLDVIQSGVEN
OMCAK:  1   MADAATIAKLEEGFKKLEAATDCKSLLKKYLTKSVFDQLKGKKTSLGATLLDVIQSGVEN  60

SCAK:   61  LDSGVGVYAPDAEAYTLFAPLFDPIIEDYHKGFKQTDKHPNKDFGDVNQFVNVDPDGKFV  120
            LDSGVGVYAPDAEAYTLFAPLFDPIIEDYHKGFKQTDKHPNKDFGDVNQFVNVDPDGKFV
OMCAK:  61  LDSGVGVYAPDAEAYTLFAPLFDPIIEDYHKGFKQTDKHPNKDFGDVNQFVNVDPDGKFV  120

SCAK:   121 ISTRVRCGRSMEGYPFNPCLTEAQYKEMEAKVSSTLKGLEGELKGSFYPLTGMAKDVQQK  180
            ISTRVRCGRSMEGYPFNPCLTEAQYKEME+KVSSTL  LEGELKG++YPLTGM KDVQQK
OMCAK:  121 ISTRVRCGRSMEGYPFNPCLTEAQYKEMESKVSSTLSNLEGELKGTYYPLTGMTKDVQQK  180

SCAK:   181 LIDDHFLFKEGDRFLQAANACRYWPSGRGIFHNDKKTFLVWCNEEDHLRIISMQMGGDLG  240
            LIDDHFLFKEGDRFLQAANACRYWP+GRGI+HND KTFLVWCNEEDHLRIISMQMGGDLG
OMCAK:  181 LIDDHFLFKEGDRFLQAANACRYWPTGRGIYHNDNKTFLVWCNEEDHLRIISMQMGGDLG  240

SCAK:   241 QVYRRLVSAVNEIEKRVPFSHMDRLGFLTFCPTNLGTTVRASVHIKLPKLAANREKLEEV  300
            QVYRRLVSAVNEIEKRVPFSHMDRLGFLTFCPTNLGTTVRASVHIKLPKLAANREKLEEV
OMCAK:  241 QVYRRLVSAVNEIEKRVPFSHMDRLGFLTFCPTNLGTTVRASVHIKLPKLAANREKLEEV  300

SCAK:   301 AGKYSLQVRGTRGEHTEAEGGVYDISNKRRMGLTEFQAVKEMQDGILELIKIEKEMQ  357
            AGKYSLQVRGTRGEHTEAEGGVYDISNKRRMGLTEFQAVKEMQDGILELIKIEKEMQ
OMCAK:  301 AGKYSLQVRGTRGEHTEAEGGVYDISNKRRMGLTEFQAVKEMQDGILELIKIEKEMQ  357
```

Figure 3. 5: Amino acid alignment between snow crab AK (SCAK) and organic mud crab AK (OMCAK) using the SIB BLAST Network Service. The match identity is 96% and the positive value is 98%

Table 3.2: '*De novo*' sequencing of the product ion spectra obtained from ESI-QqToF and MALDI-QqToF experiments of the major tryptic peptides from snow crab AK.

Start-end	Mass			Charge (z)	Peptide '*de novo*' sequence	Ion source
	Observed	Expected	Calculated			
10 - 16	425.746	849.4791	849.4596	+2	LEEGFKK	
181 - 189	574.3178	1146.6210	1146.6073	+2	LIDDHFLFK	
181 - 193	535.6201	1603.8384	1603.7995	+3	LIDDHFLFKEGDR	
194 - 202	525.7632	1049.5119	1049.5076	+2	FLQAANACR	
230 - 244	834.4453	1666.8761	1666.8171	+2	IISMQMGGDLGQVYR	
230 - 245	608.6669	1822.9788	1822.9182	+3	IISMQMGGDLGQVYRR	
246 - 255	551.3182	1100.6219	1100.6077	+2	LVSAVNEIEK	
246 - 256	629.3746	1256.7346	1256.7088	+2	LVSAVNEIEKR	
256 - 264	575.8096	1149.6047	1149.5792	+2	RVPFSHHDR	
257- 264	497.7533	993.4921	993.4781	+2	VPFSHHDR	ESI
265 - 280	898.9819	1795.9492	1795.9291	+2	LGFLTFCPTNLGTTVR	
281 - 289	497.7584	993.5022	993.4781	+2	ASVHIKLPK	
295 - 303	501.7672	1001.5198	1001.5393	+2	EKLEEVAGK	
297 - 309	746.4394	1490.8642	1490.8093	+2	LEEVAGKYSLQVR	
310 - 328	673.9952	2018.9639	2018.9294	+3	GTRGEHTEAEGGVYDISNK	
313 - 328	569.2693	1704.7859	1704.7591	+3	GEHTEAEGGVYDISNK	
313 - 329	621.3206	1860.9400	1860.8602	+3	GEHTEAEGGVYDISNKR	
330 - 340	640.3662	1278.7179	1278.6754	+2	RMGLTEFQAVK	
331 - 340	562.3232	1122.6318	1122.5743	+2	MGLTEFQAVK	
341 - 351	644.8727	1287.7308	1287.6744	+2	EMQDGILELIK	
341 - 354	829.9928	1657.9711	1657.8960	+2	EMQDGILELIKIEK	
1 - 9	923.3618	922.3545	922.4430	+1	M□ADAATISK	
16 - 24	1020.5053	1019.4980	1019.5070	+1	KLQGATDCK	
34 - 40	864.4297	863.4224	863.4389	+1	DVFDQLK	
95 - 102	967.4150	966.4077	966.4883	+1	QTDKHPNK	
103 - 118	1765.7939	1764.7866	1764.7955	+1	DFGDVNQFVNVDPDGK	
119 - 124	722.4092	721.4019	721.4123	+1	FVISTR	
119 - 126	977.4974	976.4901	976.5818	+1	FVISTRVR	
181 - 189	1147.5929	1146.5856	1146.6073	+1	LIDDHFLFK	
194 - 202	1050.5042	1049.4969	1049.5076	+1	FLQAANACR	MALDI
203 - 216	1706.7247	1705.7174	1705.7961	+1	YWPSGRGIYHNDNK	
230 - 244	1667.7269	1666.7196	1666.8171	+1	IISMQMGGDLGQVYR	
246 - 256	1257.6976	1256.6903	1256.7088	+1	LVSAVNEIEKR	
256 - 264	1150.5753	1149.5680	1149.5792	+1	RVPFSHHDR	
257 - 264	994.4757	993.4684	993.4781	+1	VPFSHHDR	
265 - 280	1796.9241	1795.9168	1795.9291	+1	LGFLTFCPTNLGTTVR	
304 - 309	765.4236	764.4163	764.4181	+1	YSLQVR	
331 - 340	1123.5787	1122.5714	1122.5743	+1	MGLTEFQAVK	
352 – 357	793.4397	792.4324	792.3687	+1	IEKEMQ	

Interestingly, the N-terminal peptide (^{1}**MADAATISK**^{9}) identified in MALDI experiments, indicates that the methionine residue is not acetylated, a structural feature which is not common in eukaryotic proteins (8, 9). The kinase active site of the snow crab AK was evaluated using the *Prosite* bioinformatics tool. Sequence hits were with Guanido kinase model (PS00112) as ATP guanido phosphotransferases, matching the AK sequence at site (^{271}**CPTNLGT**^{277}). The ATP guanido phosphotransferases active site model has a consensus pattern of (C-P-X (0, 1)-(ST)-N-(ILV)-G-T), and represents a family of structurally and functionally related enzymes that reversibly catalyze the transfer of phosphate between ATP and various phosphogens (e.g. creatine, glycocyamine, argentine kinase, etc.). The cysteine residue in the site is implicated in the catalytic activity of these enzymes, and the amino acid region near this active site residue is highly conserved and could be used as signature pattern (29).

Developing an analytical method for quantifying snow crab AK by isotopic dilution mass spectrometry in different matrices requires assigning a signature peptide as a stoichiometric surrogate for the target protein. Ideally the signature peptide should be unique only for the target protein, shorter than 15 amino acids, and free of chemically reactive residues such as methionine and cysteine (8, 27, 30). Following this criteria several candidate peptides as listed in Table 3.2, were examined and evaluated. Based on the above criteria the best signature peptide of AK was determined to be ^{246}**LVSAVNEIEK**^{255}. The ESI product spectrum for the candidate signature peptide is given in Figure 3.6. The BLAST test shows that this peptide is not only specific for AK

from snow crab, but also can be utilized as a universal signature peptide for other crustaceans' AK.

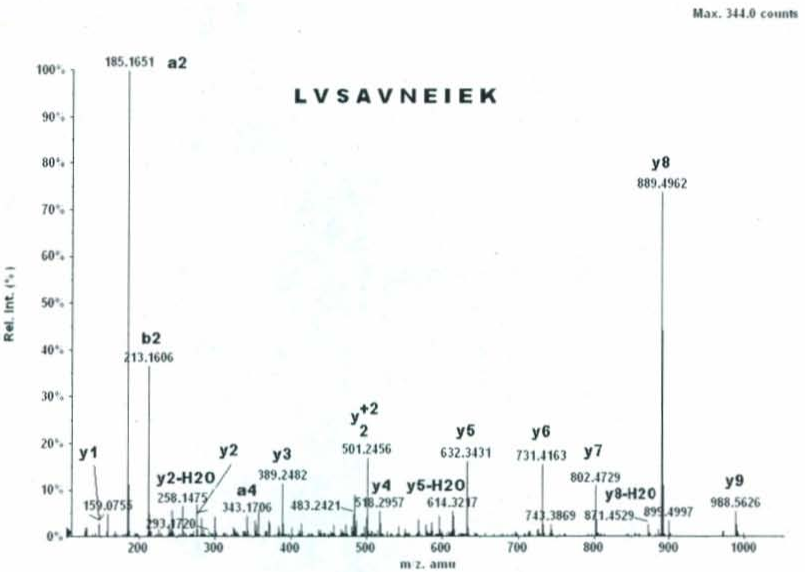


Figure 3. 6: ESI product spectrum for the snow crab AK signature peptide to be selected for quantification purposes of snow crab AK.

The AK signature peptide was chemically synthesized in light and heavy forms as analytical standards, which were subsequently used to develop a quantification method similar to a previous study (27). The MRM transitions for both the signature peptide and its deuterated form were tuned to be 551>890 and 552>892, respectively. A

chromatogram for the digested SC crude extract is shown in Figure 3.7, which is routinely used as a positive control for actual air samples analyses. Air samples were collected from different sites of a simulated processing plant, where the proteins were collected, extracted from filter papers, and tryptic digested for quantification purposes.

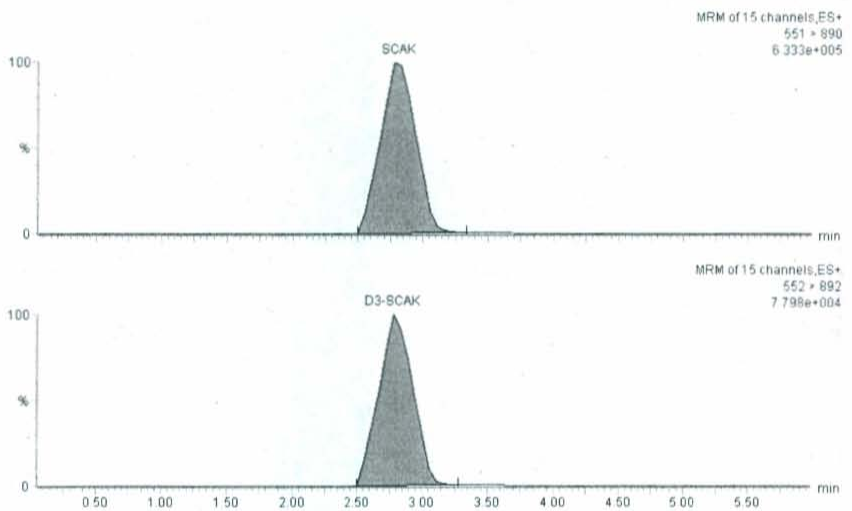


Figure 3. 7: An MRM LC-MS/MS chromatogram for the signature peptide (SCAK) and its isotopic analog (d_3-SCAK) present in the digested sample of SC crude extract.

The levels of the SCAK (as its surrogate peptide) are summarized in Table 3.3. To the best of our knowledge, it was noteworthy that arginine kinase, for the first time, is being molecularly identified as one of the aerosolized allergens in seafood processing

environment. The levels of SCAK are shown to vary among sampling locations. For example, high levels of SCAK were collected on the personal breathing zone filter during butchering, while the lowest was from a filter sample 3 m away from butchering. These experiments demonstrate the validity of the use of a signature peptide to surrogate chemically the presence of AK in crude extract and in air samples will be useful developing a very selective and sensitive quantification method for airborne proteins in the future.

Table 3. 3: Levels of snow crab AK in different work stations inside a simulated crab processing plant.

Sample Name	Conc. as solution (nM)	Amount per filter (pmol)	Volume of Air (m^3)	$\mu g/m^3$
Butchering (PBZ)	10.48	1.0484	0.045	0.93
Butchering (Area 1)	26.42	2.6419	0.175	0.60
Butchering (Area 2)	12.32	1.2324	0.190	0.26
Cooking (PBZ)	22.67	2.2675	0.150	0.60
Cooking (Area 1)	21.09	2.1095	0.130	0.65
Cooking (Area 2)	21.52	2.1515	0.175	0.49
Blank (Area 2)	<LOQ	n/a	0.200	n/a

3.4. Conclusion

Arginine kinase has recently been reported as a major seafood allergen in prawns. The present study evaluates, for the first time, the allergenicity of the snow crab AK, by examining its immunoreactivity with the sera of seven allergic patients. The primary structure of the AK was characterized through amino acid sequencing using '*bottom up*' tandem mass spectrometry. The amino acid sequence was registered in the UniProt knowledgebase. The active site of AK was determined with a consensus pattern (C-P-X (0, 1)-(ST)-N-(ILV)-G-T). The signature peptide of AK was determined and utilized as a universal signature peptide among the crustaceans' AK. An MRM tandem mass spectrometry quantification method was developed using the designated signature peptide to measure, for the first time, the levels of the aerosolized form of SCAK inside crab processing workplaces. Finally, several other allergenic proteins present in the SC crude extract were identified using the same MS strategy as that followed in the snow crab AK.

3.5. References

1. Mullowney, R., Dawe, E. Development of performance indices for the Newfoundland and Labrador snow crab (*chionoecetes opilio*) fishery using data from a vessel monitoring system. *Fisheries Research*. **2009**.100; 248-254.

2. Lopata, A. L, O'Hehir, R. E, Lehrer, S. B., Shellfish allergy. *Clin Exp Allergy*. **2010**. 40; 850-858.

3. Lopata, A. L, Lehrer, S. B. New insights into seafood allergy. *Curr Opin Allergy Clin Immunol*. **2009.** 9; 270-277.

4. Jeebhay, M. F., Robins, T. G., Lehrer, S. B., Lopata, A. L. Occupational seafood allergy: a review. *Occup Environ Med*. **2001**. 58; 553-562.

5. Chu, K. H., Tang, C. Y., Wu, A., Leung. P. S. Seafood Allergy: lessons from clinical symptoms, immunological mechanisms and molecular biology. *Adv Biochem Engin/Biotechnol*. **2005**. 97; 205-235.

