

**Bottom up proteomics of cryptocyanin protein and anti-tumor  
Thomsen–Friedenreich neoglycoconjugate vaccine**

by

**Wael Demian**

A thesis submitted to the  
School of Graduate Studies in partial fulfillment of the requirements for the degree of  
Master of Science

May 2015

Department of Biochemistry  
Memorial University of Newfoundland

St. John's

Newfoundland

## ABSTRACT

In this thesis, two different kinds of complex biomolecules were examined: cryptocyanin protein of snow crab (*Chionoecetes opilio*) and carbohydrate-protein neoglycoconjugate vaccine using mass spectrometry (MS) and tandem mass spectrometry (MS/MS).

The main objective of the first study was to determine structural changes in snow crab cryptocyanin protein. Accordingly, cryptocyanin was isolated and purified from molting and non molting snow crabs. The purified molting and non molting cryptocyanins were sequenced by using matrix assisted laser desorption/ionization mass spectrometry (MALDI-MS) and collision induced dissociation tandem mass spectrometry (CID-MS/MS). The structural differences between molting and non-molting cryptocyanin were reported using bottom-up proteomics. The molecular masses of the snow crab during molting and non-molting stages were found to be respectively 67.6 kDa and 68.1 kDa.

In the second study, two anti-tumor Thomsen-Friedenreich neoglycoconjugate vaccines were prepared by the Michael addition reaction, and their molecular composition was confirmed by MALDI-TOF-MS analysis. *De novo* peptide sequencing using electrospray ionization in the positive ion mode mass spectrometry (ESI-MS, + ion mode) and low-energy CID-MS/MS of trypsin digests using a quadrupole-quadrupole-time of flight –hybrid tandem mass spectrometer (QqTOF-MS/MS). The CID-analysis allowed us to identify three glycation sites for the glycoconjugate prepared with a TF: protein ratio of 2:1, and 14 glycation sites for the glycoconjugate prepared with a TF: protein ratio of 8:1. It should be noted that this work represents the only published report for the

determination of the exact positions of the carbohydrate haptens on this unique C-S linked glycoconjugate. In addition, this investigation constitutes a versatile example for the identification of accurate quality control necessary in commercial production of glycoconjugate vaccines. Indeed, the glycopeptides isolated were fully characterized during this work and may well represent useful reference compounds that can be used in standardization analyses.

The presented two applications of this research indicate that the cohesive and clever application of MS continue to grow due to its high sensitivity, and its ability to analyze high molecular mass biomolecules such as the two complex biomolecules.

In summary this thesis has presented a major dynamic role for the identification, characterization and sequencing of two novel complex biomolecules namely the molting and non molting and cryptocyanin proteins and the two anti-tumor Thomsen-Friedenreich neoglycoconjugate vaccines. These biomolecules were chosen as elegant applications of MS/MS in the fields of proteomics and glycomics vaccine.

## ACKNOWLEDGMENTS

I would like to express my deepest sense of gratitude and sincere respect toward my supervisor, Dr. Joseph H. Banoub for providing me the precious opportunity to conduct this research under his direction during these two years. His inspiring guidance, valuable discussion, cooperation and constant encouragement are commendable. His support gave me ability to keep a delicate balance between career and family.

I owe special gratitude to Dr. Edward Randell (supervisory committee) for his valuable support through the previous couple of years. The support and encouragement from Dr. Randell helped me to overcome many difficulties and successfully complete my study. His inspiration, and guidance, which resulted in this work, will remain a lifelong memory. His continued support and guidance has been both educational and inspiring.

I also would like to thank Dr. Robert Brown (supervisory committee) for reviewing a draft of the thesis and for his appreciated advice. Thanks are also extended to Dr. Travis Fridgen (supervisory committee) for his guidance and his support.

I greatly appreciate to the School of Graduate Studies, Biochemistry Department, Memorial University of Newfoundland and the Department of Fisheries and Oceans (DFO) for their financial support to my study program.

I am particularly grateful to my mom and my defunct dad for their endless love. Finally, I would like to thank my lovely family for their everlasting love. My deepest appreciation goes to my wife “Evon” for her patience and understanding throughout this period. My daughter “Josiane” and my upcoming son “Jason” who are the source of happiness for my life deserve many thanks.

## TABLE OF CONTENTS

<b>ABSTRACT</b>	ii
<b>ACKNOWLEDGMENTS</b>	iv
<b>LIST OF FIGURES</b>	x
<b>LIST OF SCHEMES</b>	xiii
<b>LIST OF TABLES</b>	xiv
<b>LIST OF ABBREVIATIONS</b>	xv
<b>LIST OF APPENDICES</b>	xix
<b>CHAPTER I: INTRODUCTION</b>	<b>1</b>
1.1 Mass spectrometry	1
1.1.1 Types of ionization techniques	4
1.1.2 Matrix assisted laser desorption ionization (MALDI)	5
1.1.2.1 Principle of MALDI	5
1.1.3 Electrospray ionization (ESI)	10
1.1.3.1 Principle of ESI	10
1.1.4 Mass analyzers	13
1.1.4.1 Quadrupole analyzer	13
1.1.4.2 Time of flight (TOF) analyzer	16

1.1.4.3 Hybrid quadrupole/time-of-flight (QTOF) mass spectrometer	19
1.1.5 Tandem mass spectrometry	21
1.1.5.1 Tandem in space and tandem in time	21
1.1.6 Protein mass spectrometry	23
1.1.7 Gas phase fragmentation of proteins and carbohydrates	26
1.2 Snow Crab ( <i>Chionoecetes opilio</i> )	30
1.2.1 Cryptocyanin	31
1.2.2 Cryptocyanin functions	32
1.3 Vaccines	33
1.3.1 Live, attenuated and inactivated vaccines	33
1.3.2 Toxoid and subunit vaccines	34
1.3.3 DNA and recombinant vector vaccines	34
1.3.4 Conjugate vaccines	35
1.3.5 Principle of glycoprotein vaccines	39
1.3.6 Thomsen–Friedenreich antigen (TF)	42
1.3.7 Strategy for characterization of glycoconjugate vaccines using MS	44
1.3.8 Molecular mass and carbohydrate-to-protein ratio determination	44
1.3.9 Glycation sites determination	44

1.4 Research objectives	45
1.5 Hypothesis	45
1.6 Summary	46
<b>CHAPTER II: DIFFERENTIATION BETWEEN THE CRUSTACEAN CRYPTOCCYANIN PROTEIN DURING MOLTING AND NON-MOLTING PROCESSES OF THE SNOW CRAB (CHIONOECETES OPILIO) USING MATRIX-ASSISTED LASER DESORPTION/IONIZATION MASS SPECTROMETRY AND TANDEM MASS SPECTROMETRY</b>	66
2.1 Introduction	67
2.2 Material and methods	68
2.2.1 Chemicals and reagents	68
2.2.2 Sample preparation	68
2.2.3 Extraction of snow crab hemolymph	69
2.2.4 Protein precipitation	69
2.2.5 Electrophoresis	69
2.2.6 Confirmation of the identity of the cryptocyanin gel band	70
2.2.7 Tryptic digestion of cryptocyanin	70
2.2.8 MALDI-TOF/TOF-MS analysis	71
2.2.9 MALDI-QqTOF-MS	71
2.2.10 MALDI-QqTOF-CID-MS/MS	72

2.2.11 Peptide mass fingerprint and MS/MS ion searches	73
2.3 Result and discussion	73
2.3.1 SDS-PAGE purification of the cryptocyanin	73
2.3.2 MALDI-TOF/TOF-MS determination of the molecular mass of the cryptocyanin proteins during the non-molting and molting stages	78
2.3.3 Peptide mass fingerprinting of the molting cryptocyanin	80
2.3.4 MALDI-CID-MS/MS sequencing of the digested peptides obtained from the molting cryptocyanin protein	83
2.3.5 Peptide mass fingerprinting of the non-molting cryptocyanin	85
2.3.6 MALDI-CID-MS/MS sequencing of the digested peptides obtained from the non-molting cryptocyanin protein	88
2.3.7 Structural similarities and differences between the identified peptides of the digested cryptocyanin protein in molting and non-molting snow crab	90
2.4 Conclusion	102
<b>CHAPTER III: DIRECT TARGETED GLYCATION OF THE FREE SULFHYDRYL GROUP OF CYSTEINE RESIDUE (CYS-34) OF BSA. MAPPING OF THE GLYCATION SITES OF THE ANTI-TUMOR THOMSEN-FRIEDENREICH NEOGLYCOCONJUGATE VACCINE PREPARED BY MICHAEL ADDITION REACTION</b>	<b>107</b>
3.1 Introduction	108
3.2 Material and methods	110
3.2.1 Preparation of the TF-BSA vaccine conjugate	110

3.2.2 Trypsin digestion of the glycoconjugate	110
3.2.3 MALDI-TOF-MS analysis	111
3.2.4 LC-ESI-QqTOF-CID-MS/MS analysis	112
3.2.5 MS/MS ion searches	113
3.3 Result and discussion	113
3.3.1 Synthesis of the TF-BSA vaccine conjugate by the Michael addition reaction	113
3.3.2 MALDI-TOF-MS analysis of the TF-BSA glycoconjugates	118
3.3.3 LC-ESI-MS and CID-MS/MS analyses of tryptic digest of the glycoconjugate with a TF: BSA ratio of 2:1 and 8:1	122
3.4 Conclusion	146
<b>CHAPTER IV : GENERAL CONCLUSION</b>	<b>156</b>
4.1 Future aspects	160
4.2 Conclusion	161
<b>APPENDIX A</b>	<b>165</b>
<b>APPENDIX B</b>	<b>176</b>

## List of Figures

<b>Figure 1.1</b> Components of a mass spectrometer.	3
<b>Figure 1.2</b> Schematic representation of the ionization of the matrix and the analyte in a MALDI source.	9
<b>Figure 1.3</b> Schematic representation of the production of charged droplets during ESI process.	12
<b>Figure 1.4</b> Schematic representation of a quadrupole mass filter	15
<b>Figure 1.5</b> Schematic representation of a conventional TOF analyzer (A) and a reflectron-type TOF (B)	18
<b>Figure 1.6</b> Schematic representation of QTOF instrument (QSTAR hybrid QqTOF).	20
<b>Figure 1.7</b> Schematic of tandem mass spectrometry.	22
<b>Figure 1.8</b> Workflow in bottom up proteomics.	25
<b>Figure 1.9</b> Peptide fragmentation nomenclature.	27
<b>Figure 1.10</b> Carbohydrate fragmentation nomenclature.	29
<b>Figure 1.11</b> Lipopolysaccharide of gram negative bacterial cell wall.	36
<b>Figure 1.12</b> Composition of the glycoconjugate vaccine	37
<b>Figure 1.13</b> Generation of the immune responses against the infection.	40
<b>Figure 1.14</b> Generation of the immune responses against oligosaccharide antigen-carrier protein conjugate vaccines.	41
<b>Figure 1.15</b> Typical structure of representative TF antigen bound to the hydroxyl group of serine (Ser) or threonine (Thr) on mucin glycoprotein accumulated on the cancer cell.	43
<b>Figure 2.1</b> SDS-PAGE comparison of hemolymph proteins sampled from juvenile crabs during their molting and nonmolting processes	75
<b>Figure 2.2</b> SDS-PAGE separation of molting crab hemolymph proteins stained with Coomassie Blue stain: (A): no eyestalk ablation was performed; (B): after 7	77

days ablation of eyestalk; (C): after 21 days eyestalk ablation; and (D): after 14 days eyestalk ablation

**Figure 2.3** MALDI-TOF-MS of the intact cryptocyanin protein (a) during the molting period and (b) during the non-molting period. 79

**Figure 2.4** MALDI-QqTOF-MS of 'in gel' trypsin digestion of snow crab during the molting process 81

**Figure 2.5** MALDI-QqTOF-MS of 'in gel' trypsin digestion of snow crab during the non-molting process 86

**Figure 2.6** Showing the matched peptides of cryptocyanin in different species. The matched peptides during molting process were coloured in red. The green matched peptides were sharing in both molting and non-molting stages. The underlined sequence indicates to isobaric peptides 92

**Figure 2.7** MALDI-CID-MS/MS of the  $[M+H]^+$  ions obtained from the tryptic peptides of the molting and non-molting cryptocyanins: (a) HAWAVATNKR at  $m/z$  1082.5324 and (b) FGPPFPVR at  $m/z$  916.5099 95

**Figure 2.8** MALDI-CID-MS/MS of the  $[M+H]^+$  ions of both the exclusive tryptic peptides of molting cryptocyanin: (a) DGNGAIIPFDEGR at  $m/z$  1360.6744 and (b) GVQPDKRPFGYPLDRR at  $m/z$  1744.8554 98

**Figure 2.9** MALDI-CID-MS/MS of the  $[M+H]^+$  ions of the tryptic changed and/or modified peptides obtained from nonmolting cryptocyanin (a) FDAER at  $m/z$  637.2000, (b) DPAFFR at  $m/z$  752.3000, (c) YMDNIFR at  $m/z$  958.4000, (d) RPHGYPLDR at  $m/z$  1110.5000, (e) LNHEEFSYK at  $m/z$  1166.5000, and (f) VYEDIRDPHLK at  $m/z$  1384.7000 100

**Figure 3.1** Synthesis of Thomsen–Friedenreich antigen–BSA glycoconjugate vaccines using a Michael addition 117

**Figure 3.2** a) MALDI-TOF-MS analysis of TF:BSA glycoconjugates with ratio= 2:1. b) MALDI-TOF-MS analysis of TF:BSA glycoconjugates with ratio = 8:1. 119

**Figure 3.3** CID-MS/MS spectra of identified glycated peptide GLVLIAFSQYLQQC\*PFDEHVK 5 (Cys 34) at  $m/z$  1015.4614 (+3) 129

**Figure 3.4** CID-MSMS spectra of identified glycated peptide K\*VPQVSTPTLVEVSR 9 (Lys412) at  $m/z$  731.4685 (+3) 130

**Figure 3.5** Product ions signature of the carbohydrate antigen, observed during the CID-MS/MS analysis of the glycated peptides 132

**Figure 3.6** BSA sequence where the 14 glycation sites are indicated by red asterisk 143  
= identified peptides in TF: BSA ratio = 8: 1. The three underlined aminoacids =  
identified in TF:BSA ratio= 2: 1 and in TF:BSA ratio = 8:1

**Figure 3.7** 3D structure of BSA generated from the X-ray data using the software 145  
Discovery Studio v 3.1

## List of Schemes

<b>Scheme 3.1</b> Structure of the TF antigen and its chemical transformation into a suitable Michael acceptor 3	115
<b>Scheme 3.2</b> CID-MS/MS fragmentation of the TF-glycopeptide attached with the C—S bond at Cys 34	134
<b>Scheme 3.3</b> CID-MS/MS fragmentation of the TF-glycopeptide attached with C—N bond at Lys 412	135

