UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

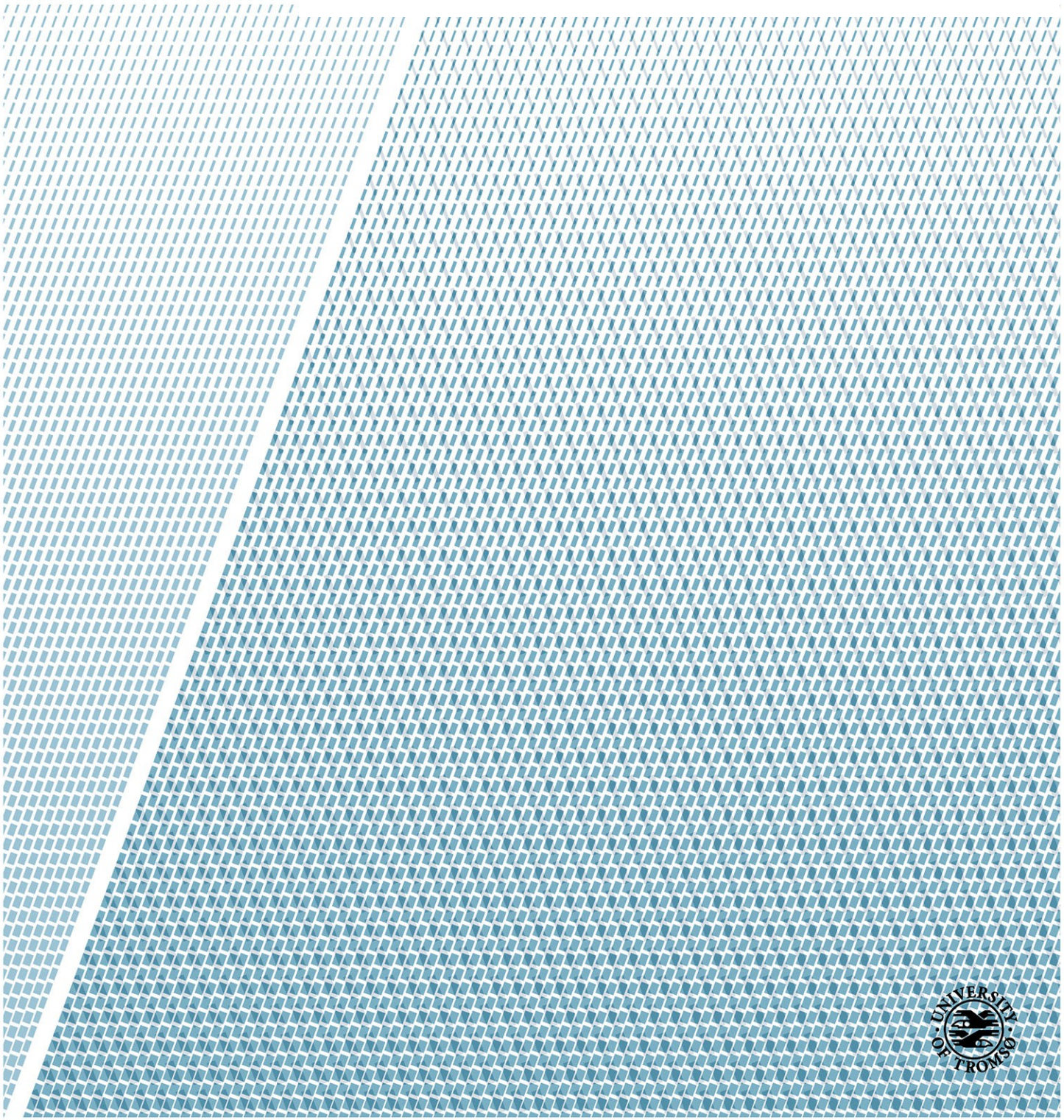
Department of Pharmacy

De Novo Sequencing of Disulphide Rich Peptides Using Tandem Mass Spectrometry

—

Regasa Chala

Master thesis in Pharmaceutical science, May 2015



Acknowledgements

First of all i thank the almighty God who helped me in all up and downs of my life, and gave me the patience and resilience in the past 5 years.

This work wouldn't be possible without the intellectual inputs of my advisors; Dr. Terje Vannskog, Dr. Terkel Hansen and Dr. Anstetesia Albert.

I thank Dr. Terje Vannskog not only for his expert guidance but also for his positivity and kindness. I thank you for making yourself available to me all the time.

My gratitude also goes to Dr. Terkel Hansen for patiently sitting down with me and sharing your expertise in *de novo* sequencing.

I am also grateful to Dr. Anastasia Albert for always available and showing me all the dos and the don'ts of peptide mass spectrometry in the last 6 months.

Last but not least, I want to thank Northern research institute for letting me use their laboratory facilities and disposing office during my stay.

Abstract

Peptide based drugs have been shown to have better target selectivity and thereby better safety profile than many small molecule drug candidates. This led many researchers and drug companies to show an increasing interest toward bioactive peptides. Developing an effective analytical method for peptide /protein structural identification is a cornerstone to meet this expanding area of research in this field.

De novo sequencing is a method of peptide structural identification without using peptide databases. Compared to the other *de novo* sequencing methods, mass spectrometric peptide sequence analysis is superior in terms of sensitivity and speed.

Peptide structural information is contained in the spectrum produced by tandem mass spectrometers MS/MS. The quality of the data contained in MS/MS spectra depends on the amino acid content of the peptide, the instrumentation and methods used.

One commonly used peptide fragmentation method is collision induced dissociation (CID) where the peptide collides with an inert gas. Depending on the amount of energy applied during fragmentation, peptides yield a wide array of product ions. Peptides of known amino acid sequence were used to determine the relationship between precursor ion mass-to-charge-ratio (*m/z)* and collision energy needed for optimal peptide fragmentation. The energy seems to increase with increasing *m/z* of the precursor ion.

A comparison of CID fragmentation methods in a quadrupole-time-of-flight (QTof) instrument revealed the merit of using a collision energy ramp compared to single collision energy.

Evaluation of *de novo* sequencing accuracy of PEAKS software showed that the program works well for smaller peptides but the accuracy deteriorates with increasing peptide length.

The amino acid sequence of an unknown peptide was determined using the PEAKS. The software has identified 84 % the amino acid sequences correctly. The rest 16 % of the residues were confirmed by Edman degradation. The peptide contained three disulphide bridges and identifying cysteine interconnectivity of the peptide is a gap to be addressed in the future.

Key words: Collision energy, ramp, PEAKS, de novo sequencing.

TABLE OF CONTENT

.

ABBREVATIONS

CE	Collision energy
CID	Collision induced dissociation
CV	Cone voltage
MS	Mass spectrometry
MS/MS	Tandem mass spectrometer
m/z	Mass to charge ratio
NCFC	Norwegian College of Fishery Science
NCM	N-Cyclohexylmaleimide
NMM	N-Methylmaleimide
NPM	N-Phenylmaleimide
% ALC	Percent average local confidence
ppt1	De novo sequenced peptide (unknown peptide)
ppt2	Another unknown peptide
PTM	Post translational modification
QToF	Quadrupole- Time-of- flight
SPE column	Solid phase extraction column
TCEP	Tris(2-carboxyethyl)phosphine

LIST OF TABLES

LIST OF FIGURES

1. Background

Peptide-based therapeutics sector has expanded at the expense of small-molecule drugs in recent years. This trend was assisted by the newly popularized availability of recombinant protein expression, better protein purification protocols and analysis tools, and the realization that peptide based drugs has exquisite potency and selectivity for their molecular targets (Góngora-Benítez *et al*. 2014, Craik *et al.* 2012).

Disulfide bridges are posttranslational modifications of peptides and proteins that play a pivotal role in the folding and stabilization of their bioactive structures (Morder, 2005). Disulphide rich peptide venoms from animals such as snakes, spiders, scorpions and certain marine snails represent one of nature's great diversity libraries of bioactive molecules (Ueberheide *et al.* 2009). This biodiversity can be exemplified by group of small peptides called conopeptides, which are widely distributed in the venom of marine mollusks, fish and worm-hunting cone snails. It is estimated that more than 50,000 varieties of conopeptides exist, out of which only less than 0,1 % of it is pharmacologically characterized (Olivera, 2006, Lewis and Garcia , 2003).

Bioactive peptides have stimulated considerable interest because of their ability to potentially alter the function of mammalian ion channels and receptors including nicotinic acetylcholine receptors, noradrenaline transporters, sodium and calcium channels (Anand *et al.* 2014, Lewis and Garcia, 2003). Synthesis and structure activity (SAR) studies of these natural peptides are presently central to the interest of medical chemistry (Moroder *et al.* 2005).

One of the barriers which limit the utilization of this rich resource of bioactive peptides has been the difficulty in elucidating their primary structure which ranges in size between 10 and 80 amino acids (Ueberheide *et a.l* 2009). Captopril, an angiotensin-converting enzyme (ACE) inhibitor, was the first venom-based drug isolated from the Brazilian viper *Buthrups jararaca*. The isolated peptide, teprotide, was *de novo* sequenced using the Edman degradation method (Ondetti, 1971). Ziconotide (calcium channel blocker), bivalirudin (thrombin inhibitor) and exenatide (GLP-1 receptor antagonist) are other success stories of peptide based drugs (King and Glenn, 2013 Craik, 2012).

Marine bioprospecting, the systematic search for novel compounds from natural sources in the marine environment, has increased rapidly in recent years (Demunshi and Chugh, 2009).

In line with this trend the Norwegian government has launched a national strategy entitled ‘Marine bioprospecting—a source of new and viable wealth creation’’ aimed at promoting research within marine natural products and drug discovery. This led to more and more Norwegian scientists to be involved in the search for bioactive peptides (Svenson, 2013).

In Tromsø, research groups working with marine bioprospecting have identified several peptides with antimicrobial activity. These peptides are usually sent to other laboratories to be sequenced by the Edman degradation method, which is relatively time consuming and needs large amounts of sample compared to the MS/MS method. Therefore in order to facilitate the growing interest in marine bioactive research, there is a need to locally develop an effective analytical method for peptide structural elucidation.

2. Mass spectrometric peptide sequence analysis

De novo sequencing is a method of elucidating peptide primary structure without using protein databases (Standing, 2003) where the mass difference between two adjacent ions is used to deduce the amino acid sequence of a peptide (Roepstorff, 1984 Tannu and Hemby, 2007). In the last two decades, most mass spectrometric based protein identification have been focusing on searching spectra from mass spectrometry (MS/MS) against protein data bases (Liu *et al*,2014, Westermeier and Naven, 2002, Perkins *et al*. 1999).

With the ever-increasing number of complete genomes published, one might think there is less need for *de novo* protein/peptide sequencing. However, protein prediction from genome is partly based on availability of genomic sequence from the organism of interest or at least from closely related species. But the fact that the genetic information of the vast majority of organism is not yet discovered and unpredictability of some post-translational modifications make *de novo* sequencing as relevant as it has been (Liu *et al*, 2014, Medzihradsky and Bohlen 2012, Zhang *et al*, 2003, Standing, 2003).

Early *de novo* sequencing relied on Edman degradation, which combines derivatisation of the N-terminal amino acid of a peptide or protein with the subsequent cleavage of the derivatised residues. These two steps repeat for each amino acid through the peptide sequence. This process is time consuming (one or two peptides per day), work-intensive and needs an unmodified N-terminal of peptides. Mass spectrometry (MS) has reduced the need for this

technique because it is fast, more sensitive and not affected by N-terminal modifications (Westermeier and Naven, 2002). However, the technique is not new since in 1986 Hunt and colleagues already described *de novo* sequencing for the first time, where tryptic peptides of apolipoprotein B were successfully sequenced using FAB-triple quadrupole mass spectrometry (Hunt *et al.* 1986). Converting the peptides to gaseous form in the mass spectrometer without decomposing the molecules has been the major analytical bottleneck (Kinter and Sherman, 2000). Until 1970s, mass spectrometric analysis of organic compounds utilized electron impact (EI) ionization methods. In this technique, a radical cation is formed from the evaporated sample by expelling an electron. EI is not well suited for analysis of polar, in volatile and thermally labile biomolecules (Baldwin, 2005).Introduction of the 'soft ionization' methods, electrospray ionization (ESI) and matrix-assisted laser desorption (MALDI) made mass spectrometry an indispensable tool for protein and peptide analysis (Seidler *et al*. 2010, Baldwin, 2005). These ground breaking ionization methods has earned the respective inventers, Koichi Tanaka and John Fenn, the Nobel Prize of 2002.

In ESI, ionized peptides are formed by spraying diluted solution of the analyte at atmospheric pressure from the tip of a fine capillary held at a high electric potential (Baldwin, 2005 Fenn, J 2002). In the high voltage capillary, an electrochemical reaction of the solvent leads to the formation of a charged droplet (Kebarle and Tang, 1993). The charged droplet will leave the nozzle by electrostatic repulsion, and the solvent evaporate while travelling towards the low pressure area of the MS to create a continuous stream of gaseous charged analyte which will enter the vacuum system (Baldwin, 2005 and Kebarle and Tang, 1993). A schematic illustration of the electrospray ionization process is shown in figure 1.

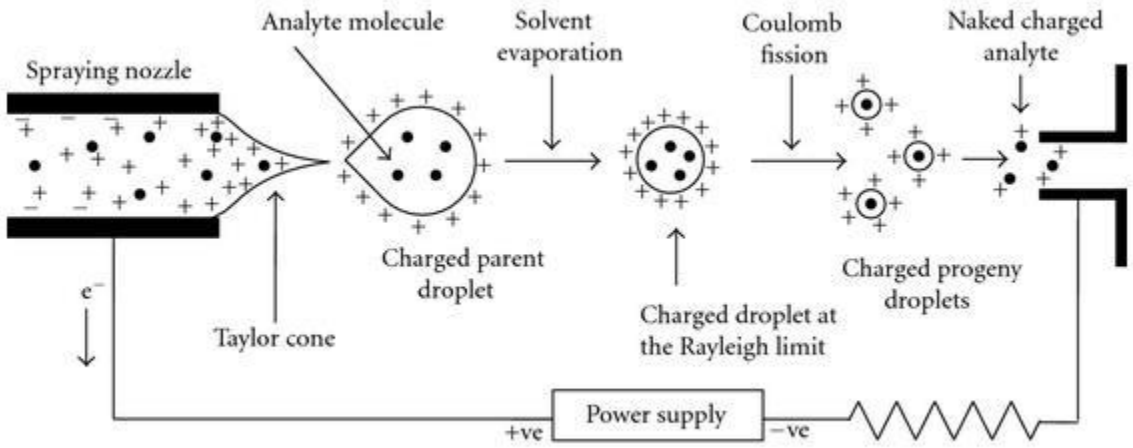


Figure 1: Schematic representation of the electrospray ionization process (Banerjee and Mazumar, 2011). While moving toward the low pressure area of MS, the ionized droplets decrease in size because of the solvent evaporation. Explosion of the droplet due to increasing charge concentration on its surface leads to formation of smaller droplets from which further solvent evaporation takes place to finally form charged analyte molecules.

Electrospray ionization process does not impart significant amount of energy that can lead to unimolecular dissossiation of the ion. But within an intermediate pressure region of the ESI is a plate with an aperture commonly referred to as the *skimmer* or *nozzle* which accelerate the ionized peptides toward low pressure area of the MS. Higher voltage applied to this plate can impart high kinetic energy to the ion. When these ions collide with molecules in the air within the intermediate pressure area of ESI, the kinetic energy will be converted to vibrational energy leading to fragmentation of the peptide (Baldwin, 2005).

The LC component, which is directly connected to MS, separates the peptides and preventing possible sample losses during off-line preparative fractionation (Zhang *et al*, 2014). For positive ionization analysis, typical solvents are acetonitrile or methanol acidified with 1% acetic acid or 0.1 % formic acid to ensure the unfolding of proteins and extensive protonation of the most basic sites (Baldwin, 2005).

The ionized peptide is directed to the mass analyzer where the ions are sorted according to their *m/z* ratios. The quadrupole mass filter (Q), the ion trap and the time-of-flight are the most commonly used mass analyzers in proteomics (Kinter and Sherman, 2000).

Tandem mass spectrometers (MS/MS) have two mass analyzers where the first mass analyzer is used to isolate an ion with a specific *m/z* (precursor ion) for further fragmentation, and the second mass analyzer determines the *m/z* of the product ions formed from the fragmentation of the precursor ion (Matthiesen, 2007 McLafferty, 1981) and can therefore used to obtain structural information.

The use of quadrupole mass analyzer as MS1 and an RF-only quadrupole collision cell in combination with an Orthogonal acceleration time-of-flight as MS2 (Figure 2) gives a very powerful combination in terms of sensitivity, resolution and mass range (Chernushevich *et al*, 2001, Baldwin, 2005).

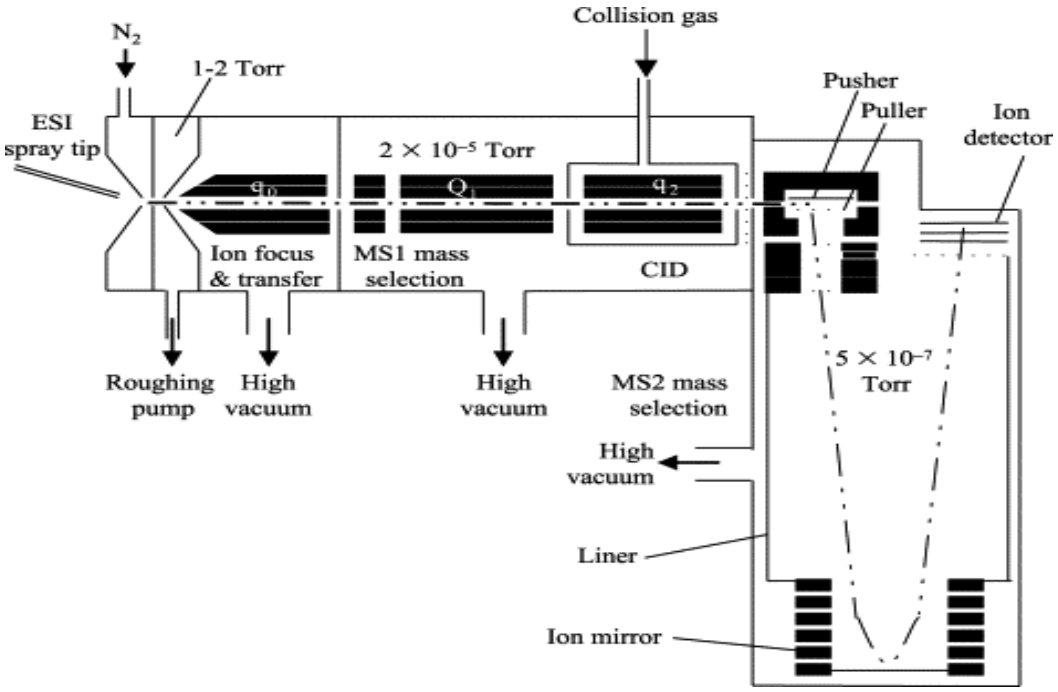


Figure-2: A Quadrupole-time–of-flight (QToF) tandem mass spectrometer (Baldwin, 2005). (q0): RF-only quadrupole which is used to focus the ionized analyte thereby hindering intermolecular collision (Q): Serves as mass filter to select precursor ion of interest for further analysis (q2): Collision cell where the ionized analyte collide with inert gases. The product ions are accelerated towards an ion mirror which in turn reflects them back to the ion detector.