6. Motoyama, K., Suma, Y., Ishizaki, S., Nagashima, Y., Shiomi, K. Molecular cloning of tropomyosins identified as allergens in six species of crustaceans. *J Agric Food Chem*. **2007**. 55; 985-991.

7. Gill, B. V., Rice, T. R., Cartier, A., Gautrin, D., Neis, B., Horth-Susin, L., Jong, M., Swanson, M., Lehrer, S. B. Identification of crab proteins that elicitIgE reactivity in snow crab–processing workers. *J Allergy Clin Immunol*. **2009**. 124; 1055-1061.

8. Abdel Rahman, A. M., Lopata, A. L, O'Hehir, R. E., Robinson, J. J., Banoub, J. H., Helleur, R. J. Characterization and '*de novo*' sequencing of snow crab tropomyosin enzymatic peptides by both electrospray ionization and matrix-assisted laser desorption ionization QqToF tandem mass spectrometry. *J. Mass. Spectrom*. **2010**. 45; 372-381.

9. Abdel Rahman, A. M., Kamath, S. D., Lopata, A. L., Helleur, R. J. Analysis of the allergenic proteins in black tiger prawn *(Penaeus monodon)* and

characterization of the major allergen tropomyosin using mass spectrometry. *Rapid Commun. Mass Spectrom.* **2010**. 24; 2462-2470.

10. Yu, C. J., Lin, Y. F., Chiang, B. L., Chow, L. P. Proteomics and immunological analysis of a novel shrimp allergen, Pen m 2. *J Immunol.* **2003**. 170; 445-453.

11. García-Orozco, K. D., Aispuro-Hernández, E., Yepiz-Plascencia, G., Calderón-de-la-Barca, A. M., Sotelo-Mundo, R. R. Molecular characterization of arginine kinase, an allergen from the shrimp Litopenaeus Vannamei. *Int Arch Allergy Immunol.* **2007**; 144: 23-28.

12. France, R. M., Sellers, D. S., Grossman, S. H. Purification, Characterization, and hydrodynamic properties of arginine kinase from gulf shrimp (*Penaeus aztecus*). *Arch Biochem Biophys.* **1997**. 345; 73-78.

13. Yao, C. L., Wu, C. G., Xiang, J. H., Dong, B. Molecular cloning and response to laminarin stimulation of arginine kinase in haemolymph in Chinese shrimp, *Fenneropenaeus chinensis*. *Fish shellfish immunol.* **2005**. 19; 317-329.

14. Ortea, I., Cañas, B., Gallardo, J. M. Mass spectrometry characterization of Species-Specific peptides from arginine kinase for the identification of commercially relevant shrimp species. *J Proteome Res.* **2009**. 8; 5356-5362.

15. Bi, X. Z., Chew, F. T. Molecular, proteomic and immunological Characterization of isoforms of arginine kinase, a cross-reactive invertebrate pan-allergen, from the house Dust Mite, dermatophagoides farinae. *J Allergy Clin Immunol.* **2004**. S226

16. Binder, M., Mahler, V., Hayek, B., Sperr, W. R., Schöller, M., Prozell, S., Wiedermann, G., Valent, P., Valenta, R., Duchěne, M. Molecular and immunological characterization of arginine kinase from the indianmeal moth, *Plodia interpunctella*, a novel cross-reactive invertebrate Pan-allergen. *J Immunol.* **2001**. 167; 5470-5477.

17. Liu, Z., Xia, L., Wu, Y., Xia, Q., Chen, J., Roux, K. H. Identification and characterization of an arginine kinase as a major allergen from Silkworm (*Bombyx mori*) larvae. *Int Arch Allergy Immunol.* **2009.** 150; 8-14.

18. Brown, A. E., France, R. M., Grossman, S. H. Purification and characterization of arginine kinase from the American cockroach (*Periplaneta americana*). *Arch. Insec. Biochem Physiol.* **2004**. 56; 51-60.

19. Sookrung, N., Chaicumpa, W., Tungrongchitr, A., Vichyanond, P., Bunnag, C., Ramasoota, P., Tongtawe, P., Sakolvaree, Y., Tapchairsri, P. Periplaneta Americana arginine kinase as a major cockroach allergen among Thai patients with major cockroach allergies. *Environ Health Perspect.* **2006**. 114; 875-880.

20. Tungtrongchitr, A., Sookrung, N., Indrawattana, N., Sae-Lim, J., Punduang, S., Phonrat, B., Wanachiwanawin, D., Chaicumpa, W. Seasonal levels of the major American cockroach allergen Per a 9 (Arginine Kinase) in Bangkok and their relevance for disease severity. *Asian pacific Journal of allergy and immunology.* **2009.** 27; 1-7.

21. Jeebhay, M. F., Cartier, A. Seafood workers and respiratory disease: an update. *Curr Opin Allergy Clin Immunol.* **2010**. 10; 104-113.

22. Lopata, A. L., Jeebhay, M. F., Reese, G., Fernandes, J., Fenemore, B., Elliott, A., Robins, T. G., Lehrer, S. B. Detection of fish antigens aerosolized during fish processing using newly developed immunoassays. *Int Arch Allergy.* **2005**. 138; 21-8.

23. Malo, J. L., Chretien, P., Mccants, M., Lehrer, S. B. Detection of snow-crab antigens by air sampling of a snow-crab production plant. *Clin. Exp. Allergy.* **1997**. 27; 75-78.

24. Kluwe, L., Maeda, K., Miegel, A., Fujita-Becker, S., Maeda, Y., Talbo, G., Houthaeve, T., Kellner, R. Rabbit skeletal muscle αα-tropomyosin expressed in baculovirus-infected insect cells possesses the authentic N-terminus structure and functions. *J Muscle Res Cell Motil.* **1995**. 16; 103-110.

25. Reese, G., Schicktanz, S., Lauer, I., Randow, S., Luttkopf, D., Vogel, L., Lehrer, S. B., Vieths, S. Structural, immunological and functional properties of natural recombinant Pen a 1, the major allergen of Brown Shrimp, Penaeus aztecus. *Clin Exp Allergy*. **2006.** 36; 517-524.

26. Rosmilah, M., Shahnaz, M., Masita, A., Noormalin, A., Jamaludin, M. Identification of major allergens of two species of local snappers: Lutjanus argentimaculatus (merah/ red snapper) and Lutjanus johnii (jenahak/ golden snapper). *Trop Biomed.* **2005**. 22; 171-177.

27. Abdel Rahman, A. M., Lopata, A. L., Randell, E. W., Helleur, R. J. Absolute quantification method and validation of airborne snow crab allergen tropomyosin using tandem mass spectrometry. *Analytica Chemica Acta.* **2010**. 681; 49-55.

28. Sergeant, K., Samyn, B., Debyser, G., Beeumen, J. V. *'De novo'* sequence analysis of N-terminal sulfonated peptides after in gel quanidination. *Proteomics.* **2005**. 5; 2369-2380.

29. Stein, D., Harn, A., David, R. A cloned ATP: guanidino kinase in the trematode Schistosoma mansoni has a novel duplicated structure. *J Biol Chem.* **1990**. 265; 6582-6588.

30. Brun, V., Masselon, C., Garin, J., Dupuis, A. Isotope dilution strategies for absolute quantitative proteomics. *J Proteomics.* **2009**. 72; 740-749.

Chapter 4: Absolute quantification method and validation of airborne snow crab allergen tropomyosin using tandem mass spectrometry [3]

[3] This Chapter has been published; Abdel Rahman, A. M., Lopata, A. L., Randell E. W., Helleur, R. J. Absolute quantification method and validation of airborne snow crab allergen tropomyosin using tandem mass spectrometry. *Analytica Chimica Acta.* **2010**. 681; 49–55.

Abstract

Measuring the levels of the major airborne allergens of snow crab in the workplace is very important in studying the prevalence of crab asthma in workers. Previously, snow crab tropomyosin (SCTM) was identified as the major aeroallergen in crab plants and a unique signature peptide was identified for this protein. The present study advances our knowledge on aeroallergens by developing a method of quantification of airborne SCTM by using isotope dilution mass spectrometry. Liquid chromatography tandem mass spectrometry was developed for separation and analysis of the signature peptides. The tryptic digestion conditions were optimized to accomplish complete digestion. The validity of the method was studied using international conference on harmonization protocol. Where 2-9% for CV (precision) and 101- 110 % for accuracy, at three different levels of quality control. Recovery of the spiked protein from PTFE and TopTip filters was measured to be 99 % and 96%, respectively. To further demonstrate the applicability and the validity of the method for real samples, 45 kg of whole snow crab were processed in an enclosed (simulated) crab processing line and air samples were collected. The levels of SCTM ranged between 0.36-3.92 $\mu g/m^3$ and 1.70- 2.31 $\mu g/m^3$ for butchering and cooking stations, respectively.

4.1. Introduction

Seafood plays an important role in nutrition worldwide, sustained by increasing international trade of a variety of new seafood products (1, 2). During seafood harvesting and processing, certain protein allergens are aerosolized into the working environment of vessels and processing plants. Allergic reactions to seafood are generated by otherwise harmless proteins which are, to some individuals, recognized as allergens. The main aeroallergens causing the shellfish allergy are believed to be tropomyosin (3-5) and possible sarcoplasmic calcium-binding protein (6), myosin light chain (7), and arginine kinase (8-9), which all can become aerosolized by seafood handling and processing. Previous studies have reported the prevalence of occupational seafood processing asthma to range from 7% to 36%, and occupational protein contact dermatitis from 3% to 11% (2). Many studies have used immunological reactivity techniques to detect TM such as enzyme-linked immunosorbent assay (ELISA) (10), radioallergosorbent test (RAST) (11), and immunoblotting (12). These techniques measure the levels of allergens, indirectly, by measuring the level of bound immunoglobulin's (IgE) to a standard antigen using radiolabels or colorimetric substrates. As a consequence, the sensitivity, specificity, and selectivity of these indirect techniques are limited in evaluating the levels of proteins recognized by these antibodies in air samples. These techniques therefore recognize various airborne proteins, maybe due to the cross-reactivity to various proteins are not only restricted to allergens and could originate from other sources besides the food processing (e.g. dust mites, microbes, etc.) (5).

Tropomyosin (TM) is a common muscle protein and mediates the interaction between the troponin complex and actin to regulate muscle contraction (10). Crustacean TM was first identified in shrimp by Hoffman *et al.* in 1981 (13). It is a water-soluble and heat-stable protein with molar masses ranging between 34 and 39 kDa (14). It has also been identified as one of the major airborne allergens present in crab processing plants (15). Developing a sensitive quantification method specifically for airborne snow crab tropomyosin (SCTM) will be very useful in detecting and controlling the worker's risk of developing allergic airway disease in the processing and harvesting workplaces (1, 4, 16). The difficulty of obtaining highly pure protein standard, such as removal of salts and surfactants, makes direct protein quantification an analytical challenge (17). The limited resolution of the sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and poor baseline separation of individual proteins by liquid chromatography are other analytical issues in protein quantification.

Traditionally the total protein concentration is measured by spectrophotometry methods such as the Lowry and Bradford techniques. These methods, however, are not protein specific, will measure contaminating proteins and have limited sensitivity. Various approaches have been developed in the last two decades for protein quantification using chromatographic methodologies coupled with tandem mass spectrometry. The combined technique's sensitivity, accuracy, and reproducibility make this an unrivaled method for protein primary structure characterization and quantification (18). Stable isotope labeling based approaches such as stable-isotope labeling by amino

acid (SILAC), isotope-coded affinity tag (ICAT), and isobaric tag for relative and absolute quantitation (iTRAQ), have been successfully developed for studies of the whole proteome (17-21). The isotopic labeling dilution concept has been implemented for quantification of proteins in biological samples. This approach takes advantage of the identical chromatographic behavior of an isotope-labeled standard analog to the native peptide, which can be differentiated by their mass-to-charge ratio in mass spectrometry (21).

Absolute quantification by isotopic dilution mass spectrometry first involves determining a signature peptide as a stoichiometric surrogate for the target protein. Ideally the signature peptide should be unique for the target protein, shorter than 15 amino acids, and free of chemically reactive residues such as methionine and cysteine (4, 18). Analytical standards of the signature peptide and its isotopic labeled analogue are then used for developing a quantification method using liquid chromatography tandem mass spectrometry (LC-MS/MS) and multi reaction monitoring (MRM) transitions. This study describes a method to quantify airborne SCTM, which is recognized as a major snow crab allergen (15, 22). This allergen has recently been characterized and amino acid '*de novo*' sequenced by mass spectrometry (Chapter 2) where the amino acid sequence of a signature tryptic peptide for SCTM which stoichiometrically surrogates this protein was reported. For this study, the signature peptide was chemically synthesized in unlabeled and labeled forms as standard and internal standard, respectively. A quantification method has been developed using LC-MS/MS (triple

quadruple mass analyzer). A comprehensive validation study was performed to determine the method's linearity, sensitivity, selectivity, and reproducibility. The linearity and the completeness of the tryptic digestion along with air filter recovery of SCTM were also determined. Finally, real air samples were collected from different processing stages of a simulated snow crab plant and assayed for SCTM using this developed method.

4.2. Materials and methods

4.2.1. Chemical and reagent

All chemicals were used without further purification. Standard signature peptide, SQLVENELDHAQEQLSAATHK (purity > 95.3%; molar mass = 2348.53 Da) and its deuterated isotopic homolog using d_3-L-alanine- (purity > 97.1%; molar mass = 2357.53 Da) were purchased from GeneMed Synthesis (San Francisco, CA, USA). Acetonitrile and formic acid were purchased from ACP (Montreal, Canada). Sodium dodecylsulfate (SDS), trypsin, ammonium bicarbonate, and ammonium formate were from Sigma ((St. Louis, MO, USA). RapiGest SF surfactant from Waters Corporation (Milford, MA, USA), TopTip filters with Poly-Hydroxyethyl (HILIC) resin from PolyLC Inc. (Columbia, MD, USA), and 33-mm polytetrafluoroethylene (PTFE) filters for air sampling was bought from SKC Inc. (Eighty Four, PA, USA).

Snow crab tropomyosin (SCTM) standard was obtained in a previous study (4) which used a series of extraction and purification techniques to purify SCTM from fresh crab tissue.

4.2.2. Peptide solutions

Working solutions (10.0 μM) of the light (TSP) and heavy (d_9-TSP) signature peptide of snow crab tropomyosin were prepared in deionised water (dH_2O) from stock solutions to be used as standard and internal standard, respectively. A linear calibration curve from 10.0 -1000.0 nM was prepared in 1.00-mL final volumes by serial dilution of the working solution of SP (10.0 μM), while a final concentration 500.0 nM of the d_9-TSP was prepared as internal standard. Three samples with different concentrations, 20.0, 250.0, and 850.0 nM, were used as quality control (QC) samples.

4.2.3. Air sample collection

Forty five kilograms of fresh snow crab were processed in a small simulated processing plant designed by the Marine Institute, Memorial University of Newfoundland. Area and personal breathing zones (PBZ) air samplers were used at a flow rate of 3.5 L/min using 33-mm PTFE filters. Three samplers were used in each processing stage and various background air samples were collected before and after

processing. The filters were stored at -80°C, to avoid proteases activities, until the protein extraction protocol.

