Research Article

Journal of MASS SPECTROMETRY

Received: 5 November 2009 Accepted: 26 January 2010 Published online in Wiley Interscience: 2 March 2010

(www.interscience.com) DOI 10.1002/jms.1721

Characterization and *de novo* sequencing of snow crab tropomyosin enzymatic peptides by both electrospary ionization and matrix-assisted laser desorption ionization QqToF tandem mass spectrometry

Anas M. Abdel Rahman,[a]* Andreas L. Lopata,[b] Robyn E. O'Hehir,[c] John J. Robinson,[d] Joseph H. Banoub[d,e] and Robert J. Helleur[a]

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 TM protein was purified from both claw and leg meats of snow crab and analyzed by electrospray ionization and matrix-assisted laser desorption/ionization (MALDI) using hybrid quadruple time-of-flight tandem mass spectrometry (QqToF-MS). The native polypeptide molecular weight of TM was determined to be 32 733 Da. The protein was further characterized using the '*bottom-up*' MS approach. A peptide mass fingerprinting was obtained by two different enzymatic digestions and *de novo* sequencing of the most abundant peptides performed. Any post-translational modifications 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 TM using MALDI-QqToF-MS will be critical for the sensitive and specific quantification of this highly allergenic protein in the work place.

Keywords: snow crab; tropomyosin; *de novo* sequencing; mass spectrometry; allergen; signature peptide; in-gel guanidation; PTM evaluation

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 across 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 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 that mediate the interaction between the troponin complex and actin, which regulates muscle contraction.[5] Crustaceans' TM was first identified in shrimp by Hoffman *et al*. in 1981.[6] It is a water-soluble and heat-stable protein with molecular weights (MWs) ranging between 34 and 39 kDa.[7] It has a highly conserved amino acid sequence among different invertebrate organisms and is present in muscle as well as non-muscle cells.[8] Differential splicing of the pre-messenger ribonucleic acid (pre-mRNA) produces isoforms of TM.[9] The TM muscle isoform contains 284 amino acids after including a highly conserved N-terminal region.[10–12]

Research has shown that TM isolated from shrimps and crabs is a major food allergen.[13] Snow crab is among the sea foods that are most frequently associated with an Immunoglobulin E (IgE) mediated type I hypersensitivity in patients with food allergy. Urticaria, asthma and diarrhea are the major clinical symptoms that are caused by type I hypersensitivity.[14]

* *Correspondence to: Anas M. Abdel Rahman, Department of Chemistry, Memorial University of Newfoundland, St John's, Newfoundland, Canada A1B 3X7. E-mail: anasar@mun.ca*

a *Department of Chemistry, Memorial University of Newfoundland, St John's, Newfoundland, Canada A1B 3X7*

b *School of Applied Science, Marine Biomedical Sciences and Health Research Group, RMIT University, Bundoora, 3083 Victoria, Australia*

c *Department of Allergy, Immunology and Respiratory Medicine, Alfred Hospital and Monash University, Melbourne, 3004 Victoria, Australia*

d *Department of Biochemistry, Memorial University of Newfoundland, St John's, Newfoundland, Canada A1B 3X9*

e *Fisheries and Oceans Canada, Science Branch, Special Projects, St John's, Newfoundland, Canada A1C 5X1*

Journal of MASS SPECTROMETRY

Monitoring the level of TM in the snow crab processing workplaces is essential in reducing the workers' 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 test (RAST),[13] and immunoblotting.[15,16] The main disadvantage in analyzing the total crab protein by conventional methods is the inclusion of additional proteins that are not necessarily involved in the allergic reactions of the exposed worker.[17] However, by targeting and quantifying the major crustacean allergen protein, one 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 sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) to yield homogeneous bands. Purified TM was used in western blot analysis to study the reactivity of the TM with specific reactivity of monoclonal and polyclonal antibodies as well as serum of patients with crustacean allergy.

The MW of the extracted TM was determined by electrospray ionization quadruple time-of-flight tandem mass spectrometry (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 matrix-assisted laser desorption/ionization (MALDI)-QqToF-MS. Peptides were also *de novo* sequenced using the '*bottom-up*' MS approach.[17–20]

It has been well established that PMF and the various outcomes of the MS approach can give complete information about the primary structure 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 the 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.

It is well known that a good score in the Basic Local Alignment Search Tool (BLAST) analysis,[20] together with good separation in high-performance liquid chromatography (HPLC) analysis, allows the best choice for the signature peptide.[18,20] The signature peptide should not have any post-translational modification (i.e. glycosylation and phosphorylation).[18] The sensitivity and specificity that MS can provide, with a high-throughput sample analysis in protein characterization, sequencing and PTM mapping, are ideal for this study.