The collision energy usually utilized in MS/MS can roughly be classified as low energy CID (< 100 eV) and high energy CID involving up to kilo electron volt kinetic energy (Seidler, 2010 Wells, 2005). Low-energy collision-induced dissociation (CID) in mass spectrometry has been used extensively in peptide sequencing and analysis of post translational modifications (PTM) (Zhang *et al*. 2014 Seidler,2010).

As shown in the equations below, collisional activation (1) and unimolecular dissociation (2) are the two phenomenon assumed to be occurring in CID.

$$Q + m_p^+ + N \rightarrow m_p^{+*} + N' \qquad (1)$$

where Q is the change in kinetic energy of the system, m_p^+ and N are the precursor ion and target in their pre-collision states, and m_p^{+*} and N' are their partners in their post-collision state.

$$m_p^{+*} \rightarrow m_a^+ + m_b \qquad (2)$$

m_a^+ and m_b are products of the unimolecular dissociation of m_p^+ (McLuckey, 1992).

3. Peptide fragmentation in MS/MS

Following collisional activation, the site of protonation directs fragmentation reaction that occurs as a means of releasing the excess internal energy added to the peptide ion by the collision (Kinter and Sherman, 2000). The fragment ions in CID fragmentation are produced primarily by cleavage of the amide bonds that join two amino acids. The analysis of these fragments provides sequence information that can be used for *de novo* sequencing (Medzihradszky, 2005). The 20 common amino acids along with their codes and mass are given in table 2.

The nomenclature of product ions was given by Roepstorff and Fohlman for the first time in 1984 (Roepstorff and Fohlman, 1984). Fragments will be detected only if they carry charge. If this charge is retained by the N- terminus fragment, the ion is classified as either a, b or c. If the charge is retained on the C- terminus, the ion type is either x, y or z as illustrated on figure 3

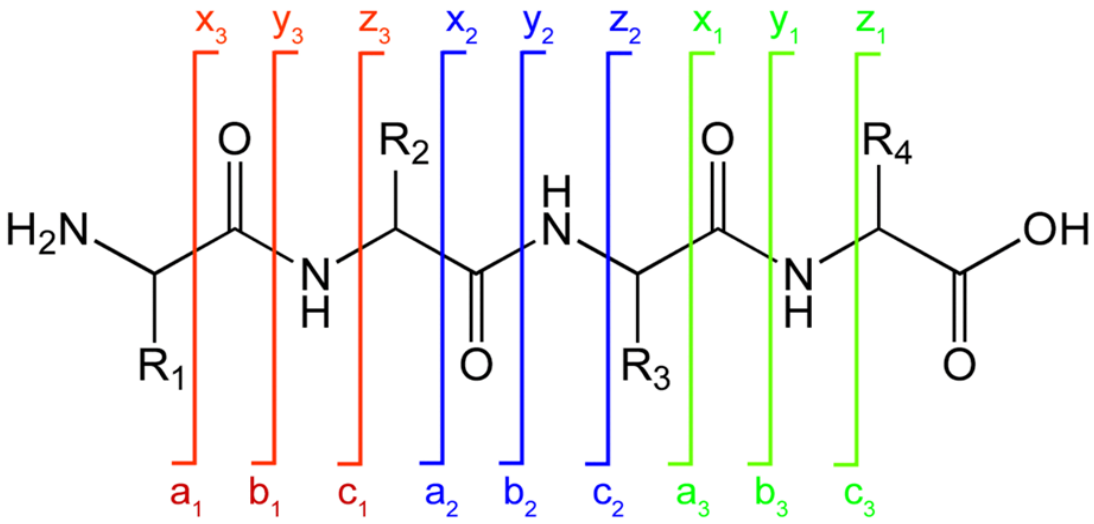


Figure 3: Nomenclature of product ions (Roepstorff and Fohlman, 1984). 'Ri' represents the amino acid side chains of the peptide. Product ions which are formed are termed a, b or c if the charge is retained on N-terminus and x, y or z if the charge is retained on the C-terminus.

Generally b, a, y and immonim ions (table-1) are more common in lower energy multistep activation spectra, while higher energy activation can additionally lead to the formation of x,y and z ions (Wysocki *et al*. 2005).). Beside these fragment ions, CID spectra contain neutral losses from certain residues of the peptide that can provide sequence information (Seidler *et al.* 2010).

In positive-ion operating conditions, electrospray ionization produces peptide ions that enter the mass spectrometer with protons attached to the basic sites. These sites include the N-terminus amine group, the amide bond and the more basic side group of lysine, arginine, or histidine residues (Kapp *et al.* 2003).

Protons associated with side chains of basic residues are strongly attached and remains fixed at that site even on collisional activation unlike protons on less basic N-terminus which may move by solvation to any of the amide linkages (Matthisen, 2007 Wysocki *et al*. 2000). This produces a heterogeneous population of the peptide where the proton is associated with an amide bond at different sites. This migration of protons along the peptide backbone makes

fragmentation reaction to occur at different amide bonds and this process is known as the mobile proton effect (Kinter and Sherman, 2000 Wysocki *et al*. 2000).

QToF mass spectrometers typically yields a series of y-ions throughout the mass range, while b-ions are thought to be less stable and fragments further. As a result b-ions are more abundant at lower masses and are often absent at higher *m/z* in the MS/MS spectrum (Katalin *et al.* 2013).

The accuracy of peptide identification using MS/MS depends on the mass resolution and mass accuracy of the instrument, the completeness of the observed fragment ion series and the extent to which the fragmentation spectrum can be correctly interpreted (Zhang *et al*. 2003).

Table 1: Immonium and related ions characteristic of the 20 common amino acids

Amino Acid	Immonium and related ion(s) masses		Comments
Ala	44		
Arg	129	59, 70, 73, 87, 100, 112	129, 73 usually weak
Asn	87	70	87 often weak, 70 weak
Asp	88		Usually weak
Cys	76		Usually weak
Gly	30		
Gln	101	84, 129	129 weak
Glu	102		Often weak if C-terminal
His	110	82, 121,123, 138, 166	110 very strong 82, 121, 123, 138 weak
Ile/Leu	86		
Lys	101	84, 112, 129	101 can be weak
Met	104	61	104 often weak
Phe	120	91	120 strong, 91 weak
Pro	70		Strong
Ser	60		
Thr	74		
Trp	159	130, 170, 171	Strong
Tyr	136	91, 107	136 strong, 107, 91 weak
Val	72		Fairly strong

Medzihradszky, 2012

Table 2: Codes and monoisotopic masses of the 20 common amino acids

Name			**Elemental composition**	**Residue mass**
Full	3 letter code	1 letter code	Neutral molecule	(Monoisotopic)
Alanine	Ala	A	C3H7NO2	71.0372
Arginine	Arg	R	C6H14N4O2	156.1011
Asparagine	Asn	N	C4H8N2O3	114.0429
Aspartic acid	Asp	D	C4H8NO4	115.0269
Cysteine	Cys	C	C3H7NO2S	103.0092
Glutamic acid	Glu	E	C5H9NO4	129.0426
Glutamine	Gln	Q	C5H10N2O3	128.0586
Glycine	Gly	G	C2H5NO2	57.0215
Histidine	His	H	C6H9N3O2	137.0589
Isoleucine	Ile	I	C6H13NO2	113.0841
Leucine	Leu	L	C6H13NO2	113.0841
Lysine	Lys	K	C6H14N2O2	128.0949
Methionine	Met	M	C5H11NO2S	131.0405
Phenylalanine	Phe	F	C9H11NO2	147.0684
Proline	Pro	P	C5H9NO2	97.0528
Serine	Ser	S	C3H7NO3	87.0320
Threonine	Thr	T	C4H9NO3	101.0477
Tryptophan	Trp	W	C11H12N2O2	186.0793
Tyrosine	Tyr	Y	C9H11NO3	163.0633
Valine	Val	V	C5H11NO2	99.0684

Source: Medzihradszky, 2012

4. PEAKS : A software for *de novo* sequencing

Manual deduction of amino acid sequences of a peptide from MS/MS spectra is tedious and time consuming. In order to address this, a number of algorithms and software packages were developed to interpret the data obtained from MS/MS experiments (Pevtsov *et al.* 2006 Ma *et al*. 2003).

PEAKS is the most popular software for *de novo* sequencing showing the best accuracy among all currently available program packages (Pevtsov *et al.* 2006). The software assigns two scores $f_y(m)$ and $f_b(m)$ for each mass m corresponding to a y-ion and a b-ion respectively. If the there is a strong intensity peak at mass m (or close to m and within the mass error tolerance), then the score is positive. In addition neutral losses that are possibly generated are also taken in to account to compute the two score functions (Hughes et al., 2010).

The software assigns local confidence score for each amino acid which is expressed as percent average local confidence (% ALC). The *de novo* sequencing results can be filtered by operator specified % ALC threshold to remove low quality sequences.

5. Aims of the thesis

Main goal

The main objective of this study is to develop methods for sequencing disulfide rich peptides from marine organisms using tandem mass spectrometer.

Sub-goals

In order to achieve this goal, we will investigate effects of some mass spectrometric parameters (cone potential and collision energy) on peptide fragmentation using model peptides of known amino acid sequences.

Peptides having various amino acid sequence length, amino acid composition and charge state will be fragmented at different collision energies to find the relationship between the effects of these variables on the choice of an optimal collision energy. The single collision energies will also be compared with collision energy ramps in terms of the sequence information obtained from the MS/MS spectra. In addition the effect of cone voltage on precursor ion intensity will also observed briefly.

The *de novo* sequencing potential of PEAKS studio 7 will also be evaluated using peptides of known amino acid sequence.

Finally an unknown bioactive peptide obtained from Norwegian College of Fishery Science will be *de novo* sequenced.

6. MATERIALS AND METHODS

6.1 Peptides

Conotoxin-α (GCCSDPRCAWRC-NH2), tertiapine (ALCNCNRIIIPHMCWKKCGKK-NH2) and neurotoxin (RSCCPCYWGGCPWGQNCYPEGCSGPKV-NH2) were produced locally while endothelin-2 (CSCSSWLDKECVYFCHLDIIW), orexin-A (pEPLPDCCRQKTCSCRLYELLHGAGNHAAGILTL-NH_2 , where pE stands for L-pyroglutamic acid) and defensin HNP-1 (ACYCRIPACIAGERRYGTCIYQGRLWAFCC were purchased from Bachem AG, Bubendorf, Switzerland.

6.2 Chemicals

The reducing agent (tris(2-carboxyethyl)phosphine(TCEP) and alkylating agents N-Cyclohexylmaleimide (NCM), N-Methylmaleimide (NMM) and N-Phenylmaleimide(NPM) were obtained from Sigma Aldrich®, Missouri, USA.

6.3 Equipment

The Waters (Milford, Connecticut, USA) Xevo™ G2 QTof Mass Spectrometer (MS) was connected to a Waters ACQUITY UPLC I-class operated by Waters® Masslynx v4.1 software was used for the MS and MS/MS experiments. The peptides were reduced and alkylated on Incubating microplate shaker from VWR®, Radnor, Pennsylvania, USA.

Analytic scale from Sartorious, Goettingen, Germany was used for weighing of samples. C18 (Octadyl) Standard density SPE column from EMpore™, Minnesota, USA was used as a reaction medium when the peptides were selectively reduced and alkylated.

An online tool, MS-product, was used to generate theoretical *m/z* of product ions http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct

De novo sequencing was done using PEAKS® studio 7 software from Bioinformatics Solutions Inc. Waterloo, Canada.

7. METHODS

7.1. LC methods

The UHPLC was run by injecting 5 µl (except for the unknown peptide where 10 µl was used) sample on to an Acquity CSH C18 (150 x 2.1 mm, 1.7 µm particle size) column (Waters, Milford, MA, USA) at a column temperature of 50 ºC. The gradient LC method used mobile phases A and B at a flow rate of 0.5 ml/min, where A was 0.1 % formic acid in water and B was 0.1 % formic acid in acetonitrile.

The mobile phase gradient was programmed as follows: 0-6.0 min: 2 – 60% B, 6.0 - 6.1 min: 60-95% B and 6.10-8.00 min, 95-95 % B.

7.2. MS methods

Leucine Enkephalin was used as lock spray (*m/z* 278.1141 Da and 556.2710 Da) with 1sec scan time at 15 sec interval. A lock spray contains a compound of known composition used by the MS as a reference to avoid a systematic drift in the mass measurement (Cox *et al*, 2011). A cone voltage of 35 V was used throughout the experiments except where the effect of cone voltage on the precursor ion intensity was tested. Collision energies used are described under respective experiments below, other MS parameters used for all experiments are described in table-3.

Table-3: MS Conditions used for all experiments.

MS Parameters	Values
Spray Capillary voltage (V)	600
Source temperature (ºC)	130
Desolvation gas flow rate (l/h)	800
Desolvation temperature (ºC)	350
Sample scan time (sec)	0.5
Collision gas	Argon
Ion source polarity	Positive
Mass range (Da)	100-5000
Extraction cone	113.2
Cone gas flow rate (l/h)	10
Analyzer mode	Resolution
Dynamic range	Normal
Data format	Centroid

7.3. Reduction of the peptides

In all cases, the model peptides were reduced by incubating 3 µl 0.5 mM peptide with 10 µl 1 M TCEP and 174 µl 50 mM ammonium formate buffer (pH=3). The reduced peptides were fragmented at collision energies ranging from 20-45eV and sequence ions (b-ions and y-ions) identified manually using the online tool, MS-Product.

The peptides were sequenced by searching the theoretical monoisotopic ions from the MS/MS spectrum of the peptide of interest. Ions of the same charge state and having *m/z* deviation less than 0.01 Da and intensity higher than $1.00e^3$ were considered as positive. Where possible, neutral losses corresponding to y-ions and b-ions were used to verify the findings.

7.4. Effect of cone potential on precursor ion intensity

In order to determine the effect of cone potential on precursor ion(s) intensity and fragmentation pattern, reduced conotoxin-α was subjected to cone voltages (CV) of 20, 25, 30 and 35 V. After comparing the spectra produced, precursor ions formed at CV 20V and CV 35V were fragmented and the peptide sequenced as described under section 7.3.

7.5. Identifying optimal collision energy (CE) for effective peptide fragmentation

Peptides of known amino acid sequences were used to establish the relationship between mass–to-charge ratio (*m/z*) of precursor ions and collision energies needed for efficient peptide fragmentation.

Conotoxin-α, tertiapine, endothelin-2, neurotoxin, defensin HNP-1 and orexin-A were used as model peptides. The precursor ions with highest charge state were chosen and fragmented at collision energies ranging from 20-45 eV. The sequence ions produced by the different CEs were manually identified using the online tool, MS-product.

For each sequence ion; *m/z*, charge state and intensity was noted. The collision energy which produced maximum number of sequence ions were considered as the optimum collision energy for that particular peptide.

7.6. Comparison of collision energy ramp with optimal single collision energy

Conotoxin-α, neurotoxin, endothelin-2 and tertiapine were fragmented at collision energy ramps of 20-30, 25-35, 25-40 and 30-40 eV. The collision energy ramp that produced the maximum number of sequence ions was considered as the optimum collision energy ramp for the peptide. In parallel, the peptides were also fragmented with their respective optimal collision energies identified in previous experiments. The optimal collision energies and optimal ramps were compared based on the number of sequence ions identified.

7.7. Sequence identification of selectively alkylated peptides with two disulphide bridges

In conotoxin-α (GCCSDPRCAWRC-NH2), C2 is connected to C12 and C3 is connected to C8. The two disulphide bridges of conotoxin-α was selectively alkylated by NMM/ NCM and NMM/NPM on SPE column following a procedure developed in our laboratory (unpublished data). The selective alkylation process has two steps; in the first step one of the disulphide bridges is reduced and subsequently alkylated by one of the alkylating agents. In the second round the other crosslinking cysteine residues will be reduced and modified by the second to alkylating agent. The final result will be the formation of two structural isomers of the modified peptide.

The peptide was fragmented and sequenced to differentiate the two structural isomers. The schematic representation of this process using NMM/NCM as an example is given in figure 4 below.

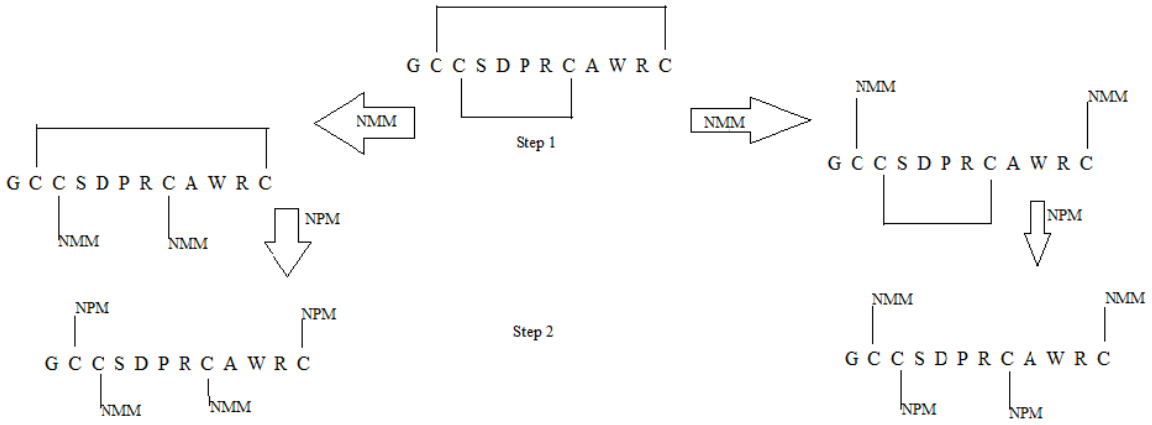


Figure 4: Schematic representation of selective alkylation of conotoxin-α alkylated by NMM and NPM. Step 1: Either C3&C8 will be alkylated by NMM or C2 and C12 will be alkylated by NMM. Step 2: C2&C12 will be alkylated by NPM and C3 &C8 will be alkylated by NPM to form the structural isomers.

In order to induce theoretical fragmentation on MS-product C2 & C12, which are connected to each other, were labeled by a letter (u) and C3&C8 are labeled by v.

7.8. Evaluating the accuracy of PEAKS Studio 7 software

In order to evaluate the accuracy of the PEAKS software, peptides of known amino acid sequence having different precursor ion *m/z*, charge state and peptide sequence length was selected. conotoxin-α (12 residues), endothelin (21 residues), neurotoxin (27 residues) and defensin HNP-1 (30-residues) were selected.