4.2.4. Protein extraction from filters and tryptic digestion

Airborne allergens were collected on 33-mm PTFE filter papers and then extracted using 1 mL of 10% SDS solution (15) containing 500 nM of the d_9-TSP as internal standard (20). The extracts were freeze dried, then subjected to trypsin digestion. The protein residues were extracted out from the freeze dried sample using 200 µL of 50 mM ammonium bicarbonate, then 10 µL 0.01% RapiGest SF solution was added followed by boiling for 5 min in a water bath at 100°C, then cooled. A 100 ng/µL trypsin solution was made up in 50mM of ammonium bicarbonate. Protein samples were spiked with the appropriate amount of typsin solution and incubated at 37°C overnight (18 hr). The digestion was quenched by adding an equivalent volume of 1% formic acid to degrade the acid labile surfactant (RapiGest SF). Samples were then vortex mixed and centrifuged for 2 min at 4500 rpm. The supernatant was removed, freeze dried, and stored at -80°C. These samples were later treated to remove SDS prior to mass spectrometric analysis.

4.2.5. SDS removal protocol

The tryptic peptide residues were reconstituted with 200 µL of 80:20 (ACN: H_2O), and then introduced to a SDS removal protocol using TopTip filters supplied with HILIC resin. The filters were conditioned twice with 50 µl of releasing solution (50 mM formic acid in 5:95 ACN: H_2O), and twice with 50 µL of binding solution (15 mM ammonium formate, pH 3.0, in 85:15 ACN: H_2O). The peptide sample was loaded onto TopTip filters using a micropipette and the resin bed washed twice with the binding solution to remove any salts and other non-retained materials. The peptides were eluted twice using 75 µL of the releasing solution. The samples were freeze dried and stored at -80 ^{0}C. The samples were reconstituted in LC mobile phase before LC-MS/MS analysis.

4.2.6. LC-MS/MS operating conditions

Peptide analysis was performed on a Waters Alliance 2795 HPLC system coupled to a Micromass Quattro Ultima (Water Corporation, Milford, MA, USA) tandem mass spectrometer (triple-quad) operated in electrospray positive-ionization mode. The peptides were separated on a reversed phase chromatography column (Kinetex C_{18}, 2.1 X 100 mm, 2.6µm particle size (Phenomenex)) at 20°C. Peptide separation was performed by the following gradient profile. The aqueous mobile phase (A) consisted of HPLC-grade water with 0.1% formic acid; the organic phase (B) was ACN with 0.1% formic acid. The gradient was started at 5% B for 0.3 min, 5-90% B in 1 min, then 90% B for 3 min after which it reverted back to the 5% in 0.5 min (6 min run time). A 20 µL injection

was used at normal draw speed. During pre- and post- injection, the injection needle was washed with 200 µl of mobile phase B followed by 600 µL of mobile phase A. The eluted peptides were desolvated at the ESI with gas flow rate of 400 l/hr and a temperature of 250 °C. The ions were accelerated through the capillary and orifice cone at 3.02 kV and 40 V, respectively. The precursor ions were fragmented at low energy CID using argon gas and the collision energy of 13 eV. The $[M+4H]^{4+}$ precursor ions of the unlabeled and labeled forms of the signature peptide were 588.15 and 590.25 m/z, respectively. Data processing was performed with Mass Lynx 4.1 software. Each MS data point represents triplicate analyses by LC/MS/MS (MRM) and measured as the peak area ratio of the selected product ions of both unlabeled and the labeled peptide.

4.2.7. Method validation

Method validation was studied according to the International Conference on Harmonization (ICH) guidelines (23) for validation of bio-analytical procedures. For linearity measurements, independent calibration curves were constructed over 4 days using six different concentrations of the standard solution (10, 50,100, 500, 750, and 1000 nM) along with blank and zero blank samples. Regression analysis was used to evaluate the linearity of the method and calculated by least squares regression.

For LOD and LOQ determinations the ICH guideline recommends calculating the lower limit of detection and quantification based on the standard deviation (SD) of the response and slope of the regression equation. LOD and LOQ were calculated using 3.3 σ/S and 10 σ/S respectively, where σ is the standard deviation of the intercept (y) and S the slope of the curve.

Following the ICH guidelines, the precision of the method was calculated using three levels of quality control samples (20, 250, and 850 nM); in terms of repeatability (intra-day), intermediate precision (inter-day), and reproducibility (mean of an inter-project trails).

Method acceptance criteria states that the precision of the calibration curve and QC samples are considered to be acceptable if CV% ≤ 15% for intra and inter day precision. Also, the accuracy compared with the nominal value needs to be within ±15% and precision at the LOQ to be ± 20%. Finally, the calibration curves must meet the above criteria and have a correlation coefficient r^2 of at least 0.99.

4.3. Results and discussions

4.3.1. LC-MS/MS

The study's core objective is to develop a reliable and efficient LC-MS/MS method suitable for measuring specific protein allergen concentrations in crab plant air samples. In a previous study, a signature tryptic peptide that represents the SCTM

allergen was assigned (SQLVENELDHAQEQLSAATHK) (4). It was chemically synthesized as an analytical MS standard and internal standard in its unlabeled and labeled forms, respectively. Using these standard peptide solutions the triple quad MS in MRM mode was optimized and the precursor and product ions were selected. The quadruply-charged precursor ions, $[M+4H]^{4+}$, for the standard and internal standard were m/z 588.15 and m/z 590.25, respectively (Fig 4.1). These precursor ions were introduced to low energy collision induced dissociation (CID).

Consequently, various product ions were observed for both forms of the signature peptide. The two most abundant product ions were m/z 712 and m/z 715 $[M-(SQ)+3H]^{3+}$ where S and Q are serine and glutamine residues) for unlabeled and labeled signature peptides, respectively. The multiple reaction monitoring (MRM) transitions 588.15→ 712 for the signature peptide standard and 590.25 → 715 for its deuterated analogue (internal standard) were used for further quantification measurements. Under optimized LC and MS conditions the signature peptide standard and internal standard eluted within ~3 min as seen in Figure 4.2. A linear calibration curve (Figure 4.3) was obtained (10-1000 nM).

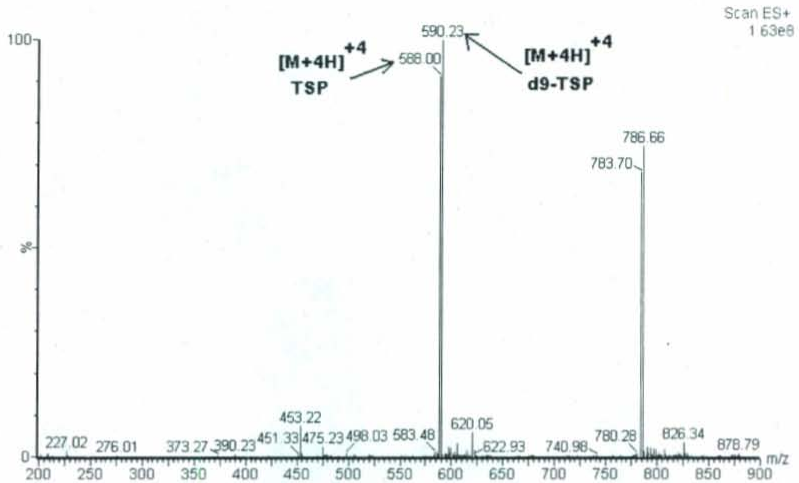


Figure 4. 1: ESI spectra for the unlabeled (TSP) and deuterated (d_9-TSP) signature peptides

4.3.2. Method validation

A comprehensive study was carried out to valid the LC-MS/MS method based on ICH guidelines (23) for linearity, accuracy, precision and sensitivity. The linearity of four different calibration curves constructed on different days was evaluated and its curve fit calculated by least square linear regression. The concentration of the calibration standards were back calculated using this regression equation. For individual calibration points the precision and the accuracy met the ICH guideline acceptance criteria. The correlation coefficient, y-intercept, and slope of the regression line are summarized in

Table 4.1 and a representative calibration curve shown in Figure 4.3. The LOQ, which is the lowest calibration point, was (10 nM) with a consistent precision and accuracy of ≤ 20% of the nominal value determined at a 10:1 signal-to-noise ratio. The LOD, defined > 3:1, signal: noise was 3 nM.

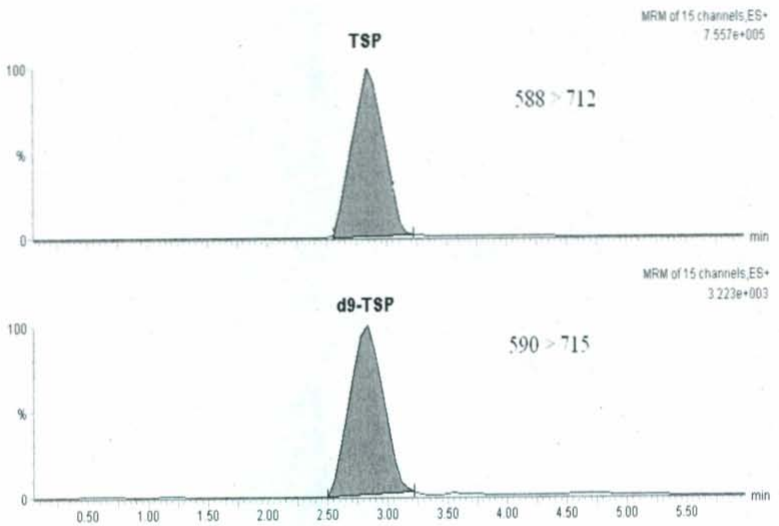


Figure 4. 2: Representative MRM chromatogram of tropomyosin signature peptide in unlabeled form (TSP) (top) and the labeled form (d_9-TSP) (bottom).

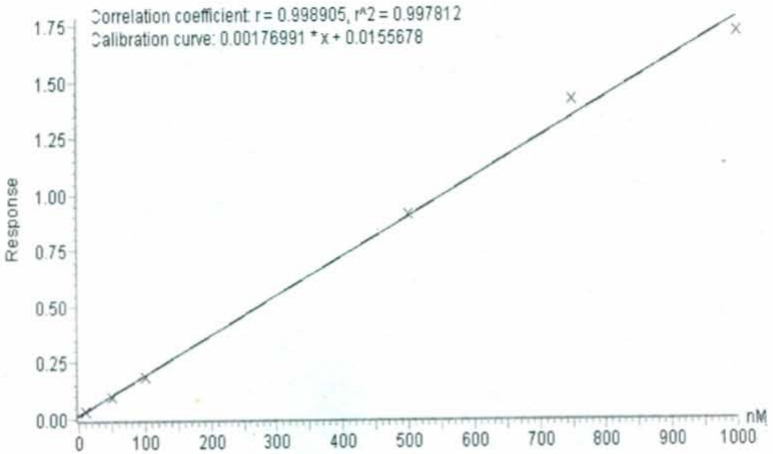


Figure 4. 3: Calibration curve for TSP concentration range of 10-1000 nM, where the response is the peak area ratio (TSP/ d_9-TSP).

Table 4. 1: Summary of linearity validation of calibration curves (Fig. 3). Average of four different curves over four days (n=4).

Range (nM)	10 –1000
Regression equation (Y)	
Slope (b)	0.12575
SD on slope (S_b)	8.1×10^{-3}
Intercept (a)	0.69933
Correlation coefficient (r)	0.9992

Accuracy and precision was evaluated and established over three days. At the intraday interval, twelve standard samples from each level of QC (low, mid, and high) were analyzed after running a valid calibration curve. The samples concentrations were back calculated and the deviation of the mean from the nominal values served as the measure of method accuracy and precision as summarized in Table 4.2. In the inter-day validation study, a total of eighteen samples were used to evaluate the method intermediate precision and reproducibility by including six of the intraday samples with the other two days (six samples) of each QC level (Table 4.2). The specificity test of the method passes due to using the MRM mass spectrometry method. Tandem MS runs were performed hourly, i.e. product ion scan and precursor ion scan, to maintain the specificity of the product-precursor relationship.

4.3.3. Optimization of tryptic digestion

The absolute quantification approach for measuring protein concentration using isotope dilution mass spectrometry (MS) involves pre-determining a signature tryptic peptide that is stoichiometrically equal to the target protein (20). Consequently, optimal and reproducible conditions of tryptic digestion must be met before any isotope MS method can be adopted. SCTM digestion parameters that should be optimized are digestion time, temperature, enzyme-to-substrate ratio, and quantity of digestion enhancer (RapiGest SF).

Table 4. 2: Summary of inter-day and intra-day validation for method sensitivity, precision and accuracy.

	Statistical Parameters	QC-L (20nM)	QC-M (250nM)	QC-H (850nM)
Intra-day	Mean*	21.01	258.99	877.88
	SD	1.89	6.22	23.46
	Precision CV (%)	9.01	2.40	2.67
	Accuracy %	105.04	103.60	109.73
Inter-day	Mean**	21.51	262.19	838.44
	SD	1.40	14.17	68.84
	Precision CV (%)	6.49	5.41	8.21
	Accuracy %	107.53	104.87	98.64

* n=12, **n= 18 (3 days x 6 samples)

All the optimization experiments were preformed in triplicate using a standard solution of SCTM. The initial parameters in the optimization experiments were those recommended by the enzyme supplier. Optimization of the enzyme-to-substrate ratio involved using 1:1 to 1:100 (w: w) trypsin to SCTM ratios and measuring the signature peptide (TSP) by LC-MS/MS. The most suitable digestion enzyme to substrate ratio was 1:1 (Fig. 4.4a).

Completeness of protein digestion is sensitive to incubation time. Therefore a set of SCTM digestion experiments were preformed with incubation times from 1 to 18 hr.