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

Experimental

Chemicals and materials

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

TM purification

Acetone powder extract

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

Purified extract

The acetone powder (10 g) was dissolved in 70 ml of buffer A (1 M KCl, 25 mM Tris–HCl, pH 8.0, 0.25 M DTT and 0.5 mM EDTA) and 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 was diluted to a concentration of 1–2 mg/ml with buffer A. With isoelectric precipitation (pI 4.6), the proteins were precipitated with 1 M HCl. The precipitant proteins were redissolved in buffer B (0.20 M KCl, 25 mM Tris–HCl, pH 8.0, 0.25 M DTT and 0.5 mM EDTA). TM precipitated with ammonium sulfate 40–70% after the other proteins have been precipitated in 0–40%. Finally, the pellets were reconstituted by deionized water, dialyzed against ammonium bicarbonate by dialysis tubes (MWCO = 12–14 000) overnight at room temperature and then lyophilized to a powder.

Sodium dodecylsulfate polyacrylamide gel electrophoresis

A pair of 12% SDS-PAGE 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 for 45 min or 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 h. After the transfer was completed, the membrane was placed in a blocking solution (5% skim milk in Tris-buffered saline (TBS)) for immunoblotting.

Immunoblotting

The rabbit antibody used in this study was generated by exposing the animals to heat-treated protein extracts from crab, prawn and lobster using 500 μg of a mixture of these three proteins sources for each injection. Each protein extract was prepared from raw crustaceans, the generated protein extract 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 analyze for antibody production, and

the final bleed was conducted at week 9.[21] The immunoblot was blocked with 5% skim milk solution for 1 h at room temperature following an overnight incubation (4 °C) with the rabbit antibody (primary antibody) using a working dilution of 1 : 40 000. The blot was washed three times with TBS–Tween. The secondary antibody used was polyclonal goat anti-rabbit antibody conjugated to HRP, with a working dilution of 1 : 20 000. The blot was incubated for 30 min in the secondary antibody solution, washed again three times with TBS–Tween and incubated with the substrate 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 as part of an ongoing survey. For immunoblotting, protein extracts were electrophoretically separated (see SDS-gels above) and the proteins transferred and incubated with human serum (diluted 1 : 10 in 1% skim milk) overnight at 4 °C. Subsequently, the blots were washed three times with PBS-T and membrane-incubated for 1 h in 5 ml of rabbit antihuman 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 A 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, and incubated with the substrate TMB, and the immunoblot membranes were analyzed for IgE reactivity using the ECL technique.[21]

In-gel digestion and guanidation

The 33-kDa protein bands were excised from the SDS-PAGE plate. The in-gel guanidation procedure was performed on all protein samples, which 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 $(NH_4)_2CO_3$ 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 50 mM $(NH_4)_2CO_3$, 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 sucked out. The gel was covered by a solution of 50 mM of $(NH_4)_2CO_3$ and incubated at 37 °C overnight to enhance protein digestion. The water-soluble peptides were extracted twice with the incubation solution, and the other remaining peptides were extracted twice with 0.15% trifluoroacetic acid (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.

Matrix-assisted laser desorption/ionization quadruple time-of-flight tandem mass spectrometry

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. The first layer matrix solution (0.5 µl) was applied to a MALDI target. The second layer matrix solution (1 µl) was mixed with 1 µl of sample. The sample/matrix mixture (1 µl) was deposited onto the first layer and allowed to dry, which was 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 s), the remaining amount of salt was removed.

MALDI-QqToF-MS and CID-MS/MS

MALDI-MS and low-energy collision (CID) analyses were carried on a QSTAR XL hybrid quadrupole–quadrupole (Qq)/ToF-MS/MS (Applied Biosystems/MDS Sciex, Foster City, CA, USA) equipped with an *o*-MALDI ion source (Applied Biosystems, Foster City, CA, USA).

Electrospray ionization quadruple time-of-flight tandem mass spectrometry

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 i.d. ×5 mm, C_{18} PepMap 100, 5 µm (LC Packing, Sunnyvale, CA)) for desalting and concentrating. Peptides were then eluted from the precolumn and separated on a nanoflow analytical column (75 µm i.d. ×15 cm, C_{18} 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% FA/0.01% TFA/2% ACN and (B) 0.08% FA/0.008% TFA/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 spectra 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, MA, USA) at a voltage of 2400 V.

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

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 (NCBInr) database 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 because the guanidation derivatization has been performed. Peptides were considered identified if the Mascot score was over 95% confidence limit.

Results 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. 1(a). The strong band at 33 kDa is presumed to be TM. The much weaker band >65 kDa, which has a high reactivity against anti-crustacean polyclonal antibody (Fig. 1(b)) along with patients' sera (Fig. 1(c) and (d)) that appeared more clearly, was presumed to either be a dimer or an aggregate form of TM.