## List of Tables

<b>Table 1.1</b> Different types of matrices used in MALDI	7
<b>Table 2.1</b> List of matching peptides obtained from the peptide mass fingerprint (PMF) obtained by MALDI-QqTOF-MS analysis of the digested molting cryptocyanin	82
<b>Table 2.2</b> CID-MS/MS sequences of the peptides obtained by the digested cryptocyanin molting protein	84
<b>Table 2.3.</b> List of matching peptides obtained from the peptide mass fingerprint (PMF) after MALDI-QqTOF-MS analysis of the digests of the gel band at 75 kDa of non-molting stage	87
<b>Table 2.4</b> CID-MS/MS sequences of the peptides obtained the digested cryptocyanin non-molting protein	89
<b>Table 3.1</b> MALDI-TOF-MS of BSA and TF: BSA glycoconjugate vaccines and the calculated carbohydrate:BSA ratios	121
<b>Table 3.2</b> CID-MS/MS analysis of the glycated peptides ions obtained by the trypsin digestion of the TF:BSA vaccine with ratio 2:1	124
<b>Table 3.3</b> CID-MS/MS analysis of the glycated peptides ions obtained by the trypsin digestion of the TF:BSA vaccine with ratio 8:1	126
<b>Table 3.4</b> Product ions obtained following the CID-MS/MS analysis of GLVLIAFSQYLQQC*PFDEHVK <b>5</b> (Cys34) ion at $m/z$ 1015.46(+3)	137
<b>Table 3.5</b> Product ions obtained following the CID-MS/MS analysis of K*VPQVSTPTLVEVSR <b>6</b> (Lys412) ion at $m/z$ 731.46(+3)	140

## List of Abbreviations

2-D	2 Dimensional
3D	3 Dimensional
$\alpha$ -CHCA	$\alpha$ -cyano-4-hydroxycinnamic acid
ACN	Acetonitrile
AP	Atmospheric pressure
APCI	Atmospheric pressure chemical ionization
APPI	Atmospheric pressure photo ionization
Ar	Argon
BCR	B cell receptor
BSA	Bovine serum albumin
CD	Cluster designation
CE	Capillary electrophoresis
CI	Chemical ionization
CID	Collision induced dissociation
CP	Current potential
CRM197	Cross-reactive material 197
Cys	Cysteine
Cys CAM	Carbamidomethyl cysteine
Da	Dalton
DESI	Desorption electrospray ionization

DP	Declustering potential
DT	Diphtheria toxoid
DTT	Dithiotheitol
ECD	Electron induced dissociation
EI	Electron ionization
ESI	Electrospray ionization
ETD	Electron transfer dissociation
FA	Formic acid
FAB	Fast atom bombardment
FP	Focusing potential
FT-ICR	Fourier transform ion cyclotron resonance
FWHM	Full width at half maximum
Gal	Galactose
GalNAc	N-acetylgalactosamine
GC	Gas chromatography
GlcN	N-acetylgalactosamin
HPLC	High pressure liquid chromatography
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IRMPD	Infrared multiphoton dissociation
KLH	Keyhole limpet hemocyanine
LC	Liquid chromatography

LIAD-MS	Laser induced acoustic desorption mass spectrometry
LPS	Lipopolysaccharide
Lys	Lysine
MCP	Microchannel plate
Met	Methionine
MIH	Molting inhibiting hormone
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MUC1	Mucin1
<i>m/z</i>	Mass to charge ratio
Nd:YAG	Neodymium doped yttrium aluminium garnet
Nd:YLF	Neodymium-doped yttrium lithium fluoride
NCBIInr	National Center for Biotechnology Information Non-redundant protein database
Oa	Orthogonal acceleration system
PMF	Peptide mass fingerprint
PTM	Post translational modifications
Q	Quadrupole
QIT	Quadrupole ion trap
QTOF	Quadrupole time of flight
RF	Radiofrequency
SA	Sinapinic acid

SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SELDI	Surface enhanced laser desorption ionization
Ser	Serine
SID	Surface induced dissociation
TACA	Tumor associated carbohydrate antigen
TCR	T cell Receptor
TF	Thomsen–Friedenreich
TFA	Trifluoroacetic acid
Thr	Threonine
TN	N-acetylglucosamine antigen
TOF	Time of flight
TT	Tetanus toxid
UV	Ultraviolet
XO-SG	X organ-sinus gland system

## List of Appendices

- Appendix A:** Additional tables for chapter 2: Differentiation between the crustacean cryptocyanin protein during molting and non-molting processes of the snow crab (*Chionoecetes opilio*) using matrix-assisted laser desorption/ionization mass spectrometry and tandem mass spectrometry 165
- Appendix B:** Additional figures and tables for chapter 3: Direct targeted glycation of the free sulfhydryl group of cysteine residue (Cys 34) of BSA. Mapping of the glycation sites of the anti-tumor Thomsen-Friedenreich neoglycoconjugate vaccine prepared by Michael addition reaction. 176

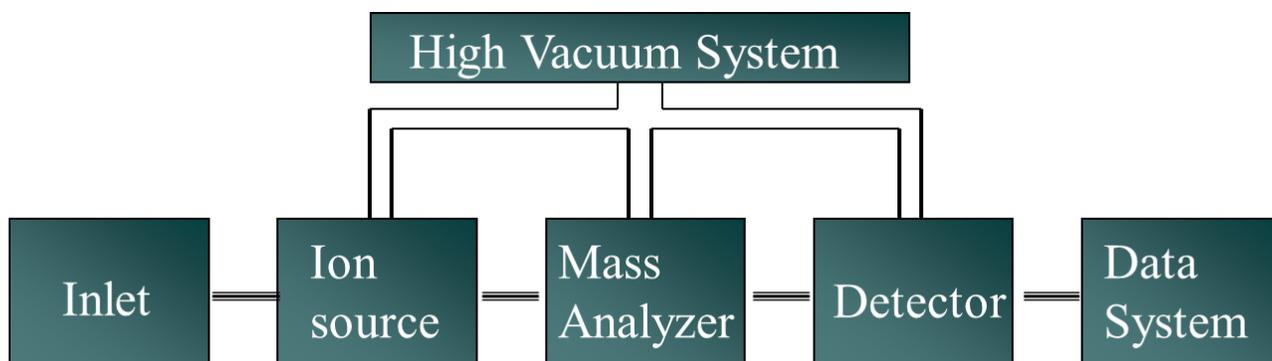
## Chapter I: Introduction

### 1.1 Mass spectrometry

Mass spectrometry (MS) is an outstanding analytical technique for the characterization of different biomolecules ranging from small (such as sugars) to larger molecules (such as proteins).<sup>[1-10]</sup> Thus, MS is extensively used in proteomics,<sup>[1]</sup> glycomics,<sup>[2]</sup> metabolomics,<sup>[3]</sup> lipidomics,<sup>[4]</sup> and in oligonucleotide analysis.<sup>[5]</sup> It is used for determining the molecular mass of chemical or biological compounds.<sup>[1-10]</sup> Additionally, it is an exceptional technique for elucidating the structure of biomolecules.<sup>[7-10]</sup> This technique holds an integral place among all analytical techniques due to its sensitivity, its low sample consumption, and its relative rapidity of analysis.<sup>[8-10]</sup>

MS determines the mass of molecules, whereby an electrical charge is placed on the molecule and the resulting ions are separated by their mass to charge ratio ( $m/z$ ).<sup>[1-10]</sup> The mass spectrometer is typically composed of three main parts: the ionization source, the analyzer, and the detector (**Figure 1.1**).<sup>[11]</sup> In the ionization source, the molecule is heated to a certain temperature, and then it undergoes ionization under atmospheric pressure in the case of electrospray ionization (ESI), atmospheric pressure photoionization (APPI) and atmospheric pressure chemical ionization (APCI). Conversely, the molecule undergoes ionization under a vacuum in the case of matrix assisted laser desorption ionization (MALDI), chemical ionization (CI), and electron ionization (EI). These ionization sources are conducted in a low pressure vacuum in order to minimize collision of ions with gas molecules. It is in the ionization source that the

ions are created. In the analyzer, the molecular and fragment ions are essentially separated according to their  $m/z$ . Finally, the detector collects ions, quantifies their intensities and amplifies their signals.<sup>[6-8]</sup> After the detector, a computer system processes the data and generates a mass spectrum, which specifies the variation of ion current observed according to the  $m/z$  ratio. The ionization sources, the analyzers, and the detectors can be oriented in a variety of configurations and accordingly can create a variety of mass spectrometer types.<sup>[6]</sup> However, the choice of ionization source and analyzer type will depend on the nature of the sample and on the data type desired (spectrum or image).<sup>[6-8]</sup>



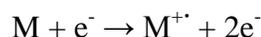
**Figure 1.1** Components of a mass spectrometer. The ion source, mass analyzer and detector are conducted under high vacuum in case of MALDI, CI and EI, but at the atmospheric pressure in case of ESI, APPI and APCI.

### 1.1.1 Types of ionization techniques

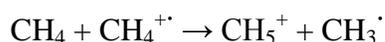
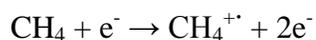
The ionization techniques are classified into hard ionization techniques and soft ionization techniques. The terminology of hard and soft is related to the internal energy used for the formation of ions from the analyte.<sup>[12-15]</sup> High internal energy is used in the hard techniques such as CI and EI.<sup>[12-14]</sup> Alternatively, a low internal energy has been used in soft techniques such as MALDI, ESI, fast atom bombardment (FAB), APPI, APCI, and more recently, desorption electrospray ionization (DESI).<sup>[14,15]</sup>

Both EI and CI are traditional techniques, working under high vacuum. In addition these techniques have been used for small molecular masses.<sup>[12-15]</sup> They can also be coupled with gas chromatography (GC) instrument, so the sample can be introduced into the source in the gas phase.

In EI, a molecule “M” is bombarded with energetic electrons that are produced by heating a tungsten wire with an internal energy of typically 70 eV.<sup>[12-14]</sup> The electrons react in the gas phase with the analyte (M) to produce a radical cation:



In CI, ions are generated by the interaction between a molecule “M” and a reactive gas like methane.<sup>[14,15]</sup> Unlike EI which generates a radical cation, CI produces protonated molecules  $[M+H]^+$ . The ion protonation occurs through two steps: the first step involves a bombardment of electrons with a neutral gas (example: methane, CH<sub>4</sub>) with an electron, and the second step is the protonation of the molecule “M” by the reactive gas which is generated from the first step.





CI is a softer method than EI, because there is generally less energy deposited into the newly formed ion, and accordingly there is generally less analyte fragmentation, although it is still considered to be a hard technique.<sup>[14,15]</sup> Because of the smaller amounts of fragmentation, the mass spectra generated by CI are generally simpler than generated by EI, and interpretation easier.<sup>[14,15]</sup> In this thesis, MALDI and ESI were used for the structural characterization of different biomolecules.

### **1.1.2 Matrix assisted laser desorption ionization (MALDI)**

MALDI is a soft ionization technique which was revolutionized analysis of large biomolecules.<sup>[16]</sup> It is widely used for the analysis of large biomolecules with conventional single stage direct MS analyzers (example: MS analysis of intact protein) and tandem mass spectrometry using collision induced dissociation (CID-MS/MS).<sup>[6-21]</sup> MALDI is well-suited for the direct analysis of biomolecules due to its high sensitivity, and high tolerance for salts and other contaminants.<sup>[18-23]</sup> MALDI can be used in the analysis of large and/or labile molecules, lipids, peptides, proteins, oligonucleotides, and synthetic polymers.<sup>[24-28]</sup> In this thesis, MALDI was used to characterize the structural changes of cryptocyanin protein during the molting and non-molting of snow crab, and for determining the ratio of hapten:protein for two synthetic vaccines.

#### **1.1.2.1 Principle of MALDI**

In typical MALDI experiments, the sample is mixed with a matrix on a plate. The matrix molecules must embed and isolate analytes by co-crystallization. MALDI matrices are typically aromatic conjugated molecules that need to be soluble in solvents, and compatible with analytes of interest.<sup>[16-21]</sup> Matrices must be able to transfer or accept

protons from non-volatile analytes.<sup>[18-20]</sup> They must be chemically inert – no reactivity with analytes and stable in the high vacuum applied in MALDI-MS.<sup>[16-21]</sup> The choice of matrix depends on the nature of both the laser used in the MALDI instrument and the analyzed sample.<sup>[16-21]</sup> Sinapinic acid (SA) is commonly used for protein, and  $\alpha$ -cyano-4-hydroxycinnamic acid ( $\alpha$ -CHCA) for peptides (see **Table 1.1**). Typically, MALDI employs ultraviolet (UV) lasers such as nitrogen lasers (337 nm) or frequency-tripled and quadrupled neodymium-doped yttrium aluminum garnet (Nd:YAG) lasers (355 nm and 266 nm, respectively).<sup>[18-20]</sup> Neodymium-doped yttrium lithium fluoride (Nd:YLF) laser (349 nm) is also used in some commercial systems.<sup>[18-20]</sup>

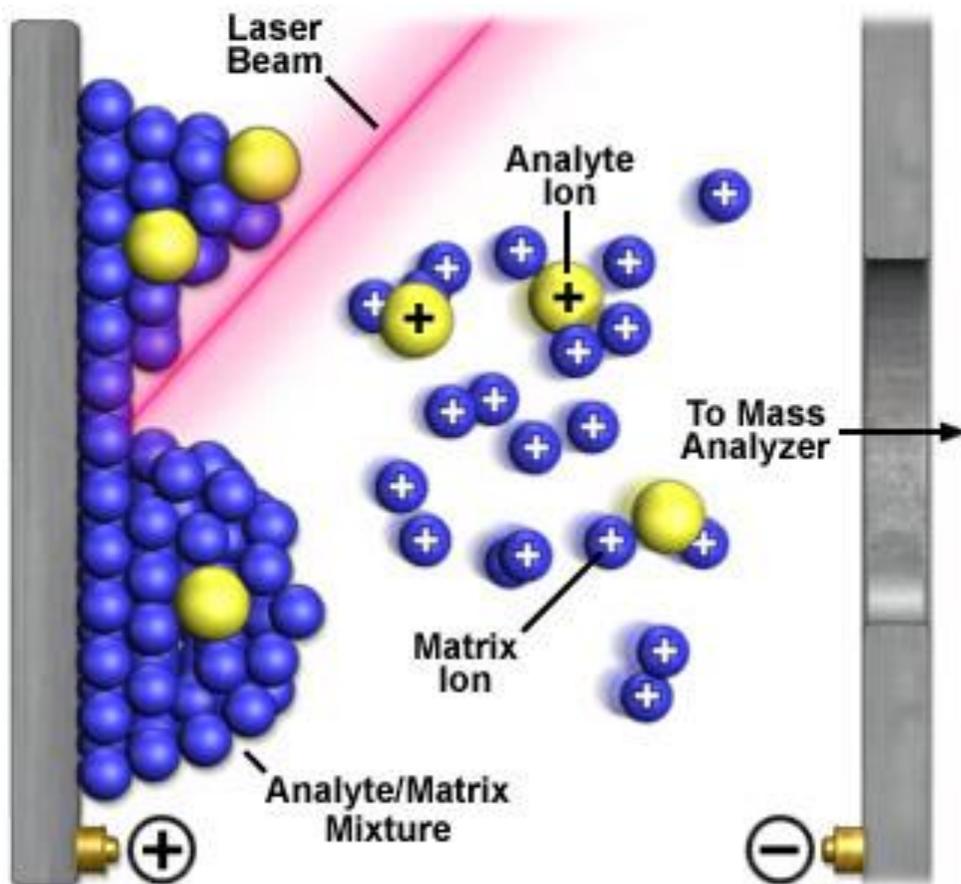
**Table 2.1** Different types of matrices used in MALDI