Prior to data entry, spectra having higher intensity of product ions in all *m/z* areas were chosen. In all the cases, precursor ion and fragment ion error tolerance of 0,005 Da was used.

After the *de novo* sequencing has run on the software, theoretical fragmentation was induced on MS-product for the suggested peptide sequence. Manual evaluation of the result was done by searching the theoretical fragment ions in the spectrum. Only amino acid sequence candidate with the highest % ALC was used for manual evaluation of the sequencing result.

7.9. *De novo* sequencing of the unknown peptide

The unknown peptide for *de novo* sequencing was obtained from The Norwegian College of Fishery Science. After dissolving the sample in 50mM ammonium formate (pH=3), the molecular weight of the peptide was identified using MS.

The peptide was incubated with freshly prepared 1 M TCEP to reduce the peptide. After the numbers of disulfide bridges were identified, a precursor ion was selected and fragmented by single collision energies of 35 and 40eV and collision energy ramps of 25-45 and 35-55 eV. The product ions of the spectra in all *m/z areas* were compared based on their intensities. The spectrum of good quality (higher product ion intensities in most of the areas) was chosen for *de novo* sequencing.

The peptide was sequenced using PEAKS by choosing precursor ion and product ion mass error tolerance of 0.005 Da. The result was carefully validated by searching theoretical fragment ions of the suggested sequence in the MS/MS spectrum.

Neutral losses and mass deviation were also taken in to consideration in verifying the correctness of the sequences.

8. RESULTS AND DISCUSSION

8.1 Effect of cone voltage on precursor ion intensity

Mass spectra of conotoxin-α at cone voltage (CV) of 35 V revealed intact peptide ion, $[M+2H]^{2+}$= 678.2632, and two fragment ions [y_7^{2+} = 445.7216 and y_{10}^{2+} = 598.2498] with significant intensities. After the CV was reduced to 20 V, the charge state of the precursor ion was shifted more towards $[M+3H]^{3+}$= 452.5224 and fragmentation of the peptide was found to be minimal as shown in figure-5 below.

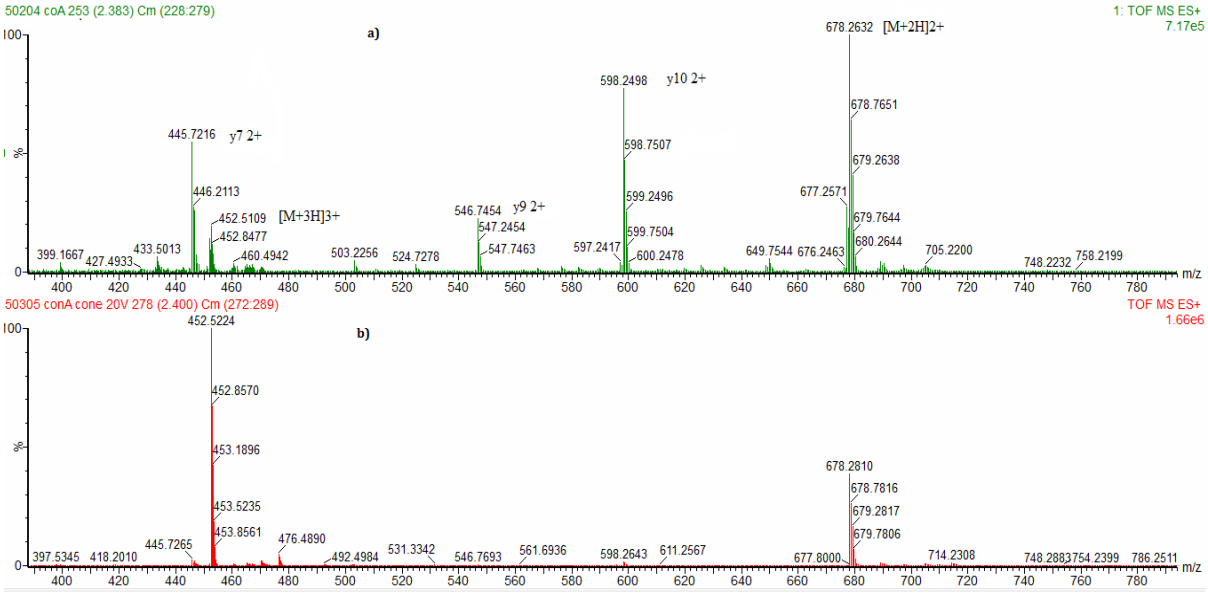


Figure 5: Mass spectra of conotoxin-α at CV 35 V and 20 V. a) Spectrum at CV 35 V showing low intensity of $[M+3H]^{3+}$ and relatively high intensity of $[M+2H]^{2+}$ and fragment ions. b) At cone voltage 20 V, the intensity of $[M+3H]^{3+}$ has significantly increased and the fragment ions disappeared.

As discussed earlier in the theory part, the potensial that accelerates the ionized peptides towards the low pressure area of the mass spectrometer can cause fragmentation of the peptide (Balwin, 2005). Similarly,the decreasing charge state of the ionized peptide wih increasing CV is probabily due the collision between the ionized peptide and molecules in the air within intermediate pressure area of the mass spectrometer which can strip off the losely bound protons.

The precursor ions, $[M+2H]^{2+}$ = 678.2632 and $[M+3H]^{3+}$ = 452.5224, obtained at CV of 35V and 20 V, respectively were fragmented at collision energies ranging from 20 eV to 40 eV. The ion $[M+3H]^{2+}$ produced fewer sequence ions at CE under 30 eV as shown in table 4.

Table-4: Observed sequence ions of Conotoxin-α fragmented at CV 20V and 35 V and CE between 20eV and 40 eV. At CV 35 and collision energies between 30 to 40 eV, sequence information about S and D was not found because of the missing y_8/b_4-ions. Reduction of the cone voltage to 20 V led to the appearance of the y_8^{2+} ion at 20eV.

Conotoxin-α $[M+2H]^{2+}$ = 678.2811 CV 35V

25eV	30 eV	35 eV	40eV	b-ions		y-ion	25eV	30 eV	35 eV	40eV
				1	G	12				
	✓(+1)	✓(+2)	✓(+1)	2	C	11				
	✓(+1)	✓(+2)	✓(+1)	3	C	10				
				4	S	9				
✓(+1)				5	D	8				
				8	P	7	✓(+2)	✓(+2)	✓(+2)	✓(+1)
				7	R	6		✓(+2)	✓(+2)	✓(+1)
				8	C	5	✓(+1)	✓(+1)	✓(+1)	✓(+1)
	✓(+1)			9	A	4		✓(+1)	✓(+1)	✓(+1)
	✓(+1)			10	W	3		✓(+1)	✓(+1)	✓(+1)
✓(+2)	✓(+2)			11	R	2		✓(+1)	✓(+1)	✓(+1)
				12	C	1		✓(+1)	✓(+1)	✓(+1)

Conotoxin-α $[M+3H]^{3+}$ = 452.5218 CV 20V

20 eV	25 eV	30 eV	b-ions		y-ions	20 eV	25 eV	30 eV
			1	G	12			
✓(+1)	✓(+1)	✓(+1)	2	C	11	✓(+2)		
✓(+1)	✓(+1)	✓(+1)	3	C	10	✓(+2)	✓(+2)	
✓(+1)			4	S	9	✓(+2)		
✓(+1)			5	D	8	✓(+2)		
			6	P	7	✓(+2)	✓(+2)	✓(+2)
			7	R	6	✓(+2)	✓(+2)	✓(+2)
			8	C	5	✓(+1)	✓(+1)	✓(+1)
			9	A	4	✓(+1)	✓(+1)	✓(+1)
			10	W	3	✓(+1)	✓(+1)	✓(+1)
			11	R	2	✓(+1)	✓(+1)	✓(+1)
			12	C	1	✓(+1)	✓(+1)	✓(+1)

In charged R containing peptides, protons are tightly bound to the basic side chain of the amino acid. In order to transfer the protons from amino acid side chain to the peptide backbone, energy will be needed (Kapp *et al*. 2003 Wysocki *et al*. 2000). In agreement with this theory, conotoxin-α of two charge state needed extra energy to mobilize the protons.

Fragment ions of $[M+3H]^{3+}$ cover almost the complete sequence ions at collision energy of 20eV this is most probably due to the lower *m/z* of the precursor ion and the extra charge which can freely move along the peptide backbone and induce charge directed dissociation. This finding indicates that the sequence information obtained from peptide fragmentation can be improved by adjusting the cone voltage.

8.2 Determination of optimum collision energy for peptide fragmentation

Mass-to -charge ratio (*m/z*) and charge state of precursor ions are the two main details obtained by an investigator at the early stage of a *de novo* sequencing experiment. Before one embarks on fragmenting the peptide of interest choosing appropriate collision energy is a crucial step. An attempt was done to establish the relationship between *m/z* and collision energy needed to effectively fragment a peptide.

Identified sequence ions of conotoxin-α, tertiapine, endothelin-2, neurotoxin, defensin HNP-1 and orexin-A fragmented at CEs between 20eV and 40eV are shown in appendix 1. Tertiapine ($z = 4$) has produced large number of sequence ions when fragmented with collision energies 20eV and 25 eV. For energies higher than 25 eV the number of sequence ions diminishes and product ions are more concentrated at the lower *m/z* range of the spectrum. The same phenomenon was observed for conotoxin-α ($z = 3$) and it is assumed that the added extra energy over the optimal collision energy may have induced internal fragmentation of the peptide and thereby reduced the intensities of the sequence ions.

Collision energies that produced the maximum number of sequence ions of the peptides (optimal collision energies) were plotted against the *m/z* as shown below in figure 6.

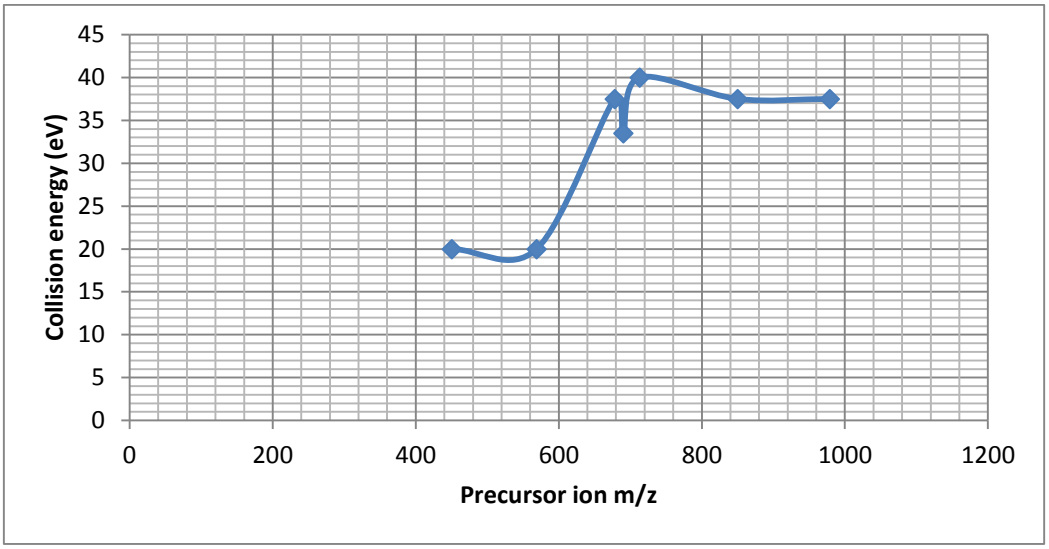


Figure 6: Optimum collision energies plotted against *m/z* of the precursor ions. Precursor ions at *m/z* under 600 tend to fragment sufficiently at collision energies between 20-30 eV. For precursor ions between m/z 600 -1000, collision energies between 30 and 40 eV seem to produce good fragmentation.

The CE needed for optimal fragmentation of the peptide increased with increasing precursor ion *m/z*, but raising the collision energy over 40 eV was associated with lower number of sequence ions. Factors such as precursor ion charge state and amino acid composition of the peptide can have a dramatic effect on the formation of sufficient fragment ions to enable subsequent identification of the peptide (Kapp *et al*, 2003). Therefore considering only *m/z* to determine an optimal collision energy might be an over simplification of the subject matter. Even if the number of tested peptides and factors considered were limited this preliminary study can give some hint as to which collision energy to choose in order to obtain a reasonable fragmentation.

8.3 Comparison of collision energy ramp with optimal single collision energy

After the peptides were fragmented by collision energy ramps of 20-30 eV, 25-35 eV and 25-40eV, a ramp that produced a maximum number of sequence ions was selected as an optimum ramp for that particular peptide.

The optimum collision energy ramps of the peptides were also compared with the corresponding optimum single collision energy based on the number of sequence ions produced as summarized in table 5.

Table-5: Summary of optimal single collision energies and optimal collision energy ramps determined for each peptide. Conotoxin-α, endothelin-2 and neurotoxin the peptides fragment well with CE between 30 and 40 and the average value was used as optimal CE.

Peptide	*m/z*	z	Optimum single CE (eV)	Optimum CE Ramp (eV)
Conotoxin-α	452.5218	3	20	20-30
Tertiapine	615.3226	4	25	25-35
Conotoxin-α	678.2739	2	35	25-40
Endothelin	850.7051	3	35	25-40
Neurotoxin	979.3962	3	35	25-40

The table clearly indicates that the choice of ramp scan is also related to the precursor ion mass-to-charge ratio.

Appendix-2 shows identified sequence ions of conotoxin-α, endothelin-2, neurotoxin and tertiapine after the peptides were fragmented by their respective optimum single collision energies and optimum collision energy ramps. In almost all of the cases, the number of sequence ions produced by the ramp modes were equal to or greater than the number of sequence ions formed by the optimum single collision energies.

Effect of collision energy ramp can also be illustrated by the fragmentation pattern of modified conotoxin-α where the cysteine residues were alkylated by the N-methylmaleimide (NMM). The identified sequence ions showed four distinct areas as indicated in table 6 where fragment ions at higher *m/z* were observed at lower CE (20-30 eV) and fragment ions at lower *m/z* were observed at higher CE (35-40eV). After fragmenting the peptide with ramp mode between 25 and 35 eV almost all sequence ions were identified.

Table-6: Identified sequence ions of Conotoxin-α alkylated by N-methylmaleimide (NMM) at different collision energies. Numbers in the bracket indicate charge state of the fragment ions. Using the ramp modus almost all the sequence ions was identified, all the ions were y-ions.

Conotoxin-α / NMM [M +3H] $^{3+}$= 600.5555													
20 eV	25eV	30eV	35eV	40eV	b-ions		y-ions	20 eV	25eV	30eV	35eV	40eV	**RAMP**
													25 - 35 eV
		✓(+1)	✓(+1)	✓ (+1)	1	G	12						
✓(+1)	✓(+1)	✓(+1)			2	C	11	✓(+2)	✓(+2)	✓(+2)			
✓(+1)		✓(+1)		✓(+1)	3	**C**	10	✓(+2)	✓(+2)	✓(+2)			✓(+1)
		✓(+1)			4	**S**	9	✓(+2)	✓(+2)	✓(+2)			✓(+1)
	✓(+1)				5	D	8		✓(+2)				
					6	P	7	✓(+2)	✓(+2)	✓(+2)	✓(+2)		✓(+1)
					7	R	6			✓(+2)	✓(+1)	✓ (+1)	✓(+1)
					8	C	5	✓(+2)		✓(+1)	✓(+1)	✓(+1)	✓+1)
					9	A	4				✓(+1)	✓(+1)	✓(+1)
					10	W	3	✓(+2)			✓(+1)	✓(+1)	✓(+1)
					11	R	2				✓ (+1)	✓(+1)	✓(+1)
					12	C	1		✓(+1)	✓(+1)	✓(+1)	✓(+1)	✓(+1)

8.4. Sequence identification of selectively alkylated peptides with two disulphide bridges

Conotoxin-α was selectively alkylated by N-Cyclohexylmaleimide (NCM) and N-Methylmaleimide (NMM) on Solid phase extraction (SPE) columns.

The MS/MS chromatogram of the selectively alkylated conotoxin-α formed two separate peaks (A, B) in both cases as shown in figure 7.

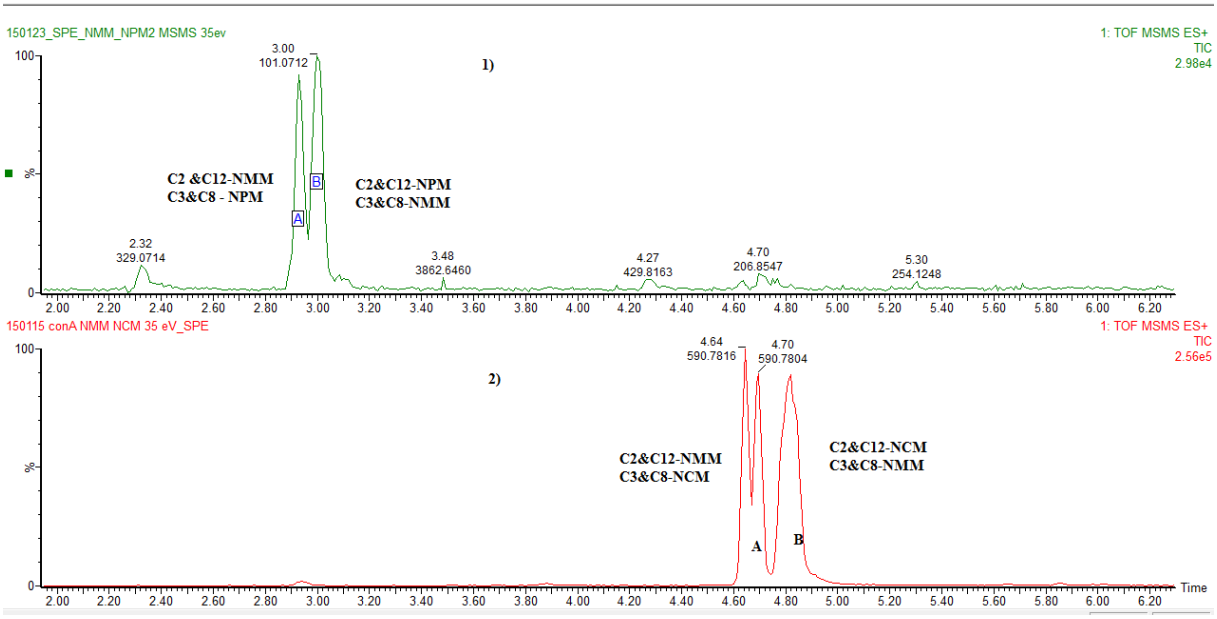


Figure 7: Chromatograms of selectively alkylated conotoxin-α. 1) The peptide was alkylated by NMM and NPM. The two isomers of conotoxin-α has formed two peaks, A and B. In A, C2&C12 were modified with NMM while C3 &C8 were by NPM. In B, C2&C12 were alkylated by NPM while C3&C8 were modified with NMM. The two peaks were not completely separated which can be due to the short retention time. 2) The peptide was alkylated by NMM and NCM. Peak A: C2&C12 are alkylated by NMM while C3&C8 are alkylated by NCM. Peak B: C2&C12 are alkylated by NCM and C3&C8 are alkylated by NMM. The two peaks of A produce similar sequence ions which may show the two peaks of A are rotamers.