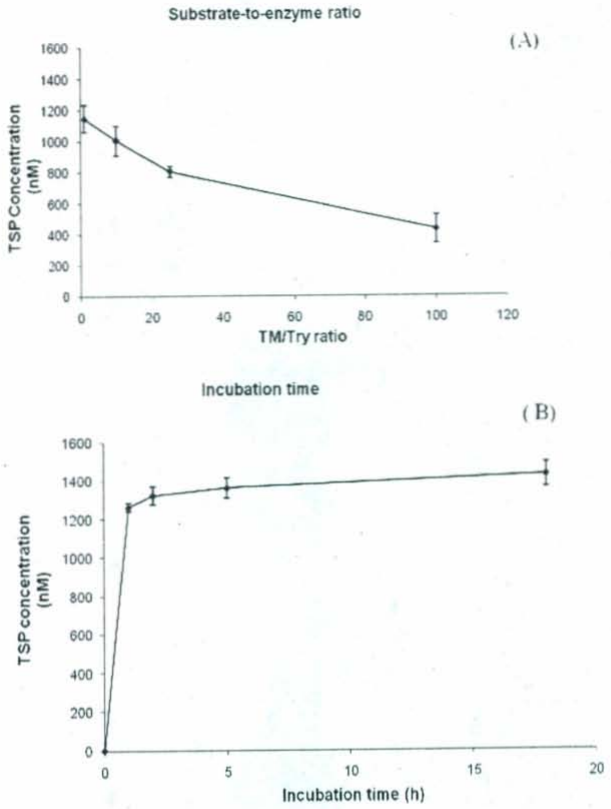
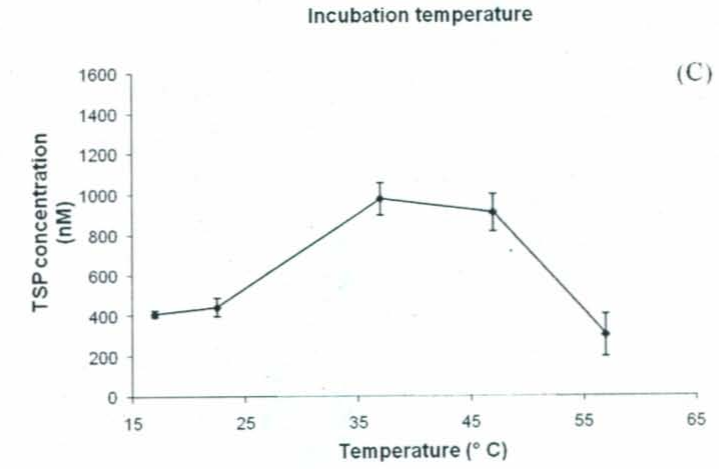
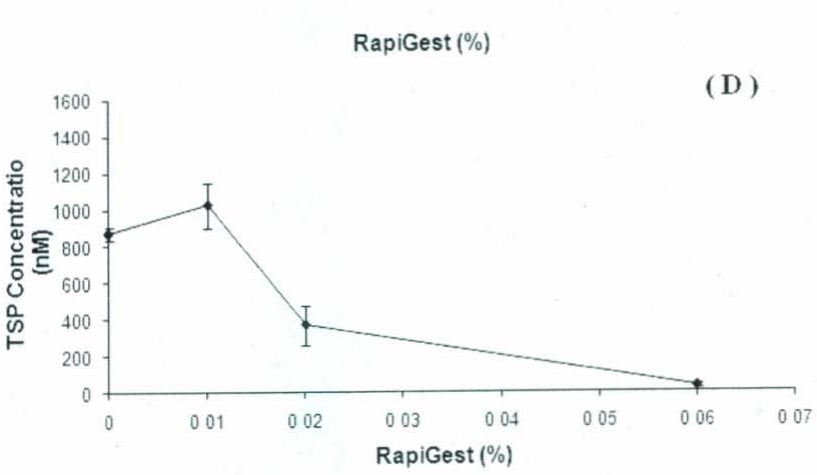
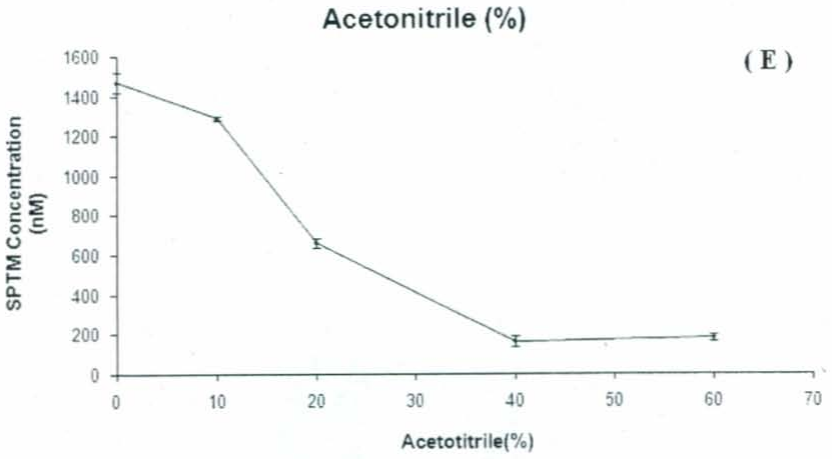


Figure 4. 4: Operational effects on trypsin digestion of tropomyosin (A) Optimum of substrate-to-enzyme ratio, (B) Incubation time, (C) Incubation temperature, (D) RapiGest %, and (E) Acetonitrile %.


Incubation temperature
(C)
TSP concentration (nM)
1600
1400
1200
1000
800
600
400
200
0
15
25
35
45
55
65
Temperature (° C)


RapiGest (%)
(D)
TSP Concentratio (nM)
1600
1400
1200
1000
800
600
400
200
0
0
0 01
0 02
0 03
0 04
0 05
0 06
0 07
RapiGest (%)



(Cont'd.) **Figure 4. 5:** Operational effects on trypsin digestion of tropomyosin (A) Optimum of substrate-to-enzyme ratio, (B) Incubation time, (C) Incubation temperature, (D) RapiGest %, and (E) Acetonitrile %.

Completeness of protein digestion is sensitive to incubation time. Therefore a set of SCTM digestion experiments were preformed with incubation times from 1 to 18 hr. The production of the signature peptide was monitored (Fig. 4.4b). Although close to maximum levels of TSP can be achieved with incubation times within 5 hr it was decided that overnight (18 hr) incubation would be more reproducible and complete. RapiGest SF surfactant is used to increase the accessibility of trypsin enzyme into the TM protein structure. A series of digestion samples were prepared with increasing amounts of RapiGest and the TSP monitored. Addition of 0.01 % RapiGest was adopted for TM

digestion (Fig. 4. 4). The linear α-helix proteins like TM seem not to need a large amount of RapiGest which is usually the case in the globular proteins (17). Digestion temperature was also studied as a means of optimizing TSP production. As observed in Figure 4.4c, a temperature of 37 °C was selected. Finally, it has been reported that the addition of acetonitrile can increase the activity of the trypsin digestion (12). Preliminary experiments showed this not to be the case for SCTM.

4.3.4. Standard addition calibration of SCTM digestion

A purified SCTM solution was standardized using the developed isotopic dilution mass spectrometric protocol and the standard addition method. Three separate samples were incubated under optimized digestion conditions. The standard addition method was performed to calculate the exact concentration of SCTM (as TSP) by measuring the level of TSP produced after complete digestion. The digested SCTM samples were further divided into five equal portions. To each solution were spiked increasing amounts of signature peptide as shown in Figure 4.5. A sufficient volume of the internal standard working solution was added to each sample to a final concentration of 1 μM. A linear regression of the standard addition curve was obtained (Fig. 4.5) and the concentration of the SCTM (as TSP) in the original sample was calculated to be 0.31 μM.

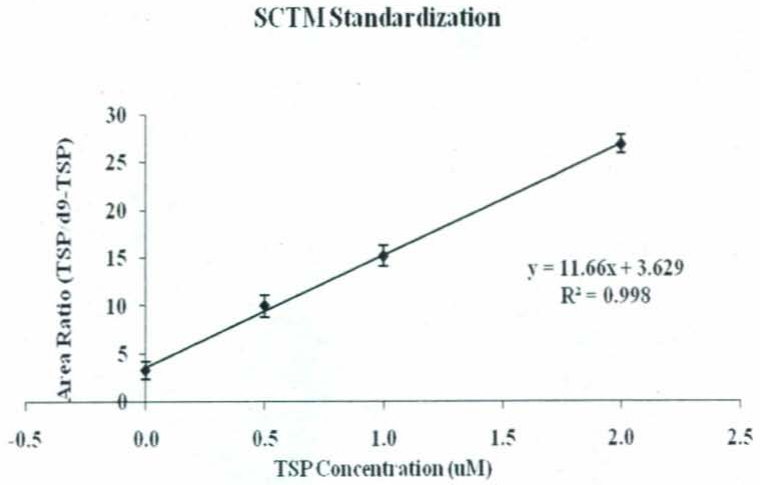


Figure 4. 6: Standard addition curve for standardization a solution of SCTM after digestion (n=3).

4.3.5. Linearity of SCTM digestion

The efficacy of the trypsin digestion of SCTM was studied to confirm the completeness of digestion over a range of SCTM concentrations and also provide additional evidence that the purified SCTM could be used as analytical standard. Four samples were prepared in triplicate with amounts of SCTM ranging from 0.31 to 6.2 pmol. Samples were digested under optimized conditions. Result shows excellent

linearity as illustrated in Figure 4.6. The curve indicates completeness of digestion at high SCTM concentrations and the suitability of SCTM solutions as reliable standards.

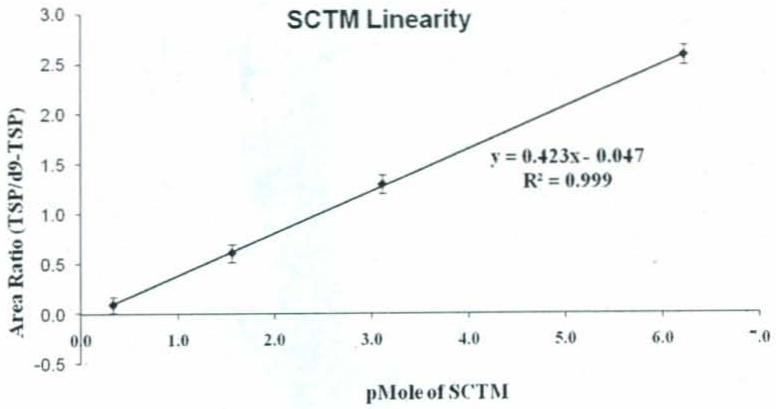


Figure 4. 7: Linearity of SCTM digestion over a wide concentration. Digestion conditions: TM/ Trypsin (1:1) along with other optimization conditions. (n=3)

4.3.6. SCTM extraction recovery from air filters

Good extraction recovery efficiency of spiked SCTM from PTFE air filters is required for developing a reliable air sampling method for airborne SCTM. A diluted solution of SCTM (31.1nM) was sprayed onto blank 33-mm PTFE filters. A total of 300

μL solution was sprayed carefully, to maintain very fine droplets of solution using a fusion pump and LC-MS electrospray apparatus (using 60 psi N_2) (Fig 4.7). Three separate PTFE filter samples were spray spiked with 9.33 pmol of SCTM. Three control samples with the same amount of the SCTM were prepared as described above (i.e., 10% SDS solution with 500 nM of d_9-TSP) (15, 24). These samples were treated using a standard protocol as described before. The SDS material was removed using TopTip filters protocol. Since all these SDS-containing samples required sample cleanup, a series of runs (n=3) was performed to determine SCTM recovery after cleanup using TopTip filters. Three other samples of TSP/d_9-TSP solution with concentrations equivalent to the medium point of the calibration curve (500 nM) were used to study the TopTip filters recoveries. The TopTip filter samples were compared with TSP/d_9-TSP standard solutions and the extraction recovery was found to be 95.6%. Using the TopTip filter recovery, the final PTFE filters recovery was calculated to be 99.5% as SCTM.

4.3.7. Airborne protein collected from a simulated crab production

To verify the applicability of the method in real crab processing plants. Plant Air samples were collected in a simulated crab processing plant. The plant had very little ventilation and small working areas; 15.9 m^3 for the butchering station and 17.8 m^3 for the cooking area. The controlled processing conditions would ensure abundant levels of airborne protein. Air samples were collected in each working area using personal

breathing zone (pbz) and area pumps. Due to the limited quantity of crab the sampling time was relatively short (50 min).

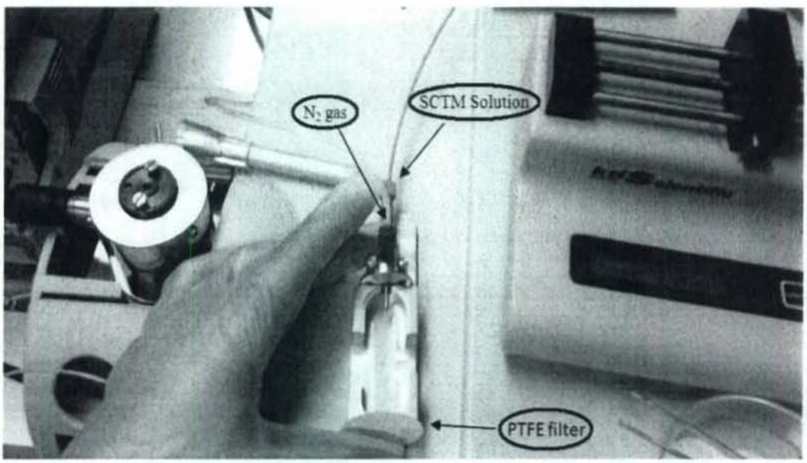


Figure 4. 8: Setup of electrospray apparatus used for spiking TM standard solution onto a PTFE filter.

The results of the air sampling are summarized in Table 4.3. Surprisingly, the levels of the SCTM (as TSP) in both sites are high for PBZ pumps, since the workers are close to the source of the crab allergen. The SCTM levels in the cooking area filters are higher than those in the butchering area. This is possibly due to high steam convection

which would produce and distribute the aeroallergens more homogenously and more effectively than at the butchering site. The method also showed subtle differences in allergen levels between the air samplers' "slightly" different locations. For example, butchering 1 sampler located above the butchering table had a SCTM level of 1.02 $\mu g/m^3$, while Butchering 2 sampler, positioned just below and beside the table had a level of only 0.36 $\mu g/m^3$. SCTM standards and field blank samples (before and after crab processing) were used as controls in the experimental design (Table 4.5). A SCTM standard was used to monitor the optimum digestion conditions. All field blanks gave values of <LOQ level. The lower limit of detection of sampling airborne TM in this experiment was 33 ng/m^3. This established analytical method has been found to be relatively sensitive to detecting, mapping, and measuring SCTM for air sampling of less than 1 hr. On the other hand, the immunological techniques such as ELISA and RAST are likely to require at least eight hours of sampling time to reach the same sensitivity as this study's chemical analysis approach (25).

Table 4. 3: Results of air sampling of a simulated processing plant. Blanks and controls samples are included.

Simulated Crab plant station	Flow rate (L/min)	Time (min)	Volume (m^3)	(TM) $\mu g/m^3$
Butchering (PBZ)	3.00	15	0.045	3.92
Butchering 1 (Area)	3.50	50	0.175	0.36
Butchering 2 (Area)	3.80	50	0.190	1.02
Cooking (PBZ)	3.00	50	0.150	2.31
Cooking 1 (Area)	2.60	50	0.130	1.87
Cooking 2 (Area)	3.50	50	0.175	1.70
Before butchering (Area)	2.60	50	0.130	<LOQ
After butchering before cooking (Area)	3.50	50	0.175	<LOQ
After cooking (Area)	3.50	50	0.175	<LOQ
Extraction Blank	N/A	N/A	N/A	<LOQ
Digestion control		Pass		

PBZ= personal breathing zone LOQ= lower limit of quantification

4.4. Conclusion

For evaluating the levels of the airborne SCTM in the seafood processing workplaces, an approach was developed and optimized using isotopic labeled mass spectrometry (LC-MS/MS). The validity of the instrumental method was studied in terms

of linearity, selectivity, accuracy, and precision using ICH guidelines. Tryptic digestion conditions were optimized for SCTM protein (α helices). The linearity of digestion of SCTM and subsequent TSP analysis was verified thus showing the methods ability to use SCTM as a standard for studying the SCTM extraction recovery from PTFE filters. Ultimately, the developed methodology was successful in the analysis of real air samples collected in a simulated crab processing plant. The method sensitivity and specificity allows for measurement of significant differences in allergen concentrations between small spatial variations in sampling sites and also requires only short sampling times (< 1 hr) to quantify accurately airborne allergen levels.

4.5. References

1. Jeebhay, M. F., Cartier, A. Seafood workers and respiratory disease: an update. *Curr. Opin Allergy Clin Immunol.* **2010.** 10; 104-113.

2. Lopata, A. L., O'Hehir, R. E., Lehrer, S. B., Shellfish allergy. *Clinical and Experim Allergy.* **2010**. 40; 850-858.

3. Motoyama, K., Suma, Y., Ishizaki, S., Nagashima, Y., Shiomi, K. Molecular cloning of tropomyosins identified as allergens in six species of crustaceans. *J Agric Food Chem.* **2007.** 5; 985–991.