<b>Matrix</b>	<b>Wavelength</b>	<b>Typical Applications</b>
2,5-Dihydroxybenzoic acid (2,5-DHB)	337 / 355 nm	Proteins and Oligosaccharides
Sinapinic acid (SA)	337 / 355 nm	Proteins
$\alpha$ -Cyano-4-hydroxycinnamic acid (HCCA)	337 / 355 nm	Peptides and proteins
3-Hydroxycinnamic acid (3-HPA)	337 / 355 nm	Oligonucleic acids
Picolinic acid (PA)	266 nm	Nucleic acids
2,4,6-Trihydroxyacetophenone (2,4,6-THAP)	337 / 355 nm	Oligonucleic acids and Acidic oligosaccharides
6-Aza-2-thiothymine	266, 337, 355 nm	Oligonucleic acids and Acidic oligosaccharides
2-(4'-Hydroxybenzeneazo) benzoic acid (HABA)	337 / 355 nm	Proteins & carbohydrates
2,6-Dihydroxyacetophenone (2,6-DHAP)	337 / 355 nm	Oligonucleic acids
3-Aminoquinoline	337 nm	Oligosaccharides
3-Hydroxy picolinic acid	337 nm	Nucleic acids
Nicotinic acids	266 nm	Proteins, peptides and Adduct formation
Thiourea	266 nm	Large protein
1,5-Diaminonaphthalene (DAN)	349 nm	Lipids

The matrix molecules (MH) absorb the laser energy ( $h\nu$ ), and they become excited according to the following mechanism:  $MH + h\nu \rightarrow [MH]^*$ . The excited matrix molecules  $[MH]^*$  transfer a proton to the analyte molecule, or they accept a proton through a collision with the analyte “M”. Accordingly, they form a quasimolecular ion, for example  $[M+H]^+$  in the case of an added proton,  $[M+Na]^+$  in the case of an added sodium ion, or  $[M-H]^-$  in the case of a removed proton. The formation of  $[M+H]^+$  and  $[M-H]^-$  ions can be controlled by changing the polarity in the ionization chamber of the MALDI instrument. MALDI predominantly yields singly charged ions which are referred to be as “lucky survivors”,<sup>[23]</sup> but multiply charged ions ( $[M+nH]^{n+}$ ) can also be created (**Figure 1.2**).<sup>[18-23]</sup> The premise of lucky survivor principle is that all ions exist in the sample solution, however most ions are removed during MALDI process by recombination with counter ions. Then few ions “lucky ones” escape from this fate, and are detected.<sup>[23]</sup>

Creation of singly charged ions or multiply charged ions depends on the nature of the matrix, the laser intensity and the applied voltage.<sup>[18-23]</sup> For example, using SA as a matrix with an intact protein,  $[M+H]^+$  and  $[M+2H]^{2+}$  are created.<sup>[18-23]</sup>

MALDI has also been developed to work under atmospheric pressure (AP).<sup>[29,30]</sup> Newer sources for atmospheric pressure (AP-MALDI) applications include Desorption/Ionization on Silicon (AP-DIOS) and Laser-Induced Acoustic Desorption Mass Spectrometry (LIAD-MS). An advantage, of AP-MALDI is that it does not need matrix.<sup>[29,30]</sup>



**Figure 1.2** Schematic representation of the ionization of the matrix and the analyte in a MALDI source.<sup>[18]</sup> [Reproduced with the permission of the author].

### 1.1.3 Electrospray ionization (ESI)

ESI is a soft ionization technique that was developed by John Fenn.<sup>[31]</sup> In ESI, the gas phase-charged ions are created from the macromolecules dissolved in a polar solvent. The connection of an ESI source to a quadrupole mass analyzer allowed for the development of the first mass spectrometer capable of analyzing proteins with a molecular mass up to 40 kDa.<sup>[32,33]</sup> Furthermore, an ESI source is able to handle flow rates up to 1000  $\mu\text{l min}^{-1}$ , which makes it connectable to high pressure liquid chromatography (HPLC) system.<sup>[34]</sup>

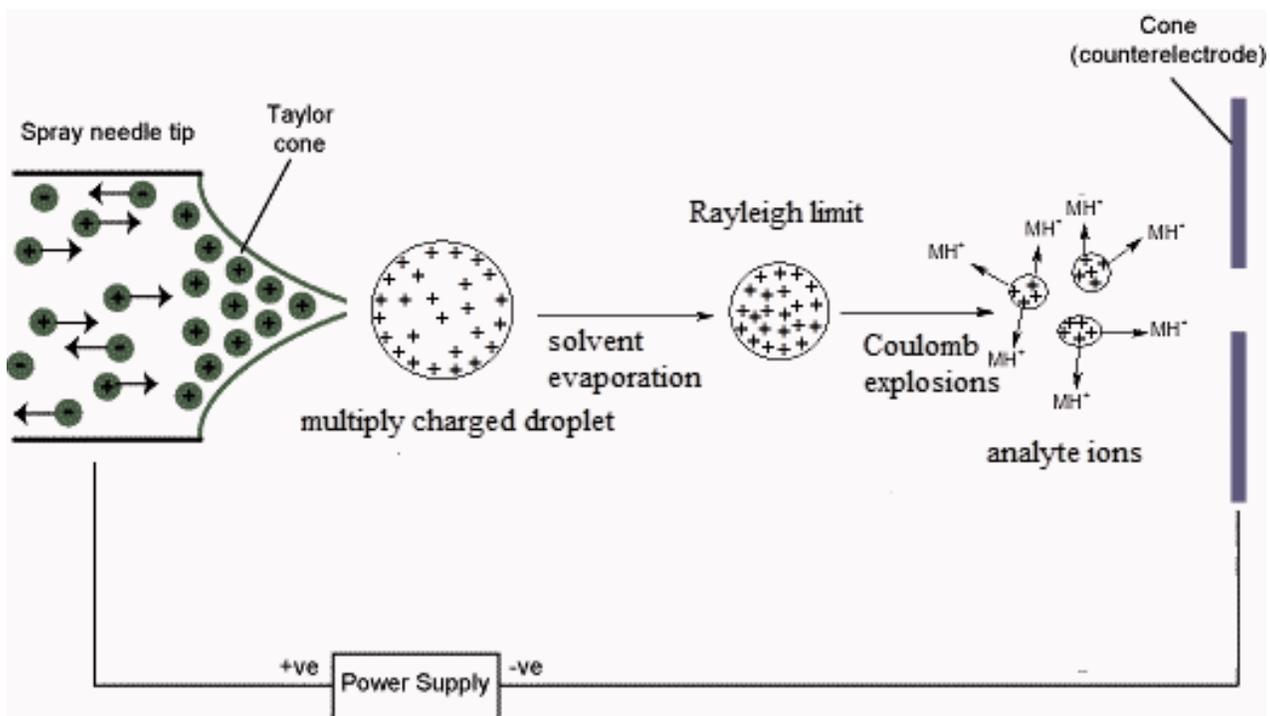
#### 1.1.3.1 Principle of ESI

In ESI experiments, a sample is dissolved in a polar organic solvent and infused into the source through a capillary, under atmospheric pressure. The source of ESI instruments consist of two main parts: a spray needle tip and a cone (counter electrode) connected to a power supply.<sup>[35-37]</sup> A high electric potential energy (2-6 kV) is applied at the spray needle tip to generate charged droplets in an effect known as the Taylor cone.<sup>[35-37]</sup> The droplets migrate with the aid of an electrical field. The droplets are reduced continuously in size as solvent evaporates. At a certain limit called the “Rayleigh limit”, the droplet is too small and is not able to carry the charges on its surface.<sup>[37]</sup> Accordingly, the droplets explode into small droplets, and then they are dried by warm nitrogen gas.<sup>[35-37]</sup> Finally, at the end of the ionization process, single and multi-charged ions are formed (**Figure 1.3**). A typical mass spectrum obtained by ESI-MS of a biomolecule correspond usually to a statistical distribution of multiply charged ions produced through protonation  $[\text{M} + n\text{H}]^{n+}$ , or deprotonation  $[\text{M} - n\text{H}]^{n-}$ .<sup>[35-41]</sup> The formation of  $[\text{M} + n\text{H}]^{n+}$  or  $[\text{M} - n\text{H}]^{n-}$  can be controlled by changing the polarity of the ESI ion source. During an electrospray

experiment, the analyte molecules acquire a charge through proton transfer reactions between the electrospray solution and the analyte.<sup>[39-41]</sup> However, electrochemical and photoelectrochemical reactions which take place at the electrode interfaces of the source have an influence on the ionization process.<sup>[39-41]</sup>

In APCI, the ionization of analytes occurs by primary ions such as  $O_2^{+}$  or  $N_2^{+}$  which are formed by corona discharge electrode. APCI is mainly applied for polar compounds, and it can be used for less polar samples.<sup>[39]</sup> However, complete non-polar samples cannot be analyzed by either APCI or ESI.<sup>[39-41]</sup>

A newer soft ionization technique, namely the APPI, has been developed. It is extensively applied for the solid and non-polar compounds. The ionization process takes place by interacting of a vaporized sample with photons which are emitted from krypton discharge lamp at 10 eV. These photons ionize compounds, which have ionization energies below 10 eV.<sup>[39]</sup> Gases and solvents are not ionized by APPI, however it can be applied for polar and non polar molecules.<sup>[39]</sup>



**Figure 1.3** Schematic representation of the production of charged droplets during ESI process.  $\text{MH}^+$  is the protonated molecule.

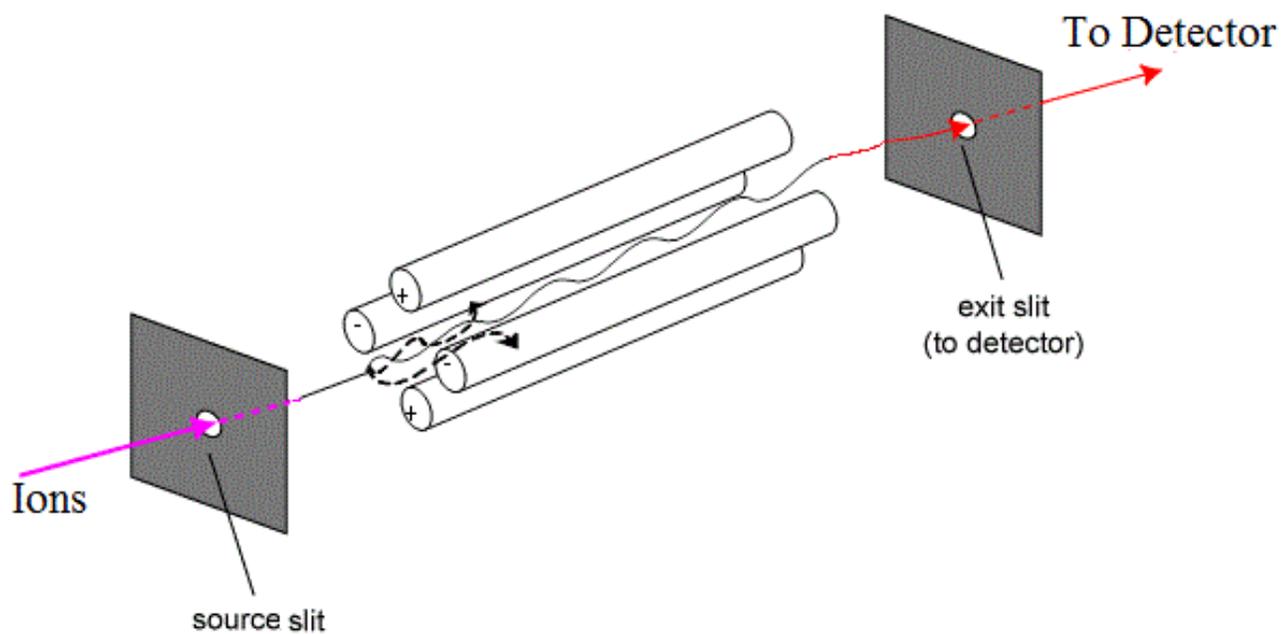
### 1.1.4 Mass analyzers

A mass analyzer is the second part of a mass spectrometer, which is used to scan and separate charged ions based on their  $m/z$  values.<sup>[11,42-49]</sup> In the mass analyzers, the isolation of ions is usually electrically-driven, although traditional analyzers, namely magnetic sector instruments, employ a magnetic field that influences ion separation. There are four common mass analyzers: quadrupole (Q), quadrupole ion trap (QIT), time of flight (TOF) and Fourier transform ion cyclotron resonance (FT-ICR) instruments. In this section, I will focus mostly on Q, TOF, and the coupling QTOF analyzers, a type of hybrid mass instrument.<sup>[42-56]</sup> These analyzers vary in size, price, resolution, mass range, and the ability to perform MS/MS experiments.<sup>[11,42-49]</sup> However, the choices of mass analyzer will eventually depend on the mass range, the accuracy of the mass measurement and the capacity to distinguish ions with similar  $m/z$  values (resolution). Both FT-ICR and QIT are capable of doing multiple mass spectrometric experiments ( $MS^n$ ), FT-ICR is undoubtedly, the most powerful in terms of accurate mass measurements with resolution of full width at half maximum (FWHM) 1000,000.<sup>[44]</sup>

#### 1.1.4.1 Quadrupole analyzer

The quadrupole mass analyzer “quadrupole mass filter” was first described by Paul Wolfgang in the 1950s, for which he received the Nobel Prize in 1989.<sup>[42-43]</sup> The principle of this analyzer is based on changing ion trajectories by oscillating electric fields to separate these ions according to their  $m/z$  ratios.<sup>[42-43]</sup> A quadrupole analyzer is made up of four circular rods, and they are perfectly parallel rods: two of them are negatively charged and the other two are positively charged (**Figure 1.4**).<sup>[42-43]</sup> Simultaneously, direct current potential (DC, or called “ $U$ ”) and a radiofrequency

potential (RF, or called “ $V$ ”) are applied between the two pairs of rods, . The operation of both DC and RF potentials causes the ions to oscillate according to the equation  $\Phi_0 = \pm (U \pm V \cos \omega t)$ , where  $\Phi_0$  is the voltage applied to the rods,  $\omega$  is the angular frequency,  $U$  is the DC voltage and  $V$  is the RF voltage amplitude.<sup>[6,10,42-43]</sup> Then, the  $U$ ,  $V$  and  $\omega$  values will be changed to make an ion of a certain  $m/z$  value having stable trajectory inside the analyzer, and this ion can easily reach the detector “the survived ion is called resonant ion”.<sup>[6,10,42-43]</sup> For example, the positively charged ions created from the ion source will move in the space between the rods. Then, they will be drawn towards the negatively charged rods. However, once the polarity of these negative rods is changed, the ion will switch their trajectories before striking the rod.<sup>[42-50]</sup> The remaining ions will possess different  $m/z$  values and will eventually collide with one of the rods and be lost “these ions are called non resonant ions”.<sup>[6,10,42-43]</sup> The quadrupole mass analyzer presents several advantages: excellent transmission efficiency, suitability for GC, liquid chromatography (LC) and capillary electrophoresis (CE), low cost, robustness, and easy to use.<sup>[42-43]</sup> However, the mass range of the quadrupole (which is less than 4000 Da), its low resolution, and the need to couple more than two quadrupoles to achieve MS/MS analysis, are the major limitations of this analyzer.<sup>[42-49,61]</sup>