The two peaks of the MS/MS chromatogram of conotoxin-α alkylated by NMM and NPM were sequenced separately. In the first peak (A) C2 and C12 were found to be alkylated by NMM while C3 and C8 were alkylated by NPM. In the second peak (B), the MS/MS spectrum revealed C2 and C12 are bound to NPM while C3 and C8 are to NMM. The identified sequence ions are presented in appendix 3.

Similarly structural isomers of conotoxin-α were identified using MS/MS spectrum obtained from A and B of the second chromatogram in above figure.

In addition sequencing the two peaks of A (2), yielded similar residue sequence which may indicate that the two peaks of A are most likely rotamers.

Sequence ions between y1-y4 and y10 were used to differentiate the two structural isomers as indicated in appendix 3.

8.5. **Evaluation of the potential of Peaks Studio 7 software for *de novo* sequencing**

According to the manufacturer of the software, the performance of PEAKS is dependent on the quality of the spectrum selected for the *de novo* sequencing. In order to minimize the confounding effect of the spectral quality, MS/MS data produced at optimum collision energies/ramps from the previous experiments were used.

Sequencing of conotoxin-α, endothelin-2, neurotoxin and defensin using PEAKS has shown various accuracy levels.

As previously discussed under section 8.1, manual sequencing of conotoxin-α fragmented at CV 20 V and CE 20eV has identified almost all the sequence ions of the peptide. Automated sequencing using PEAKS has also identified the sequence of the peptide with % ALC of 91 as shown in figure 7 below.

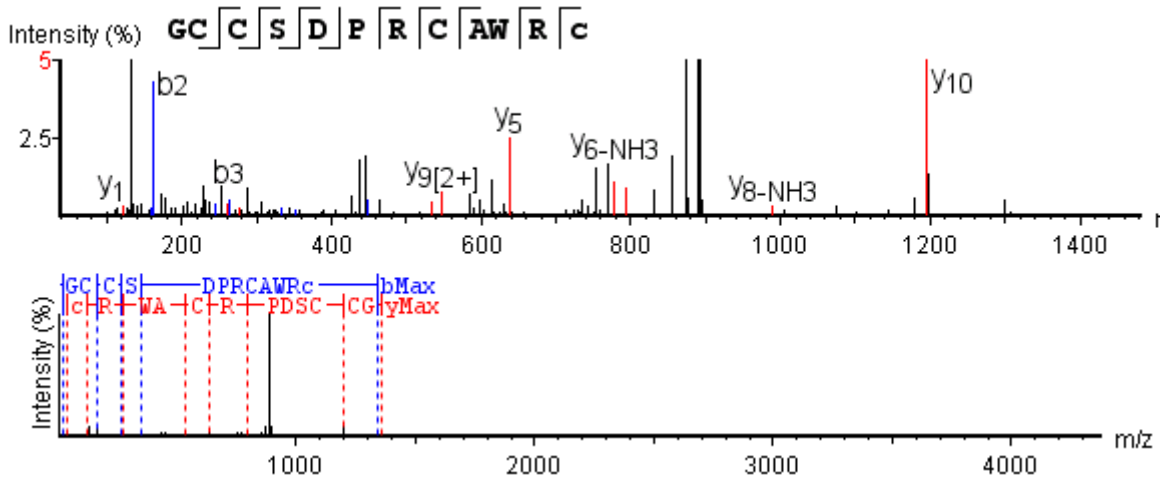
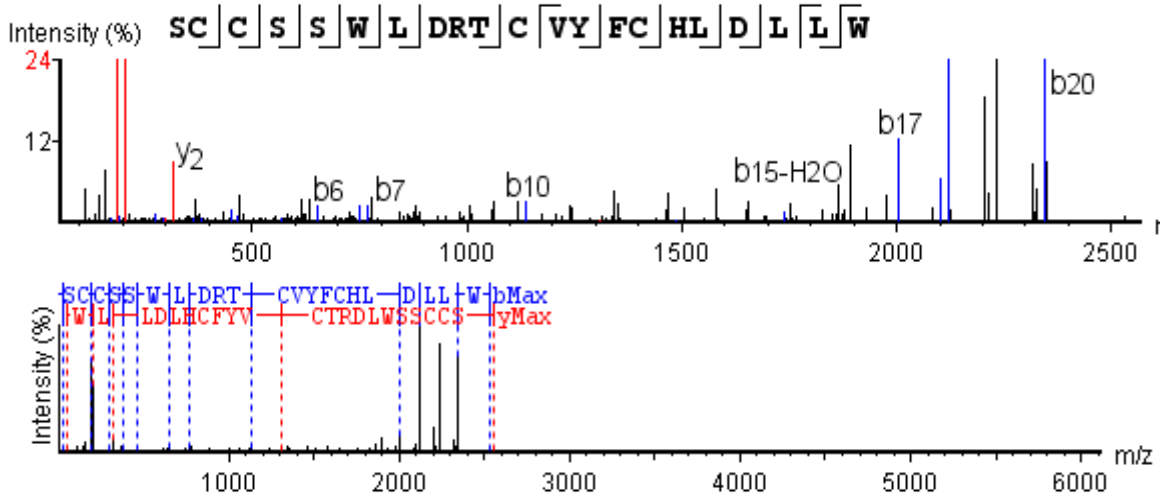


Figure 8: Annotated MS/MS spectrum of conotoxin-α. The PEAKS software has correctly identified the amino acid sequence of the peptide.

The identified amino acids at the N-terminal, GC, are ambiguous because of the missing y_{11}-ions from the spectrum. In the same way, the software didn't clearly determine whether y3 is –WRC or -ARC. This can be confirmed by looking for the theoretical values of y_3/b_9 of sequences, GCCSDPRCAWRC-NH2 and GCCSDPRCWARC-NH2 in the MS/MS spectrum. In this case, only the y3-ion (*m/z* = 463.2234) of the former sequence was identified in the spectrum proving its correctness.

The other peptide that has been sequenced by PEAKS was endothelin-2 fragmented at collision energy ramp between 25eV and 40 eV. With manual sequencing, y7/b14 was missing in the MS/MS spectrum as shown in appendix 2.The sequencing result from the PEAKS software shows two areas of the peptide which were incorrectly assigned, the two residues of the N-terminus and the three residues DRT in the middle of the peptide as shown in figure 8.

A)



B)

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	88.0399	70.0293	71.0129	44.5199	S					21
2	191.0480	173.0360	174.0220	96.0245	C	2463.0752	2445.0647	2446.0481	1232.0376	20
3	294.0570	276.0490	277.0312	147.5291	C	2360.0662	2342.0557	2343.0391	1180.5331	19
4	381.0880	363.0810	364.0633	191.0480	S	2257.0569	2239.0464	2240.0298	1129.0284	18
5	468.1220	450.1120	451.0953	234.5611	S	2170.0249	2152.0144	2152.9978	1085.5125	17
6	654.2040	636.1910	637.1746	327.6008	W	2082.9929	2064.9824	2065.9658	1041.9965	16
7	767.2900	749.2740	750.2587	384.1428	L	1896.9135	1878.9030	1879.8865	948.9567	15
8	882.3126	864.3021	865.2856	441.6563	D	1783.8295	1765.8190	1766.8025	892.4147	14
9	1038.4137	1020.4031	1021.3867	519.7068	R	1668.8025	1650.7920	1651.7755	834.9012	13
10	1139.4570	1121.4509	1122.4344	570.2280	T	1512.7014	1494.6909	1495.6744	756.8507	12
11	1242.4706	1224.4601	1225.4436	621.7353	C	1411.6537	1393.6432	1394.6267	706.3268	11
12	1341.5389	1323.5284	1324.5120	671.2695	V	1308.6420	1290.6340	1291.6176	654.8223	10
13	1504.6023	1486.5959	1487.5753	752.8011	Y	1209.5760	1191.5656	1192.5491	605.2880	9
14	1651.6708	1633.6603	1634.6438	826.3354	F	1046.5128	1028.5023	1029.4858	523.7564	8
15	1754.6799	1736.6650	1737.6530	877.8400	C	899.4443	881.4338	882.4174	450.2222	7
16	1891.7388	1873.7283	1874.7118	946.3694	H	796.4352	778.4246	779.4082	398.7176	6
17	2004.8240	1986.8124	1987.7959	1002.9114	L	659.3763	641.3657	642.3493	330.1881	5
18	2119.8491	2101.8440	2102.8228	1060.4249	D	546.2922	528.2816	529.2652	273.6461	4
19	2232.9338	2214.9233	2215.9067	1116.9669	L	431.2653	413.2547	414.2383	216.1326	3
20	2346.0229	2328.0076	2328.9910	1173.5090	L	318.1840	300.1690	301.1542	159.5906	2
21					W	205.0980	187.0866	188.0710	103.0486	1

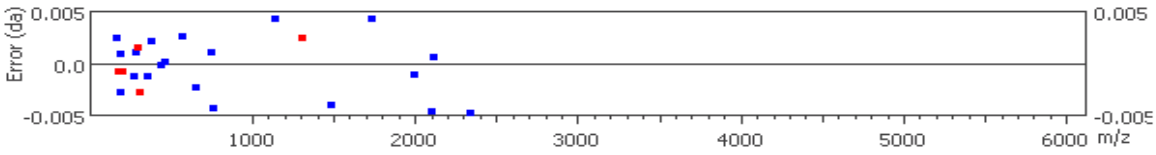


Figure 9: Amino acid sequence (A) and ion table (B) of endothelin-2 as assigned by PEAKS. The correct sequence of endothelin-2 is CSCSSWLDKECVYFCHLDIIW. The sequences were correctly identified except for the two residues at the N-terminal and the segment DRT.

Assuming the peptide sequence is unknown, it can be challenging to identify the correct order of the two N-terminal residues since both b_1 and y_{20} are missing from the spectrum. For the other incorrectly labelled residues, DRT, it is possible to confirm whether the suggested sequence is correct or not by going through the spectrum manually.

The three residues can be arranged in 6 different ways:-DRT- , -DTR- , -RDT- , -RTD- , -TRD- and TDR.

As a result, b_8 can have three different values: SCCSSWLD- (b_8^{1+}, m/z = 882.3121), SCCSSWLR-(b_8^{1+} = 923.3862 or b_8^{2+} = 462.1967) and SCCSSWLT- (b_8^{1+} = 868.3328). By searching these theoretical *m/z* values in the MS/MS spectrum we can see a high intensity ion at *m/z* 882.3130 with mass deviation below 0.01 Da for the first alternative. But for the other two possible sequences, both the b-ions and their corresponding y-ions are missing from the MS/MS spectrum. Still this may not prove the correctness of the sequence since two amino acids can be isobaric (e.g. I =L) or two amino acids can have the same mass as a single amino acid (e.g. GG = N) as shown in table-7. Because there are no residue(s) which have the same mass as D, one can positively conclude the presence of D at that site.

This leaves only two possibilities for b_9: SCCSSWLDRTCVYFCHLDIIW

(b_9^{1+}=1038.4132/b_9^{2+}=519.7102 or y_{12}^{1+}=1512.7014/y_{12}^{2+}=756.854) and SCCSSWLDTRCVYFCHLDIIW (b^{1+} = 983.3597, y_{12}^{1+} =1567.7548/ y_{12}^{2+} = 784.3811). There are no *m/z* values in the spectrum that correspond to these sequence ions but b_{10}^{1+} is present in the spectrum at *m/z* =1139.4572. Therefore we cannot confirm whether R and T are found in the peptide or not. According to table 7 RT has a mass of 257.1488 and other residues which have a mass close to this value (257.1376) is KE. Therefore if RT is replaced by KE, $b9^{1+}$ can be found at *m/z* = 1010. 4070 with fairly high intensity.

In this case even if the ion is found in the MS/MS spectrum, the software didn't recognized the ion. This kind of error can only be discovered by manual evaluation of the sequencing output therefore it is very important to go through the actual spectrum confirm the presence of all the suggested residues.

PEAKS has also correctly identified the sequences of tertiapine which has 21 residues. But for neurotoxin (27 residues) fragmented by collision energy ramps between 25eV and 40 eV, only residues between y_1 and y_{17} were identified correctly (data not shown). As shown in appendix 2, manual sequencing has identified almost all the sequence ions except b5/y22. This might indicate the *de novo* sequencing accuracy of the software is dependent also on the length of the peptide. Therefore for larger peptides, one has to consider digesting the peptide in order to obtain accurate amino acid sequence.

The software has also failed to sequence defensin-HNP1 which has 31 residues but in this case even with manual sequencing few sequence ions were recovered from the MS/MS spectrum. Therefore the failure of the software to sequence defensin- HNP1 can attributed to the poor quality of the spectrum and the amino acid length of the peptide.

Table 7 : Masses of amino acid residues in a peptide sequence.

Seq.	Mass	Seq.	Mass	Seq.	Mass	Seq.	Mass
A	71.0371	DT	216.0746	HH	274.1178	NR	270.1440
AA	142.0742	DV	214.0954	HK	265.1539	NS	201.0750
AC	174.0463	DW	301.1063	HL	250.1430	NT	215.0906
AD	186.0641	DY	278.0903	HM	268.0994	NV	213.1113
AE	200.0797	E	129.0426	HN	251.1018	NW	300.1223
AF	218.1055	EE	258.0852	HP	234.1117	NY	277.1063
AG	128.0586	EF	276.1110	HQ	265.1175	P	97.0528
AH	208.0960	EG	186.0641	HR	293.1600	PP	194.1055
AK	199.1321	EH	266.1015	HS	224.0909	PQ	225.1113
AL	184.1212	EK	257.1375	HT	238.1066	PR	253.1539
AM	202.0776	EL	242.1267	HV	236.1273	PS	184.0848
AN	185.0801	EM	260.0831	HW	323.1382	PT	198.1004
AP	168.0899	EN	243.0855	HY	300.1223	PV	196.1212
AQ	199.0957	EP	226.0953	K	128.0950	PW	283.1321
AR	227.1382	EQ	257.1012	KK	256.1899	PY	260.1161
AS	158.0691	ER	285.1437	KL	241.1790	Q	128.0586
AT	172.0848	ES	216.0746	KM	259.1354	QQ	256.1172
AV	170.1055	ET	230.0903	KN	242.1379	QR	284.1597
AW	257.1164	EV	228.1110	KP	225.1477	QS	215.0906
AY	234.1004	EW	315.1219	KQ	256.1535	QT	229.1063
C	103.0092	EY	292.1059	KR	284.1960	QV	227.1270
CC	206.0184	F	147.0684	KS	215.1270	QW	314.1379
CD	218.0361	FF	294.1368	KT	229.1426	QY	291.1219
CE	232.0518	FG	204.0899	KV	227.1634	R	156.1011
CF	250.0776	FH	284.1273	KW	314.1743	RR	312.2022
CG	160.0307	FK	275.1634	KY	291.1583	RS	243.1331
CH	240.0681	FL	260.1525	L	113.0841	RT	257.1488
CK	231.1042	FM	278.1089	LL	226.1681	RV	255.1695
CL	216.0933	FN	261.1113	LM	244.1246	RW	342.1804
CM	234.0497	FP	244.1212	LN	227.1270	RY	319.1644
CN	217.0521	FQ	275.1270	LP	210.1368	S	87.0320
CP	200.0620	FR	303.1695	LQ	241.1426	SS	174.0641
CQ	231.0678	FS	234.1004	LR	269.1852	ST	188.0797
CR	259.1103	FT	248.1161	LS	200.1161	SV	186.1004
CS	190.0412	FV	246.1368	LT	214.1317	SW	273.1113
CT	204.0569	FW	333.1477	LV	212.1525	SY	250.0954
CV	202.0776	FY	310.1317	LW	299.1634	T	101.0477
CW	289.0885	G	57.0215	LY	276.1474	TT	202.0954
CY	266.0725	GG	114.0429	M	131.0405	TV	200.1161
D	115.0269	GH	194.0804	MM	262.0810	TW	287.1270
DD	230.0539	GK	185.1164	MN	245.0834	TY	264.1110
DE	244.0695	GL	170.1055	MP	228.0933	V	99.0684
DF	262.0953	GM	188.0620	MQ	259.0991	VV	198.1368
DG	172.0484	GN	171.0644	MR	287.1416	VW	285.1477
DH	252.0859	GP	154.0742	MS	218.0725	VY	262.1317
DK	243.1219	GQ	185.0801	MT	232.0882	W	186.0793
DL	228.1110	GR	213.1226	MV	230.1089	WW	372.1586
DM	246.0674	GS	144.0535	MW	317.1198	WY	349.1426
DN	229.0699	GT	158.0692	MY	294.1038	Y	163.0633
DP	212.0797	GV	156.0899	N	114.0429	YY	326.1267
DQ	243.0855	GW	243.1008	NN	228.0859		
DR	271.1281	GY	220.0848	NP	211.0957		
DS	202.0590	H	137.0589	NQ	242.1015		

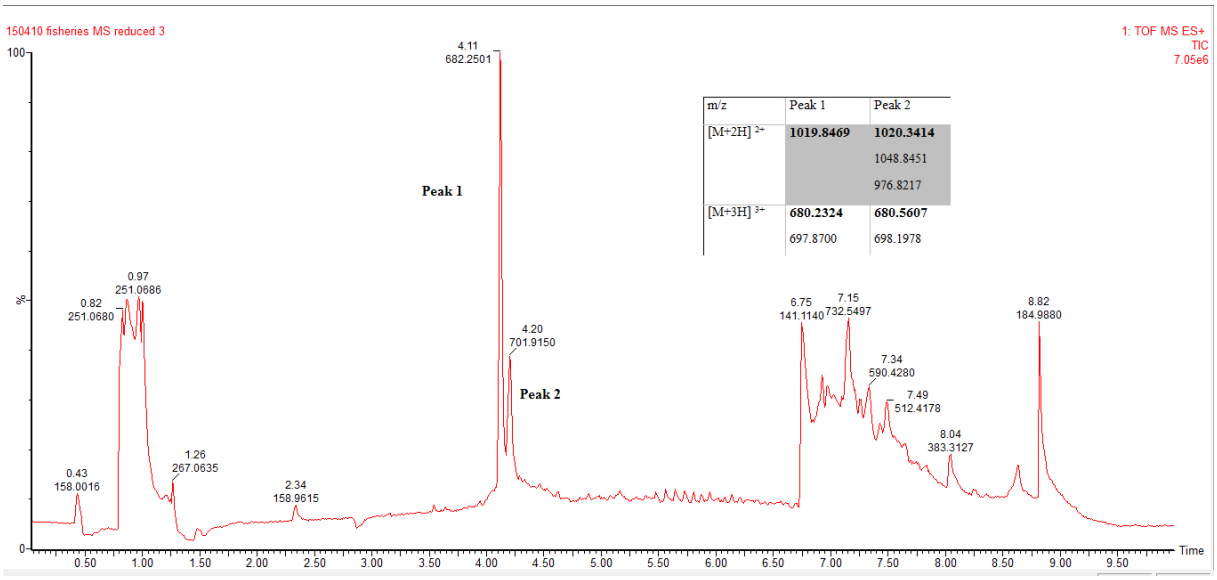
PEAKS® complete software proteomics.

http://www.bioinfor.com/peaks/downloads/masstable.html

8.6 *De novo* dequencing of a bioactive peptide

8.6.1. Mass spectrometry of the unknown peptide (ppt1)

The MS chromatogram of the smaple obtained from Norwegian College of Fishery Science (NCFC) revealed the presence of two major peaks labeled 1 and 2 in figure 10 and *m/z* of the major peptides found under each peak is given in the table.



m/z	Peak 1	Peak 2
$[M+2H]^{2+}$	**1019.8469**	**1020.3414**
		1048.8451
		976.8217
$[M+3H]^{3+}$	**680.2324**	**680.5607**
	697.8700	698.1978

Figure 10: Mass spectrometery of the sample obtained from NCFS reveals at two major peaks (1 and 2). The table contains high intensity ions from the two peaks.