The purified protein for both the claw and leg extracts was further investigated using direct flow injection into ESI source of the QqToF instrument. The multiply charged ion spectrum of the purified TM gave a series of multicharged ions, which were rationalized as $[M + 45H]^{45+}$ *m/z* 729.6096; $[M + 44H]^{44+}$ *m/z*

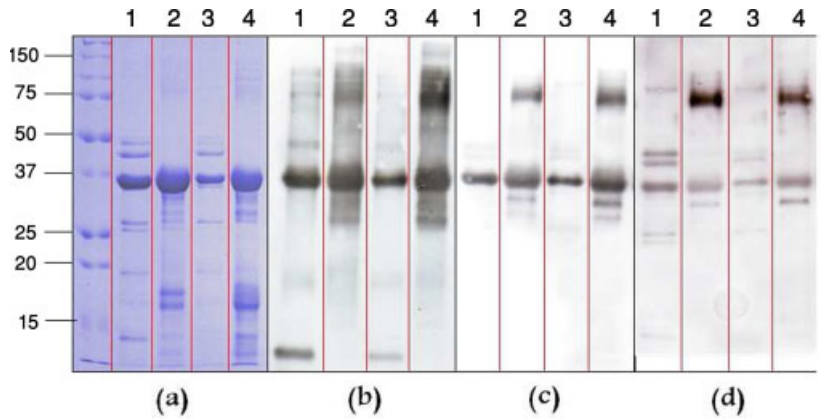

Figure 1. Snow crab extract profiling by (a) SDS-PAGE, (b) immunoblotting anti-crustacean polyclonal antibody, (c) immunoblotting patient serum A and (d) immunoblotting patient serum B, where (1) is claw acetone powder extract, (2) is claw purified TM extract, (3) is leg acetone powder extract and (4)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

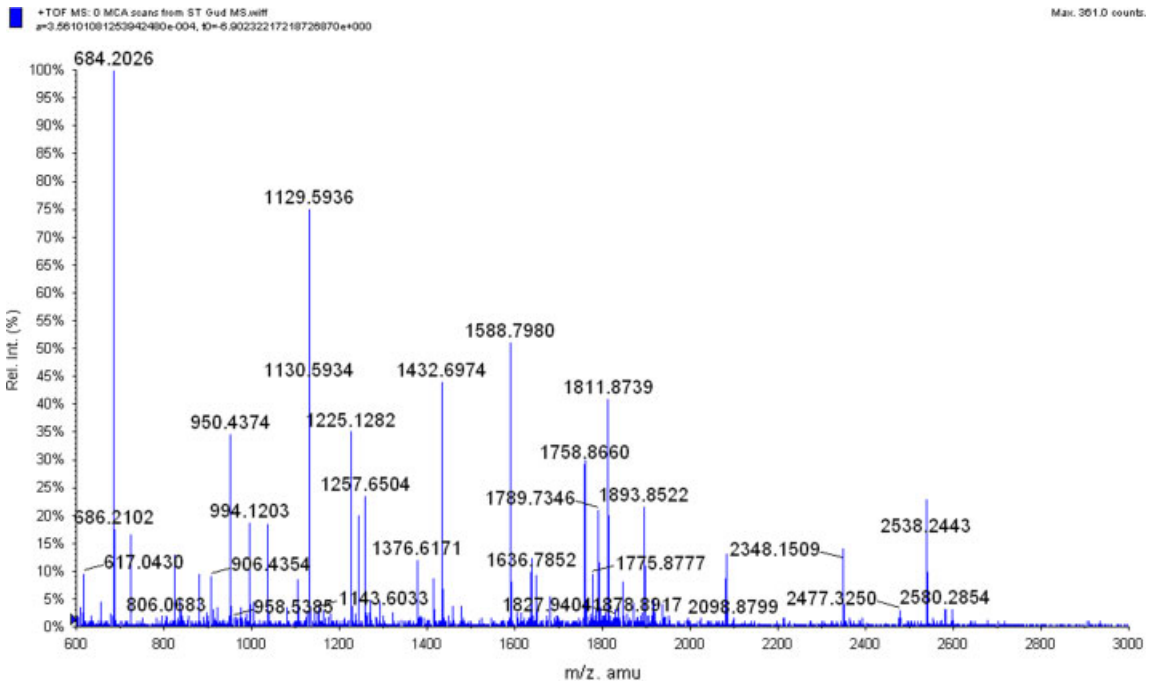

Figure 2. MALDI-QqToF mass spectrum of 'in-gel' trypsin digestion of snow crab TM. Each peak in the spectrum represents a tryptic peptide. Their masses in this spectrum are used for PMF characterization.