**Figure 1.4** Schematic representation of a quadrupole mass filter.

#### 1.1.4.2 Time of flight (TOF) analyzer

The TOF mass analyzer was firstly described by Stephens in 1946, and it is considered the simplest mass analyzer.<sup>[50,62]</sup> The TOF analyzer is compatible with many ionization techniques, specifically the pulsed nature of laser in MALDI.<sup>[50,62]</sup> TOF analyzers have been used in different applications for biomolecules such as proteins, oligonucleotides, and synthetic polymers.<sup>[24-28]</sup>

Unlike quadrupole and magnetic sector analyzers, the flight of a charged ion inside the TOF analyzer to the detector does not require any electrical or magnetic fields.<sup>[50,62]</sup> However, the travelling of charged ions from an ion source to the TOF analyzer is simply done under the effect of the applied voltage “Vs”.<sup>[62]</sup> A charged ion travels through a TOF tube, also called, field free area or flight free region, which is 1-2 m in length.<sup>[62, 63]</sup>

Ions travel from the ion source in bundles by laser desorption.<sup>[62-64]</sup> Then, they are accelerated to the TOF tube by applying a voltage between an electrode and an extraction grid (see **Figure 1.5**). All accelerated ions gain the same kinetic energy “ $E_c$ ” =  $1/2mv^2 = qVs$  (m is mass of the ion, v = velocity of the ion, q is charge of the ion, Vs = the applied voltage at the acceleration region). In addition, the travelled ion has the charge “q” = Ze (Z = charge number, e = elementary charge) and mass “m”.<sup>[62-64]</sup> So:

$$E_c = 1/2mv^2 = qVs = ZeVs$$

$$v = (2 Ze Vs/m)^{1/2}$$

Thus, the ion travels from the accelerated region of the ion source to TOF tube by a velocity,  $v = (2 Ze Vs/m)^{1/2}$ .<sup>[62-64]</sup> Then, the ion moves in straight line inside the flight free

region at constant velocity “v” to the detector. The time required for travelling the field-free region with a length  $L$  before reaching the detector is given by  $t = L/v$ .<sup>[62-64]</sup>

Replacing  $v$  by its value from the previous equation gives:

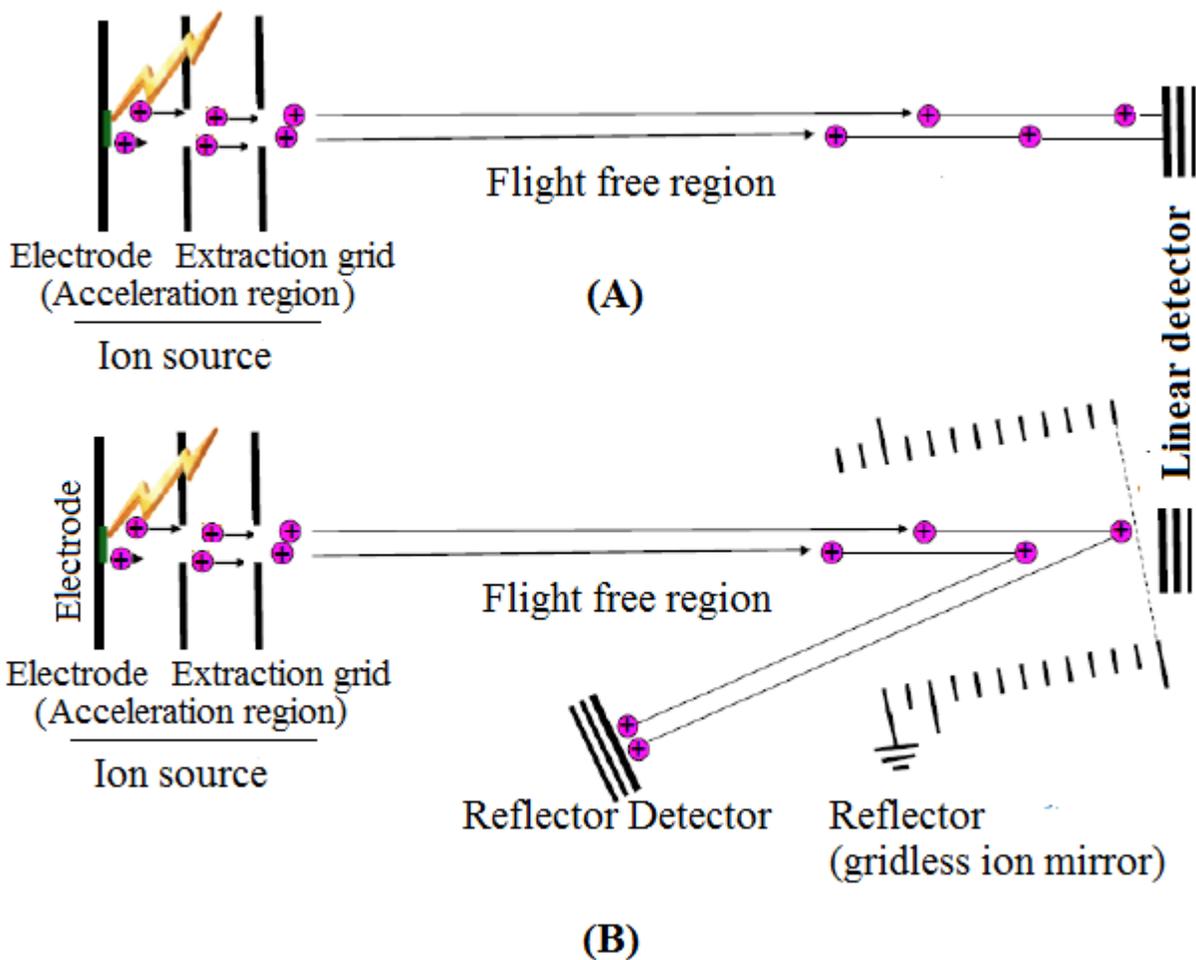
$$t = L / (2 Ze Vs/m)^{1/2}$$

$$t^2 = m/z (L^2/2e Vs)$$

Thus, the separated ions will be measured based on their  $m/z$  values, since the other factors  $L$ ,  $V_s$  and  $e$  are constant.<sup>[62-64]</sup> Unlike with quadrupole or magnetic sector analyzers, in a TOF instrument, all formed ions will eventually reach the detector.<sup>[62-71]</sup>

The linear mode of a conventional TOF analyzer (**Figure 1.5A**) has been refined because its resolving power is limited. The linear mode has been modified into a reflector mode, in which an electrostatic ion mirror “reflectron”, is incorporated to increase the resolving power as shown in **Figure 1.5B**.<sup>[65]</sup>

Currently, all modern MALDI-TOF-MS operate with delayed pulsed extraction (also known as time-lag focusing).<sup>[51]</sup> Delayed extraction is a method used in TOF to increase the resolution of TOF analyzer by reducing the kinetic energy of some ions which have the same  $m/z$  ratio. A voltage pulse is applied to decelerate the ions outside the source with typically the same  $m/z$  ratio.<sup>[51]</sup> In contrast to continuous extraction used in conventional instruments, delayed extraction increases the mass resolution with high sensitivity.<sup>[51]</sup>

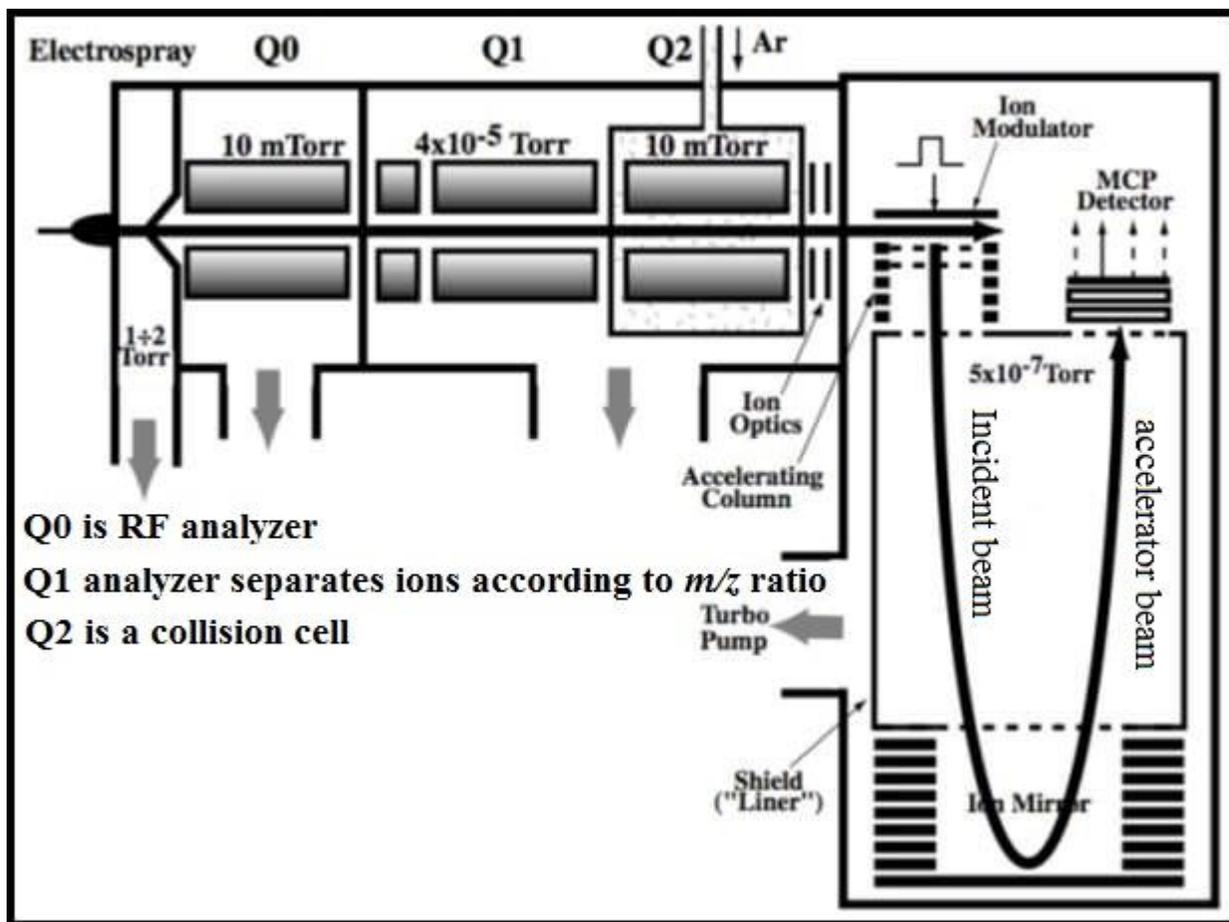


**Figure 1.5** Schematic representation of a conventional TOF analyzer (A) and a reflectron-type TOF (B).

### 1.1.4.3. Hybrid quadrupole/time-of-flight (QTOF) mass spectrometer

Hybrid mass spectrometers consist of different types of mass analyzers. Accordingly, a QTOF instrument consists of two different types of analyzer: three quadrupole analyzers and one time of flight analyzer (**Figure 1.6**).<sup>[51-71]</sup> The purpose of combining two or more different analyzers is to increase the mass resolution, and to provide high-quality, informative, simple, one-stage MS and tandem MS/MS spectra.

Despite the coupling of the TOF analyzer with MALDI source is common because the TOF system works on a pulsed process (pulsed nature of the laser).<sup>[51-65]</sup> However the coupling of TOF analyzer with an ESI or any other API sources is difficult because these sources produce a continuous ion beam.<sup>[51-71]</sup> Thus, the TOF was not suitable to be coupled with ESI until the development of the orthogonal injection system (also called orthogonal acceleration system “oa system”). In the “oa” system, the ion is accelerated with a direction which is perpendicular to its initial direction, and then the continuous ion beam will be transformed into a pulsed process.<sup>[51-71]</sup> It was reported that the properties of ESI combined to orthogonal TOF “reflector TOF” have shown great potential.<sup>[59,66]</sup> ESI-QqTOF hybrid instruments use lower internal energies for the fragmentation of molecular ions. Thus, the spectra of ESI-QqTOF hybrid instruments are relatively easy to interpret. Additionally, ESI-QqTOF shows a resolving power of 10,000 FWHM. Another advantage of LC-ESI-QTOF mass spectrometry is its adaptability to protein identification and sequencing.<sup>[52-66]</sup>



**Figure 1.6** Schematic representation of QTOF instrument (QSTAR hybrid QqTOF). QTOF consists of three quadrupole Q0 is the first quadrupole, Q1 is the second quadrupole, and Q2 is the third quadrupole. Ar, argon; MCP, microchannel plate; RF, radiofrequency. [Reproduced with the permission from the publisher of Ref. 11]

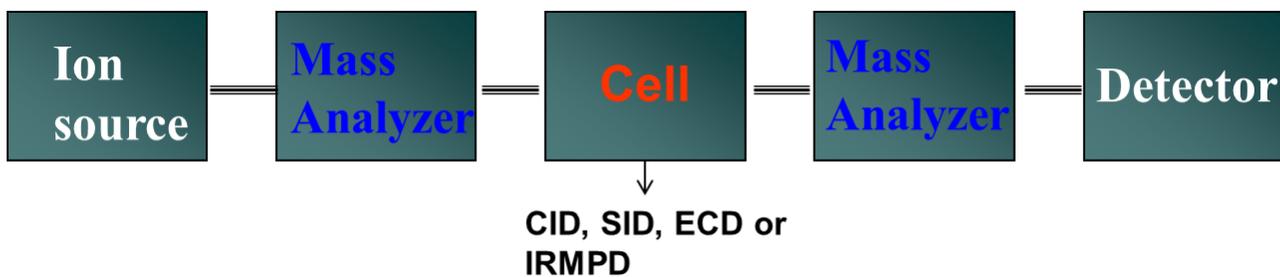
### 1.1.5 Tandem mass spectrometry

Tandem mass spectrometry is often abbreviated as MS/MS or MS<sup>n</sup>, typically consists of at least two analyzers coupled together.<sup>[66-71]</sup> The first analyzer is used for the selection of a precursor ion (parent ion which is desired to be fragmented). The second analyzer is used for the scanning of product ions (fragments ions of the precursor ion). The precursor ion fragmentation can be done by using a neutral gas such as argon or helium in the case of collision induced dissociation (CID), by collision with the collision cell surface in the case of surface induced dissociation (SID), by using electrons in the case of electron capture dissociation (ECD) or by laser in infrared multiphoton dissociation (IRMPD), see **Figure 1.7**. There are two different types of tandem mass spectrometry; tandem in space and tandem in time.<sup>[66-73]</sup>

#### 1.1.5.1 Tandem in space and tandem in time

With tandem in space, the mass spectrometer consists of more than one analyzer and a collision cell. The first mass analyzer is used to select an ion “precursor ion” for analysis. The precursor ion is then fragmented inside the collision cell to produce the product ions. Then, the second analyzer will analyze the product ions.<sup>[71-73]</sup>

With tandem in time, a selected precursor ion is trapped for a specific time in the first analyzer. The product ions are formed inside the first analyzer. Then, the second analyzer will scan for product ions. This type of tandem MS is more common in QIT, the FTICR, and the Orbitrap instruments.<sup>[71]</sup>



**Figure 1.7** Schematic of tandem mass spectrometry. CID, collision induced dissociation; SID, surface induced dissociation; ECD, electron capture dissociation; IRMPD, infrared multiphoton dissociation.