After the peptide is reduced, the MS chromatogram showed two peaks labled A and B which are not completely separated as shown in figure 11.

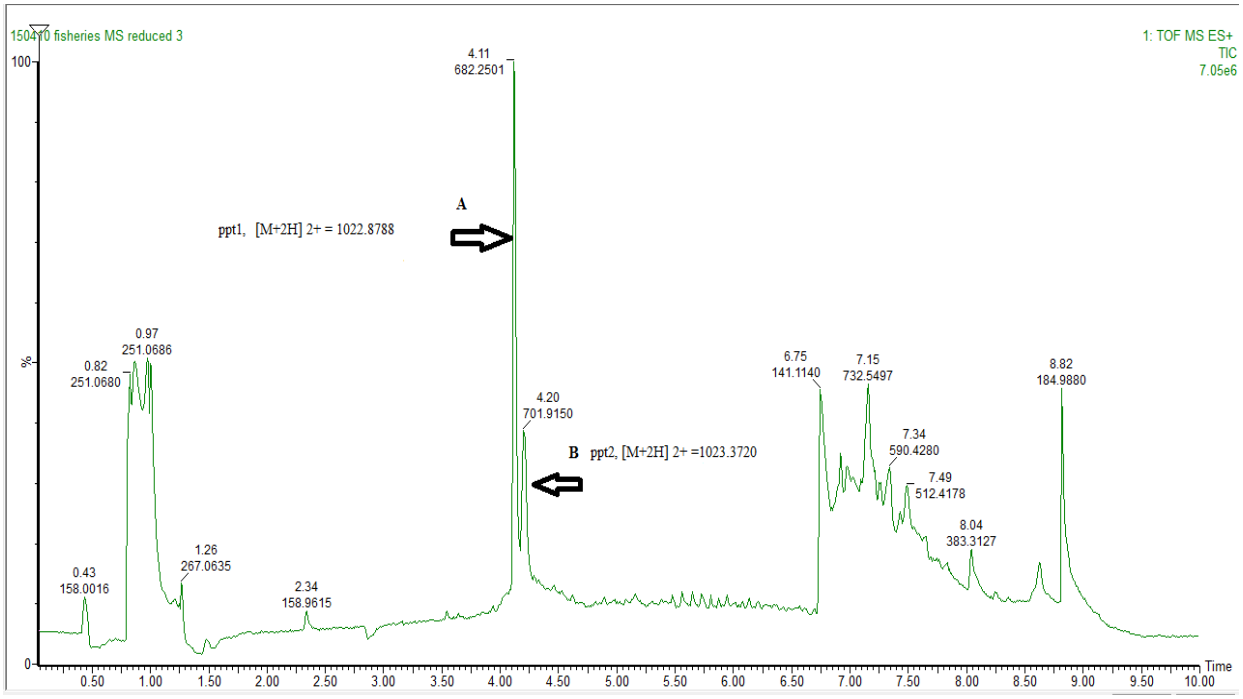


Figure 11: Chromatogram of the reduced peptide displaying two peaks which are not completely separated, A &B. (A) mainly contains a peptide, $[M+2H]^{2+}$ = 1022.8788, which will be referred to as ppt1 in the text. (B) Also contains a relatively intense peptide, $[M+2H]^{2+}$ = 1023.3720, labeled ppt2. The *m/z* difference between ppt1 and ppt 2 is 0.5.

The first peak of the above chromatogram contains an abundant peptide (ppt1) with *m/z* = 1022.8788 and this peptide had *m/z* value of 1019.8369 in its oxidized form. Similarly *m/z* of pp2 has also increased from 1020.3416 to 1023.3720 when reduced. This indicates both ppt1 and ppt2 have three disulphide bridges.

8.6.2. Tandem mass spectrometry of ppt1

Because of its high abundance in the mixture and prior information we have from NCFC about its biological activity, ppt1 was selected and fragmented for *de novo* sequencing.

The MS/MS chromatogram of ppt1 shows two peaks, C and D. Tailing of ppt1 (C) comes out with peak of ppt2 (D) as indicated in figure 12 below. This is due to the fact that the *m/z* difference between ppt1 and ppt2 with two charge state is only 0.5 and the instrument cannot discriminate the ions during ion selection because of the close *m/z* values.

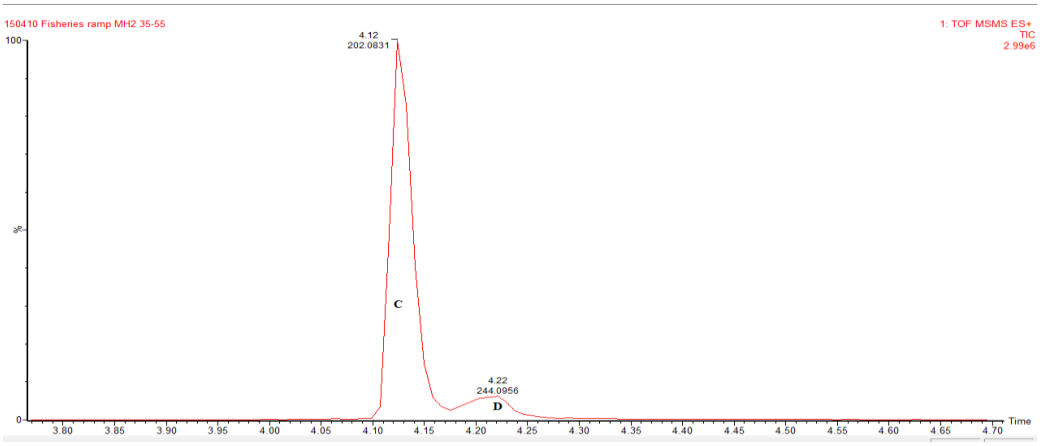


Figure 12: The MS/MS chromatogram of ppt1 showing two peaks, labelled C and D. Tailing from peaks C comes out with peaks of D.

The fact that ppt1 and ppt2 have so close retention time may suggest that they have some common structural features. In addition, gross comparison of the MS/MS spectra extracted from the two peaks (C&D) of the chromatograms shows very similar product ions in all areas of the spectra as shown in figure 13.

One structural commonality which can lead to this phenomenon is if ppt1 and ppt2 have the same amino acid sequence and the C-terminal of ppt1 is amidated.

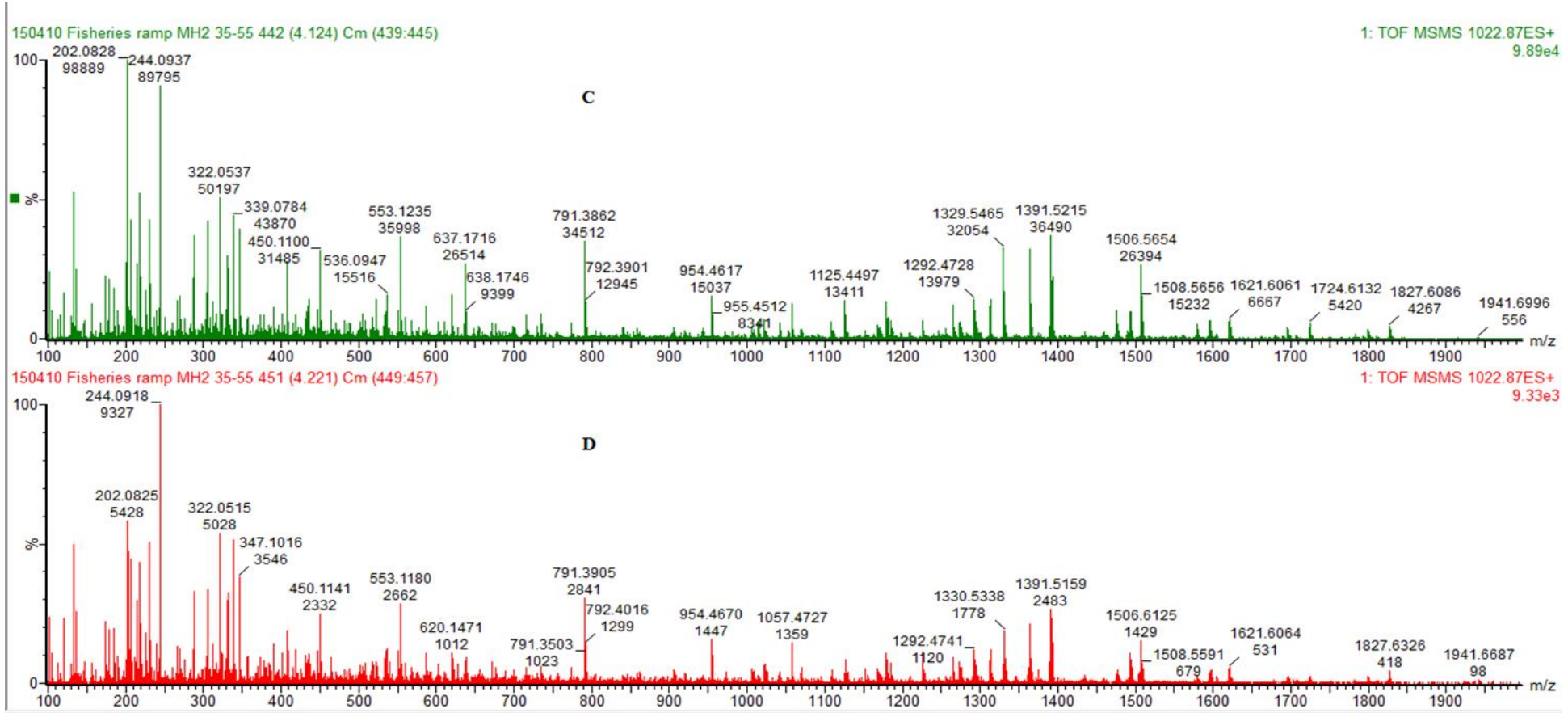


Figure 13: Spectra showing product ions extracted from the two peaks in the MS/MS chromatogram. (C) & (D) refers to the peaks of the chromatogram they extracted from (See also figure-12 above).

8.6.3 De *novo* sequencing of ppt1 using PEAKS Studio 7 software.

The MS/MS data of ppt1 in figure-13C shows ions at *m/z* 120.0809, 129.0647 and 136.0766 which indicates the presence of Phe, Arg and Tyr, respectively.

Because of the mass difference (1Da), close retention time and similarity in MS/MS product ions of ppt1 and ppt2, it was assumed that ppt1 is an amidated form of ppt2.

The data entry for the *de novo* sequencing was done by choosing retention times between 4.09 – 4.16 min, post translational modification (PTM) parameter 'amidated' and ion mass error of 0.005 Da.

The *de novo* sequencing of ppt1 at percent average local confidence (% ALC) of 86 is shown in figure 14.

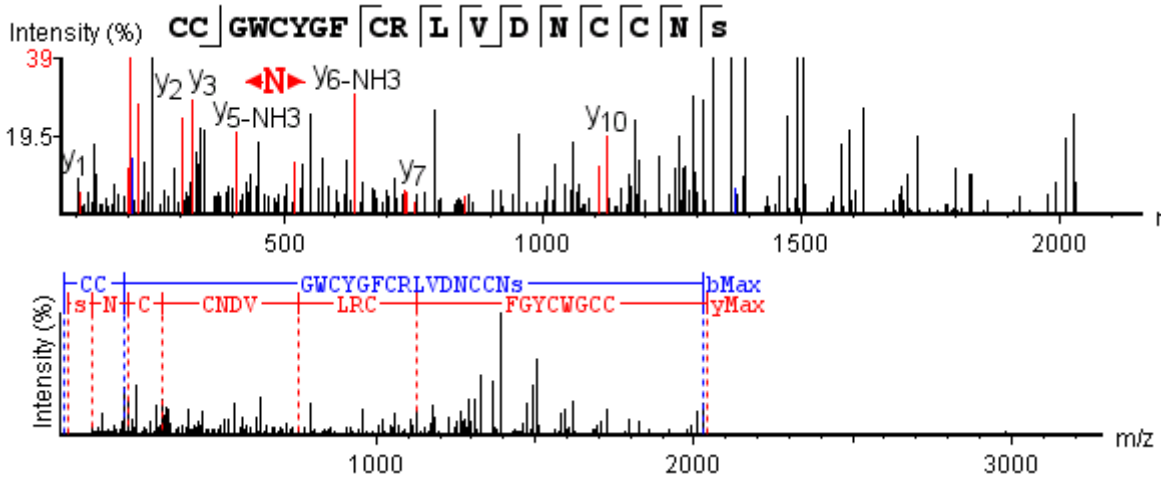


Figure 14: Automated *de novo* sequencing output of ppt1. The amino acids y1- y8 were identified with high confidence while the orders of CR and –GWCYGF- were not precisely determined.

The other candidates suggested by PEAKS having ALC greater than 50 % were:

CCGWCYGFCRLTNAQCCD (% ALC 69)

CCWGCYGFFARPNDCCNS (%ALC 52)

QDCCCYGFTSALVWECCD (% ALC 52).

For the suggested amino acid sequence of ppt1 at % ALC of 86, theoretical fragmentation was induced on Mass-product. The theoretical and observed *m/z* of sequence ions, mass deviation and intensity of the sequence ions are presented in table-8.

Table 8: Theoretical and observed *m/z* of fragment ions from the *de novo* sequencing output of ppt1. Many of the b-ions deviate from the theoretically expected value by more than 0.01 unlike the y-ions.

Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity	b-ion	Amino acid	y-ion	Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity
---	104.0177		357	1	C	18	---			
207.0256	207.026	-0.0004	4.28e4	2	C	17	1941.7321			
264.0471	264.0529	-0.0058	1.06e3	3	G	16	1838.7230			
450.1264	450.11	0.0164	3.23e4	4	W	15	1781.7			
553.1356	553.1234	0.0122	3.67e4	5	C	14	1595.62			
716.1989	716.1804	0.0185	8.59e3	6	Y	13	1492.61			
773.2204	773.2045	0.0159	5.60e3	7	G	12	1329.55	1329.5465	0.0032	3.22e4
920.2888	920.2645	0.0243	2.77e3	8	F	11	1272.53	1272.5194	0.0088	5.34e3
1023.298	1023.3062	-0.0082	1.42e3	9	C	10	1125.46	1125.4495	0.0103	1.34e4
1179.3991	1179.3936	0.0055	1.35e4	10	R	9	1022.45	1022.454	-0.0034	6.91e3
1292.4832	1292.4722	0.011	1.46e4	11	L	8	866.35			
1391.5516	1391.5215	0.0301	3.65e4	12	V	7	753.265	753.2664	-0.001	1.52e3
1506,5785	1506.5652	0.0133	2.67e4	13	D	6	654,197	654.1927	0.0043	4.37e3
1620,6215	1620.5947	0.0268	6.07e3	14	N	5	539,17	539.1656	0.0045	3.80e3
1723,6306	1723.618	0.0126	5.71e3	15	C	4	425,127	425.1257	0.0015	4.43e3
1826,6398	1826.6534	-0.0136	4.20e3	16	C	3	322,118	322.1192	-0.0012	8.2e3
1940,6828	1940.6692	0.0136	317	17	N	2	219,109	219.1096	-0.0008	2.25e4
---				18	S	1	105,066	105.0654	0.0005	1.04e4

All the sequence ions are found in the spectrum with reasonably high intensity. But many of the b-ions show fragment ion mass errors greater than 0.01 unlike the y-ions. We can also see that the deviation starts from b_4 where the observed value is less than the expected value. This might suggest W might be incorporated incorrectly.

According to table 7, the mass of W is 186.0793 Da, which is close to the combined mass of AD and EG (186.0641 Da and 186.0640 Da, respectively). In addition, W which usually produces a strong immonium ion (Medzihradszky, 2005) is not present in the lower mass range of the MS/MS spectrum. Absence of an immonium ion from a spectrum doesn't mean the residue is not present in the peptide (Medzihradszky, 2005) but it can build on the previous suspicion that one or more incorrect amino acids are present in the sequence.

After replacing W by AD and inducing theoretical fragmentation on Mass-product, *m/z* deviations from the theoretical values for the identified b-ions fall below 0.01 as shown in table 9.

Table 9: Theoretical and observed *m/z* sequence identified by PEAKS at % ALC 86 after W is replaced by -AD- .*m/z* deviation of the observed b-ions fall below 0.01

Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity	b-ion		y-ion	Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity
---	104.0177		357	1	C	19	---			
207.0256	207.026	0.0004	4.28e4	2	C	18	1941.7169			
264.0471	264.0529	0.0058	1.06e3	3	G	17	1838.7077			
335.0842				4	A	16	1781.6862			
450.1112	450.11	-0.0012	3.23e4	5	D	15	1710.6491			
553.1204	553.1234	0.003	3.67e4	6	C	14	1595.6222			
716.1837	716.1804	-0.0033	8.59e3	7	Y	13	1492.613			
773.2051	773.2045	-0.0006	5.60e3	8	G	12	1329.5497	1329.5465	0.0032	3.22e4
920.2736	920.2645	-0.0091	2.77e3	9	F	11	1272.5282	1272.5194	0.0088	5.34e3
1023.2827				10	C	10	1125.4598	1125.4497	0.010	1.34e4
1179.3839	1179.3936	0.0097	1.35e4	11	R	9	1022.4506	1022.454	-0.0034	6.91e3
1292.4679	1292.4722	0.0043	1.46e4	12	L	8	866.3495			
1391.5363				13	V	7	753.2654	753.2664	-0.0010	1.52e3
1506.5633	1506.5652	0.0019	2.67e4	14	D	6	654.197	654.1927	0.0043	4.37e3
1620.6062				15	N	5	539.1701	539.1656	0.0045	3.80e3
1723.6154	1723.618	0.0026	5.71e3	16	C	4	425.1272	425.1257	0.0015	4.43e3
1826.6246				17	C	3	322.118	322.1192	-0.0012	8.2e3
1940.6675	1940.6692	0.0017	317	18	N	2	219.1088	219.1096	-0.0008	2.25e4
---				19	S	1	105.0659	105.0654	0.0005	1.04e4

Because of the low mass error AD/EG can be the most likely residues at that site instead of W. Since the ion b_5 of CCGADCYGFCRLVDNCCNS-NH2 and CCGDA CYGFCRLVDNCCNS-NH2 have the same *m/z* we can searched for the b_4-ion in the spectrum to identify the correct sequence.