4. Abdel Rahman, A. M., Lopata, A. L., O'Hehir, R. E., Robinson, J. J., Banoub, J. H., Helleur,R. J. Characterization and '*de novo*' sequencing of snow crab tropomyosin enzymatic peptides by both electrospray ionization and matrix-assisted laser desorption ionization QqToF tandem mass spectrometry. *J Mass Spectrom.* 2010. 45; 372-381.

5. Abdel Rahman, A. M., Kamath, S., Lopata, A. L., Helleur, R. J. Analysis of the allergenic proteins in black tiger prawn (Penaeus monodon) and characterization of the major allergen tropomyosin using mass spectrometry. *Rapid Commun Mass Spectrom.* **2010.** 24; 2462–2470.

6. Shiomi, K., Sato, Y., Hamamoto, S., Mita, H., Shimakura, K., Sarcoplasmic calcium-binding protein: identification as a new allergen of the black tiger shrimp Penaeus monodon. *Int Arch Allergy Immunol.* **2008**. 146; 91-98.

7. Ayuso, R., Grishina, G., Bardina, L., Carrillo, T., Blanco, C., Ibáñez, M. D., Sampson, H.A., Beyer, K., Myosin light chain is a novel shrimp allergen, Lit v 3. *J Allergy Clin Immunol.* **2008.** 122; 795-802.

8. García-Orozco, K. D., Aispuro-Hernández, E., Yepiz-Plascencia, G., Calderón-de-la-Barca, A. M., Sotelo-Mundo, R. R. Molecular characterization of arginine kinase, an allergen from the shrimp Litopenaeus vannamei. *Int Arch Allergy Immunol.* **2007.** 144; 23-28.

9. France, R. M., Sellers, D. S., Grossman, S. H., Purification, characterization, and hydrodynamic properties of arginine kinase from gulf shrimp (Penaeus aztecus). *Arch Biochem Biophys*. **1997.** 345; 73-78.

10. Lu, Y., Ohshima, T., Ushio, H., Hamada, Y., Shiomi, K. Immunological characteristics of monoclonal antibodies against shellfish majorallergen tropomyosin. *Food Chemistry.* **2007.** 100; 1093-1099.

11. Reese, G., Schicktanz, S., Lauer, I., Randow, S., Luttkopf, D., Vogel, L., Lehrer, S. B., Vieths, S. Structural, immunological and functional properties of natural recombinant Pen a 1, the major allergen of Brown Shrimp, Penaeus aztecus. *Clin Exp Allergy.* **2006.** 36; 517-524.

12. Rosmilah, M., Shahnaz, M., Masita, A., Noormalin, A., Jamaludin, M. Identification of major allergens of two species of local snappers: Lutjanus argentimaculatus (merah/ red snapper) and Lutjanus johnii (jenahak/ golden snapper). *Trop. Biomed.* **2005.** 22; 171-177.

13. Hoffman, D. R., Day, E. D., Miller, J. S. The major heat stable allergen of shrimp. *Ann Allergy.* **1981.** 47; 17-22.

14. Liang, Y. L., Cao, M. J., Su, W. J., Zhang, L. J., Huang, Y. Y., Liu, G.M. Identification and characterisation of the major allergen of Chinese mitten crab (*Eriocheir sinensis*). *Food Chemistry.* **2008.** 111; 998-1003.

15. Gill, B. V., Rice, T. R., Cartier, A., Gautrin, D., Neis, B., Horth-Susin, L., Jong, M., Swanson, M., Lehrer, S. B. Identification of crab proteins that elicit IgE reactivity in snow crab-processing workers. *J Allergy Clin Immunol.* **2009.** 124; 1055-1061.

16. Gautrin, D., Cartier, A., Howse, D., Horth-Susin, L., Jong, M., Swanson, M., Lehrer, S., Fox, G., Neis, B. Occupational asthma and allergy in snow crab processing in Newfoundland and Labrador. *Occup. Environ. Med.* **2010.** 67; 17-23.

17. Norrgran, J., Williams, T. L., Woolfitt, A. R., Solano, M. I., Pirkle, J. L., Barr, J. R. Optimization of digestion parameters for protein quantification. *Anal Biochem.* **2009.** 393; 48-55.

18. Brun, V., Masselon, C., Garin, J., Dupuis, A. Isotope dilution strategies for absolute quantitative proteomics. *J Proteomics.* **2009.** 72; 740-749.

19. Ong, S. E., Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol.* **2005.**1; 252–62.

20. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem.* **2007.** 389; 1017–31.

21. Iliuk, A., Galan, J., Tao, W. Playing tag with quantitative proteomics. *Anal Bioanal Chem.* **2009.** 393; 503–13.

22. Cartier, A., Malo, J. L., Forest, F., Lafrance, M., Pineau, L., St-Aubin, J. J., Dubois, J. Y. Occupational asthma in snow crab-processing workers. *J Allergy Clin Immunol*. **1984.** 74; 261-269.

23. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Guidance for Industry Q2B Validation of Analytical Procedures: Methodology. **1996** ICH.

24. Gordon, S., Tee, R., Lowson, D., Taylor, A. J. N. Comparison and Optimization of Filter Elution Methods for the Measurement of Airborne Allergen. *Ann occup Hyg.* **1992**. 36; 575-587.

25. Malo, J. L., Chrétien, P., Mc Cants, M., Lehrer, S. Detection of snow-crab antigens by air sampling of a snow-crab production plant. *Clin Exp Allergy.* **1997.** 27; 75-78.

Chapter 5: Analysis of the allergenic proteins in black tiger prawn *(Penaeus monodon)* and characterization of the major allergen tropomyosin using mass spectrometry [4]

[4] This Chapter has been published; Abdel Rahman, A. M., Kamath, S., Lopata, A. L., Helleur, R. J. Analysis of the allergenic proteins in black tiger prawn (*Penaeus monodon*) and characterization of the major allergen tropomyosin using mass spectrometry. *Rapid Commun Mass Spectrom.* **2010**. 24; 2462–2470.

Abstract

Crustaceans are the third most prevalent cause of food induced anaphylaxis after peanuts and tree nuts. The severity of the allergenic proteins depends mainly on the amino acid sequence that induces production of IgE antibodies. In black tiger prawn *(Penaeus monodon)*, the crude protein extract was profiled and its allergenic potency was examined against patient's sera. Proteins having strong immunoreactivity with patient's IgE were characterized using peptide mass fingerprinting (PMF). Tropomyosin (TM) (33kDa), myosin light chain (20 kDa), and arginine kinase (40 kDa) were identified as allergenic proteins. Tropomyosin, the most abundant and potent allergen, was purified using ion-exchange chromatography for '*de novo*' sequencing experiments. Using '*bottom up*' tandem mass spectrometry, the full amino acid sequence was achieved by a combination of MALDI and ESI tandem mass spectrometry (QqToF). Myosin light chain and arginine kinase were also characterized, and their related peptides were '*de novo*' sequenced using the same approach. The immunological reactivity of the crude prawn extracts and purified TM samples were analyzed using a large number of patients' sera. A signature peptide was assigned for the TM protein for future quantification work of black tiger prawn TM levels in different matrices (i.e. water, air, food) in the seafood industry.

5.1. Introduction

Black tiger prawn (BTP) (*Penaeus monodon*) is one of the most important commercial products of the shrimp industry in Southeast Asia and is widely distributed in the Indo-West-Pacific region (1). However, it is also recognized as one of the most common causes of Type I (IgE)-mediated food hypersensitivity (2). Ingesting black tiger prawn may cause hypersensitivity reactions such as urticaria, angioedema, asthma, diarrhea and anaphylaxis allergic patients (3). Immediate hypersensitivity reactions to seafood have become an important issue since consumption of seafood and subsequent processing of the shellfish has increased worldwide. The seafood processing industry has experienced tremendous growth in recent years with over 42 million workers worldwide engaged in various activities of seafood production (i.e. processing, harvesting, etc), and these workers are being continuously exposed to seafood allergens (4-6).

Tropomyosin (2) and arginine kinase (4, 7-10) have been identified as immunologically reactive proteins from black tiger prawn (BTP) in previous studies using molecular genetics approaches. Tropomyosin, which is a well-known shellfish allergen, was first identified by Hoffman *et al.* in 1981(11) and recently characterized as a major allergen in other prawns. (12, 13) Tropomyosin is a water-soluble and heat-stable protein with the molecular weight (MW) ranging between 34 and 39 kDa (14). It has a highly conserved amino acid sequence among different invertebrate organisms and is present in muscle as well as non-muscle cells (15, 16). Furthermore, differential splicing of the pre-messenger ribonucleic acid (pre-mRNA) produce isoforms of TM (17).

Previous studies have evaluated some of these seafood allergens and reported the prevalence for occupational protein contact dermatitis range from 3% to 11% and for occupational asthma from 7% to 36% (18). Immunochemical techniques used in previous studies to identify reactive proteins in air samples include enzyme-linked immunosorbent assay (ELISA) (16,19), radioallergosorbent test (RAST) (20), and immunoblotting (16, 21) But these techniques are generally not sensitive, selective, or specific enough to evaluate the levels of the individual and most harmful protein allergens in sample matrices such as air samples. These techniques only recognize the total airborne proteins, which could include both allergenic and non-allergenic from the seafood and may also include allergenic proteins from non-seafood sources (e.g. dust mites, microbes, etc.). Non-specificity of the antibody-antigen interactions may have lead to uncertainty in previous characterization studies.

Quantifying the major BTP allergens to enable the correlation of severity with specific allergen load requires a highly sensitive, specific, and reproducible analytical technique. Isotopic dilution mass spectrometry has played a crucial role in the protein quantification in the last two decades and is the method of choice for our further study in quantification of protein allergens. The current study targets the meat of BTP as the chief source of potential airborne allergens in the workplace. The allergenicity of purified TM from BTP was analyzed using immunoblotting against specific patients' sera and specific polyclonal antibodies. A very specific and selective method has being developed in our laboratory to both characterize and quantify these allergens using mass spectrometry

whereby specific allergenic proteins have been targeted rather than a group of reactive but uncharacterized proteins. This will assist in identifying the actual allergen in the environment and in quantification using isotopic dilution mass spectrometry technique.

Black tiger prawn tropomyosin is, for the first time, '*de novo*' sequenced using '*bottom up*' tandem mass spectrometry. Multiple types of enzyme digestion, ion sources, and derivatization protocols are used to cover the whole amino acid sequence. Furthermore, the mass spectra of the peptide mass fingerprinting (PMF) and the peptide fragment fingerprinting (PFF) were uploaded to the Mascot search engine which reports the amino acid sequence of the target allergen as tropomyosin. The post translational modifications motifs of TM were evaluated theoretically using bioinformatics tools and by visually screening through the mass spectra.

5.2. Experimental

5.2.1. Chemicals and reagents

All chemicals were used without further purification. Acetonitrile, hydrochloric acid, and methanol were supplied from ACP (Montreal, Canada). RapiGest SF surfactant was purchased from Water Corporation (Milford, MA, USA) and trypsin, Asp-N and endoproteinase (Glu-C V8) sequencing grade enzymes from Sigma-Aldrich (St. Louis, MO, USA). Ammonium bicarbonate, o-methylisourea hemisulfate, ammonium

hydroxide, horseradish peroxidase (HRP), and α-cyano-4-hydroxycinamic acid (HCCA) matrix were purchased from Sigma-Aldrich (St. Louis, MO, USA). The Bradford assay kit was purchased from BioRad (Hercules, CA, USA). The dialysis bags were from Fischer Scientific (Roncho Dominguez, CA, USA). For desalting purposes, ZipTip C_{18} filters were purchased from Millipore Corporation (Bedford, MA, USA). For TM purification steps, phosphate buffer saline (PBS) and Tris buffer saline (TBS) tablets were purchased from Amresco, USA. Acetic acid, sodium acetate, sodium chloride, and Tween-20, used for the washing steps, were purchased from Sigma-Aldrich (St. Louis, MO, USA). The Amicon spin filters used for fraction concentration were from Millipore Corporation (Bedford, MA, USA). The tracker dye (Coomassie stain R250) and PVDF membrane for the immunoblotting were from BioRad (Hercules, CA, USA). Skimmed milk for the immunoblotting procedure was purchased from a local supermarket.

5.2.2. Protein extracts

Fresh black tiger prawns were purchased from the local market (Victoria, Australia) and transported to Dr. Lopata's laboratory on ice. For the preparation of raw protein extract the prawn muscles were shredded into pieces and homogenized in phosphate buffered saline (PBS), pH 7.2, using an Ultraturrax homogenizer (IKA). The slurry was then centrifuged at 8000 rpm for 20 minutes and the supernatant was filter sterilized. This PBS protein extract was stored in aliquots at -80° C until further

experiments. A cooked prawn extract was prepared by heating whole prawns at 100° C in PBS, pH 7.2 for 20 minutes and then a similar procedure was followed as described for the raw prawn protein extract.

5.2.3. Tropomyosin purification

The tropomyosin from BTP was purified from the crude (raw and cooked) extract using a strong ion exchange chromatographic column on a Biologic LP purification system (BioRad, USA).

Before loading the proteins onto the column, the crude extract was exchanged into the chromatographic starting buffer (30 mM acetate buffer, pH 5.5) using Amikon spin filters of 3 kDa (molecular weight cut off (MWCO)). After equilibrating the column with the starting buffer, approximately 10 mg of crude proteins were loaded onto the column. The column was then washed with 5 column volumes of the starting buffer. It was then further washed with 250 mM NaCl in 30 mM acetate buffer, pH 5.5 to elute unwanted proteins. The tropomyosin (TM) eluted at approximately 47 min. The collected fraction was then concentrated using an Amikon spin filter with a 3 kDa MWCO.

5.2.4. Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE)

The BTP crude extracts and the purified TM were profiled using 12 % SDS-PAGE. A protein solution (10 µg) was added to each of the wells, and electrophoresis was run at a voltage of 170 V for 1 hr, or until the tracker dye was seen at the base of the gel. One gel was stained with Coomassie brilliant blue using a standard protocol. For the second gel, the separated proteins were transferred to a nitrocellulose membrane at 100 V for 1 hr. After the transfer was completed, the membrane was placed in a blocking solution (5% Skim milk in Tris-buffered saline (TBS)) for immunoblotting.

5.2.5. Immunoblotting

The in-house rabbit polyclonal anti-crustacean antibody was generated by injecting a mixture of heated protein extracts from black tiger prawn (*Penaeus monodon*), king prawn (*Melicertus latisulcatus*), mud crab (*Scylla serrata*), and slipper lobster (*Thenus orientalis*) into rabbits. The final bleed was performed on week 9 and the sensitivity of the serum was analyzed using ELISA.

After separating the proteins using SDS-PAGE, samples were immobilized on a PVDF membrane using a semi-dry immunoblot apparatus (Biorad, USA). The membranes were then blocked with 5% Skim milk solution for 1 hr at room temperature following an hour incubation with the rabbit antibody (primary antibody) using a working dilution of 1:40000. The blot was washed three times with TBS-Tween. The secondary antibody used was Polyclonal Goat anti-rabbit antibody conjugated to horse radish

peroxidase (HRP) (Promega, USA), with a working dilution of 1:20000. The blot was incubated for 30 min in the secondary antibody solution, washed again three times with TBS-Tween and incubated with chemiluminescent substrate (Sigma Aldrich, USA). The bands were then visualized using the enhanced chemiluminescent (ECL) technique (22).