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 the Analyst QS 1.1 software, to give an MW of claw TM extract of 32 783 ± 0.02% Da. Another series of multicharged ions was also deconvoluted using the same tools to give 32 776 ± 0.02% Da as the MW of the leg TM extract. The differences in mass between the calculated and the experimental values are related to the post-translational modification or to intramolecular modifications (i.e. N-terminal acetylation (42 Da)).[23]

The separated proteins in SDS-PAGE were transferred to the nitrocellulose membrane and incubated with anti-TM mAb, anti-crustacean TM polyclonal antibody and sera of two sensitized patients with different symptoms, as seen in Fig. 1(b)–(d). This permitted the examination of the bioreactivity 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 one sample was 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-MS (Fig. 2). The combination of these two enzymes generates two sets of peptides with different termini and MS 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 enzymatically 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

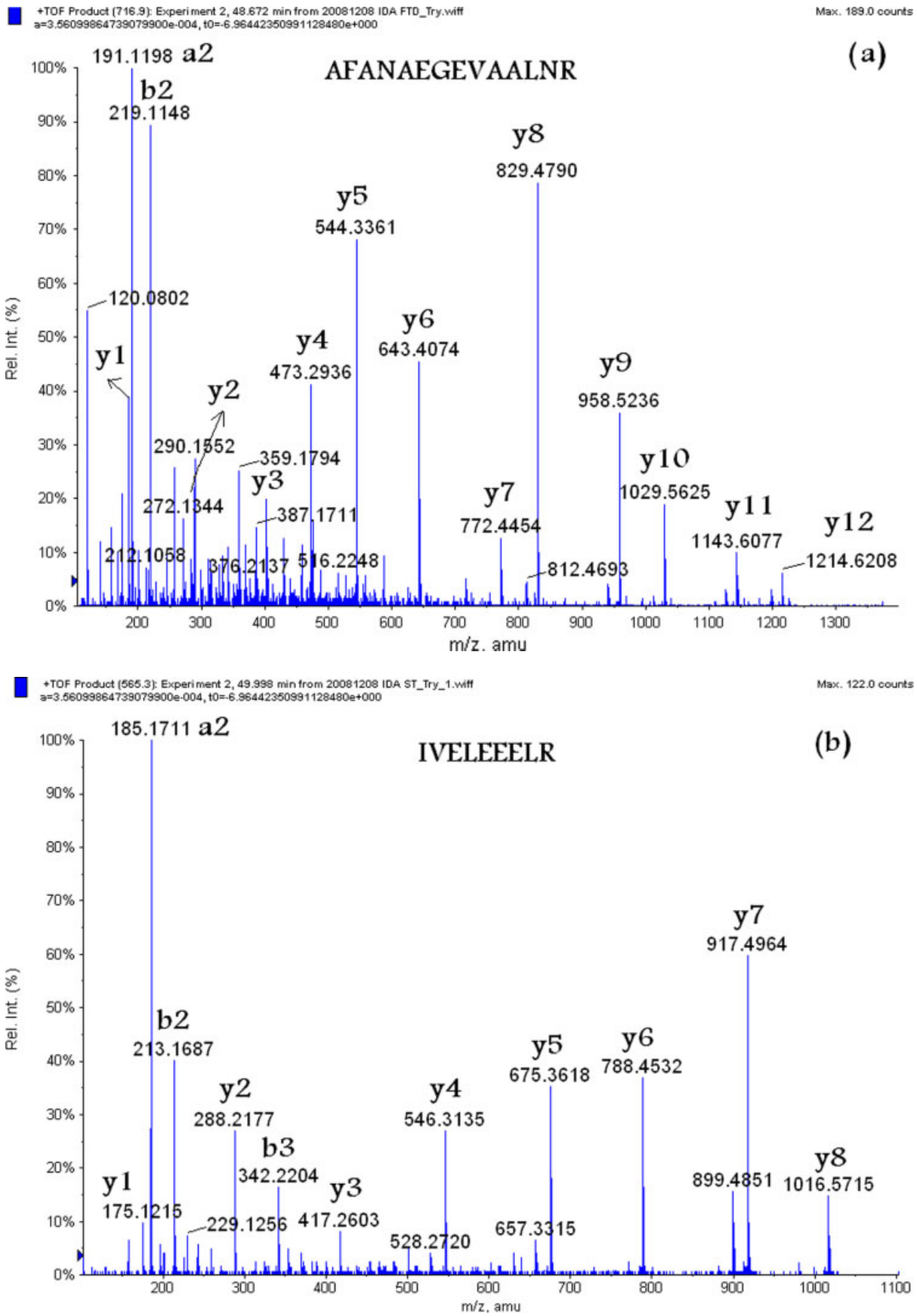

Figure 3. Representative ESI product ion spectra of selected precursor ions of (a) $[M + 2H]^{2+} = m/z$ 716.9 and (b) $[M + 2H]^{2+} = m/z$ 565.3 of two selected peptides obtained from the tryptic digestion of SC TM having sequences AFANAEGEVAALNR and IVELEEELR, respectively.