If the second sequence is correct, the value of b_4^{1+} = 379.0741 or y_{15}^{1+} = 1666.6593/ y_{15}^{2+} = 833.8333 will be found in the spectrum, but the search gave no result. Following the same procedure, no ion corresponding to *m/z* of b_4 or the corresponding y-ions were found using the first sequence option, CCGADCYGFCRLVDNCCNS-NH2.

The other residues which have combined mass similar to W is -EG-. After following the same procedure to find the value of b4 in the spectrum for sequences; CCGEG- and CCGGE, none of the values were found. This leaves us with an unconfirmed residue sequence, W.

Looking at the local confidence level of the residues on PEAKS, G has a value of 34 % which is much lower than the values assigned to other amino acids (where all the residues got confidence score >70%).

Based on the low confidence score assigned to G and the uncertainty associated with W, one can certainly suspect that –GW- in the sequence might be incorrectly assigned.

Continuing with the rest of the sequence,

The next unconfirmed sequences will be CCGADCYGF-, here there are 4 possibilities of b6: CCGADC- (b_6^{1+}=553.1204/y_{13}^{1+}=1492.6130/y_{13}^{2+} = 746.8101), CCGADY- (b_6^{1+} = 613.1745 or y_{13}^{1+}=1432.5589/ y_{13}^{2+}= 716.7831), CCGADG- (b_6^{1+} = 507.1326 or y_{13}^{1+} =1538.6007/ y_{13}^{2+} = 769.8040) and CCGADF- (b_6^{1+} = 597.1796 or y_{13}^{1+} = 1448.5538/y_{13}^{2+} =724.7805).

Since we know b_6^{1+} of CCGADC is present, we will look for ions of the other possible sequences. After searching for all the ions none of them were found, this confirms the correctness of the sequence up to b_6. In the same way b_7 will have three possibilities, b_8 has two possibilities. Repeating the same process existence of these residues was manually confirmed.

PEAKS also didn't identify how –C- and -R- are arranged relative to each other (CR or RC) and we will manually go through the spectrum to find the possible correct sequence. If we start from CCGADCYGFCRLVDNCCNS-NH2, the *m/z* value b_{10}^{1+} = 1022.4540 is found in

the MS/MS spectrum unlike b_{10} of the other alternative. Therefore we finally come to the conclusion that the most likely sequence of ppt1 is as shown in figure-15.

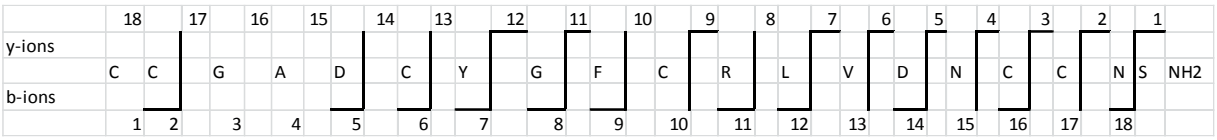


Figure 15: Possible amino acid sequences of ppt 1 after manual and automated sequencing. The residue G was manually found in MS/ MS spectrum but the local confidence score assigned to it very low (34%). Existence of –AD- not confirmed.

The peptide ppt1 was also sequenced by removing the post translational modification (PTM) parameter, amidation, during the data entry keeping all the other parameters (ion error, retention time) constant. The obtained sequence with % ALC of 72 is shown below in figure-14.

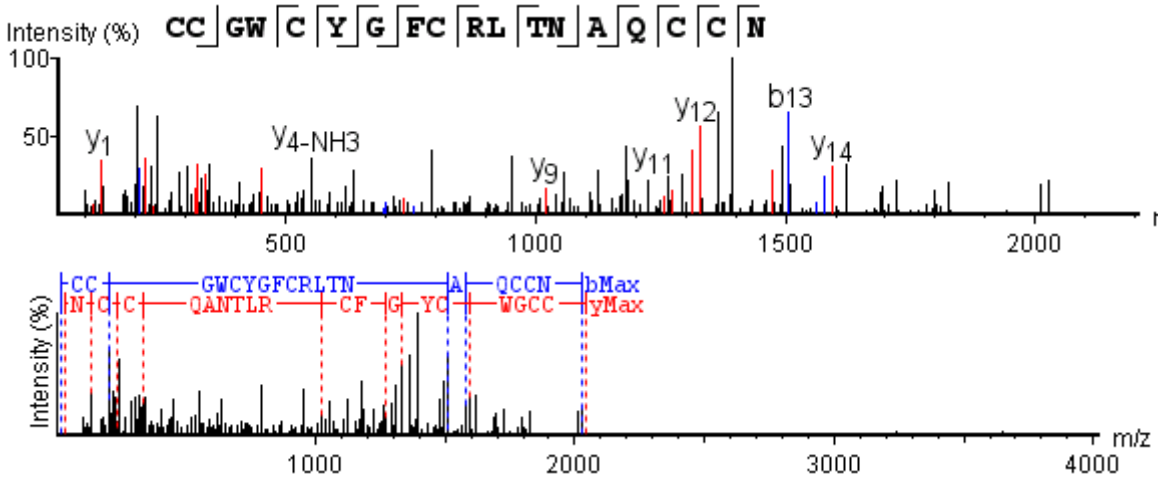


Figure 16: Annotated MS/MS spectrum and alignment of ppt1 after the PTM parameter was removed holding all the other parameters constant.

This amino acid sequence was identified as the second candidate having % ALC 69 when PTM parameter, amidated, was used as indicated earlier.

The theoretical *m/z*, the observed masses, mass deviation and intensities of the identified ions are given in table 10.

Table-10: Identified sequence ions of ppt1when no PTM parameter was used

In silico m/z	Identified m/Z	Mass error	Intensity				In silico m/z	Identified m/Z	Mass error	Intensity
---				1	C	19	---			
207.0256	207.026	-0.0004	4.25e4	2	C	18	1941.7169			
264.0471	264.0527	-0.0056	1.00e3	3	G	17	1838.7077			
335.0842				4	A	16	1781.6862			
450.1112	450.11	0.0012	3.22e4	5	D	15	1710.6491			
553.1204	553.1235	-0.0031	3.63e4	6	C	14	1595.6222			
716.1837	716.1805	0.0032	8.54e3	7	Y	13	1492.613			
773.2051	773.2045	0.0006	5.60e3	8	G	12	1329.5497	1329.5465	0.0032	3.20e4
920.2736	920.2643	0.0093	2.75e3	9	F	11	1272.5282	1272.5193	0.0089	5.29e3
1023.2827				10	C	10	1125.4598	1125.4495	0.0103	1.36e4
1179.3839	1179.3934	-0.0095	1.34e4	11	R	9	1022.4506	1022.4545	-0.0039	6.83e3
1292.4679	1292.4722	-0.0043	1.46e4	12	L	8	866.3495			
1393.5156				13	T	7	753.2654	753.2664	-0.001	1.52e3
1507.5585	1507.551	0.0075	1.46e4	14	N	6	652.2178			
1578.5956				15	A	5	538.1748			
1706.6542				16	Q	4	467.1377	467.1388	-0.0011	1.21e3
1809.6634				17	C	3	339.0791	339.0784	0.0007	4.47e4
1912.6726				18	C	2	236.07	236.0689	0.0011	7.50e3
---				19	N	1	133.0608	133.0611	-0.0003	5.31e4

Absence of b_{13}/ y_6 leads to unconfirmed sequence of TN, in addition to GAD. Both *de novo* sequencing results of ppt1 show similar amino acid sequences between b_1-b_{12} and the C-terminal – CCN- part. These facts may indicate the N-terminus of ppt1 is actually amidated. The MS/MS spectrum of ppt1 and the annotated ions are given i figure-17 and figure-18.

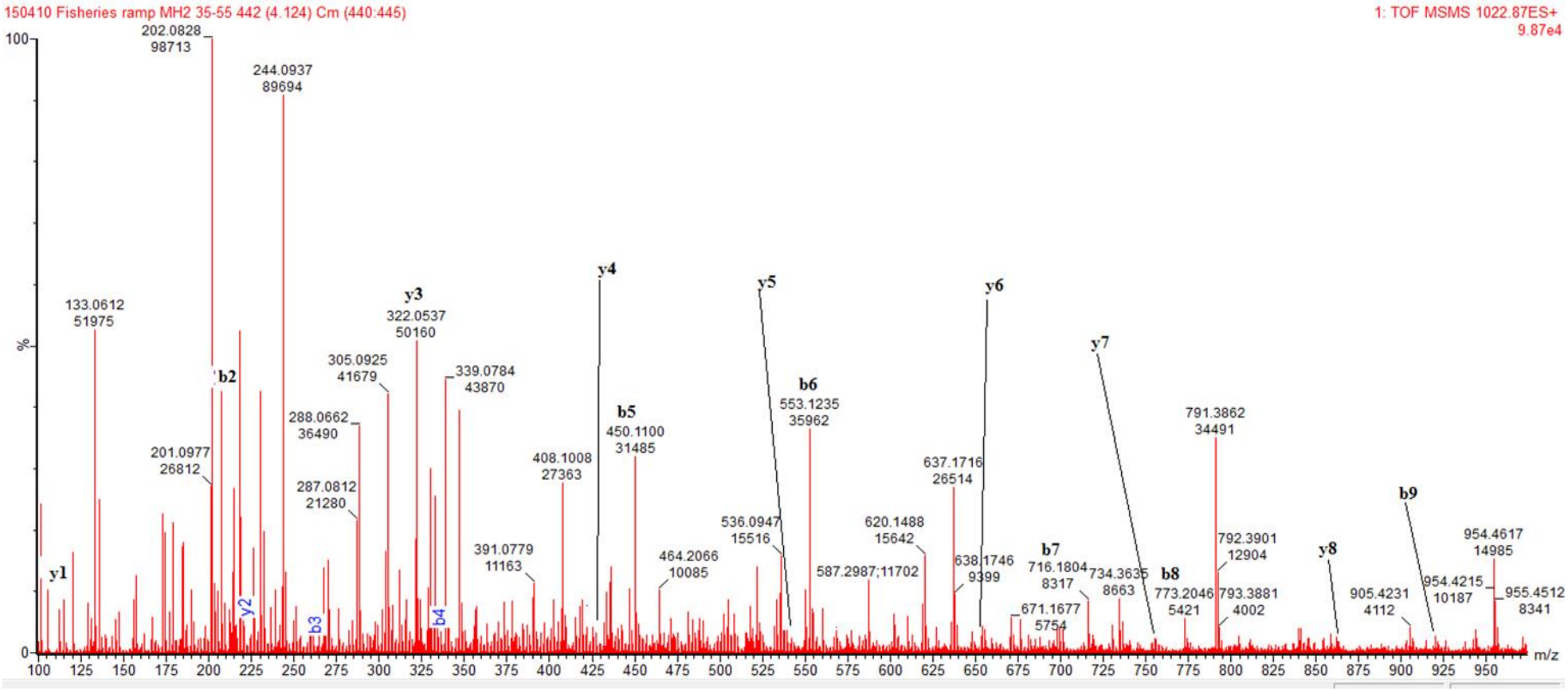


Figure -17: MS/MS spectrum of ppt1 and identified sequence ions (*m/z* =1022.8733 , z= +2) for values below 950. Corresponding to amino acid sequence CCGADCYGFCRLVDNCCNS-NH2

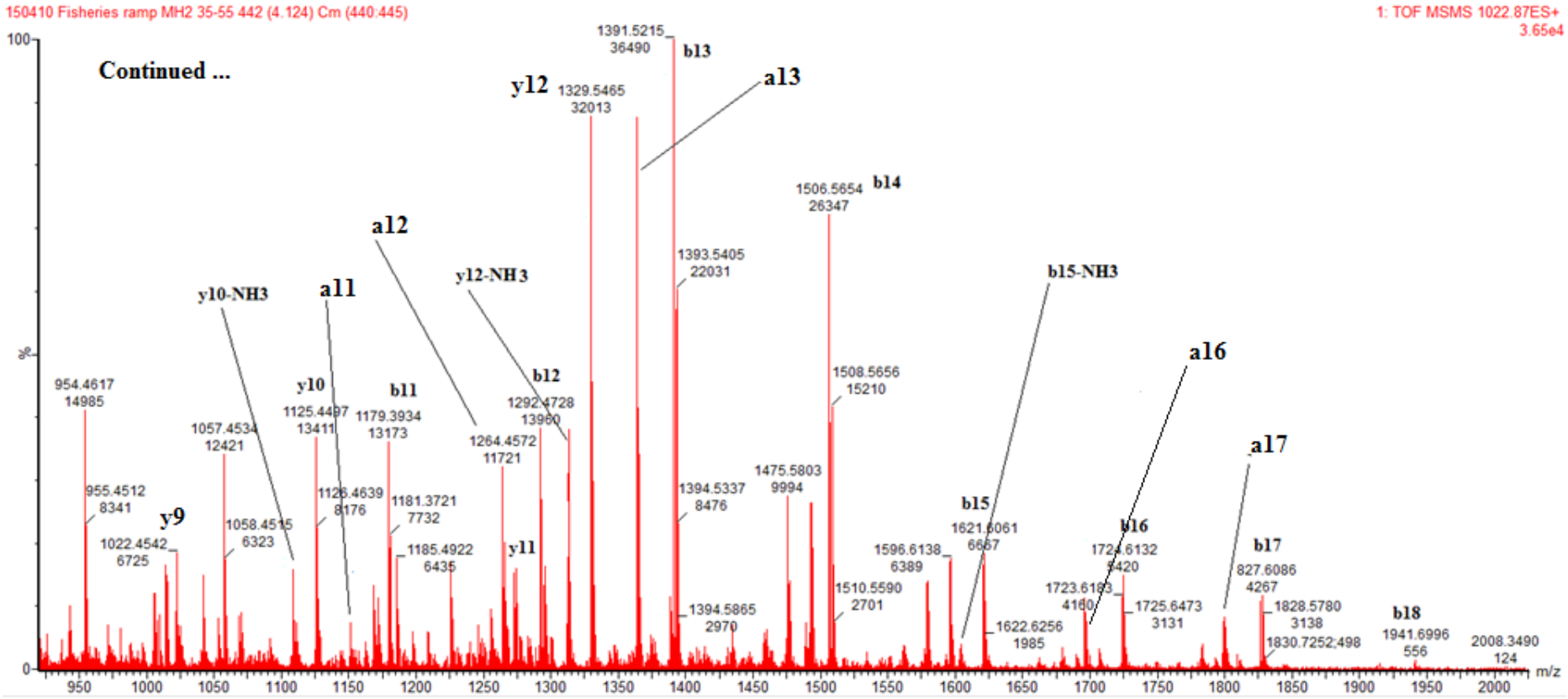


Figure -18: MS/MS spectrum of ppt1 (*m/z* =1022.8733, z= +2) showing sequence ions and neutral losses. The corresponding amino acid sequence is CCGADCYGFCRLVDNCCNS-NH2.

The correct sequence of the peptide after Edman degradation was found to be CCDQCYGFCRLVDNCCNS-NH2, and 84 % of the amino acid sequence was correctly assigned by PEAKS software. Ions corresponding to –DQ-, b_3^{1+} (322.0535, at intensity of 5.11e4) and b_4^{1+} (450.11 at intensity of 3.23e4) are found in the spectrum.

As mentioned in the previous discussion, residues –GAD- were not confirmed positively under the manual evaluation. Other residues having the same mass as –GAD- include NE, DQ and EGG. These residues were not considered under the manual evaluation by erroneously thinking G was the correct residue.

The correct sequence could have been confirmed by searching the b_3 values of: CCNE- (b_3^{1+} =321.0686), CCEN- (b_3^{1+} = 336.0682), CCQD- (335.0842), CCDQ- (b_3^{1+}= 322.0526). The first three are not found in the MS/MS spectrum this implies, it is DQ which is found at the site instead of W.

Therefore when using PEAKS as a *de novo* sequencing tool, it is very important to see the results in terms of the residue local confidence scores. In addition, for ambiguously assigned sequence segments, it is mandatory to consider each and every amino acid combination to look for the correct residue at the site.

Neutral losses obtained from MS/MS for the correct sequence of ppt1 are presented in table 11.

Table 11: Neutral losses from the sequence ions of ppt1

ion	*m/z*	ion	*m/z*	ion	*m/z*	*m/z*	b		y	*m/z*	ion	*m/z*	ion	*m/z*	ion	*m/z*
						---	1	C	18	---						
				a_2	179.0307	207.0256	2	C	17	1941.7169						
				b_3-H_2O	304.0420	322.0526	3	D	16	1838.7077						
b_4-NH_3	433.0846	b_4-H_2O	432.1006	a_4-NH_3	405.0897	450.1112	4	Q	15	1723.6808						
b_5-H_2O	535.1098	b_5-NH_3	536.0938	a_5-NH_3	508.0989	553.1204	5	C	14	1595.6222						
		a_6-NH_3	671.1622	a_6	688.1888	716.1837	6	Y	13	1492.6130	y_{13}-H_2O	1474.6024				
b_7-NH_3	756.1786	b_7-H_2O	755.1946	a_7-NH_3	728.1837	773.2051	7	G	12	1329.5497						
b_8-H_2O	902.2630	b_8-NH_3	903.2470	a_8	892.2786	920.2736	8	F	11	1272.5282						
						1023.2827	9	C	10	1125.4598	y_{10}-NH_3	1108.4332	a_{10}	1151.3889		
				b_{10}-NH_3	1162.3573	1179.3839	10	R	9	1022.4506	y_9-NH_3	1005.4241				
				a_{11}	1264.4730	1292.4679	11	L	8	866.3495						
		b_{12}-NH_3	1374.5098	a_{12}	1363.5414	1391.5363	12	V	7	753.2654	y_7-NH_3	736.2389				
				b_{13}-NH_3	1489.5367	1506.5633	13	D	6	654.1970	y_6-H_2O	636.1865	b_6-H_2O	698.1731	y_6-NH_3	637.1705
						1620.6062	14	N	5	539.1701	y_5-NH_3	522.1435				
		b_{15}-NH_3	1706.5888	a_{15}	1695.6205	1723.6154	15	C	4	425.1272	y_4-NH_3	408.1006	y_4-H_2O	407.1166		
		a_{16}-NH_3	1781.6031	b_{16}-NH_3	1809.5980	1826.6246	16	C	3	322.1180	y_3-NH_3	305.0914	y_3-H_2O	304.1074		
						1940.6675	17	N	2	219.1088	y_2-NH_3	202.0822	y_2-H_2O	201.0982		
						---	18	S	1	105.0659						

9. Conclusion

This work has addressed the effect of different MS parameters on peptide fragmentation. One of the parameters that can be manipulated by the operator is the cone potential, which can affect the precursor ion charge state and fragmentation prior to the arrival of the charged analyte in the collision cell. By changing the cone voltage we managed to change the fragmentation pattern of a model peptide. But it is assumed that, its benefit is more pronounced when the fragmentation characteristics of the peptide is known to the operator.