### 1.1.6 Protein mass spectrometry

Mass spectrometry of proteins is a way for characterizing and sequencing a protein through two common approaches: a top down approach and a bottom up approach.<sup>[1,74]</sup> A single stage MALDI, or a de-convolution process of ESI are used for determining the protein molecular mass in its intact form.<sup>[1,74,75]</sup> In the single stage MALDI of an intact protein, two major ions  $[M+H]^+$  and  $[M+2H]^{2+}$  are usually generated.<sup>[1,74,75]</sup> Then, the protein molecular mass is measured by removing a proton from the  $[M+H]^+$  ion. Thus, in the de-convolution process of ESI, the MS software transforms group of peak ions based on their intensities and charges into a single ion peak, which symbolizes the protein molecular mass.<sup>[1,74,75]</sup> Single stage MALDI is preferred due to its low salt tolerance and its lower fragment spectrum.<sup>[1,74,75]</sup>

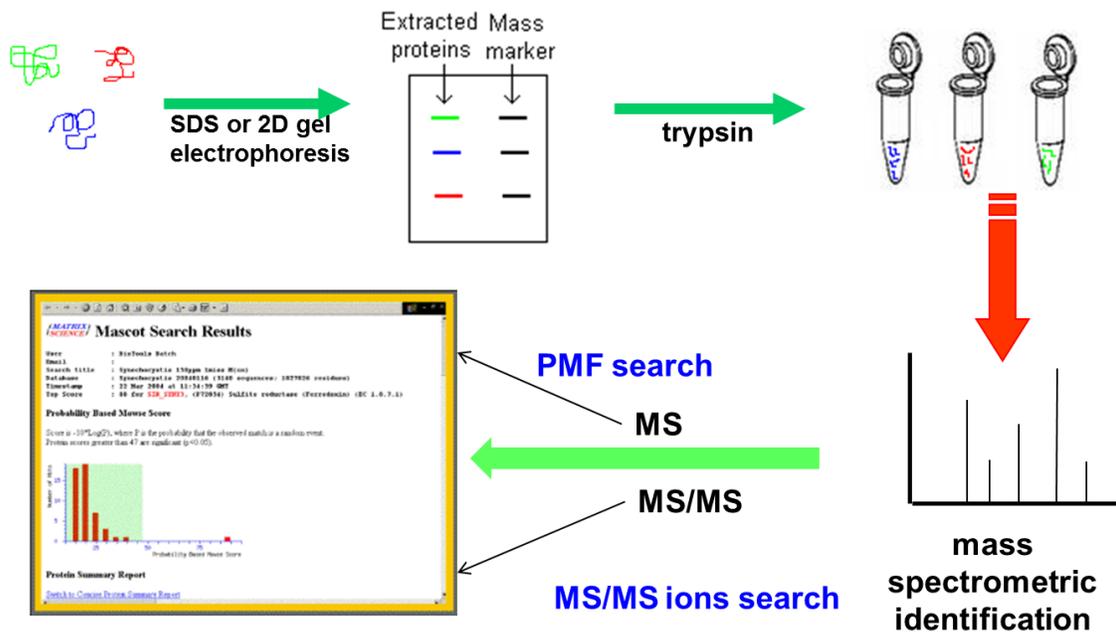
Top down proteomics and bottom up proteomics are powerful methods for the identification of proteins based on mass spectrometric analysis and database homology searches.<sup>[1,76-81]</sup> The first step in both approaches involves the extraction of a protein by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gels, LC separation, immunopurification or two dimensional (2-D) gels.<sup>[1,76-81]</sup> The principle of both approaches is similar, but the form in which the protein is introduced to the mass spectrometer is different in the two approaches.<sup>[1,76-81]</sup> In the top down approach, the isolated protein is introduced to an ESI mass spectrometer in its intact form.<sup>[76-80]</sup> This protein will then be degraded into product ions during the MS/MS gas phase fragmentation.<sup>[76-80]</sup>

In bottom up proteomics, proteolytic digestion of the protein must be done prior to introducing the protein into the mass spectrometer.<sup>[1,81,82]</sup> The proteolytic digestion is done

by enzymes which cleave at specific sites on the protein (e.g. trypsin cleaves at the carboxylic end of lysine and arginine or Gluc V8 cleaves at the carboxylic end of glutamic acid and aspartic acid).<sup>[1,81-85]</sup> This approach consists of two major stages: peptide mass fingerprinting (PMF) and tandem MS/MS ions searches.<sup>[1,81-85]</sup>

PMF and MS/MS ion searches are performed individually by processing the MS and MS/MS spectra, respectively, into a peak list. By introducing the peak list into a MS database, a report will be generated, which will provide the similarities between your protein and other proteins recorded in the database. The high similarities (matching parts) are represented by high scores.

In PMF and MS/MS ion searches, the identity of the original protein is determined by comparing the peptide mass spectra with theoretical peptide masses calculated from a proteomic or genomic database.<sup>[1,81-85]</sup> Both PMF and MS/MS ion searches are complementary to each other.<sup>[1,81-85]</sup> However, PMF results include unmatched masses from MS spectra due to post translational modifications causing mass shifts, partial proteolysis or non-specific cleavages.<sup>[1,81-85]</sup> PMF does not work for identifying mutant forms of a protein. For that reason, MS/MS data are necessary for identifying peptide sequences and for avoiding false positive results. **Figure 1.8** illustrates the major steps of bottom up proteomics approach, which also referred to be as shotgun proteomics.



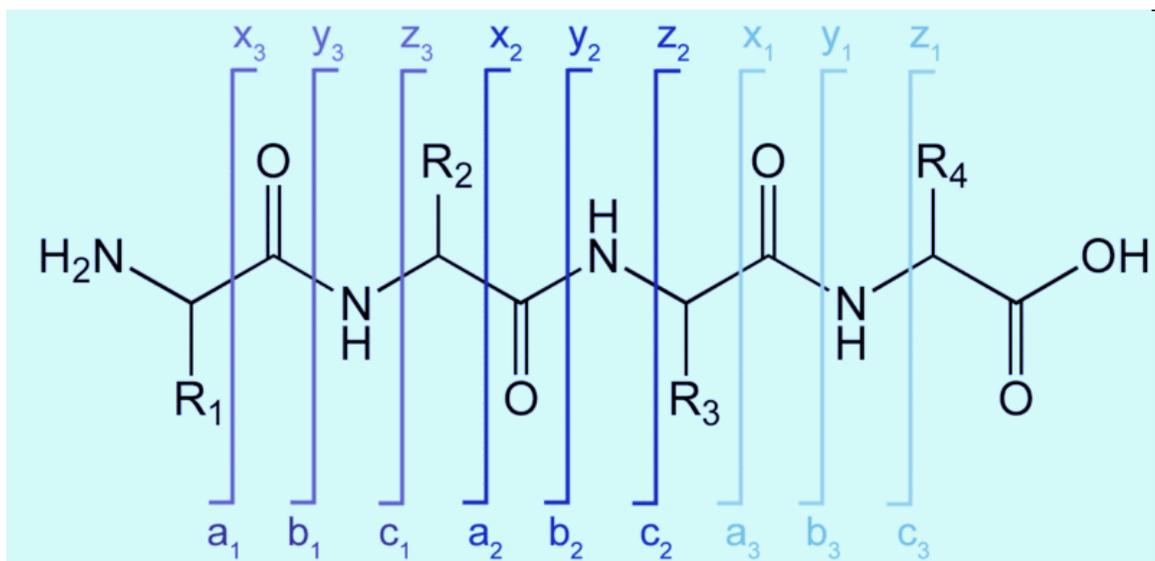
**Figure 1.8** Workflow in bottom up proteomics. Protein is extracted from a mixture, separated, digested, and analyzed by tandem mass spectrometry. The PMF and MS/MS ions search is done through the Mascot protein database. SDS, sodium dodecyl sulfate; 2D, 2 dimensional; PMF, peptide mass fingerprint; MS, mass spectrometry; MS/MS, tandem mass spectrometry.

### 1.1.7 Gas phase fragmentation of proteins and carbohydrates

Gas phase ions are formed in the ionization source. It was observed that during the ionization process, the energy gained by an ionized molecule plays a role in the formation of in-source fragment ions. Then, the precursor ions are fragmented into product ions via MS/MS. In the case of a protein, the product ions are amino acid fragments. Roepstorff and Fohlman proposed the conventional nomenclature of amino acid fragments, which was subsequently modified by Johnson in 1987.<sup>[86, 87]</sup> **Figure 1.9** shows the nomenclature of the peptide fragments, which are generated by CID-MS/MS. There are two types of the peptide fragments: fragments of amino terminus (a, b, c) and fragments of carboxyl terminus (x, y, z). Sometimes, internal fragment with a single side chain formed by the combination of a- and y- fragments is called an immonium ion.<sup>[86,87]</sup> However, my focus will be on y- and b- fragments because these fragments represent a complete amino acid or a complete peptide.<sup>[86,87]</sup>

*De novo* peptide sequencing using mass spectrometry can be achieved by the aforementioned bottom up proteomics approach, in which the databases have been used for calculating the y- and b- ions via MS/MS ion search. Since not all y- and b- values are shown on a spectrum, calculating y- and b- values manually becomes a big challenge.<sup>[1,81-</sup>

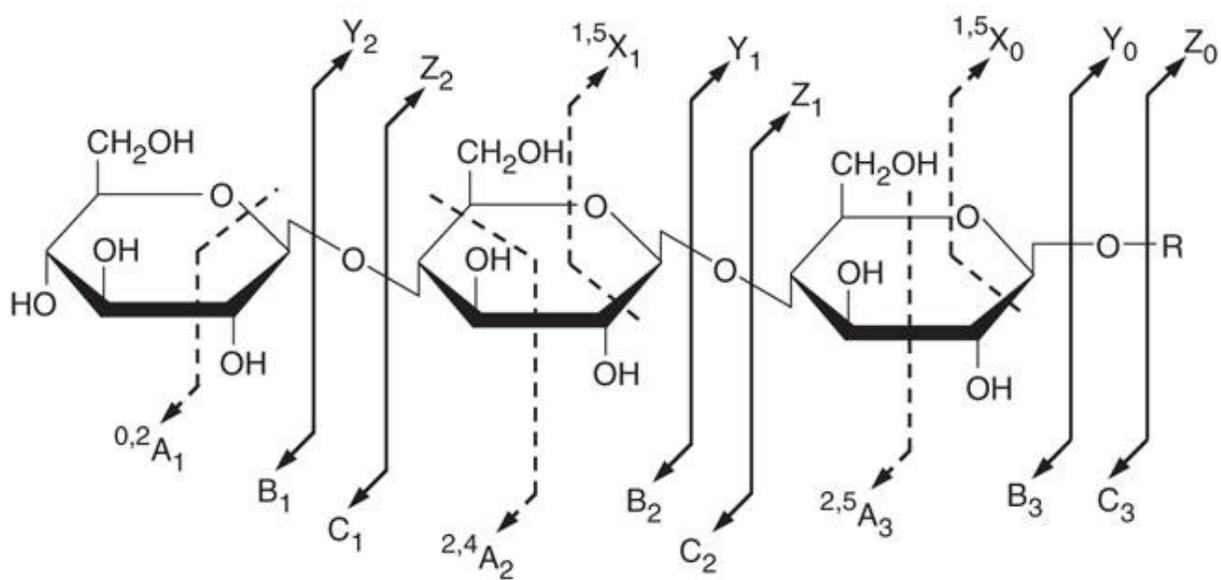
85]



**Figure 1.9** Peptide fragmentation nomenclature. This figure is adapted from the Ref. [86].

In 1988, Domon and Costello reported the carbohydrate fragmentation nomenclature (**Figure 1.10**).<sup>[89]</sup> The idea of carbohydrate fragmentation nomenclature is similar to that of protein nomenclature. According to Domon and Costello, there are two types of carbohydrate fragments: X, Y, Z fragments from the reducing end and A, B, C fragments from the non-reducing end (X or A fragments come from the fission through the oxonium ion ring).<sup>[89]</sup>

In **Figure 1.10**, R stands for a non-carbohydrate group (protein or lipid), which is bound to the carbohydrate part through a spacer (linker). So, a resulting spectrum should contain fragments of the carbohydrate part, the non-carbohydrate part and/or the spacer linked to either carbohydrate moiety or R.



**Figure 1.10** Carbohydrate fragmentation nomenclature. R, reducing end. [Reproduced with the permission from the publisher of the Ref. 84]

## 1.2 Snow crab (*Chionoecetes opilio*)

Snow crab (*Chionoecetes opilio*) is a Brachyuran crab of the family Majidae.<sup>[89]</sup> The fishery of this species occupies a large area of Newfoundland's southeast coast since it is found from the Gulf of Marine to West Greenland. It is the most valuable economic fishery on the Grand Bank of Newfoundland's southeast coast.<sup>[89]</sup>

Snow crabs are invertebrates and have immune system that lack immunological specificity.<sup>[90]</sup> They lack memory cells and lymphocytes, and accordingly the immune system does not involve antigen-antibody reactions.<sup>[90]</sup> Instead, it relies mainly on antimicrobial proteins.<sup>[90]</sup> It was reported that there are different antimicrobial proteins found in different species of crabs. These proteins are secreted directly into the blue blood of the crabs (crab hemolymph).<sup>[90]</sup> Proline peptide was the first antimicrobial peptide characterized in shore crab (*Carcinus maenas*).<sup>[90,91]</sup> Both a lectin and callenactin peptide have antimicrobial activity in blue crab (*Callinectes sapidus*).<sup>[92,93]</sup> Mud crabs (*Scylla serrata*) have two antimicrobial proteins in its hemolymph: scygonadin of 43 kDa and a lectin of 4 kDa, which is called scyllin.<sup>[94]</sup> Additionally, lectin is bound non-covalently to sialic acid to act as an antimicrobial complex in Indian horseshoe crab (*Carcinoscorpus rotunda*).<sup>[94,95]</sup> Furthermore, a complex of crustin protein with characteristic four disulfide core- containing whey acidic protein (WAP) domain is the key player in the immune system of spider crabs (*Hyas araneus*).<sup>[96,97]</sup> In this thesis, I have focused on an antimicrobial protein from snow crab (*Chionoecetes opilio*), which is called cryptocyanin.