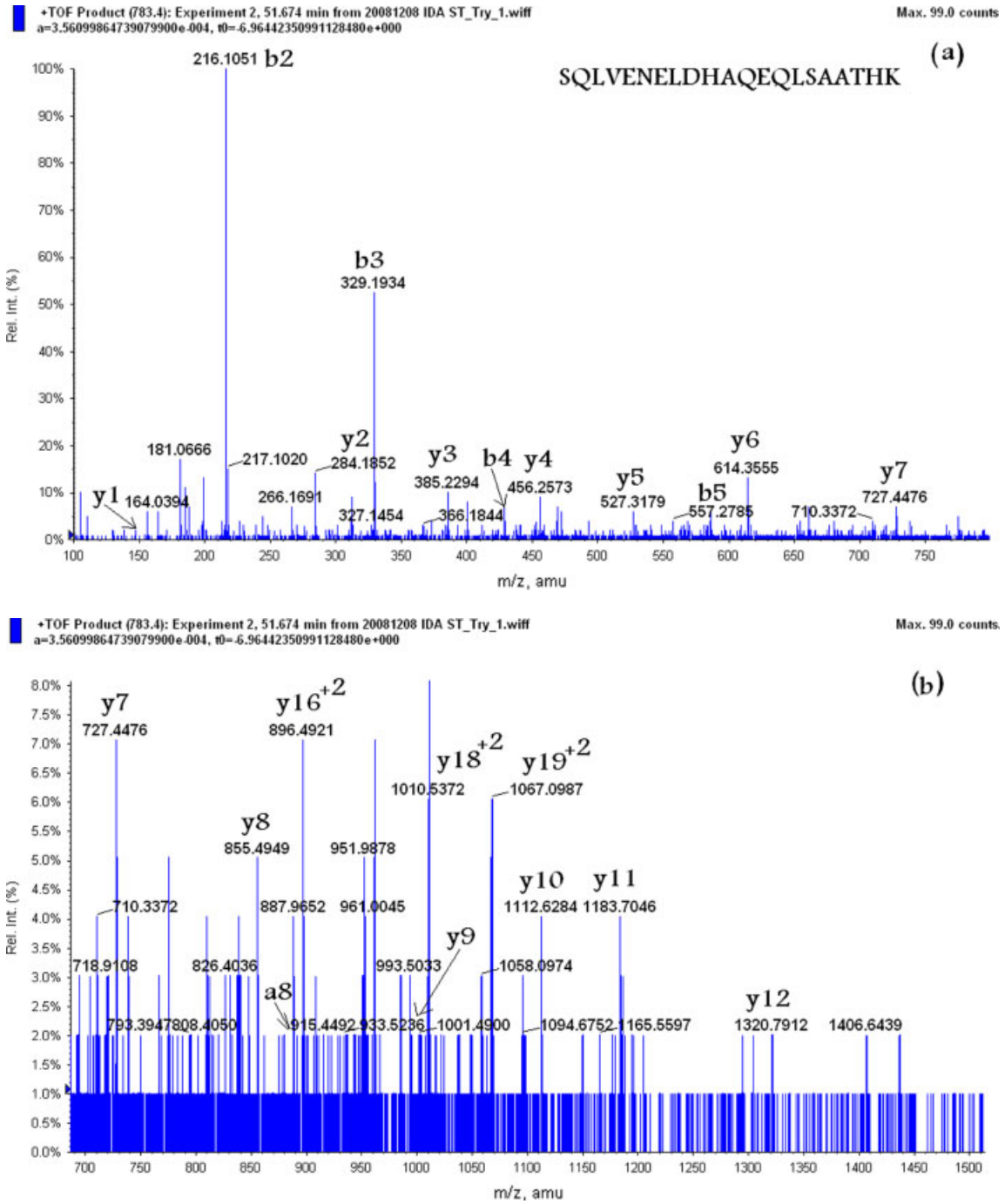

Figure 4. ESI product ion spectrum of snow crab TM signature peptide $[M + 3H]^{3+} = m/z$ 783.4 Da. The axes have been cut to improve peak visualization. The top (a) and bottom (b) correspond to m/z ranges of 100–800 and 700–1500, respectively.

digestion, which is the top probability based on Mowse scores of matching with the snow crab TM protein (complementary deoxyribonucleic acid (cDNA) based Library (NCBInr)). The above scores are matched with the Mascot algorithms' criterion: 'Individual ions scores >82 indicate identity or extensive homology ($p < 0.05$)'.[25] For improving the sensitivity of lysine-containing peptides in the MALDI source and for increasing 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. 1(a)).