The most informative collision energy was found to be directly related to the *m/z* of the precursor ion. This optimal collision energy was also compared with the most informative collision energy ramps. In the present work, the results points to the superiority of collision energy ramp mode compared to single collision energy.

Considering the complexity of peptide fragmentation characteristics in MS/MS, larger data sets are required to establish how the different factors affect the quality of a spectrum for *de novo* sequencing. It would be important to continue the project with a wide variety of peptides (having different *m/z* ratio, mass, charge state, amino acid sequence) to determine how the different MS parameters are related to the fragmentation of peptides.

A software package, PEAKS studio 7, used for automated *de novo* sequencing was also tested for accuracy using model peptides having different characteristics (precursor ion charge, *m/z* and chain length). The software is found to be fairly accurate for smaller peptides, considering a good spectrum is chosen. But it was found out that its precision diminishes with increasing peptide length.

An unknown bioactive peptide was also sequenced using the software and 84% of the residues were correctly sequenced.

Peaks can be an effective tool for the laborious and time consuming de *novo* sequencing work, especially for smaller peptides. But for larger peptides, breaking it down by enzymes should be considered.

Determining the cysteine residue crosslinking of the peptide which was not addresses in this paper due to time constrain is the next step toward understanding the structure of the sequenced peptide.

10. REFERENCES

1 Anand, P., Grigoryan, A., Bhuiyan, M. H., Ueberheide, B., Russell, V., Quinonez, J., Moy, P., Chait, B. T., Poget, S. F. and Holford. 2014. Sample limited characterization of a novel disulfide-rich venom peptide toxin from terebrid marine snail Terebra variegate. PLoS One 9 (4): e94122.

2 Baldwin, M. A.2005. Mass spectrometers for the analysis of biomolecules. Methods Enzymol 402: 4-48.

3 Banerjee, S. Mazumdar, S. 2011. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. Int J Anal Chem 2012: 1-41.

4 Chernushevich I. V., Loboda, A. V. and Thomson B. A. 2001. An introduction to quadrupole-time-of-flight mass spectrometry. J Mass Spectrom 36(8): 849–86.

5 Cox, J., Michalski, A. and Mann, M. 2011. Software lock mass by two-dimensional minimization of peptide mass errors. J Am Soc Mass Spectrom 22 (8) 1373-8

6 Craik, D. J., Fairlie, D. P., Liras, S. and Price, D. 2013. The Future of Peptide-based Drugs. Chemical Biology & Drug 81(1): 136-147.

7 Demunshi, Y. Chugh, A. 2009. Patenting trend in Marine Bioprospecting based Pharmaceutical sector. Journal of Intellectual Property Rights 14: 122-130.

8 Fenn, J. 2002. Electrospray ionization mass spectrometry: How it all began. J Biomol Tech 13(3): 101–118.

9 Gongora-Benítez, M. Tulla-Puche, J. and Albericio, F. 2014. Multifaceted Roles of Disulfide Bonds. Peptides as Therapeutics. Chem.Rev 114: 901−926.

10 Hughes, C., Ma, B., Lajoie, G. A. 2010. De novo sequencing methods in proteomics. Methods Mol Biol. 604: 105-21.

11 Hunt, D.F., Yates, J.R., Shabanowitz, J., Winston, S., Hauer, C.R. 1986. Protein sequencing by tandem mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America 83(17):6233-7.

12 Kapp, E.A, Schutz, F., Reid, G. E., Eddes, J. S., Moritz, R. L., O'Hair, R. A. J., Speed, T. P.Simpson, R. J. 2003. Mining a tandem mass spectrometry database to determine the trends and global factors influencing peptide fragmentation. Analytical Chemistry 75 (22): 6251-6264.

13 Kebarle, P., Tang, L.1993. From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry. Anal Chem 65 53-77.

14 King, Glenn, .F. 2013. Venoms as a platform for human drugs: translating toxins into therapeutics. Biol. Ther 11 (11): 1469-148.

15 Kinter, M., Sherman, E.N. 2000. Protein sequencing and identification using tandem mass spectrometry. John Wiley & Sons, Inc., 605 Third Avenue, New York.

16 Lewis RJ, Garcia ML.2003. Therapeutic potential of venom peptides. Nature reviews Drug discovery. 2(10):790-802.

17 Liu, X., Dekker, L. J., Wu, S., Vanduijn, M. M., Luider, T. M., Tolic, N., Kou, Q., Dvorkin, M., Alexandrova, S., Vyatkina, K., Pasa-Tolic, L. and Pevzner, P. A. 2014. *De novo* protein sequencing by combining top-down and bottom-up tandem mass spectra. J Proteome Res 13 (7) :3241-8.

18 Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., Lajoie, G.2003. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 17(20): 2337-42.

19 Matthisen, R. Mass spectrometry Data analysis in proteomics. 2007: Humana Press Inc. Totowa, New Jersey 07512.

20 McLuckey, A. S.1992. Principles of Collisional Activation in Analytical Mass Spectrometry. Journal of the American Society for Mass Spectrometry 2 (6):599–614.

21 McLafferty, F. W.1981. Tandem mass spectrometry.1981; Science 214(4518):280-7.

22 Medzihradszky, K.F.2005. Peptide sequence analysis. Methods Enzymol. 402:209-44.

23 Medzihradszky, K. F.2005. Peptide sequence analysis. Methods Enzymol 402: 209-44

24 Medzihradszky, K.F, Bohlen, C.J.2012. Partial *de novo* sequencing and unusual CID fragmentation of a 7 kDa, disulfide-bridged toxin. Journal of the American Society for Mass Spectrometry. 23(5):923-34.

25 Moroder, L., Musiol, H. J., Gotz, M. and Renner, C. 2005. Synthesis of single- and multiple-stranded cystine-rich peptides. Biopolymers 80 (2-3):85-97.

26 Olivera, B. M.2006. Conus peptides: biodiversity-based discovery and exogenomics. J Biol Chem281(42): 31173-7.

27 Ondetti, A.M, Williams, J.N., Sabo, F.E., Pluscec, J., Weaver, R.E. and Kocy, O.1971. Angiotensin-converting enzyme inhibitors from the venom of *Bothrops*

jararaca. Isolation, elucidation of structure and synthesis.Biochemistry 19: 4033–4039.

28 PEAKS® complete software for proteomics. http://www.bioinfor.com/peaks/downloads/masstable.html. Waterloo, Canada. [Cited 2015 May 11].

29 Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20(18): 3551-67.

30 Pevtsov, S., Fedulova, I., Mirzaei, H., Buck, C., Zhang, X.2006. Performance Evaluation of Existing *De Novo* Sequencing Algorithms. J Proteome Res 5(11):3018-28.

31 Roepstorff, P., Fohlman, J.1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed. Mass Spectrom 11 (11): 601.

32 Seidler, J., Zinn, N., Boehm, M.E. and Lehmann, W.D.2010. *De novo* sequencing of peptides by MS/MS. Proteomics 10(4):634-49.

33 Standing, K. G. 2003. Peptide and protein de novo sequencing by mass spectrometry. Current opinion in structural biology 13(5):595-601.

34 Svenson, J. 2013.MabCent: Arctic marine bioprospecting in Norway. Phytochem Rev 12: 567-578.

35 Tannu, N. S. and Hemby, S. E. 2007.De novo protein sequence analysis of *Macaca mulatta.* BMC Genomics 8 (270).

36 Wells, J. M. 2005. Collision-Induced dissociation (CID) of Peptides and Proteins. Methods in Enzymology.402:148 - 85.

37 Westermeier, R. and Naven, T.2002. Proteomics in practice: A laboratory manual of proteome Analysis. pp 125-54. Wiley –VCH Verlag GmbH; Freiburg, Germany.

38 Wysocki, V. H., Tsaprailis, G., Smith, L. L. and Breci, L. A. 2000. Mobile and localized protons: a framework for understanding peptide dissociation. J Mass Spectrom 35(12): 1399-406

39 Wysocki, V. H., Resing, K. A., Zhang, Q. F. and Cheng, G. L. 2005. Mass spectrometry of peptides and proteins. Methods 35(3):211-222.

40 Ueberheide BM, Fenyo D, Alewood PF, Chait BT.2009. Rapid sensitive analysis of cysteine rich peptide venom components. Proceedings of the National Academy of Sciences of the United States of America 106(17):6910-5.

41 Zhang W, Krutchinsky AN, Chait BT.2003. "*De novo*" peptide sequencing by MALDI-quadrupole-ion trap mass spectrometry: a preliminary study. Journal of the American Society for Mass Spectrometry 14(9):1012-21.

42 Zhang, G., Annan, R.S., Carr, S.A. and Neubert, T.A. 2014. Overview of Peptide and protein analysis by mass spectrometry. Current protocols in molecular biology 108:10 21 1-10 21 30.

10. Appendix 1: Determining the effect of precursor ion *m/z* on collision energy needed for effective peptide fragmentation

Tables a-e : Show sequence ions identified after Tertiapine, Endothelin, Neurotoxin, Defensing and Orexin-A were fragmented at collision energies ranging from 20 - 45 eV. The numbers in the table indicate the charge state of the fragment ions.

Table a: Tertiapine $[M+4H]^{4+}$ = 615.3226

The peptide fragments well at collision energies of 20 eV and 25 eV. The peptide loses the first two amino acids Ala-Leu probably in the ion source at a cone voltage of 35 V and produces an intense y 19^{4+} (*m/z*= 569.2928, z=4), this fragment was also dissociated best at this collision energy range

20eV	25 eV	30eV	y-ion		b-ion	20eV	25 eV	30eV
			21	**A**	**1**			
			20	**L**	**2**	✓ (+1)	✓ (+1)	
			19	**C**	**3**			
✓ (+3)	✓ (+3)		**18**	**N**	4			✓ (+1)
✓ (+3)			17	C	5	✓ (+1,2)	✓ (+1,2)	
✓ (+3)			16	N	6	✓ (+1,2)	✓ (+1,2)	
			15	R	7	✓ (+1,2)	✓ (+1,2)	✓ (+1,2)
	✓ (+2)		14	I	8	✓ (+1,2)	✓ (+1, +2)	✓ (+2)
✓ (+2)			13	I	9	✓ (+2)	✓ (+1, +2)	✓ (+1,2)
✓ (+2,3)	✓ (+2)		12	I	10	✓ (+1,2)	✓ (+2)	
✓ (+2,3)	✓ (+1,2,3)		11	P	11	✓ (+2)	✓ (+2)	
✓ (+2,3)	✓ (+1,2,3)	✓ (+2,3)	10	H	12	✓ (+2)	✓ (+2)	
✓ (+1,2)	✓ (+2)	✓ (+2)	9	M	13	✓ (+,2)		
✓ (+1,2)	✓ (+1,2)		8	C	14	✓ (+2)	✓ (+2)	
✓ (+1)	✓ (+1)		7	W	15	✓ (+3)		
	(+1)		6	K	16	✓ (+3)	✓ (+2)	
✓ (+2)	✓ (+1,2)	✓ (+2)	5	K	17	✓ (+2)		
✓ (+1)	✓ (+1)		4	C	18			
	✓ (+1)		3	G	19			
	✓ (+1)		2	K	20	✓ (+3)	(+2)	
✓ (+1)	✓ (+1)	✓ (+1)	1	K	21			

Table b: Endothelin $[M+3H]^{3+}$ = 850.7051.

The peptide fragments well between 30 and 40 eV and 35 eV appears to be the optimum collision energy. The sequence ions are dominated by b-ions.

20 eV	25	30	35	40	b-ions		y-ions	20	25	30	35	40
✓ (+1)					1	C	21					
✓ (+1)		✓ (+1)	✓ (+1)	✓ (+1)	2	S	20					
✓ (+1)			✓ (+1)	✓ (+1)	3	C	19					
	✓ (+1)		✓ (+1)	✓ (+1)	4	S	18			✓ (+2)		
	✓ (+1)	✓ (+1)		✓ (+1)	5	S	17					
		✓ (+1)	✓ (+1)	✓ (+1)	6	W	16					
	✓ (+1)	✓ (+1	✓ (+1)	✓ (+1)	7	L	15					
	✓ (+1)	✓ +1)	✓ (+1)	✓ (+1)	8	D	14			✓ (+2)	✓ (+2)	
✓ (+1)		✓ (+1)	✓ (+1)	✓ (+1)	9	K	13					
	✓ (+1)	✓ (+1,2)	✓ (+1)	✓ (+1,2)	10	E	12					
✓ (+1)	✓ (+1,2)	✓ (+1,2)	✓ (+2)	✓ (+1,2)	11	C	11		✓ (+1)	✓ (+1)		
	✓ (+1,2)	✓ (+1,2)	✓ (+1,2)	✓ (+1,2)	12	V	10					
✓ (+2)	✓ (+1,2)	✓ (+1,2)	✓ (+1,2)	✓ (+1)	13	Y	9	✓ (+2)	✓ (+1)	✓ (+1)		
	✓ (+1,2)		✓ (+1,2)	✓ (+1,2)	14	F	8		✓ (+1)	✓ (+1)		
	✓ (+1)	✓ (+1)	✓ (+1,2)	✓ (+1,2)	15	C	7		✓ (+1)			
	✓ +1,2	✓ (+1,2)	✓ (+2)	✓ (+2)	16	H	6		✓ (+1)			
	✓ (+2)	✓ (+2)	✓ (+2)	✓ (+2)	17	L	5					
✓ (+2)	✓ (+3)	✓ (+2,3)	✓ (+2)	✓ (+2)	18	D	4					
✓ (+2,3)	✓ (+2)	✓ (+2)	✓ (+2)		19	I	3	✓ (+1)				
✓ (+2,3)	✓ (+2)	✓ (+2)	✓ (+2)		20	I	2	✓ (+1)	✓ (+1)	✓ (+1)	✓ (+1)	✓ (+1)
					21	W	1	✓ (+1)	✓ (+1)	✓ (+1)	✓ (+1)	✓ (+1)

Table-c: Neurotoxin [M+3H] $^{3+}$ = 979.3961

Neurotoxin produces very intense y_9 and y_{16}^{2+} ions and low intensity ions in the other *m/z* ranges with the collision energies under 30 eV (not indicated). All the other collision energies (30,35 and 40eV) have produced almost complete sequence ions. 35 eV was taken as optimum CE.

30eV	35 V	40 eV	b-ions		y-ions	30 eV	35eV	40eV
			1	**R**	27			
			2	**S**	26			
✓ +1	✓ +1	✓ +1	3	**C**	25			
✓ +1	✓ +1	✓ +1	4	**C**	24			
			5	**P**	23	✓ +2		
✓ +1	✓ +1	✓ +1	6	**C**	22			
✓ +1,2	✓ +1,2	✓ +2	7	**Y**	21	✓ +2		
✓ +1,2	✓ +1,2	✓ +1, 2	8	**W**	20	✓ +2		
✓ +1,2	✓ +1,2	✓ +2	9	**G**	19	✓ +2	✓ +2	
✓ +2	✓ +1,2	✓ +2	10	**G**	18	✓ +2	✓ +2	
	✓ +1,2	✓ +1,2	11	**C**	17			✓ +2
✓ +2	✓ +2	✓ +2	12	**P**	16	✓ +1,2	✓ +2	✓ +2
✓ +1,2	✓ +1,2	✓ +2	13	**W**	15		✓ +1,2	✓ +1
✓ +2	✓ +2	✓ +2	14	**G**	14	✓ +1,2	✓ +2	✓ +1
✓ +1,2	✓ +1,2	✓ +1,2	15	**Q**	13	✓ +1,2	✓ +2	
✓ +2	✓ +1,2		16	**N**	12	✓ +2	✓ +1,2	✓ +1
✓ +2	✓ +2	✓ +1	17	**C**	11	✓ +1	✓ +1	✓ +1
	✓ +2		18	**Y**	10	✓ +1,2	✓ +1,2	✓ +1, 2
			19	**P**	9	✓ +1,2	✓ +1,2	✓ +1,2
			20	**E**	8	✓ +1	✓ +1	✓ +1
			21	**G**	7	✓ +1	✓ +1	✓ +1
✓ +2			22	**C**	6	✓ +1	✓ +1	✓ +1
			23	**S**	5	✓ +1	✓ +1	✓ +1
			24	**G**	4	✓ +1	✓ +1	✓ +1
			24	**P**	3	✓ +1	✓ +1	✓ +1
			26	**K**	2	✓ +1	✓ +1	✓ +1
			27	**V**	1	✓ +1	✓ +1	✓ +1

Table-d : Defensin $[M+5H]^{5+}$ = 690.1198

Identified sequence ions of defensin. Collision energy of 35eV appears to be optimal.

20eV	25eV	30eV	35eV	40eV	b-ions		y-ions	20eV	25eV	30eV	35eV	40eV
					1	A	30					
✓(+2	✓+1		✓ +1	✓ +1	2	C	29					
		✓+1	✓ +1	✓ +1	3	Y	28					
				✓ +1	4	C	27					
✓+1,2	✓+1,2	✓+1	✓+1,2	✓ +1	5	R	26			✓ +3	✓ +4	
✓+1,2	✓+2			✓ +1	6	I	25	✓ +3		✓ +3		
✓+2	✓+2	✓+2	✓ +1,2	✓ +1	7	P	24	✓ +3		✓ +3		
✓+2		✓+1		✓ +1	8	A	23	✓ +3	✓ +3	✓ +3	✓ +3	
✓+2		✓+2	✓+1	✓ +1	9	C	22	✓ +3	✓ +3	✓ +3		
✓+2		✓+1	✓ +1		10	I	21			✓ +2	✓ +3	
		✓+1,2			11	A	20				✓ +2	
				✓ +2	12	G	19					
	✓+1				13	E	18					
			✓+2		14	R	17	✓ +3	✓ +2		✓ +3	
		✓+3	✓ +3		15	R	16					
			✓+2		16	Y	15					
		✓+2	✓ +3		17	G	14					
		✓+2	✓+2		18	T	13					
			✓+3		19	C	12					
					20	I	11			✓ +1	✓ +1	
					21	Y	10		✓ +1	✓ +1		
			✓+3		22	Q	9		✓ +1	✓ +1,2	✓ +1	
		✓+3	✓+3		23	G	8	✓ +2			✓ +1	
✓+3		✓+3			24	R	7	✓ +2				
✓+3	✓+3	✓+3	✓+4		25	L	6					
	✓+3	✓+3	✓ +4		26	W	5					
	✓+3	✓+3			27	A	4	✓ +1				
	✓+3				28	F	3	✓ +1	✓ +1			
					29	C	2	✓ +1	✓ +1		✓ +1	✓+1
					30	C	1	✓ +1	✓ +1		✓ +1	✓+1

Table-e: Orexin-A $[M+5H]^{5+}$ = 713.5572

The peptide contains a modified glutamic acid at the N-terminus, pyroglutamic acid. Labeled as 'u'. Most of the observed sequence ions were at the higher m/z range of the spectrum. This might be because of the absence of the amino group at the N-terminus. 40eV seems to be the best collision energy.