To demonstrate the allergenicity of the isolated prawn proteins, different extracts were analyzed for IgE antibody binding from allergic patients. The human sera were collected from patients with strong allergic reactivity to shellfish. Ethical approval for this study was acquired by Monash University, Melbourne, Australia. For immunoblotting, protein extracts were separated by electrophoresis (see SDS-gels above), proteins transferred and incubated with human serum (diluted 1:20 in 1% skim milk) overnight at 4˚C. Subsequently blots were washed three times with PBS-T and the membrane incubated for 1 hr with rabbit antihuman IgE antibody (DAKO, USA) (diluted 1:1000) in PBS-T containing 1% skimmed milk. After washing the membrane with PBS-T three times, it was incubated for 30 min with HRP tagged Goat Anti-Rabbit polyclonal antibody (DAKO, USA) (diluted 1:1000) in PBS-T containing 1% skim milk. Finally, the membrane was washed with PBS three times, incubated with the chemiluminescent substrate, and the immunoblot membranes analyzed for IgE reactivity using the ECL technique (22).

5.2.6. Enzymatic digestion and guanidation

The examined protein bands were excised from the SDS-PAGE plate. To increase the sensitivity of the lysine-containing peptides in the MALDI experiments, the *in-gel* guanidation procedure was performed on all protein samples using the protocol developed by Sergeant *et al.* (23). The gel pieces were destained by washing three times with 200 mM of $NH_4H\,CO_3$ in a solution of 50% acetonitrile in dH_2O, at 30°C for 30 min. The destained pieces were dried under a stream of N_2, and then covered by a solution of 50 mM NH_4HCO_3, pH 7.8 containing 5 ng/μL trypsin, Asp-N, and endoproteinase Glu-C V8 in ice for 30 min for rehydration. After rehydration, any excess solution was removed. The gel was then covered by a solution of 50 mM of NH_4HCO_3 and incubated at 37°C overnight to enhance protein digestion. The water soluble peptides were extracted twice with the incubation solution and other remaining peptides extracted twice with 0.15%TFA in 50% ACN after a 2 min vortex mixing. The samples were freeze dried, and reconstituted prior to analysis with 10 μL of 0.1% TFA and desalted with C_{18} ZipTips (10 μL bed).

An *in-solution* digestion for the pure extract of the BTP TM was performed using RapiGest SF surfactant, (24) which enhances the digestion efficiency. Subsequently, the solution was incubated with proper buffers as *in-gel* digestion protocol overnight with a concentration 20ng/ml of enzyme. The digestion was quenched and the surfactant was precipitated by 1% formic acid at room temperature.

5.2.7. MALDI plate preparation

The protein samples were prepared for MALDI analysis using a protocol described in Chapter 2, section 2.2.6.1. Two-layer sample/matrix preparation was employed for plate spotting. The first layer solution consisted of 20 mg/ml HCCA in (1/9) methanol/acetone. The second layer of solution consisted of saturated HCCA in 40% ACN. A 0.5 μL sample of the first layer of matrix solution was applied to a MALDI target. A 1 μL sample of the second layer matrix solution was mixed with 1 μL of sample. Finally, a 1 μL of the sample/matrix mixture was deposited onto the first layer and allowed to dry, followed by an on-target wash step (25).

5.2.8. MALDI- and ESI- QqToF MS

MALDI-MS and low-energy (CID) analyses were carried out on a QSTAR XL hybrid quadrupole-quadrupole/time-of-flight tandem mass spectrometer (QqToF-MS/MS) (Applied Biosystems/MDS Sciex, Foster City, USA) equipped with an o-MALDI ion source (Applied Biosystems, Foster City, CA).

Peptide separation was conducted using a DIONEX UltiMate3000 Nano LC System (Germering, Germany). A 250 fmol sample of protein digest dissolved in 0.1% TFA was loaded onto a precolumn (300 μm ID X 5 mm, C_{18} PepMap 100, 5 μm (LC Packing, Sunnyvale, CA)) for desalting and concentrating. Peptides were then eluted from the pre-column and separated on a nanoflow analytical column (75 μm ID X 150

mm, C_{18} PepMap 100, 3 μm, 100 °A, (LC Packing, Sunnyvale, CA) at 180 nL/min using the following gradient regime. The aqueous mobile phases consisted of (A) 0.1% formic acid/ 0.01% trifluoroacetic acid/ 2% ACN and (B) 0.08% formic acid/ 0.008% trifluoroacetic acid / 98% ACN. A gradient of 0% B for 10 min, 0-60% B in 55 min, 60-90% in 3 min, 90% B for 5 min was applied. Including a regeneration step one run was 106 min long.

The ESI-MS of the LC-eluting peptides were measured with the same hybrid QqToF-MS/MS system equipped with a nanoelectrospray source (Protana XYZ manipulator). The nanoelectrospray were generated from a PicoTip needle (10 μm i.d., New Objectives, Wobum, USA) at a voltage of 2400 V. Individual target protein was further analyzed by CID-MS/MS, the resulting peptide were '*de novo*' sequenced and the results confirmed by using the National Center for Biotechnology Information non-redundant database (NCBInr) with the Matrix Science (Mascot) search engine (precursor and product ion mass tolerance set at 0.2 Da). Methionine oxidation was allowed as a variable modification and guanidinyl (K) as a fixed modification since the guanidation derivatization has been performed in MALDI experiments. Peptides were considered identified if the Mascot score was over 95% confidence limit.

5.3. Results and discussion

The crude protein extracts were isolated from fresh BTP tissues then profiled by SDS-PAGE (Fig. 5.1 A) then transferred to nitrocellulose membranes for Western blotting. Subsequently the blots were incubated with rabbit polyclonal anti-crustacean antibody (Fig. 5.1 B) or sensitized patients' sera. This permitted the examination of the immunoreactivity of the extracted proteins, before and after heat treatments (band I vs band II). The strong IgE reactive band with a molecular weight ~33 kDa was further characterized using peptide mass finger printing (PMF).

The 33 kDa band was excised and exposed to trypsin digestion. The generated peptides were introduced to MALDI- and LC-ESI mass spectrometry. The precursor ions spectra (Fig. 5.2) were uploaded to the Mascot search engine NCBInr databases, where the proteins' sequences are *in-silico* generated. The Mascot search engine scored 112 identifying this band as a *Penaeus monodon* tropomyosin. This represents a top probability based on Mowse scores matching with the BTP TM protein (complementary deoxyribonucleic acid (cDNA) based Library (NCBInr)). The above scores were matched with the Mascot algorithms' criterion; individual ions scores >82 indicate identity or extensive homology ($p < 0.05$).

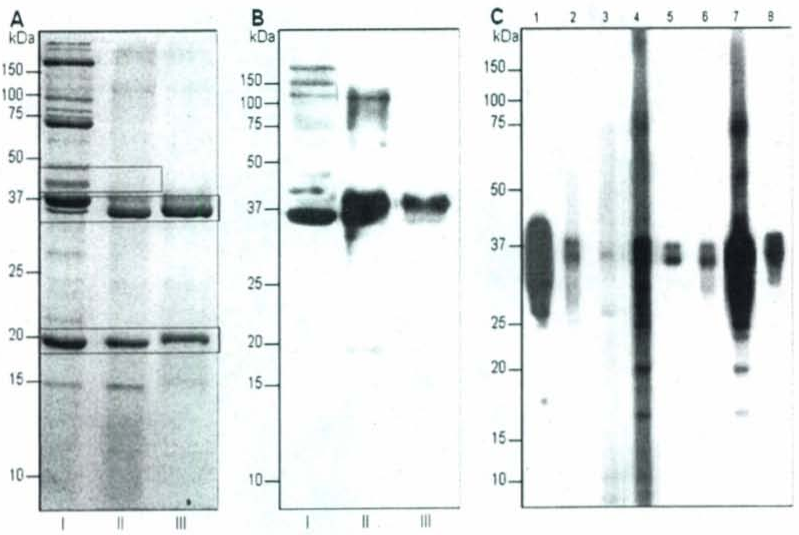


Figure 5. 1: Immunochemical profile of the various protein extracts of the black tiger prawn, I- Raw muscle extract, II- Heat treated muscle extract, III- Purified fraction of TM. A- SDS-PAGE Coomassie profile of the prawn extracts, B- Immunoblot using rabbit polyclonal anti-crustacean antibody, C- IgE antibody Immunoblot against BTP TM using eight shellfish allergic patient sera. The red box in (A) indicates the position of (top to bottom) Arginine kinase, Tropomyosin and Myosin light chain respectively.

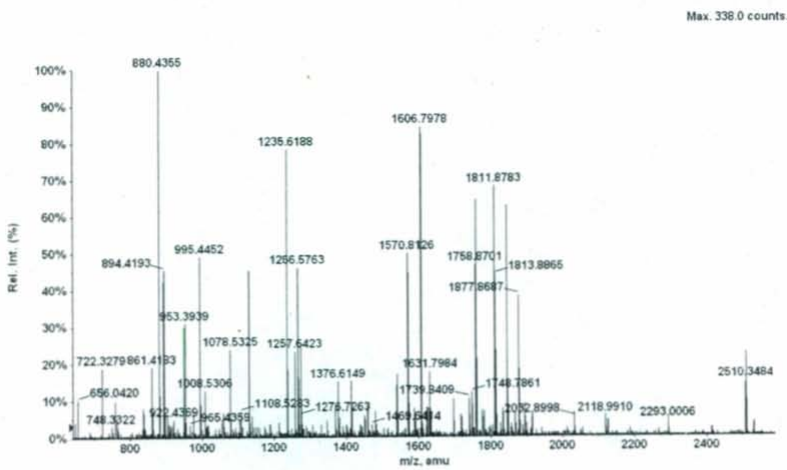


Figure 5. 2: The MS spectrum generated after tryptic digestion of BTP TM. This spectrum scored 112 in Mascot's algorithm as TM for black tiger prawn.

As discussed above, the tropomyosin protein was isolated from fresh BTP and purified using an effective ion-exchange chromatography protocol (experimental section see Fig. 5.3). The allergenicity of the pure extract was further studied using eight allergic patients' sera as shown in Figure 5.1 C. The IgE levels in the patient's serum vary with respect to the severity of the hypersensitivity. Nevertheless, all of them have a strong reactivity with the purified BTP TM, which confirmed TM to be the major allergen in this prawn species.

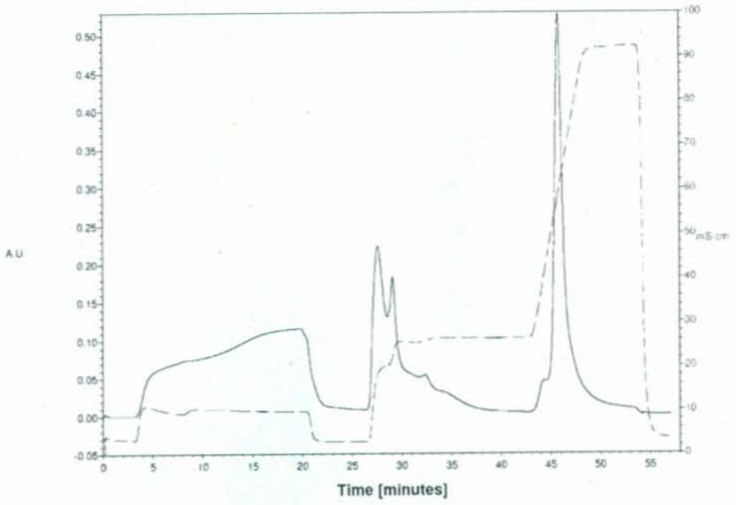


Figure 5. 3: Ion-exchange chromatogram represents the purification of tropomyosin from BTP. The increase in the absorption (in A.U) at 280nm (solid line) at 47 minutes represents the elution of the tropomyosin fraction with increasing salt concentration (dash line; conductivity in mS/cm).

Further characterization is necessary to understand the structural biochemistry of this potent allergen. The purified TM extract was '*de novo*' sequenced using '*bottom up*' approach peptide fragment fingerprinting. Various sets of enzymatic peptides were produced by trypsin, Glu-C V8, or Asp-N digestions and individually introduced to both MALDI and ESI ion sources of a tandem MS. In the precursor ion spectra, the most abundant peptides ions (i. e., m/z 721.2927 $[M+H]^{1+}$, m/z 879.3927 $[M+H]^{1+}$m/z

949.3927 $[M+H]^{1+}$, m/z 1060.4927 $[M+H]^{1+}$, m/z 1107.4927 $[M+H]^{1+}$, m/z 1128.4927 $[M+H]^{1+}$, m/z 1136.4927 $[M+H]^{1+}$, m/z 1145.5927 $[M+H]^{1+}$, m/z 1156.5927 $[M+H]^{1+}$, m/z 1211.4927 $[M+H]^{1+}$, etc) were selected sequentially in the gas phase and exposed to low-energetic CID. The peptide fragment ions were further separated by the TOF analyzer with respect to their mass-to-charge ratio.

In LC-ESI-QqToF analyses, the peptides were initially separated with respect to their polarity on the nano-HPLC column (C_{18} PepMap), using LC separation and a suitable solvent minimizes ion suppression and increases the abundance of ionized peptides (26). Therefore, the multiple-charged precursor ions (i.e. *m/z* 457.7712 $[M+2H]^{2+}$, *m/z* 539.7791 $[M+2H]^{2+}$, *m/z* 565.3142 $[M+2H]^{2+}$, *m/z* 606.7909 $[M+2H]^{2+}$, *m/z* 629.3417 $[M+2H]^{2+}$, *m/z* 688.8189 $[M+2H]^{2+}$, *m/z* 695.3667 $[M+2H]^{2+}$, *m/z* 514.2560 $[M+3H]^{3+}$, *m/z* 1016.0235 $[M+2H]^{2+}$, etc) were selected and isolated for low-energy CID experiments. The fragments of these peptides were separated by the ToF mass analyzer.

The product spectra generated by both ionization mass spectrometry techniques were uploaded to the Mascot MS/MS search engine against the NCBInr database. The

Table 5. 1: A list of generated peptides of TM extracted from BTP that have been generated by this study using different enzymes and MS ion sources.