Both MALDI and ESI ionization 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), some of which are shown in Fig. 2, were selected in the first quadruple and fragmented in the low-energy CID collision cell, and the product


```
1         11        21        31        41        51
MDAIKKKMQAMKLEKDNAMDKADTLEQQNKEANLRAEKTEEEIRANQKKSQLVENELDHA
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 5. Full amino acid sequence of snow crab TM with the calculated molecular weight 32 655 Da.

ions were scanned by the ToF analyzer. The full sequenced peptides are summarized in the Table 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 of peptides' precursor ions and product ions were uploaded to the Mascot MS/MS ion search engine against the 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 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. 3.

The *de novo* sequences 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

Table 1. *De novo* sequencing of the product ion spectra obtained from the MALDI-QqToF and ESI-QqToF experiments of the major peptides from snow crab TM 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	5070	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	+	–
	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	+	–

Journal of MASS SPECTROMETRY

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

	NetPhos 2.0 phosphorylation prediction				This study: mass spectrometric data			
Peptide (charge states)	Position	Score[a]	Predicted AA residue	*m/z*	Observed	MW (expt.)	MW (calc.)	Delta
DNAMDKAD**T**LEQQNK (+3)	24	0.449	–	600.9205	574.2832	1719.8278	1719.7733	0.0544
AEK**T**EEEIR (+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
SQLVENELDHAQEQLSAA**T**HK (+3)	68	0.598	T	810.0429	783.4223	2347.2450	2347.1404	0.1046
SQLVENELDHAQEQL**S**AATHK (+3)	65	0.434	–	810.0429	783.4223	2347.2450	2347.1404	0.1046
IQLLEEDLER**S**EER (+3)	102	0.936	S	613.6226	586.9802	1757.9187	1757.8795	0.0392
RLN**T**ATTKLAE (+2)	108	0.149	–	649.3292	609.3691	1216.7236	1216.6775	0.0461
RLNTA**T**TKLAE (+2)	110	0.364	–	649.3292	609.3691	1216.7236	1216.6775	0.0461
RLNTAT**T**KLAE (+2)	111	0.569	T	649.3292	609.3691	1216.7236	1216.6775	0.0461
LAEA**S**QAADESER (+2)	117	0.203	–	728.8012	688.8408	1375.6670	1375.6215	0.0454
LAEASQAADE**S**ER (+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
SL**S**DEERMDALENQLK (+3)	136	0.997	S	653.2906	626.6616	1876.9628	1876.8836	0.0792
K**Y**DEVAR (+2)	162	0.925	Y	480.7130	440.7463	879.4781	879.4450	0.0331
RAE**S**GESKIVELEEE (+2)	185	0.996	S	892.9011	852.9547	1703.8948	1703.8213	0.0735
RAESGE**S**KIVELEEE (+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
SLEV**S**EEK (+2)	210	0.943	S	500.7154	460.7420	919.4694	919.4498	0.0196
EE**T**YKEQIK (+2)	220	0.985	T	624.2814	584.3200	1166.6255	1166.5819	0.0436
EET**Y**KEQIK (+2)	221	0.936	Y	624.2814	584.3200	1166.6255	1166.5819	0.0436
QIK**T**LANKLKAAEARAE (+3)	227	0.056	–	645.6856	619.0606	1854.1601	1854.0686	0.0915
FAER**S**VQKLQKE (+2)	245	0.009	–	771.8874	731.9433	1461.8720	1461.7939	0.0781
KEK**Y**KNIADE (+2)	267	0.993	Y	659.3080	619.3475	1236.6804	1236.6350	0.0454
YKNIADEMDQAF**S**ELSGF (+2)	279	0.646	S	1072.9478	1033.0053	2063.9961	2063.9146	0.0816
YKNIADEMDQAFSEL**S**GF (+2)	282	0.017	–	1072.9478	1033.0053	2063.9961	2063.9146	0.0816

[a] The NetPhos 2.0 threshold is >0.500.

either of the ionization sources or by two different digestion enzymes.

Analyzing the results in Table 1, it is interesting to note that the N-terminus peptide (residues 1–7, with sequence MDAIKKK) as a major ion was difficult to observe (ionize) and to produce a product spectrum. This is an indication that it was blocked with the 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 many missed cleavages up to five, whereas the tryptic peptides were with a maximum of one missed cleavage (Table 1).