Snow crabs are crustaceans which undergo an ecdysis (molting) process, which allows the snow crab to grow.<sup>[98]</sup> It involves shedding off the exoskeleton, and for a short time afterward the crab is unable to move until it regains muscle control and the new

exoskeleton hardens.<sup>[98]</sup> At this point the crab will begin eating the old exoskeleton which helps to recycle the calcium carbonate and other minerals necessary for the new exoskeleton's health and rigidity.<sup>[98]</sup> Because crabs are vulnerable during this time, they tend to bury underground in the ocean's muddy floor for protection, insulation, and darkness, which is necessary to allow the molting hormone (MH) to be secreted so that the shedding process can begin.<sup>[98-101]</sup> Molting is an important sign of growth as the snow crab matures and gets ready to reproduce.<sup>[98]</sup> During molting, the crab is more susceptible to microbial infection as the new exoskeleton becomes soft.<sup>[98-101]</sup> Hence, the crab cryptocyanin is elevated five-fold to adapt its function as anti-microbial protein.<sup>[98-101]</sup>

### 1.2.1 Cryptocyanin

Cryptocyanin is a hexameric protein which exists in different crab species including Dungeness crab (*Metacarcinus magister*) and blue crab (*Portunus pelagicus*).<sup>[99-104]</sup> Cryptocyanin and hemocyanin are the most abundant proteins found in the snow crab hemolymph.<sup>[99-104]</sup> They are synthesized in the hepatopancreas and are directly secreted in the crab hemolymph.<sup>[99-104]</sup> Both proteins are similar in size, sequence and structure.<sup>[99-104]</sup> The hemocyanin structure has copper sites, and cavities to hold oxygen.<sup>[99-104]</sup> Thus, it acts as a respiratory protein "oxygen transporter".<sup>[99-104]</sup> Unlike hemocyanin, cryptocyanin is a copper free protein, and it is a non-respiratory protein. However, cryptocyanin is expressed from the same gene of hemocyanin, which is called the hemocyanin gene duplication or fusion.<sup>[99-104]</sup>

Cryptocyanin expression is controlled by a hormone called molting inhibiting hormone (MIH).<sup>[99-104]</sup> MIH is secreted from an endocrine organ called the X organ-sinus

gland system (XO-SG), which is found in the crab eyestalk. Consequently, ablation of eyestalks leads to the suppression of cryptocyanin.<sup>[99-104]</sup>

### **1.2.2 Cryptocyanin functions**

The snow crab fishing season takes place after the crabs have molted and have new shells. The ability to predict the proportion of adolescent crabs that will molt and appear as soft-shelled individuals will allow industry to avoid fishing in areas where high incidences of soft crabs are likely. Cryptocyanin can be used as a predictor tool for the fishing season since its concentration is elevated five times during the premolt stage.

More interestingly, during the non-molting stage, crab cryptocyanin has many functions. It works as a carrier protein for steroids, hormones and fatty acids.<sup>[105]</sup> This function is similar to serum albumin which constitutes 50% of human serum protein.<sup>[105]</sup> Additionally, it has a putative osmolyte role.<sup>[105]</sup> Moreover, cryptocyanin has divalent cation binding sites for  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  ions. Besides the non-molting functions, cryptocyanin carries the  $\text{Ca}^{2+}$  ions from the old exoskeleton to the hemolymph, preparing for the molting process.<sup>[105]</sup> Somewhere  $\text{Ca}^{2+}$  ions have been stored on the loading sites of cryptocyanin until their use for hardening the new exoskeleton during the molting process. At the time of ecdysis, cryptocyanin moves from the hemolymph in an elongated epidermal cell to the new shell in order to deliver the  $\text{Ca}^{2+}$  ions. At such time, cryptocyanin levels are elevated to achieve its immune function because the crab's new shell is soft, and the snow crab is more susceptible to microbial infection.<sup>[105]</sup>

### **1.3 Vaccines**

A vaccine is a substance used to confer protection by stimulating the immune system to secrete antibodies against diseases.<sup>[106,107]</sup> In 1796, Edward Jenner discovered that the human inoculation with cowpox was able to protect against smallpox infection.<sup>[106,107]</sup> Since then, different vaccines were developed to prevent infectious disease.<sup>[106,107]</sup> There are different forms of vaccines: attenuated vaccines, inactivated vaccines, toxoid vaccines, subunit vaccines, DNA vaccines, recombinant vector vaccines and conjugate vaccines.<sup>[108]</sup>

#### **1.3.1 Live, attenuated and inactivated vaccines**

Live, attenuated vaccines confer protection by natural infection with a “living microbe” that has been manipulated to suppress the virulent effect.<sup>[108,109]</sup> Live attenuated vaccines provoke a strong immune response but can also cause disease in patients with a flagging immune system such as those with HIV infection or receiving chemotherapy.<sup>[108,109]</sup> Another drawback of this vaccine, is that it is not specific for many bacteria because they have different species or genes. Additionally it needs to be refrigerated to stay potent. Examples include vaccines against measles, cholera, mumps and chickenpox.<sup>[108,109]</sup>

An inactivated vaccine is created to avoid the drawbacks of the attenuated one. This type is produced after killing the microbe by chemicals like formaldehyde, by irradiating or through heat. An inactivated vaccine cannot cause the disease. However, it tends to elicit a weak immune response.<sup>[108,109]</sup> Example includes a vaccine against hepatitis A.<sup>[108,109]</sup>

### **1.3.2 Toxoid and subunit vaccines**

An antigenic part of a microbe or its secreted toxin can cause infection. Accordingly, scientists use these parts of microbes for designing vaccines. When preparing vaccines against bacterial toxins, the toxin is treated by formalin to make it harmless, called the “detoxified form”. These “detoxified” toxins are used as toxoid vaccines.<sup>[108,109]</sup> When the toxoid vaccine is administered, the immune system generates a strong immune response to the neutralized toxin, and accordingly it produces enough antibodies to block the toxin sites. Examples of toxoid vaccines include those against diphtheria and tetanus.<sup>[108,109]</sup>

As an alternative, parts of the microbe “different epitopes” are used to synthesize the subunit vaccines. The drawback with such type of vaccine, is its non-specificity.<sup>[108,109]</sup> Accordingly, one vaccine needs 1-20 epitopes to be more specific<sup>[108,109]</sup> Important epitopes can be picked up after chemical degradation of the microbe, so it is time consuming process.<sup>[108,109]</sup> These epitopes can be manufactured using recombinant DNA technology. The hepatitis B vaccine is an example for this type of vaccine.<sup>[108,109]</sup>

### **1.3.3 DNA and recombinant vector vaccines**

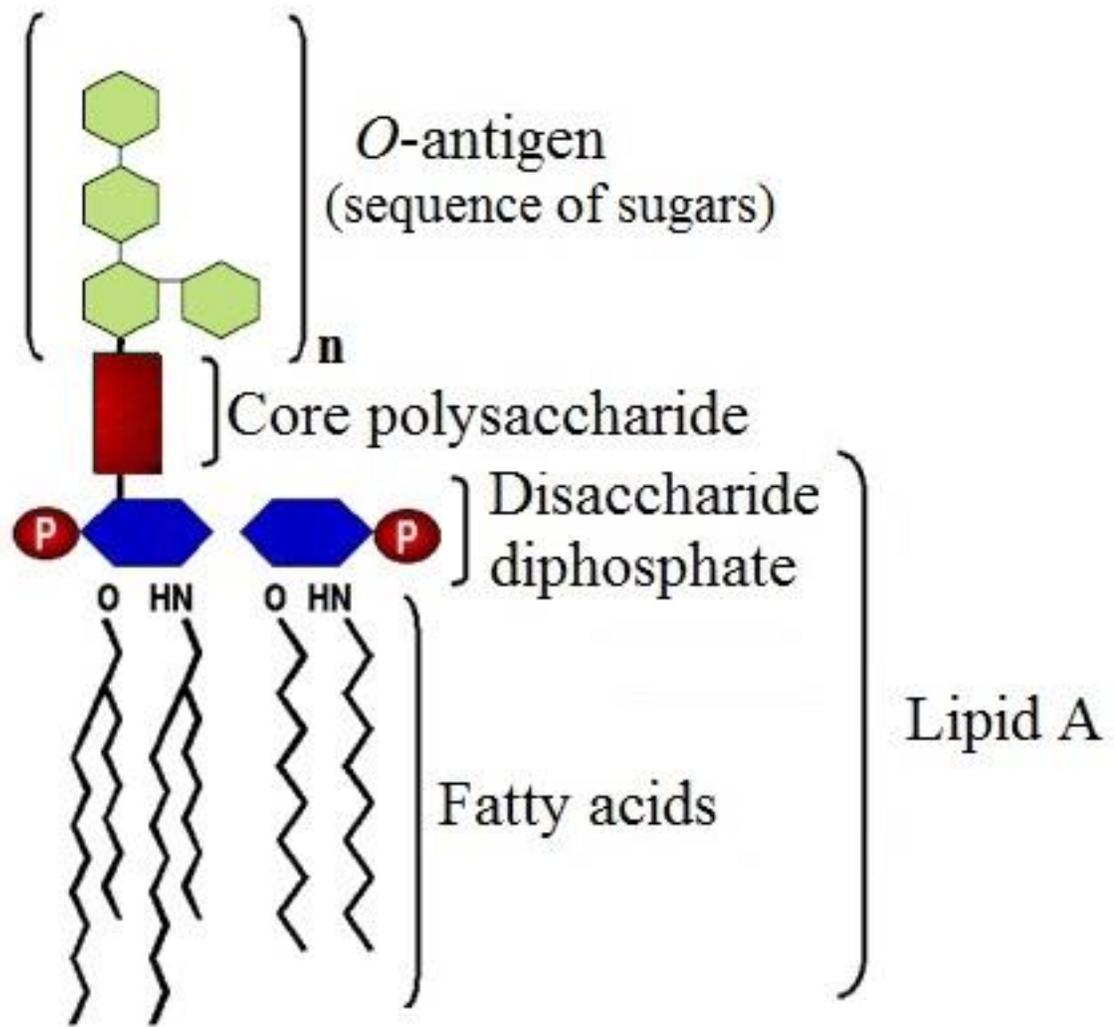
DNA vaccines are prepared by making use of microbial genes that encode vital antigens to make a vaccine against the microbe. DNA vaccines, containing these genes, can be picked up by competent host cells after administration by using a syringe.<sup>[108-111]</sup> The host cell works as a factory to translate these genes and to produce antigens which stimulate a strong immune response. Examples of DNA vaccines include those used for herpes and influenza vaccines.<sup>[108-111]</sup>

Since, not all host cells are capable picking up the naked DNA vaccine. Scientists have tried to increase the efficiency of the naked DNA vaccine by making a coat around the genes using gold molecules, to facilitate its uptake by the host cells.<sup>[108-111]</sup> Accordingly, recombinant vector vaccines have been developed in which vectors (viruses or bacteria, which work as gene carriers) are used for introducing the microbial genes to the host cells.<sup>[108-111]</sup>

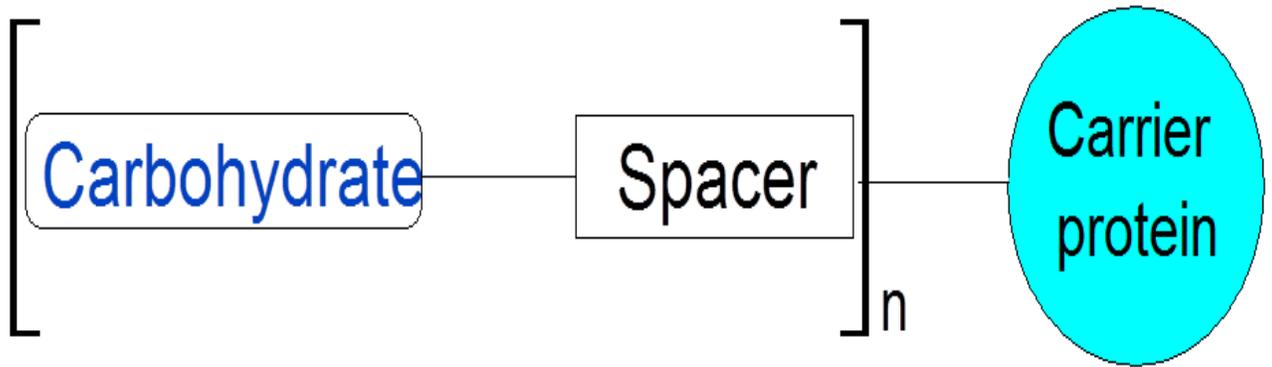
### 1.3.4 Conjugate vaccines

Gram negative bacteria cell walls are coated with lipopolysaccharides (LPS).<sup>[112,113]</sup> LPS consists of three main parts: an external part called the *O*-antigen, which is composed of repeats of identical sugar oligosaccharide units, and an internal core which is covalently linked to the third part, “lipid A” (**Figure 1.11**).<sup>[112,113]</sup> Lipid A represents the toxic part of LPS, and it consists of a  $\beta$ -D-(1 $\rightarrow$ 6) glucosamine (GlcN) disaccharide in which O-3, O-3', O-4', C<sub>2</sub>-N and C<sub>2</sub>-N' are acylated with different C:12 and C:14 fatty acids.<sup>[112,113]</sup> The carbohydrate part of the lipid A or the *O*-antigen part can be conjugated with biomolecules such as proteins or dendrimers (elegant branched molecules consist of cascade molecules) to create a glycoconjugate vaccine.<sup>[112-115]</sup>

A carbohydrate antigen neoglycoconjugate vaccine is composed of the carbohydrate part called a “hapten” linked by a spacer to a biomolecule (which is usually a protein), see **Figure 1.12**. The synthesis of efficient glycoconjugate vaccines depends on many factors, and their efficacy relies on the saccharide size, the average number of saccharide chains per conjugate molecule, the nature of the carrier and the distance between the saccharide and the protein in the formed glycoconjugate.<sup>[116-119]</sup>



**Figure 1.11** Lipopolysaccharide of gram negative bacterial cell wall. P, Phosphorous group



**Figure 1.12** Composition of the glycoconjugate vaccine.

It is worthy to mention that antigenic carbohydrate-protein vaccines are not limited to fighting bacteria, as they can be developed as anticancer drugs by conjugation of a tumor associated carbohydrate antigen (TACA) to a carrier protein.<sup>[120]</sup> The first carbohydrate-protein vaccine was reported by Landsteiner's group,<sup>[121,122]</sup> and later, it was discovered that they can induce strong antibody reaction.<sup>[123]</sup>

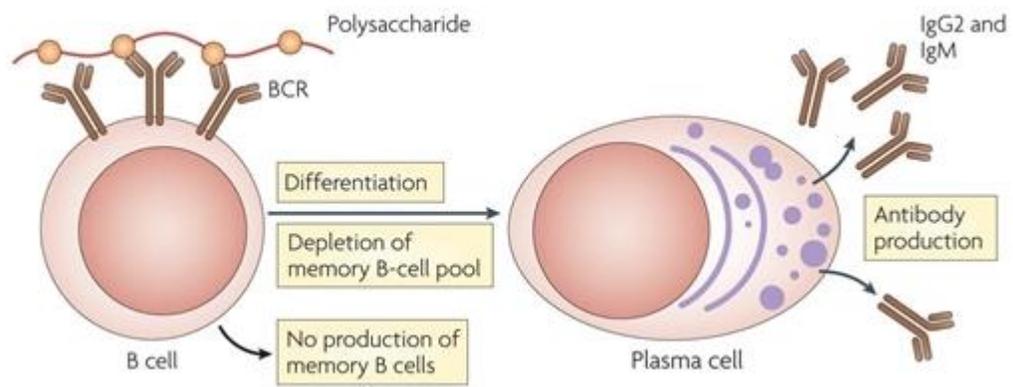
Different methods have been used for the synthesis of carbohydrate antigen-protein neoglycoconjugates. One of these methods consisted of the use of squaric acid chemistry for the single-point attachment of carbohydrates to proteins.<sup>[123-126]</sup> Tietze *et al.* used squaric acid diethyl esters for conjugation.<sup>[124]</sup> It has also been reported that squaric acid dimethyl esters,<sup>[127,128]</sup> as well as didecyl squarate<sup>[129]</sup> can be used for the single point attachment of carbohydrates to proteins. In addition, Kamath *et al.* used squaric acid amide ethyl esters for the conjugation of oligosaccharides to protein and monitored the conjugation using MALDI-TOF-MS.<sup>[130]</sup>