Protease	Residues position	Peptide sequence (Z)	Molecular weight		Ion Source
			Calculated	Expected	
Trypsin	8 - 15	MQAMKLEK (2)	977.5038	977.4839	ESI
	16 - 30	DNAMDRADTLEQQNK (3)	1747.7737	1747.7795	
	22 - 35	ADTLEQQNKEANNR (2)	1747.7795	1747.7958	
	36 - 48	AEKSEEEVHNLQK (3)	1539.7529	1539.7462	
	39 - 48	SEEEVHNLQK (2)	1211.5782	1211.5673	
	39 - 49	SEEEVHNLQKR (2)	1367.6793	1367.6832	
	50 - 66	MQQLENDLDQVQESLLK (2)	2029.9990	2030.0026	
	67- 76	ANIQLVEKDK (2)	1156.6452	1156.6454	
	77 - 90	ALSNAEGEVAALNR (2)	1413.7212	1413.7342	
	77 - 91	ALSNAEGEVAALNRR (2)	1569.8223	1569.8431	
	91 - 101	RIQLLEEDLER (2)	1412.7623	1412.7934	
	102 - 112	SEERLNTATTK (2)	1248.6310	1248.6226	
	106 - 125	LNTATTKLAEASQAADESER (3)	2105.0236	2105.0447	
	113 - 125	LAEASQAADESER (2)	1375.6215	1375.6232	
	134 - 149	SLSDEERMDALENQLK	1876.8836	1876.8873	
	141 - 149	MDALENQLK (2)	1060. 5223	1060.5236	
	153 - 160	FLAEEADR (2)	949.4505	949.4376	
	168 - 178	KLAMVEADLER (2)	1273.6700	1273.6709	
	169 - 182	LAMVEADLERAEER (2)	1630.7984	1630.8204	
	190 - 198	IVELEEELR (2)	1128.6026	1128.6139	
	206 - 217	SLEVSEEKANQR	1388.6895	1388.7188	
	218 - 226	EEAYKEQIK (2)	1136.5713	1136.5687	
	252 - 264	EVDRLEDELVNEK (2)	1586.7787	1586.7966	
	267 - 284	YKSITDELDQTFSELSGY (2)	2094.9633	2094.9916	
	269 - 284	SITDELDQTFSELSGY	1803.8051	1803.8475	
Glu-C (V8)	3 - 14	AIKKKMQAMKLE (2)	1417.8148	1417.8501	
	43 - 54	VHNLQKRMQQLE (2)	1522.8038	1522.8268	
	63 - 73	SLLKANIQLVE (2)	1226.7234	1226.7528	
	85 - 96	VAALNRRIQLLE (2)	1394.8357	1394.8680	
	85 - 97	VAALNRRIQLLEE (2)	1523.8783	1523.8982	

(Cont'd.) **Table 5. 2:** A list of generated peptides of TM extracted from BTP that have been generated by this study using different enzymes and MS ion sources.

Protease	Residues position	Peptide sequence (Z)	Molecular weight		Ion Source
			Calculated	Expected	
Glu-C (V8)	105 - 115	RLNTATTKLAE (2)	1216.6775	1216.6927	ESI
	125 - 131	RMRKVLE (2)	930.5433	930.5576	
	151 - 157	ARFLAEE (2)	834.4235	834.4310	
	165 - 173	VARKLAMVE (2)	1015.5848	1015.5959	
	197 - 208	LRVVGNNLKSLE (2)	1340.7776	1340.8081	
	224 - 236	QIKTLTNKLKAAE (2)	1456.8613	1456.8909	
	244 - 252	RSVQKLQKE (2)	1114.6458	1114.6608	
Trypsin	16 - 30	DNAMDRADTLEQQNK	1747.7795	1747.7927	MALDI
	22 - 30	ADTLEQQNK	1747.7795	1747.7788	
	22 - 35	ADTLEQQNKEANNR	1629.7706	1629.7785	
	36 - 48	AEKSEEEVHNLQK	1539.7529	1539.6927	
	39 - 48	SEEEVHNLQK	1211.5782	1211.4927	
	39 - 49	SEEEVHNLQKR	1367.6793	1367.5927	
	50 - 66	MQQLENDLDQVQESLLK	2029.9990	2029.9927	
	67 - 74	ANIQLVEK	913.5232	913.4396	
	67 - 76	ANIQLVEKDK	1156.6452	1156.5927	
	77 - 90	ALSNAEGEVAALNR	14137212	1413.6927	
	91 - 101	RIQLLEEDLER	1412.7623	1412.6927	
	92 - 105	IQLLEEDLERSEER	1757.8795	1757.7927	
	102 - 112	SEERLNTATTK	1248.6310	1248.5582	
	113 - 125	LAEASQAADESER	1375.6215	1375.5927	
	134 - 149	SLSDEERMDALENQLK	1876.8836	1876.7927	
	141 - 149	MDALENQLK	1060.5223	1060.4927	
	153 - 160	FLAEEADR	949.4505	949.3927	
	153 - 161	FLAEEADRK	1077.5454	1077.5252	
	162 - 168	YDEVARK	879.4450	879.4282	
	168 - 178	KLAMVEADLER	1273.6700	1273.5927	
	169 - 178	LAMVEADLER	1145.5750	1145.5927	
	169 - 182	LAMVEADLERAEER	1630.7984	1630.7927	
	190 - 198	IVELEEELR	1128.6026	1128.4927	
	206 - 217	SLEVSEEKANQR	1388.6895	1388.5927	
	214 - 222	ANQREEAYK	1107.5308	1107.4927	

(Cont'd.) **Table 5. 3:** A list of generated peptides of TM extracted from BTP that have been generated by this study using different enzymes and MS ion sources.

Protease	Residues position	Peptide sequence (Z)	Molecular weight		Ion Source
			Calculated	Expected	
Trypsin	218 - 226	EEAYKEQIK	1136.5713	1136.4927	MALDI
	232 - 238	LKAAEAR	757.4446	757.4322	
	239 - 244	AEFAER	721.3395	721.2927	
	252 - 264	EVDRLEDELVNEK	1586.7787	1586.6927	
	269 - 284	SITDELDQTFSELSGY	1803.8051	1803.7927	
ASP-N	2 - 15	DAIKKKMQAMKLEK (3)	1660.9368	1660.9815	ESI
	58 - 74	DQVQESLLKANIQLVEK (3)	1954.0734	1954.1163	
	103-114	EERLNTATTKLA(2)	1345.7201	1345.7476	
	121-130	DESERMRKVL (2)	1261.6448	1261.6673	
	150-158	EARFLAEEA (2)	1034.5032	1034.5286	
	163-174	DEVARKLAMVEA (2)	1330.6915	1330.7083	
	184-194	ETGESKIVELE (2)	1232.6136	1232.6384	
	195-207	EELRVVGNNLKSL (2)	1469.8202	1469.8368	
	223-235	EQIKTLTNKLKAA (2)	1456.8613	1456.8808	
	243-253	ERSVQKLQKEV (2)	1342.7568	1342.7718	
	275 - 284	DQTFSELSGY(2)	1145.4877	1145.5052	
	2 - 15	DAIKKKMQA MKLEK	1660.9368	1660.8927	MALDI
	58 - 74	DQVQESLLKANIQLVEK	1954.0735	1953.9927	
	62 - 74	ESLLKANIQLVEK	1483.8609	1483.7927	
	75 - 95	DKALSNAEGEVAALNRRIQLL	2280.2549	2280.1927	
	82 - 97	EGEVAALNRRIQLLEE	1838.9850	1838.8927	
	103 - 114	EERLNTATTKLA	1345.7201	1345.6927	
	142 - 149	DALENQLK	929.4818	929.3927	
	150 - 158	EARFLAEEA	1034.5032	1034.3927	
	163 - 174	DEVARKLAMVEA	1330.6915	1330.5927	
	195 - 207	EELRVVGNNLKS	1346.6864	1346.5927	
	223 - 235	EQIKTLTNKLKAA	1456.8613	1456.7927	
	236 - 242	EARAEFA	792.3766	792.2927	
	243 - 253	ERSVQKLQKEV	1342.7568	1342.6927	
	258 - 271	DELVNEKEKYKSIT	1694.7826	1694.7927	
	275 - 284	DQTFSELSGY	1145.4877	145.3927	

Mowse scores for these runs were on average 1461, 481, and 187 as BTPTM identification for the *in-solution* generated peptides by trypsin, Glu-C V8, and Asp-N digestion, respectively. Table 5.1 reports the only peptides that match the Mascot criteria for the individual ions along with expected and calculated molecular weight. Elucidated representative MS/MS spectra for two peptides are shown in Figure 5.4, where the most dominant peptide fragment ions (y and b ions) are highlighted.

Multiple types of enzymatic digestions, ion sources, and derivatization protocols were applied to maximize the amino acid coverage of the TM protein (27). As stated, three different sets of enzymatic peptides were generated from BTP TM by trypsin, Asp-N, and Glu-C (V8) proteases. These different peptides with different termini enlarge the amino acid sequence coverage probability of the entire protein. A list of those peptides was organized in order as listed in table 5.1. The full amino acid '*de novo*' sequence for BTP TM is shown in figure 5.5.

In MS studies, peptides containing the N-terminal methionine are difficult to be ionized by ESI ion sources due to PTM acetylation on the amino acid. The evaluation of this PTM in shrimp TM by the *NetAcet 1.0* server scored a high value of 0.471 (28). Acetylation modifications are quite common for Eukaryotic TM, as reported for bovine, chicken (P04268) and human (P09493) cases in the UniProtKB/Swiss-Prot databases (25).

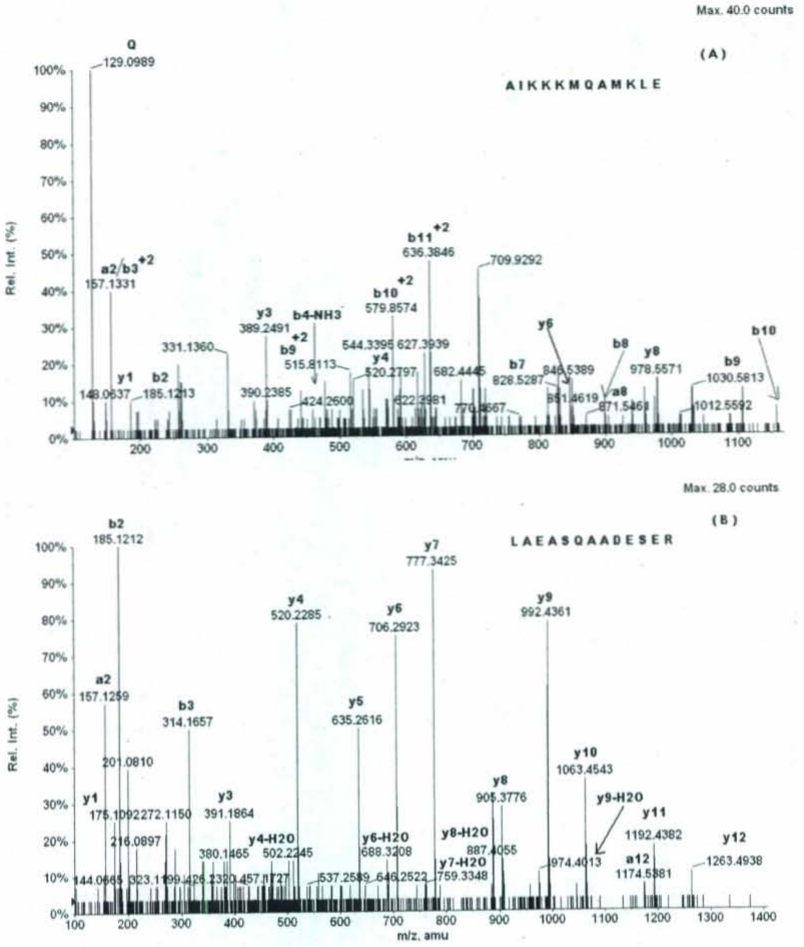


Figure 5. 4: A representative product spectra of two selected peptides A) generated from Glu-C V8 digestion and ESI ions sources for m/z 709.9323 $[M+2H]^{2+}$ with sequence AIKKKMQAMKLE and B) generated from tryptic digestion and MALDI ion source for m/z 1375.6215 $[M+H]^{1+}$ with sequence LAEASQAADESER.

```
  1 MDAIKKKMQA MKLEKDNAMD RADTLEQQNK EANNRAEKSE EEVHNLQKRM
 51 QQLENDLDQV QESLLKANIQ LVEKDKALSN AEGEVAALNR RIQLLEEDLE
101 RSEERLNTAT TKLAEASQAA DESERMRKVL ENRSLSDEER MDALENQLKE
151 ARFLAEEADR KYDEVARKLA MVEADLERAE ERAETGESKI VELEEELRVV
201 GNNLKSLEVS EEKANQREEA YKEQIKTLTN KLKAAEARAE FAERSVQKLQ
251 KEVDRLEDEL VNEKEKYKSI TDELDQTFSE LSGY
```

Figure 5. 5: BTP TM full amino acid sequence by '*bottom up*' approach

In another case with MS sensitivity, the MALDI ion source is highly selective for the arginine-containing peptides. Therefore, a guanidation reaction was performed to increase the sensitivity of the lysine-containing peptides in MALDI work. The guanidation reaction chemically modifies the lysine side chain to homoarginine, which has a proton affinity equivalent to the arginine residues resulting in better sensitivity using MALDI source (results are not shown).

Additional BTP proteins have been characterized in this study using mass spectrometry. A number of their enzymatic peptides were sequenced using the same MS strategy as described for TM. The 20kDa band as shown in Figure 5.1A and which reacts strongly with three patients' IgE (Fig. 5.1 C) was excised and characterized. Two relevant and abundant peptides of this band were '*de novo*' sequenced using PFF. The amino acid sequences for these peptides were EGFQLMDR and GTFDEIGR and searched against the NCBInr databank using the Mascot search engine. These peptide matched with

myosin light chain, a recently reported an allergen in whiteleg shrimp (*Litopenaeus vannamei* (Lit v3)) (29).

The BTP arginine kinase was also characterized as the 40-KDa protein which also reacted with seven patients' IgE. The band was excised, tryptic digested, and analyzed by MS. The most abundant peptides were '*de novo*' sequenced (i.e AVFDQLKEK, VSSTLSSLEGELK, GTYYPLTGMSK, LIDDHFLFK, IISMQMGGDLGQVFRR, LTSAVNEIEKR, IPFSHHDR, GTRGEHTEAEGGIYDISNK).

Additional muscle proteins were profiled by mass spectrometry in the same manner as the arginine kinase and myosin light chain. These proteins are troponin C, myosin heavy chain, and calmodulin.

As stated above, another major objective of this study is to identify a suitable signature peptide as a surrogate for the BTP TM protein in quantitative measurements of TM in the seafood workplace. A very important factor for selecting a signature peptide is the absence of PTM groups (i. e. phosphorylation and glycosylation) (30). Therefore, the precursor ions and intensity data generated from PMF experiments were uploaded on the ExPASy FindMod tool to check if there was any potential peptide having any PTM motif(s). The report indicated the absence of any type of modifications. Further confirmation was obtained by manual searching for the calculated molecular ions of PTM motifs in the precursor spectra. Additional assessment was performed using the *NetPhos 2.0* server (31). Next, peptides that scored >0.50 in abundance rating were excluded as

the signature peptide nomination for more suitable candidate sensitivity in quantification. All resultant peptides from the tryptic digestion without any missing cleavage (Table 1) were examined for signature peptide criteria. The uniqueness of peptides was recommended by the Mascot search engine along with PTM evaluation from the *NetPhos 2.0* server (31). Those selected peptides were then introduced to the protein BLAST test. The NCBI BLAST test, which is used to find regions of local similarity between sequences of the NCBI database and calculates the statistical significance of matches, reported that the peptide located at 67-74 (ANIQLVEK) would be a suitable signature peptide for black tiger prawn TM (with 100% identity, score=28.2 bits (590), and expected= 94). The product ion spectrum of this peptide, m/z 457.7712 $[M+2H]^{2+}$, was collected using an ESI source (Fig. 5.6), where the most abundant y and b ions were assigned.