In addition, a study was undertaken to identify the chemical nature of the higher molecular bands, which have strong antibody reactivity to both rabbit and patients IgE. The 65-kDa band of the SDS-PAGE was excised, destained and digested using the same protocols described above. Remarkably, the same amino acid sequence was obtained by the *de novo* sequencing, but with higher relative intensities of the peptide signals. This is the first study to conclusively demonstrate that the higher molecular IgE antibody binding protein is a dimer of the allergen TM. This is of significant importance, as several aerosolized allergenic proteins have been demonstrated in the crustacean processing industry since the 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 TM protein, another important factor in selecting a signature peptide is 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 a MW of 32 783 and 32 776 Da for both claw and leg extracts, respectively. The experimental MW is compared with the calculated molecular weight of TM (32 655 Da), which is based on the cDNA sequence,[28] and 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) cases in the UniProtKB/Swiss-Prot data bases.

The precursor ion 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 obtained report indicated the 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). 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 match between the predicted phosphorylated molecular ions and the precursor experimental molecular ion in PMF analyses. This result was further confirmed by the lack of any phosphorylated immonium ions on the product ion spectra.

All the resultant peptides produced by both trypsin and V8 enzymes were introduced to the BLAST test to find out which of the 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 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 50–70 (SQLVENELDHAQEQLSAATHK) is an ideal signature peptide (Figs 4 and 5) for TM (*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 is shown below:

SQLVENELDHAQEQLSAATHK

Conclusion

It was demonstrated for the first time that the full length of a major crustacean allergen, TM, can be sequenced utilizing a combination of ESI and MALDI using a hybrid QqToF-MS (Fig. 5).

Using a '*bottom-up*' MS approach,[31] the sequence of TM from legs and claws of the snow crab (*C. opilio)* demonstrated to have the same amino acid sequence and an average molecular weight of 32 733 Da, which is slightly higher than 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 MW difference (stated above). In addition, a higher MW protein, which is immunologically recognized by patient sera, was identified as the TM dimer. This finding is of significant importance, as previous studies on environmental allergen exposure among snow crab processors were not able to identify the nature of the generated protein allergens during processing. Based on molecular data obtained, a signature peptide of 21 amino acids was chosen to surrogate chemically the full-length SC TM. The identification of a signature peptide for the allergen TM 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 spectroscopy in characterizing minute amounts of unknown allergens, which will be vital in the identification of yet unknown snow crab allergens responsible for occupational sensitization.

Acknowledgements

This research was partially funded by a Canadian Institutes of Health Research Interdisciplinary Capacity Enhancement Grant (ICH 62328) through the SafetyNet Centre for Occupational Health & Safety Research, Memorial University of Newfoundland. We would like to acknowledge Dr Barbara Neis (SafstyNet), Memorial University of Newfoundland, and Department of Chemistry and NSERC for the financial support. Finally, the Genomic and Proteomics Facility and D. Jackman (Biochemistry) are acknowledged.

References

[1] A. L. Lopata, S. B. Lehrer. New insights into seafood allergy. *Curr. Opin. Allergy Clin. Immunol.* **2009**, *9*, 270.
[2] Food and Agriculture Organization of the United Nations (FAO). *The State of World Fisheries and Aquaculture 2006*. Fisheries and Aquaculture Department: Rome, **2007**.
[3] M. F. Jeebhay, T. G. Robins, M. E. Miller, E. Bateman, M. Smuts, R. Baatjies, A. L. Lopata. Occupational allergy and asthma among salt water fish processing workers. *Am. J. Ind. Med.* **2008**, *51*, 899.
[4] M. F. Jeebhay, T. G. Robins, S. B. Lehrer, A. L. Lopata. Occupational seafood allergy – a review. *Occup. Env. Med.* **2001**, *58*, 553.
[5] Y. Lu, T. Ohshima, H. Ushio, Y. Hamada, K. Shiomi. Immuological characteristics of monoclonal antibodies against shellfish major allergen tropomyosin. *Food Chem.* **2007**, *100*, 1093.
[6] D. R. Hoffman, E. D. Day, J. S. Miller. The major heat stable allergen of shrimp. *Ann. Allergy* **1981**, *47*, 17.