More recently, the group of Kováč utilized squaric acid chemistry to conjugate different carbohydrate antigens to a protein carrier.<sup>[131-134]</sup> They used this strategy to conjugate the synthetic tetrasaccharide side chain of the *Bacillus anthracis* exosporium to bovine serum albumin (BSA) protein.<sup>[133]</sup> Michael addition is a conjugation addition reaction, which has been used for targeting the sulfhydryl group of the cysteine residues, and the  $\epsilon$ - amino groups of lysine residues.<sup>[135-138]</sup> In this study, Michael addition was used to target the free sulfhydryl group of cysteine 34 (Cys 34) on BSA. BSA is rich in potential glycation sites because it contains 60 lysine residues and 35 cysteine residues, 34 of which form 17 disulfide groups, leaving one free sulfhydryl group on the Cys 34.<sup>[139-140]</sup>

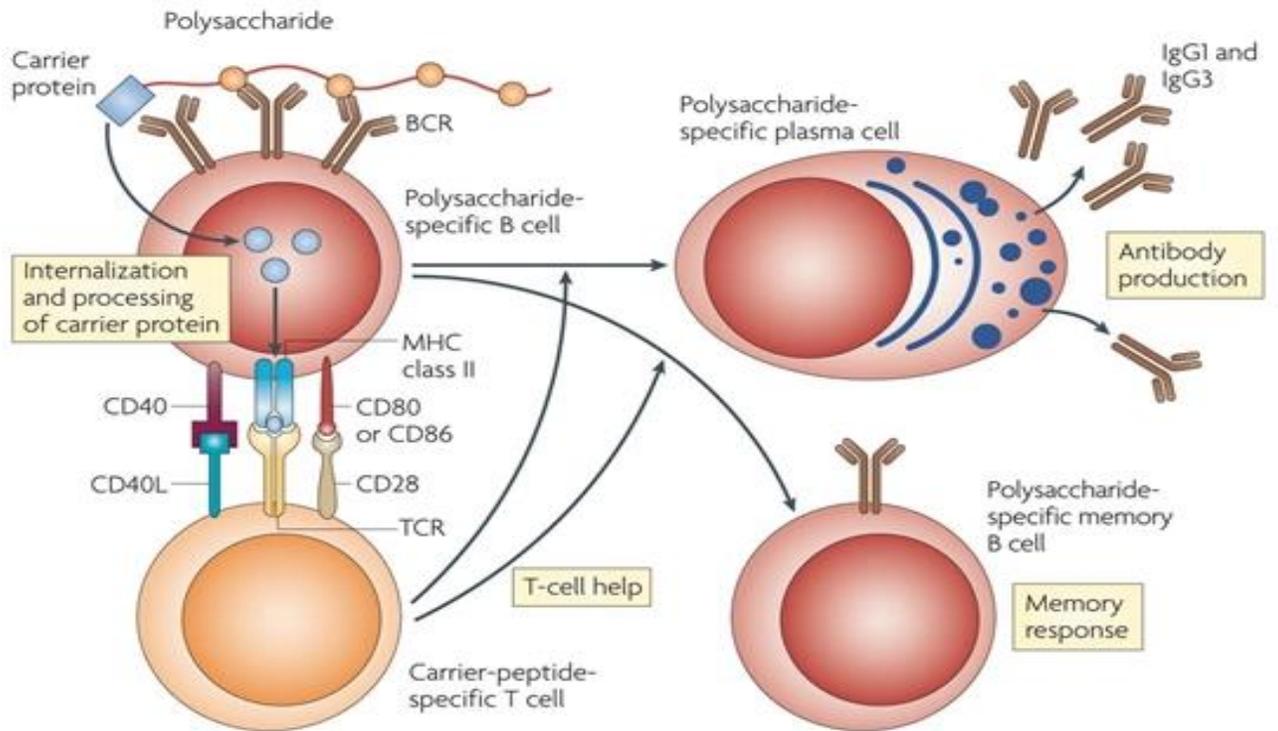
In essence, the carrier protein used in the glycoconjugate vaccine synthesis is very important, since this part is responsible for production of the memory cells which confers protection.<sup>[141]</sup> Recently, Roy's group had provided evidence for the role of the carrier protein.<sup>[142]</sup> The carrier protein had been replaced with repetitively branched molecules called dendrimers. Conversely, dendrimers do not elicit the memory B cells, although they provoke plasma B cells to secrete enough antibodies.<sup>[142]</sup>

### **1.3.5 Principle of glycoprotein vaccines**

The major role of glycoprotein vaccine is to confer protection through the production of memory cells. In case of infection, the immune system produces antibodies to kill the infected cells. However the memory pool is depleted and the memory cells are not secreted **Figure 1.13.**<sup>[141]</sup> **Figure 1.14** shows the molecular mechanisms of the generation of immune responses against protein-polysaccharide conjugate vaccines.<sup>[141]</sup>



**Figure 1.13** Generation of the immune responses against the infection.<sup>[141]</sup> BCR, B cell receptor; IgG2, immunoglobulin G2; IgM, immunoglobulin M. [Reproduced with the permission from the publisher of Ref. 141].

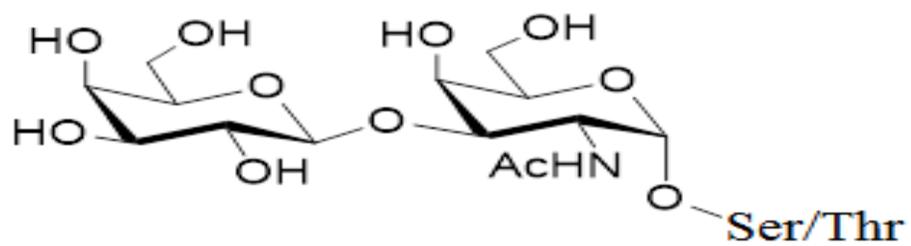


**Figure 1.14** Generation of the immune responses against oligosaccharide antigen-carrier protein conjugate vaccines <sup>[141]</sup> BCR, B cell receptor; MHC, Major histocompatibility protein; CD, cluster designation; TCR, T cell receptor; IgG1, immunoglobulin1; IgG3, immunoglobulin3. [Reproduced with the permission from the publisher of Ref. 141].

### 1.3.6 Thomsen–Friedenreich antigen (TF)

Thomsen–Friedenreich antigen (TF) is a TACA recognized by major histocompatibility protein class I.<sup>[142-147]</sup> TF is a disaccharide ( $\beta$ -galactose and  $\alpha$ -galactoseamine) bound to the hydroxyl group of threonine and serine residues on the core of mucin protein (**Figure 1.15**).<sup>[142-147]</sup> It helps different cancers such as breast, colon, prostate or bladder to metastasize, and it causes aberrant *O*-glycosylation of the mucin protein. In case of cancer, 50% of mucin *O*-sites gain *O*-glycosylation by TF and T<sub>N</sub> (N-acetylglucoseamine) antigens.<sup>[142-147]</sup> Additionally, sialic acid glycoproteins are overexpressed by cancer cells. So, in case of cancer, sialyl TF and sialyl T<sub>N</sub> have been formed, causing aberrant glycosylation.<sup>[142-147]</sup>

Several studies were reported for a drug design against the aforementioned antigen “TF”. Monoclonal antibodies had been produced from mouse such as JAA-F11 (IgG3) and C5 (IgM). More recently, Hoffmann-Röder *et al.* prepared several synthetic antitumor vaccines composed of BSA conjugates of mucin 1 (MUC1) glycopeptides with the TF (TF -MUC1-BSA conjugate) and a fluorine-substituted analogue (difluoro- TF -MUC1-BSA conjugate).<sup>[148,149]</sup>



**Figure 1.15** Typical structure of representative TF antigen bound to the hydroxyl group of serine (Ser) or threonine (Thr) on mucin glycoprotein accumulated on the cancer cell.

### **1.3.7 Strategy for characterization of glycoconjugate vaccines using MS**

Initially, application of MALDI-TOF-MS to the different hapten-BSA glycoconjugate vaccines allowed us to determine the hapten-to-BSA ratios. Then the glycoconjugate vaccine samples were digested and analyzed by MALDI-TOF/TOF-MS/MS and LC-ESI-QqTOF-MS/MS for the determination of glycation sites. The digestion is done by two different enzymes; trypsin which will not be able to digest or react with the glycated lysine residues of the protein and the other enzyme is GluC V8 endoproteinase which is known to digest proteins at C-terminus of the aspartic acid and glutamic acid residues. Finally the MS/MS spectra are submitted to Mascot library to get the matched and non-matched peptides.

### **1.3.8 Molecular mass and carbohydrate-to-protein ratio determination**

The purpose of determining the molecular mass and the carbohydrate-to-protein ratio of a carbohydrate-protein neoglycoconjugate is to define the number of carbohydrates that are incorporated in the protein carrier, as a result of conjugation. Two main methods are currently used for the molecular mass determination of carbohydrate-protein glycoconjugates: MALDI-TOF-MS<sup>[150,151]</sup> and surface enhanced laser desorption ionization time-of-flight mass spectrometry (SELDI-TOF-MS).<sup>[131-133,152-154]</sup>

Both of these methods allow the carbohydrate-to-protein ratio of the neoglycoconjugates to be determined by comparing the molecular mass of the protein before and after the conjugation to that of the neoglycoconjugate.

### **1.3.9 Glycation sites determination**

The glycation sites determination of carbohydrate-protein neoglycoconjugates is usually carried out by first digesting the neoglycoconjugate using a protease, such as

trypsin or GluC V8 endoproteinsase, followed by MALDI-TOF-MS/MS or LC-MS/MS.<sup>[83-85,155]</sup>

It has to be noted that during the tandem mass spectrometry analysis of the glycoconjugate digests, the identification of the glycated peptides are confirmed by the presence of diagnostic product ions of the carbohydrate in the mass spectrum. In addition, the tandem mass spectrometry analysis also reveals the sequence of the peptide through diagnostic product ions of the peptide moiety of the glycated peptide. The combined information allows the unambiguous characterization of the carbohydrate-peptide and glycation site identification. It has to be noted that during tandem mass spectrometry analyses of the glycated peptides, the product ions corresponding to the fragmentation of the peptide portion are identified using the nomenclature established by Roepstorff *et al.* and lately is modified by Johnson and coworkers,<sup>[86,87]</sup> and the product ions resulting from the fragmentation of the carbohydrate moiety are assigned using the nomenclature introduced by Domon and Costello, as: A, B, C, X, Y and Z.<sup>[88]</sup>

#### **1.4 Research objectives**

The objectives of this work are to use the mass spectrometric bottom-up proteomics approach for characterizing cryptocyanin protein in snow crab. Additionally, a proteomics strategy is used for mapping the glycation sites of an anti-cancer Thomsen–Friedenreich vaccine.

#### **1.5 Hypotheses**

MALDI-CID-MS/MS and ESI-CID-MS/MS can be used as effective proteomic approaches to sequence cryptocyanin and for glycation site mapping of the anti-tumor TF neoglycoconjugate vaccine. Gel electrophoresis is not an accurate tool for determining the

molecular mass of unknown proteins such as my protein of interest “cryptocyanin”. I hypothesize that single stage MALDI will prove to be an accurate means of measuring the molecular mass of protein, particularly cryptocyanin. In addition, cryptocyanin protein of snow crab in the non-molting stage is expected to change its structure to adapt its function as a key player in hardening the new shell of the snow crab, and its role as antimicrobial protein. I hypothesize that these changes can be identified by single stage MALDI technique.

In the second part of this work, I further hypothesised that Cys 34 residue on BSA is more reactive than the lysine residues, accordingly two ratios (one with high glycation sites and the other with low glycation sites) are prepared using Michael addition, which is specific for targeting the cysteine and lysine residues.<sup>[139-140]</sup>

## **1.6 Summary**

Bottom up proteomics is an exceptional approach that uses MS/MS for protein sequencing. In this thesis, examples of the uses of MS/MS for different biomolecules will be presented. Two key examples include: MALDI and MS/MS for the determination of changes in snow crab cryptocyanin protein, and ESI-QqTOF MS/MS is used for mapping of the glycation sites of the anti-tumor TF neoglycoconjugate vaccine, as well as the glycation of the free sulfhydryl group of Cys 34 of BSA.

## References

- 1- R. Aebersold, M. Mann. Mass spectrometry-based proteomics. *Nature* **2003**, 422, 198.
- 2- W. Morelle, J.C. Michalski. Glycomics and mass spectrometry. *Curr Pharm Des.* **2005**, 11, 2615.
- 3- K. Dettmer, P.A. Aronov, B.D. Hammock. Mass spectrometry-based metabolomics. *Mass Spectrom Rev.* **2007**, 26, 51.
- 4- S.J. Blanksby, T.W. Mitchell. Advances in mass spectrometry for lipidomics. *Ann. Rev. Anal. Chem.* **2010**, 3, 433.
- 5- J.H. Banoub, R.P. Newton, E. Esmans, D.F. Ewing, G. Mackenzie. Recent developments in mass spectrometry for the characterization of nucleosides, nucleotides, oligonucleotides, and nucleic acids. *Chem Rev.* **2005**, 105, 1869.
- 6- K. Downard. Mass spectrometry: a foundation course, TJ International Ltd, Cornwall, **2004**.
- 7- M.A. Baldwin. Mass spectrometers for the analysis of biomolecules. *Methods Enzymol* **2005**, 402, 3.
- 8- F.W. Aston. Mass spectra and isotopes, Longmans, Green and Co, New York **1942**.
- 9- T.W. Hutchens, T.T. Yip. New desorption strategies for mass spectrometric analysis of macromolecules. *Rapid Commun Mass Spectrom* **1993**, 7, 576.

- 10- W. Henderson, J.S. McIndoe. Mass spectrometry of inorganic and organometallic compounds. John Wiley & Sons Ltd., Chichester **2005**.
- 11- I.V. Chernushevich, A.V. Loboda, B.A. Thomson. An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom* **2001**, *36*, 849.
- 12- B. Soltmann, C.C. Sweeley, J.F. Holland. Electron impact ionization mass spectrometry using field desorption activated emitters as solid sample probes. *Anal. Chem.* **1977**, *49*, 1164.
- 13- R. Gohlke, F.W. McLafferty. Early gas chromatography/mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 367.
- 14- E. De Hoffmann, V. Stroobant. In mass spectrometry: principles and applications (second ed.). Toronto: John Wiley & Sons, Ltd. p. 14, **2003**.
- 15- M.S.B. Munson, F.H. Field. Chemical ionization mass spectrometry. I. general introduction. *J. Am. Chem. Soc.* **1966**, *88*, 2621.
- 16- K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida. Protein and polymer analysis up to  $m/z$  100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.***1988**, *2*, 151
- 17- K.L. Busch. Desorption Ionization Mass Spectrometry. *J. Mass Spectrom.* **1995**, *30*, 233.
- 18- Tools of the Trade; Matrix assisted laser desorption ionization (MALDI). (n.d.), from Magnet Lab; National High Magnetic Field Laboratory; FSU: Retrieved May 5, **2009**. [http://www.magnet.fsu.edu/education/tutorials/tools/ionization\\_maldi.html](http://www.magnet.fsu.edu/education/tutorials/tools/ionization_maldi.html)

- 19- R.C. Beavis, B.T. Chait. Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Commun. Mass Spectrom.* **1989**, 3, 432.
- 20- R.C. Beavis, B.T. Chait.  $\alpha$ -cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry. *Org. Mass Spectrom.* **1992**, 27, 156.
- 21- R.C. Beavis, B.T. Chait. Velocity distributions of intact high mass polypeptide molecule ions produced by matrix assisted laser desorption. *Chem. Phys. Lett.* **1991**, 181, 479.
- 22- R. Knochenmuss, R. Zenobi. MALDI ionization: the role of in-plume processes. *Chem. Rev.* **2003**, 103, 441.
- 23- M. Karas, M. Glückmann, J. Schäfer. Ionization in MALDI: singly charged molecular ions are the lucky survivors", *J. Mass Spectrom.* **2000**,35, 1.
- 24- K. Strupat, M. Karas, F. Hillenkamp. 2,5-Dihydroxybenzoic acid: a new matrix for laser desorption-ionization mass spectrometry. *Int. J. Mass Spectrom. Ion Processes* **1991**, 72, 89.
- 25- B. Guo, S. Wang, Y. Fan. Improving the performance of MALDI-TOF in oligonucleotide analysis using a new SDIFA technology. *Anal. Chem.*, **2000**, 72, 5792.
- 26- R.G. Keck, J.B. Briggs, A.J.S. Jones. Oligosaccharide release and MALDI-TOF MS analysis of N-linked carbohydrate structures from glycoproteins. *Meth. Mol. Biol.* **2005**, 308, 381.
- 27- B. Fuchs, J. Schiller. MALDI-TOF MS analysis of lipids from cells, tissues and body fluids. *Subcell. Biochem.* **2008**, 49, 541.