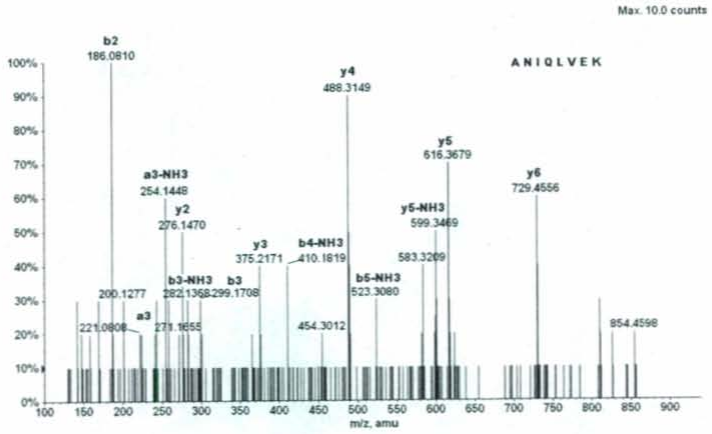


Figure 5. 6: The product ion spectrum of, m/z 457.7712 $[M+2H]^{2+}$ for the selected signature peptide with sequence ANIQLVEK.

5.4. Conclusion

A general strategy was demonstrated to characterize the important allergenic proteins present in black tiger prawn. Tropomyosin, myosin light chain, and arginine kinase were identified using PMF, and some of their highly abundant peptides were '*de novo*' sequenced using PFF. Tropomyosin as the major allergen was completely sequenced using different enzymatic digestion strategies, derivatization protocols, and ion sources for mass spectrometry. The allergenicity of these three proteins was confirmed in this study by the immunoblotting of these proteins with patient's sera. For BTP TM quantification purposes, the selection rules for a recommended signature

peptide have been followed. The experimental data for the resultant peptides was evaluated by bioinformatics approaches to determine the PTM-free peptides. The nominated peptides were introduced to the BLAST test to acquire the highest scoring peptide as signature peptide. This peptide (ANIQLVEK) will be chemically synthesized (labeled and unlabeled forms) and used in further studies to develop a quantification method for BTP TM in different environments in seafood industry workplace.

5.5. References

1. Shiau, S. Y., Chen, Y. Estimation of the dietary vitamin A requirement of juvenile grass shrimp, Penaeus monodon. *J. Nutr.* **2000.** *130; 9-94.*

2. Motoyama, K., Suma, Y., Ishizaki, S., Nagashima, Y., Shiomi, K. Molecular Cloning of Tropomyosins Identified as Allergens in Six Species of Crustaceans. *J Agric Food Chem.* **2007.** 55; 985-991.

3. Lopata, A. L., Lehrer, S. B. New insights into seafood allergy. *Curr Opin Allergy Clin Immunol.* **2009.** 9; 270-277.

4. Yu, C. J., Lin, Y. F., Chiang, B. L., Chow, L. P. Proteomics and immunological analysis of a novel shrimp allergen, Pen m 2. *J. Immunol.* **2003.** 170; 445-453.

5. Food and Agriculture Organization of the United Nations (FAO). The State of World Fisheries and Aquaculture. *Fisheries and Aquaculture Department: Rome,* **2007.**

6. Jeebhay, M. F., Robins, T. G., Lehrer, S. B., Lopata, A. L. Occupational seafood allergy: a review. *Occup Environ Med.* **2001.** 58: 553-652.

7. García-Orozco, K. D., Aispuro-Hernández, E., Yepiz-Plascencia, G., Calderón-de-la-Barca, A. M., Sotelo-Mundo, R. R. Molecular characterization of arginine kinase, an allergen from the shrimp Litopenaeus vannamei. *Int Arch Allergy Immunol*. **2007.** 144; 23-28.

8. France, R. M., Sellers, D. S., Grossman, S. H. Purification, characterization, and hydrodynamic properties of arginine kinase from gulf shrimp (Penaeus aztecus). *Arch Biochem Biophys*. **1997**. 345; 73-78.

9. Yao, C. L., Wu, C. G., Xiang, J. H., Dong, B. Molecular cloning and response to laminarin stimulation of arginine kinase in haemolymph in Chinese shrimp, Fenneropenaeus chinensis. *Fish shellfish immunol.* **2005**. 19; 317-329.

10. Ortea, I., Cañas, B., Gallardo, J. M. Mass spectrometry characterization of species-specific peptides from arginine kinase for the identification of commercially relevant shrimp species. *J Proteome Res*. **2009.** 8; 5356-5362.

11. Hoffman, D. R., Day, E. D., Miller, J. S. The major heat stable allergen of shrimp. *Ann Allergy*. **1981**. 47; 17-22.

12. Daul, C. B., Slattery, M., Reese, G., Lehrer, S. B. Identification of the Major Brown Shrimp (Penaeus aztecus) Allergen as the Muscle Protein Tropomyosin. *Int Arch Allergy Immunol.* **1994.** 105; 49-55.

13. Shanti, K. N., Martin, B. M., Nagpal, S., Metcalfe, D. D., Subba Rao, P. V. Identification of tropomyosin as the major shrimp allergen and characterization of its IgE-binding epitopes. *J Immunol.* **1993.** 151; 5354-5363.

14. Liang, Y. L., Cao, M. J., Su, W. J., Zhang, L. J., Huang, Y. Y., Liu, G. M. Identification and characterisation of the major allergen of Chinese mitten crab (Eriocheir sinensis). *Food Chem.* **2008.** 111; 998-1003.

15. Jeebhay, M. F., Robins, T. G., Seixas, N., Baatjies, R., George, D. A., Rusford, E., Lehrer, S. B., Lopata, A. L. Environmental exposure characterization of fish processing workers. *Ann Occup Hyg.* **2005**. 49; 423-437.

16. Lopata, A. L., Jeebhay, M. F., Reese, G., Fernandes, J., Swoboda, I., Robins, T. G., Lehrer, S. B. Detection of fish antigens aerosolized during fish processing using newly developed immunoassays. *Int Arch Allergy Immunol.* **2005.**138; 21-28.

17. Malo, J. L., Chretien, P., Mccants, M., Lehrer, S. B. Detection of snow-crab antigens by air sampling of a snow-crab production plant. *Clin Exp Allergy.* **1997.** 27; 75-78.

18. Kluwe, L., Maeda, K., Miegel, A., Fujita-Becker, S., Maeda, Y., Talbo, G., Houthaeve, T., Kellner, R. Rabbit skeletal muscle alpha alpha-tropomyosin

expressed in baculovirus-infected insect cells possesses the authentic N-terminus structure and functions. *J. Muscle Res Cell Motil.* **1995.** 16; 103-110.

19. Reese, G., Schicktanz, S., Lauer, I., Randow, S., Luttkopf, D., Vogel, L., Lehrer, S. B., Vieths, S. Structural, immunological and functional properties of natural recombinant Pen a 1, the major allergen of Brown Shrimp, Penaeus aztecus. *Clin Exp Allergy.* **2006**. 36; 517-524.

20. Rosmilah, M., Shahnaz, M., Masita, A., Noormalin, A., Jamaludin, M. Identification of major allergens of two species of local snappers: Lutjanus argentimaculatus (merah/ red snapper) and Lutjanus johnii (jenahak/ golden snapper). *Trop Biomed.* **2005.** 22; 171-177.

21. Lopata, A. L., Jeebhay, M. F., Reese, G., Fernandes, J., Fenemore, B., Elliott, A., Robins, T. G., Lehrer, S. B. Detection of fish antigens aerosolized during fish processing using newly developed immunoassays. *Int Arch Allergy Immunol.* **2005.** 138; 21-28.

22. Sergeant, K., Samyn, B., Debyser, G., Beeumen, J. V. '*De novo* ' sequence analysis of *N*-terminal sulfonated peptides after in-gel guanidination. *Proteomics.* **2005**. 5; 2369-2380.

23. Norrgran, J., Williams, T. L., Woolfitt, A.R., Solano, M. I., Pirkle, J. L., Barr, J. R. Optimization of digestion parameters for protein quantification. *Analytical Biochemistry.* **2009.** 393; 48-55.

24. Abdel Rahman, A. M., Lopata, A. L., O'Hehir, R. E., Robinson, J. J., Banoub, J. H., Helleur, R. J. Characterization and '*de novo* ' sequencing of snow crab tropomyosin enzymatic peptides by both electrospray ionization and matrix-assisted laser desorption ionization QqToF tandem mass spectrometry. *J Mass Spectrom.* **2010.** 45; 372-381.

25. De Hoffmann, E., Stroobant V. Mass Spectrometry: Principles and Applications. John Wiley; New York, 2007. 3rd edition.

26. Kinter, M., Sherman, N. E. Protein Sequencing and I dentification Using Tandem Mass Spectrometry. *John Wiley; New York,* **2000.**

27. Kiemer, L., Bendtsen, J. D., Blom, N. NetAcet: prediction of N-terminal acetylation sites. *Bioinformatics.* **2005.** 21. 1269-1270.

28. Ayuso, R., Grishina, G., Bardina, L., Carrillo, T., Blanco, C., Ibáñez, M. D., Sampson, H. A., Beyer, K. Myosin light chain is a novel shrimp allergen, Lit v 3. *J Allergy Clin Immunol.* **2008.** 122; 795-802.

29. Brun, V., Masselon, C., Garin, J., Dupuis, A. Isotope dilution strategies for absolute quantitative proteomics. *Proteomics.* **2009.** 72; 740-749.

30. Blom, N., Gammeltoft, S., Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **1999.** 294; 1351-1362.

Chapter 6: Summary and future work.

The snow crab fishery is one of the most important fisheries in Newfoundland and Labrador but the province also has a high prevalence of crab asthma in the workplace. Previously, Memorial University's SafetyNet Centre for Occupational Health and Safety Research performed a comprehensive study on the snow crab asthma across three years (2004-2007). This study showed the importance of the crab asthma and its influence on the workers' health and safety, particularly, the huge number of workers who most of them are seasonal with limited alternative income. As well, most of actual crab plants were designed for ground fish processing (with its poor ventilation), which switched to crab processing in the early 1990s. These facts provoked us to develop a comprehensive strategy for multi-allergen analysis in air samples using mass spectrometry to identify the main allergenic proteins and to measure their levels in the air work environment. The two species of interest were snow crab and black tiger prawn.

This study began with isolating and purifying the important allergenic proteins from snow crab meat. The allergenicity of both crude meat extracts and purified proteins was examined by immunoblotting with patients' sera. Tropomyosin (TM) was conclusively identified as one of the major allergens and was '*de novo*' sequenced using tandem mass spectrometry. The '*bottom up*' approach of protein sequencing was optimized using different enzymatic digestions, ESI and MALDI ion sources, and various derivatization regimes. The post translational modification sites were evaluated using bioinformatics tools to assist in selecting the candidate signature peptides for

tropomyosin allergen. A signature peptide was carefully selected using specific criteria to represent the molecular surrogate of target protein.

Besides tropomyosin several other extracted proteins showed strong reactivity with sensitized patients' IgE. These proteins were characterized using tandem mass spectrometry followed by bioinformatics evaluations. Sarcoplasmic Ca-binding, troponin, actine, smooth endoplasmic reticulum Ca^{2+} ATPase, and arginine kinase were identified as other snow crab allergens along with tropomyosin. It was noteworthy that 49% of the participant patients reacted to arginine kinase (AK). Consequently, the AK protein was purified and characterized using '*bottom up*' tandem mass spectrometry. The gene sequence of snow crab arginine kinase was not available in the data bank. Therefore, the protein MS data of AK was manipulated manually and the sequence was confirmed by the AK's gene sequence of the closest living species (organ mud crab). The active site of snow crab AK was also evaluated by searching for consensus active motif. For analytical purposes, the signature peptide of AK was determined by meeting the criteria of the signature peptide.

An airborne allergen quantification method was developed using multi-reaction monitoring tandem mass spectrometry. The signature peptides of snow crab's tropomyosin and arginine kinase were chemically synthesized as analytical standards in both light and heavy forms. These standards were used to optimize the analytical method

using ESI triple quad tandem mass spectrometry. The analytical method was validated using FDA guidelines in terms of accuracy, sensitivity, linearity, and specificity.

To simulate the collection of allergens on test air filters, known amounts of target allergens were placed on filter papers. Sample handling methods were developed and optimized to maintain a reproducible and sensitive assay approach. The completeness of the tryptic digestion was optimized under different conditions to ensure the stoichiometric ratio between the signature peptide and its protein. The spiked proteins were eluted from filter papers using a detergent. This detergent was removed prior to MS analysis by TopTip filters. The extraction and recovery of the protein and its signature peptide measured by MS was found to be highly effective.

Real air samples for allergen analyses were collected using a simulated snow crab processing plant. The bound proteins on the filters were extracted and quantified. The analytical signals of both allergens (TM and AK) were high, this showing not just the selectivity but also the sensitivity of the developed quantitation method for airborne allergens. Of notable significance, arginine kinase has been identified as an aeroallergenic protein. This is a clear signal to occupational health researchers that they should also focus on the aerosolized forms of allergens in future studies.

The same comprehensive strategy can be applied to any other seafood species. Dr. Lopata and his co-workers in Australia are working with black tiger prawn. They are interested in potential allergenic proteins in this species since it is one of the major

seafood products in Southeast Asia. The TM and crude extracts were analyzed in the same manner as described for the snow crab. The allergenicity of the extracts was evaluated using different patients' sera. The complete primary structure of TM was determined using tandem mass spectrometry. Post translational modifications were evaluated to exclude any modified peptides as signature peptides. A signature peptide was determined and recommended for quantitation purposes. Myosin light chain, arginine kinase, troponin C, myosin heavy chain, and calcium modulated protein (Calmoduli) were also identified as allergens as shown by their reactivity with patients' sera.

Suggested future work of this project is divided to three parts; the allergens identification, improved allergen quantification, and the implementation of these approaches in the future of the seafood occupational health and safety. Allergen evaluation and identification should be expedited by performing the allergeomics strategy, which utilizes the proteomics approach to protein identification and allergenicity evalution using tandem mass spectrometry and immunoblotting, respectively. The crude extract samples should be profiled in 2D-gel electrophoresis, which produces a highly resolved and sensitive profile. For allergenicity evaluation, these 2D-gels will be blotted and incubated with pool of patients' sera. The reactive proteins would then be excised, digested and characterized using mass spectrometry. Consequently, signature peptides for these allergens will be assigned for developing an absolute quantification method.

It is highly recommended, as a future work, to have an allergeomics profile and absolute multi-allergens quantitation method. Therefore, proteomics quantitation methods such as iTRAQ or TMT can be used to give a relative quantitation profile for all the detected allergens in each sample in very specific and sensitive manner. All candidate allergens that have been shown up in the iTRAQ profile will be considered for developing an absolute quantitation method. This can be performed by designing and then synthesizing an artificial polypeptide consisting of all the signature peptides (of all target allergens) as a standard and internal standard in light and heavy forms, respectively. The tryptic digestion of this artificial polypeptide will yield a group of signature peptides related to the target allergenic proteins. The heavy form standard polypeptide will be added to sample filters, as internal standard right before the extraction and digestion. This process will control the completeness of the digestion and sample loss as well as give an absolute quantification for each single allergen in the sample in a single run. This approach could be performed for air samples as well as for controlling the levels of allergens in food as a quality control test (food gradients) for the sake of reporting the nutritional facts.

Such a future study would introduce accurate, sensitive, and specific advanced monitoring approaches to minimize allergens exposure among seafood processing workers in the shellfish industry. Having information about the identity and quantity of these allergens will motivate occupational health and safety researcher to improve workplace health management. Monitoring workplaces by this real-time method will

improve the regulations of the workstation, ventilation systems, and machine designs, which will help minimize the exposure. Using this approach the Department of Fisheries and Aquaculture, who issues the processing plants' operational licenses, would be able to monitor the ventilation system of fish plants, and determine the threshold of allergen levels in the workplaces environment.