25eV	30 eV	35eV	40eV	b-ions		y-ions	25eV	30 eV	35eV	40eV
				1	u	33				
				2	P	32				
				3	L	31				
				4	P	30				
		✓ (+2)		5	D	29				
				6	C	28				
				7	C	27				
			✓ (+2)	8	R	26				
				9	Q	25				
				10	K	24				
				11	T	23				
			✓ (+2)	12	C	22				
			✓ (+2)	13	S	21				
			✓ (+2)	14	C	20				
			✓ (+2)	15	R	19				
			✓ (+2)	16	L	18				
		✓ (+2)		17	Y	17				
✓(+2)	✓(+2)	✓ (+2)	✓ (+2,3)	18	E	16				
	✓(+2)	✓ (+2,3)	✓ (+2,3)	19	L	15				
✓(+2)	✓(+2)	✓ (+2,3)	✓ (+2,3)	20	L	14				
✓(+3)	✓(+2,3)	✓ (+2,3)	✓ (+2,3)	21	H	13				
✓(+3)	✓(+2,3)	✓ (+2,3)	✓ (+2)	22	G	12				
✓(+3)	✓(+2,3)	✓ (+2,3)	✓ (+2)	23	A	11				
✓(+3)	✓(+2,3)	✓ (+2,3)	✓ (+2)	24	G	10				
	✓(+2,3)	✓ (+2,3)	✓ (+2,3)	25	N	9				
✓(+3)	✓(+3)			26	H	8				
✓(+3)	✓(+3)			27	A	7				
✓(+3)	✓(+3)			28	A	6				
✓(+3)	✓(+3)	✓ (+3)		29	G	5				
✓(+3)	✓(+3)			30	I	4				
✓(+3)				31	L	3	✓ (+1)	✓ (+1)		
		✓ (+3)	✓ (+3)	32	T	2	✓ (+1)	✓ (+1)	✓(+1)	✓(+1)
				33	L	1	✓ (+1)	✓ (+1)	✓(+1)	✓(+1)

Appendix-2

After the peptides were fragmented by collision energy ramps of 20-30eV, 25-35eV, 25-40eV and 30-40 eV, the collision energy ramps which produced larger number of sequence ions were selected as optimum collision energy ramp for that particular peptide. The peptides were fragmented at by their respective single collision energy (obtained from the above table) and the collision energy ramps. Comparison of optimum single collision energy (CE) and optimum collision energy ramps are given tables, f-l.

Table-f: Identified sequence ions of endothelin at optimum CE, 35 eV.

Insilico m/z	Observed m/z	m/z error	Intensity	b-ion		y-ion	Insilico m/z	Observed m/z	m/z error	Intensity
				1	**C**	21				
191,0485	191,0475	0,001	1.43e4	2	**S**	20				
294,0577	294,0588	-0,0011	2.66e3	3	**C**	19				
381,0897	381,088	0,0017	8.00e3	4	**S**	18				
468,1217	468,1122	0,0095	7.75e3	5	**S**	17				
654,201	654,2035	-0,0025	1.14e4	6	**W**	16				
767,2851	767,2867	-0,0016	8.12e3	7	**L**	15				
882,3121	882,3179	-0,0058	8.91e3	8	**D**	14	892,4128	892,4055	0,0073	1.12e4
1010,407	1010,3972	0,0098	3.82e3	9	**K**	13				
1139,4496	1139,4484	0,0012	8.89e3	10	**E**	12				
621,733	621,7317	0,0013	1.79e3	11	**C**	11				
671,2672	671,2706	-0,0034	3.06e3	12	**V**	10				
1504,5905	1504,5885	0,002	5.23e3	13	**Y**	9				
				14	**F**	8				
1754,6681	1754,6859	-0,0178	3.26e3	15	**C**	7				
946,3672	946,372	-0,0048	3.55e4	16	**H**	6				
1002,9092	1002,9144	-0,0052	2.00e4	17	**L**	5				
1060,4227	1060,4259	-0,0032	3.28e5	18	**D**	4				
1116,9647	1116,9736	-0,0089	2.19e4	19	**I**	3				
1173,5067	1173,5079	-0,0012	3.51e3	20	**I**	2	318,1812	318,1811	0,00	5.9e45
				21	**W**	1	205,0972	205,10	0,0006	1.14e5

Table-g: Sequence ions of endothelin at optimum CE ramp (25-40 eV). The identified ions correspond to the sequence ions identified at the optimum CE.

Insilico *m/z*	Observed *m/z*	*m/z* error	Intensity	**b-ions**		y-ions	Insilico *m/z*	Observed *m/z*	*m/z* error	Intensity
				1	**C**	21				Intensity
191,0485	191,0495	-0,001	9.21e3	2	**S**	20				
294,0577	294,0554	0,0023	4.233e3	3	**C**	19				
381,0897	381,0898	-0,0001	5.58e3	4	**S**	18				
468,1217	468,1239	-0,0022	5.42e3	5	**S**	17				
654,201	654,2029	-0,0019	8.03e3	6	**W**	16				
767,2851	767,2867	-0,0016	6.79e3	7	**L**	15				
882,3121	882,3129	-0,0008	7.47e3	8	**D**	14	892,4128	892,4088	0,004	9.01e3
1010,407	1010,4009	0,0061	3.39e3	9	**K**	13				
1139,4496	1139,4573	-0,0077	4.66e3	10	**E**	12				
621,733	621,7331	-0,0001	2.26e3	11	**C**	11				
1341,5272	1341,5264	0,0008	7.96e3	12	**V**	10				
752,7989	752,8015	-0,0026	2.20e3	13	**Y**	9				
				14	**F**	8				
1754,6681	1754,6775	-0,0094	2.95e3	15	**C**	7				
946,3672	946,3615	0,0057	2.23e4	16	**H**	6				
1002,9092	1002,9039	0,0053	2.13e4	17	**L**	5				
1060,4227	1060,4177	0,005	2.16e5	18	**D**	4				
1116,9647	1116,955	0,0097	1.45e5	19	**I**	3				
				20	**I**	2	318,1812	318,1808	0,0004	1.05e5
				21	**W**	1	205,0972	205,10	0,0009	6.42e5

Table-g: Identified sequence ion of conotoxin-α (z=2) at optimal CE, 35 eV. The ion y7 shows very high intensity even at higher CE. This can be due to bond cleavage enhancement effect of proline (Xa-Pro).

Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity	b-ions		y-ions	Theoretical *m/z*	Observed m/z	Mass deviation	Intensity
				1	C	12				
161,0379	161,038	-0,0002	9.42e3	2	R	11				
264,0471	264,048	-0,0011	2.43e3	3	W	10				
				4	A	9				
				5	C	8				
				6	R	7	445,72	445,715	0,0004	1.36e5
				7	P	6	397,19	397,1861	0,003	3.87e3
				8	D	5	637,27	637,2726	-0,0029	2.28e4
				9	S	4	534,26	534,2645	-0,004	4.78e3
				10	C	3	463,22	463,2282	-0,0048	2.95e3
				11	C	2	277,14	277,1473	-0,0032	3.12e3
				12	G	1	121,04	121,0427	0,0003	4.89e3

Table-h: Identified sequence ion of conotoxin-α (z=2) at its optimal collision energy ramp, 25-40 eV. The ramp scan has produced two extra sequence ions (b5 and y_{10}) compared to the optimum CE.

Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity	b-ion		y-ion	Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity
				1	**C**	**12**				
161,0379	161,0386	-0,0007	7.80e3	2	**R**	**11**				
264,0471	264,0463	0,0008	3.93e3	3	**W**	**10**	598,2495	598,2557	-0,0062	1.11e3
				4	**A**	**9**				
466,1061	466,1129	-0,0068	1.74e3	5	**C**	**8**				
				6	**R**	**7**	445,7154	445,7152	0,0002	2.27e5
				7	**P**	**6**	397,1891	397,1865	0,0026	5.19e3
				8	**D**	**5**	637,2697	637,2685	0,0012	2.52e4
				9	**S**	**4**	534,2605	534,2632	-0,0027	3.73e3
				10	**C**	**3**	463,2234	463,2263	-0,0029	3.28e3
				11	**C**	**2**	277,1441	277,148	-0,0039	2.94e3
				12	**G**	**1**	121,043	121,0416	0,0014	5.00e3

Table-I : Identified sequence ions of neurotoxin at its optimum fixed collision energy, 35 eV

Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity	b-ions		y-ions	Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity
				1	R	27				
				2	S	26				
347,15	347,1493	0,0003	4.75e3	3	C	25				
450,159	450,1584	0,0004	1.84e4	4	C	24				
				5	P	23				
650,221	650,2261	-0,0054	1.78e4	6	C	22				
813,284	813,2932	-0,0091	6.66e4	7	Y	21				
999,363	999,3729	-0,0095	5.03e4	8	W	20				
1056,38	1056,389	-0,0042	1.70e4	9	G	19	969,41	969,4163	-0,0081	3.54e3
1113,41	1113,4131	-0,0068	1.45e4	10	G	18	940,9	940,905	-0,0075	1.23e3
1216,42	1216,4248	-0,0093	1.34e5	11	C	17			0	
657,238	657,2416	-0,0038	1.78e4	12	P	16	860,88	860,8881	-0,006	1.03e5
750,277	750,2791	-0,0017	8.60e3	13	W	15			0	
778,788	778,7959	-0,0077	1.14e4	14	G	14	719,32	719,3231	-0,007	2.83e3
1684,63	1684,6316	-0,004	3.71e3	15	Q	13	690,81	690,801	0,0044	2.73e3
				16	N	12	1252,5	1252,5558	-0,0109	4.18e4
951,344	951,3513	-0,0078	1.35e4	17	C	11	1138,5	1138,5048	-0,0028	3.09e4
1032,88	1032,8824	-0,0072	5.92e3	18	Y	10	518,25	518,252	-0,002	1.08e4
				19	P	9	872,43	872,4369	-0,0074	1.92e5
				20	E	8	775,38	775,3805	-0,0038	7.85e3
				21	G	7	646,33	646,34	-0,0059	2.61e4
				22	C	6	589,31	589,3135	-0,0009	5.07e3
				23	S	5	486,3	486,3056	-0,0021	1.43e4
				24	G	4	399,27	399,2715	0,0014	1.36e4
				25	P	3	342,25	342,251	-0,001	1.25e4
				26	K	2	245,2	245,1968	0,0004	1.76e3
				27	V	1	117,1	117,1031	-0,0009	2.18e3

Table-J: Sequence ions of neurotoxin at its optimum ramp (25-40 eV). More sequence ions were identified compared to the optimum fixed CE.

Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity	b-ion		y-ion	Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity
				1	R	27				
244,1404	244,1371	0,003	4.07e3	2	S	26				
347,1496	347,148	0,002	4.55e3	3	C	25				
450,1588	450,1582	0,001	4.66e4	4	C	24				
				5	P	23				
650,2207	650,2248	-0,004	1.53e3	6	C	22				
813,2841	813,2825	0,002	6.66e4	7	Y	21				
500,1853	500,1878	-0,002	9.93e3	8	W	20	1062,4	1062,4435	0,0044	1.05e4
528,6961	528,6915	0,005	1.56e4	9	G	19	969,41	969,4011	0,0071	1.86e4
557,2068	557,2049	0,002	2.40e4	10	G	18	940,9	940,8849	0,0126	5.22e3
608,7114	608,7059	0,005	1.93e4	11	C	17	912,39	912,3826	0,0041	1.09e4
657,2378	657,2333	0,005	1.74e4	12	P	16	860,88	860,8774	0,0047	1.52e5
750,2774	750,2742	0,003	1.70e4	13	W	15			0	
778,7882	778,7852	0,003	1.82e4	14	G	14	719,32	719,312	0,0041	1.45e4
842,8174	842,8091	0,008	1.12e4	15	Q	13	690,81	690,8033	0,0021	8.12e3
899,8389	899,8371	0,002	1.67e4	16	N	12	626,78	626,7766	-0,0005	5.54e3
951,3435	951,3354	0,008	3.17e4	17	C	11	1138,5	1138,4988	0,0032	3.12e4
1032,875	1032,871	0,005	5.92e3	18	Y	10	518,25	518,2491	0,0009	1.42e4
1081,402	1081,412	-0,010	1.12e3	19	P	9	436,72	436,7204	-0,002	3.65e4
1145,923	1145,911	0,012	3.28e3	20	E	8	775,38	775,3733	0,0034	5.98e3
1174,434	1174,43	0,003	2.17e3	21	G	7	646,33	646,329	0,0051	3.26e4
				22	C	6	589,31	589,3105	0,0021	5.21e3
				23	S	5	486,3	486,3022	0,0013	1.97e4
				24	G	4	399,27	399,2707	0,0007	1.47e4
				25	P	3	342,25	342,2502	-0,0002	1.70e4
				26	K	2	245,2	245,1986	-0,0014	2.39e3
				27	V	1	117,1	117,1027	-0,0005	2.18e3

Table-K: Identified sequence ions of tertiapine at optimum CE of 25 eV

Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity	b-ion		y-ion	Theoretical *m/z*	Observed *m/z*	Mass deviation	Intensity
				1	C	19				
218,059	218,0593	-0,0003	3.19e4	2	N	18	724,3826	724,3787	0,0039	6.60e3
				3	C	17				
				4	N	16				
591,2126	591,2142	-0,0016	8.54e4	5	R	15				
704,2967	704,2991	-0,0024	1.40e5	6	I	14				
817,3807	817,3848	-0,0041	1.89e5	7	I	13				
930,4648	930,4698	-0,005	8.23e4	8	I	12	729,3882	729,3805	0,0077	1.09e4
514,2624	514,2628	-0,0004	6.06e3	9	H	11	672,8461	672,8477	-0,0016	1.05e5
582,7919	582,794	-0,0021	1.12e4	10	P	10	624,3197	624,3216	-0,0019	7.41e4
648,3121	648,3141	-0,002	5.63e3	11	M	9	1110,573	1110,583	-0,0095	3.39e4
699,8167	699,8221	-0,0054	6.23e3	12	C	8	979,5328	979,5342	-0,0014	3.13e4
				13	W	7	876,5236	876,5258	-0,0022	9.17e3
				14	K	6	690,4443	690,4473	-0,003	1.28e4
				15	K	5	562,3494	562,3489	0,000	3.80e4
				16	C	4	434,2544	434,2548	0,000	1.13e4
				17	G	3	331,2452	331,2446	0,001	2.21e4
				18	K	2	274,2238	274,223	0,001	6.27e3
				19	K	1	146,1288	146,1284	0,000	5.73e4

Table-L: Identified sequence ions of tertiapine at optimum CE ramp between 25 and 35 eV. Ion b17 was not identified when the peptide was fragmented at its optimal single collision energy.

In silico	Observed	Deviation	Intensity	b-ion		y-ion	In silico	Observed	Deviation	Intensity
				1	C	19				
218,0594	218,0596	-0,0002	3.33e4	2	N	18	724,3826	724,3906	-0,0080	1.34e3
				3	C	17				
				4	N	16				
591,2126	591,2143	-0,0017	1.29e5	5	R	15				
704,2967	704,2986	-0,0019	9.18e4	6	I	14				
817,3807	817,3839	-0,0032	7.81e4	7	I	13				
930,4648	930,4665	-0,0017	3.06e4	8	I	12	729,3882	729,3828	0,0054	1.15e3
514,2624	514,2628	-0,0004	2.19e3	9	H	11	672,8461	672,845	0,0011	1.58e4
582,7919	582,7923	-0,0004	3.99e3	10	P	10	624,3197	624,3214	-0,0017	1.62e4
				11	M	9	555,7903	555,7932	-0,0029	6.04e3
699,8167	699,8068	0,0099	1.07e3	12	C	8	979,5328	979,5291	0,0037	9.32e3
				13	W	7	876,5236	876,5176	0,0060	3.29e3
				14	K	6	690,4443	690,4464	-0,0021	6.91e3
				15	K	5	562,3494	562,3488	0,0006	1.81e3
				16	C	4	434,2544	434,2558	-0,0014	2.33e4
448,8998	448,9001	-0,0003	5.23e3	17	G	3	331,2452	331,2448	0,0004	6.66e3
				18	K	2	274,2238	274,2258	-0,0020	5.64e3
				19	K	1	146,1288	146,1284	0,0004	4.31e4

Appendix 3:

Table-M: Identified sequence ions (colored) of selectively alkylated conotoxin-α by NMM and NCM. The two structural isomers are presented side by side. (Left) C2 and C12 of the peptide are bound to NMM. (right) C2 and C12 of the peptide are bound to NCM and in the same manner C3 and C12 are bound to NMM. y1, y2, y3, y10 and b2 and b8 are signature ions that can help us to differentiate the isomers. b2, y1 and y10 were identified.

C2 & C12 -NMM , C3 &C8 - NCM							**C2 & C12-NCM, C3 & C12- NMM**						
b	**b^{+2}**				**y**	**y^{+2}**	**b**	**b^{+2}**				**y**	**y^{+2}**
---	---	1	**G**	12	---	---	---	---	1	**G**	12	---	---
272.0700	---	2	**u**	11	1878.7543	939.8808	**340.1326**	---	2	**u**	11	1878.7543	939.8808
554.1738	---	3	**v**	10	1664.7130	**832.8602**	554.1738	---	3	**v**	10	1596.6504	**798.8289**
641.2058	---	4	**S**	9	1382.6092	691.8083	641.2058	---	4	**S**	9	1382.6092	691.8083
756.2327	---	5	**D**	8	1295.5772	648.2922	756.2327	---	5	**D**	8	1295.5772	648.2922
853.2855	---	6	**P**	7	1180.5503	590.7788	853.2855	---	6	**P**	7	1180.5503	590.7788
1009.3866	505.1969	7	**R**	6	1083.4975	542.2524	1009.3866	505.1969	7	**R**	6	1083.4975	542.2524
1291.4904	646.2489	8	**v**	5	927.3964	464.2018	1223.4278	612.2176	8	**v**	5	927.3964	464.2018
1362.5275	681.7674	9	**A**	4	645.2926	323.1499	1294.4649	647.7361	9	**A**	4	**713.3552**	357.1812
1548.6069	774.8071	10	**W**	3	574.2555	287.6314	1480.5443	740.7758	10	**W**	3	**642.3181**	321.6627
1704.7080	852.8576	11	**R**	2	**388.1761**	194.5917	1636.6454	818.8263	11	**R**	2	**456.2387**	228.6230
---	---	12	**u**	1	**232.0750**	---	---	---	12	**u**	1	**300.1376**	---