[7] Y. L. Liang, M. J. Cao, W. J. Su, L. J. Zhang, Y. Y. Huang, G. M. Liu. Identification and characterization of the major allergen of Chinese mitten crab (*Eriocheir sinensis*). *Food Chem.* **2008**, *111*, 998.
[8] M. F. Jeebahay, T. G. Robins, N. Seixas, R. Baatjies, D. A. George, E. Rusford, S. B. Lehrer, A. L. Lopata. Environmental exposure characterization of fish processing workers. *Ann. Occup. Hyg.* **2005**, *49*, 423.
[9] J. L. Malo, P. Chretien, M. Mccants, S. B. Lehrer. Detection of Snow crab antigens by air sampling of a snow crab production plant. *Clin. Exp. Allergy* **1997**, *27*, 75.
[10] B. M. Wolska, D. F. Wieczorek. The role of tropomyosin in the regulation of myocardial contraction and relaxation. *Pflugers Arch.* **2003**, *446*, 1.
[11] W. G. Lewis, L. B. Smillie. The amino acid sequence of rabbit cardiac tropomyosin. *J. Biol. Chem.* **1980**, *255*, 6854.
[12] D. H. Heeley, T. Bieger, D. M. Waddleton, C. Hong, D. M. Jackman, C. McGowan, W. S. Davidson, R. C. Beavis. Characterisation of fast, slow and cardiac muscle tropomyosins from salmonid fish. *Eur. J. Biochem.* **1995**, *232*, 226.
[13] G. Reese, S. Schicktanz, I. Lauer, S. Randow, D. Luttkopf, L. Vogel, S. B. Lehrer, S. Vieths. 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.
[14] M. Geng, J. Ji, F. E. Regnier. Signature-peptide approach to detecting protein in complex mixtures. *J. Chromatogr. A* **2000**, *870*, 295.
[15] L. Kluwe, K. Maeda, A. Miegel, S. Fujita-Becker, Y. Maeda, G. Talbo, T. Houthaeve, R. Kellner. Rabbit skeletal muscle $\alpha\alpha$-tropomyosin expressed in baculovirus-infected insect cells possessed the authentic N-terminus structure and function. *J. Muscle Res. Cell Motil.* **1995**, *16*, 103.
[16] M. Rosmilah, M. Shahnaz, A. Masita, A. Noormalin, M. Jamaludin. 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.
[17] A. R. Lorenz, S. Scheurer, D. Haustein, S. Vieths. Recombinant food allergens. *J. Chromatogr. B* **2001**, *756*, 255.
[18] A. M. Cohen, A. A. Mansour, J. H. Banoub. '*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.
[19] J. H. Banoub, A. M. Cohen, A. A. Mansour, P. Thibault. 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.
[20] M. Kinter, N. E. Sherman. *Protein Sequencing and Identification Using Tandem Mass Spectrometry*. John Wiley: New York, **2000**.

[21] A. L. Lopata, M. F. Jeebhay, G. Reese, J. Fernandes, B. Fenemore, A. Elliott, T. G. Robins, S. B. Lehrer. Detection of fish antigens aerosolized during fish processing using newly developed immunoassays. *Int. Arch. Allergy* **2005**, *138*, 21.
[22] K. Sergeant, B. Samyn, G. Debyser, J. V. Beeumen. '*De novo*' sequence analysis of N-terminal sulfonated peptides after in gel quanidination. *Proteomics* **2005**, *5*, 2369.
[23] M. C. Huang, Y. Ochiai. Fish fast skeletal muscle tropomyosins show species-specific thermal stability. *Comp. Biochem. Physiol. B* **2005**, *141*, 461.
[24] E. Hoffmann, V. Stroobant. *Mass Spectrometry: Principles and Applications*. John Wiley: New York, **2007**.
[25] D. J. Pappin, P. Hojrup, A. J. Bleasby. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **1993**, *3*, 487.
[26] A. Cartier, J. L. Malo, F. Forest, M. Lafrance, L. Pineau, J. J. St-Aubin, J. Y. Dubois. Occupational asthma in snow crab-processing workers. *J. Allergy Clin. Immunol.* **1984**, *74*, 261.
[27] B. V. Gill, T. R. Rice, A. Cartier, D. Gautrin, B. Neis, L. Horth-Susin, M. Jong, M. Swanson, S. B. Lehrer. Identification of crab proteins that elicit IgE reactivity in snow crab – processing workers. *J. Allergy. Clin. Immunol.* **2009**, *124*, 1055.
[28] K. Motoyama, Y. Suma, S. Ishizaki, Y. Nagashima, K. Shiomi. Molecular cloning of tropomyosins identified as allergens in six species of crustaceans. *J. Agric. Food Chem.* **2007**, *55*, 985.
[29] L. Kiemer, J. D. Bendtsen, N. Blom. NetAcet: Prediction of N-terminal acetylation sites. *Bioinformatics* **2005**, *21*, 1269.
[30] N. Blom, S. Gammeltoft, S. Brunak. Sequence and structure-based prediction of Eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **1999**, *294*, 1351.
[31] A. M. Cohen, F. Jahouh, S. Sioud, R. M. Rideout, M. J. Morgan, J. H. Banoub. Quantification of Greenland halibut serum vitellogenin: a trip from the deep sea to the mass spectrometer. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 1049.