- 28- A.Walch, S. Rauser, S.O. Deininger, H. Höfler. MALDI imaging mass spectrometry for direct tissue analysis: a new frontier for molecular histology. *Histochem. Cell Biol.* **2008**, *130*, 421.
- 29- V.V. Laiko, M.A. Baldwin, A.L. Burlingame. Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **2000**, *72*, 652.
- 30- V.V. Laiko, N.I. Taranenko, V.D. Berkout, B.D. Musselman, V.M. Doroshenko. Atmospheric pressure laser desorption/ionization on porous silicon. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1737.
- 31- C.M. Whitehouse, R.N. Dreyer, M. Yamashita, J.B. Fenn. Electrospray interface for liquid chromatographs and mass spectrometers. *Anal Chem* **1985**, *57*, 675.
- 32- M. Dole, L. Mach, R.L. Hines, R.C. Mobley, L.D. Ferguson, M.B. Alice. Molecular beams of macroions. *J Chem Phys* **1968**, *49*, 2240.
- 33- C.K. Meng, M. Mann, J.B. Fenn. Of protons or proteins. A beam's a beam for a' that. (Burns OS). *Z. Phys. D. Atoms, Molecules and Clusters.* **1988**, *10*, 361.
- 34- C.S. Ho, C.W.K. Lam, M.H.M. Chan, R.C.K. Cheung, L.K. Law, L.C.W. Lit, K.F. Ng, M.W.M. Suen, H.L. Tai. Electrospray ionization mass spectrometry: principles and clinical applications. *Clin. Biochem. Rev.* **2003**, *24*, 3.
- 35- N.B. Cech, C.G. Enke. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.* **2002**, *20*, 362.
- 36- A.P. Bruins. Mechanistic aspects of electrospray ionization. *J. Chromatogr. A.* **1998**, *794*, 345.

- 37- L. Rayleigh. On the equilibrium of liquid conducting masses charged with electricity. *Philos. Mag.* **1882**, *14*, 184.
- 38- M. Abonnenc, L. Qiao, B.H. Liu, H.H. Girault. Electrochemical aspects of electrospray and laser desorption/ionization for mass spectrometry. *Annu. Rev. Anal. Chem.* **2010**, *3*, 231.
- 39- E. De Hoffmann, V. Stroobant. In mass spectrometry: principles and applications. Ed. John Wiley and Sons, p.46, **2007**.
- 40- M. Yamashita, J.B. Fenn. Electrospray ion source. Another variation on the free-Jet theme. *J. Phys. Chem.* **1984**, *88*, 4451.
- 41- H.Y. Kim, N.S. Jr. Application of thermospray high performance liquid chromatography/mass spectrometry for the determination of phospholipids and related compounds. *Anal. Chem.* **1987**, *59*, 722.
- 42- P.H. Dawson. Quadrupole mass spectrometry and its applications. Elsevier, Amsterdam, **1976**. (reissued by AIP Press: Woodbury, NY, **1995**).
- 43- M.A. Grayson., Ed. Measuring mass: from positive rays to proteins, Chemical Heritage Press, Philadelphia, **2002**.
- 44- L. Pasa-Tolic, C. Masselon, R.C. Barry, Y. Shen, R.D. Smith. Proteomic analyses using an accurate mass and time tag strategy. *Biotechniques* **2004**, *37*, 62.
- 45- L. Yang, T.D. Mann, D. Little, N. Wu, R.P. Clement, P.J. Rudewicz. Evaluation of a four-channel multiplexed electrospray triple quadrupole mass spectrometer for the simultaneous validation of LC/MS/MS methods in four different preclinical matrixes. *Anal. Chem.* **2001**, *73*, 1740.

- 46- O. Fiehn, J. Kopka, R.N. Trethewey, L. Willmitzer. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.* **2000**, *72*, 3573.
- 47- P. Schmitt-Kopplin, M. Frommberger. Capillary electrophoresis – mass spectrometry: 15 years of developments and applications. *Electrophor.* **2003**, *24*, 3837.
- 48- S.A. McLuckey, J.M. Wells. Mass analysis at the advent of the 21st century. *Chem. Rev.* **2001**, *101*, 571.
- 49- [http://huygenscms.gsfc.nasa.gov/MS\\_Analyzer\\_1.htm](http://huygenscms.gsfc.nasa.gov/MS_Analyzer_1.htm)
- 50- W.E. Stephens. A pulsed mass spectrometer with time dispersion. *Phys. Rev.* **1946**, *69*, 691.
- 51- W.L. Wiley, I.H. McLaren. Time-of-Flight mass spectrometer with improved resolution. *Rev. Sci. Instrum.* **1955**, *16*, 1150.
- 52- A. Ingendoh, M. Karas, F. Hillenkamp, U. Giessmann. Factors affecting the mass resolution in matrix-assisted laser desorption-ionization mass spectrometry. *Int. J. Mass Spectrom. Ion Process.* **1994**, *131*, 345.
- 53- G.H. Herbert, R.A.W. Johnstone. In mass spectrometry basics. Boca Raton: CRC Press LLC, p. 189. **2003**.
- 54- W. Pusch, M. Kostrzewa. Application of MALDI-TOF mass spectrometry in screening and diagnostic research. *Curr. Pharm. des.* **2005**, *11*, 2577.
- 55- H.R. Morris, T. Paxton, M. Panico, R. McDowell, A. Dell. A Novel geometry mass spectrometer, the Q-TOF, for low-femtomole/attomole-range biopolymer sequencing. *J. Prot. Chem.* **1997**, *16*, 469.

- 56- H. Oberacher, F. Pitterl. On the use of ESI-QqTOF-MS/MS for the comparative sequencing of nucleic acids. *Biopol.* **2009**, *91*, 401.
- 57- N .Tachon, F. Jahouh, M. Delmas, J.H. Banoub. Structural determination by atmospheric pressure photoionization tandem mass spectrometry of some compounds isolated from the SARA fractions obtained from bitumen. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 2657.
- 58- A. Wattenberg, A.J. Organ, K. Schneider, R. Tyldesley, R. Bordoli, R.H. Bateman. Sequence dependent fragmentation of peptides generated by MALDI quadrupole time-of-flight (MALDI Q-TOF) mass spectrometry and its implications for protein identification. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 772.
- 59- J.W. Flora, A.P. Null, D.C. Muddiman. Dual-micro-ESI source for precise mass determination on a quadrupole time-of-flight mass spectrometer for genomic and proteomic applications. *Anal. Bioanal. Chem.* **2002**, *373*, 538.
- 60- P. Hufnagel, R. Rabus. Mass spectrometric identification of proteins in complex post-genomic projects. *J. Mol. Microbiol. Biotechnol.* **2006**, *11*, 53.
- 61- R. E. March, R. J. Hughes. Quadrupole storage mass spectrometry, John Wiley & Sons, Inc, New York **1989**.
- 62- M. Merchant, S.R. Weinberger. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophor.* **2000**, *21*, 1164.
- 63- R.S. Brown, J.J. Lennon. Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer. *Anal Chem* **1995**, *67*, 1998.

- 64- R.D. Edmondson, D.H. Russell. Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass measurement accuracy by using delayed extraction. *J Am Soc Mass Spectrom* **1996**, 7, 995.
- 65- B.A. Mamyrin, D.V. Shmikk. The linear mass reflectron. *Sov Phys. JETP* **1979**, 49, 762.
- 66- A. Shevchenko, I. Chernushevich, W. Ens, K.G. Standing, B. Thomson, M. Wilm, M. Mann. Rapid *de novo* peptide sequencing by a combination of nanoelectrospray, isotopic labelling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **1997**, 11, 1015.
- 67- E. De Hoffmann. Tandem mass spectrometry: a primer. *J. Mass Spectrom.* **1996**, 31, 129.
- 68- F.W. McLafferty. In tandem mass spectrometry, John Wiley and Sons, New York, **1983**.
- 69- W.J. Griffiths, A.P. Jonsson, S. Liu, D.K. Rai, Y. Wang. Electrospray and tandem mass spectrometry in biochemistry. *Biochem. J.* **2001**, 355, 545.
- 70- J. Li, S.M. Assmann. Mass spectrometry. An essential tool in proteome analysis. *Plant Physiol.* **2000**, 123, 807.
- 71- J.V. Johnson, R.A. Yost, P.E. Kelley, D.C. Bradford. Tandem-in-space and tandem-in-time mass spectrometry: triple quadrupoles and quadrupole ion traps. *Anal. Chem.* **1990**, 62, 2162.

- 72- A.W. Purcell, J.J. Gorman. The use of post-source decay in matrix-assisted laser desorption/ionisation mass spectrometry to delineate T cell determinants. *J Immunol Methods* **2001**, 249, 17.
- 73- N. Joly, A. El-Aneed, P. Martin, R. Cecchelli, J.H. Banoub. Structural determination of the novel fragmentation routes of morphine opiate receptor antagonists using electrospray ionization quadrupole time-of-flight tandem mass spectrometry. *Rapid Commun Mass Spectrom* **2005**, 19, 3119.
- 74- P. Hernandez, M. Müller, R. D. Appel. Automated protein identification by tandem mass spectrometry: issues and strategies. *Mass Spectrometry Reviews* **2006**, 25, 235.
- 75- F. Jahouh , W.L.L. Demian, R. Sakksena, H. Shu-Jie, R.J. Brown, P. Kovac, R. Roy, J.H. Banoub. Glycoconjugate vaccines used for prevention from biological agents: tandem mass spectrometric analysis. Detection of chemical, biological, radiological and nuclear agents for the prevention of terrorism NATO science for peace and security series A: Chemistry and Biology, pp 233-274, **2014**.
- 76- B.T. Chait. Chemistry. Mass spectrometry: bottom-up or top-down?. *Science* **2006**, 314, 65.
- 77- S.K. Sze, Y. Ge, H. Oh, F.W. McLafferty. Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 1774.

- 78- J. Whitelegge, F. Halgand, P. Souda, V. Zabrouskov. Top-down mass spectrometry of integral membrane proteins. *Expert review of proteomics* **2006**, *3*, 585.
- 79- R. Lakshmanan, J.J. Wolff, R. Alvarado, J.A. Loo. Top-down protein identification of proteasome proteins with nanoLC-FT-ICR-MS employing data-independent fragmentation methods. *Proteomics* **2014**,*14*,1271.
- 80- A. Kiss, D.F. Smith, B.R. Reschke, M.J. Powell, R.M. Heeren. Top-down mass spectrometry imaging of intact proteins by laser ablation ESI FT-ICR MS. *Proteomics* **2014**, *14*, 1283.
- 81- D.A. Wolters, M.P. Washburn, J.R. Yates. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, *73*, 5683.
- 82- J.R. Yates, C.I. Ruse, A. Nakorchevsky. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **2009**, *11*, 49.
- 83- F. Jahouh, R. Saksena, D. Aiello, A. Napoli, G. Sindona, P. Kovác, J.H. Banoub. Glycation sites in neoglycoconjugates from the terminal monosaccharide antigen of the O-PS of *Vibrio cholerae* O1, serotype Ogawa, and BSA revealed by matrix-assisted laser desorption-ionization tandem mass spectrometry. *J. Mass Spectrom.* **2010**, *10*, 1148.
- 84- W. Morelle, J.C. Michalski. Analysis of protein glycosylation by mass spectrometry. *Nature Protocols* **2007**, *2*, 1585.
- 85- F. Jahouh, S. Hou, P. Kovác, J.H. Banoub. Determination of glycation sites by tandem mass spectrometry in a synthetic lactose-bovine serum albumin conjugate, a

vaccine model prepared by dialkyl squarate chemistry. *Rapid Commun. Mass Spectrom.* **2012**, 26, 1.

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

87- R.S. Johnson, S.A. Martin, K. Biemann, J.T. Stults, J.T. Watson. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* **1987**, 59, 2621.

88- B. Domon, C. Costello C. A systematic nomenclature of carbohydrate fragmentation in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **1988**, 5, 397.

89- D.R.J. Mallowney, E.M. Hynick, E.G. Dawe, W.A.Coffey. In distribution and habitat of cold water species on the Grand Bank of Newfoundland, (Eds: K. Sasruwatari, M. Nisshimura), NOVA, New York, **2012**.

90- W.S. Fredrick, S.Ravichandran. Hemolymph proteins in marine crustaceans. *Asian Pac J Trop Biomed.* **2012**, 496, 502.

91- H.S. Ai, J.X. Liao, X.D. Huang, Z.X. Yin, S.P. Weng, Z.Y. Zhao. Anovel prophenoxidase 2 exists in shrimp hemocytes. *Dev Comp Immunol* **2009**, 33, 59

92- M. Jimbo, R. Usui, R. Sakai, K. Muramoto, H. Kamiya. Purification, cloning and characterization of egg lectins from the teleost *Tribolodon brandti*. *Comp Biochem Physiol B Biochem Mol Biol.* **2007**, 147, 164.

- 93- Z.M. Darnell, D. Rittschof. Role of larval release pheromones and peptide mimics in abdominal pumping and swimming behavior of ovigerous blue crabs, *Callinectes sapidus*. *J Exp Mar Biol Ecol* **2010**, 391, 112.
- 94- C. Imjongjirak, P. Amparyup, A. Tassanakajon. Two novel antimicrobial peptides, arasin-likeSp and GRPSp, from the mud crab *Scylla paramamosain*, exhibit the activity against some crustacean pathogenic bacteria. *Fish Shellfish Immunol*. **2011**, 30,706.
- 95- G. Rameshkumar, S. Ravichandran. Antimicrobial peptide from the crab, *Thalamita crenata* (Latreille, 1829). *World J Fish Mar Sci* **2009**,1,74.
- 96- V.J. Smith, J.M.O. Fernandes, G.D. Kemp, C. Hauton. Crustins: enigmatic WAP domain-containing antibacterial proteins from crustaceans. *Dev Comp Immunol* **2008**, 32, 758.
- 97- S.V. Sperstad, T. Haug, T. Vasskog, K. Stensvåg. Hyastatin, a glycine-rich multi-domain antimicrobial peptide isolated from the spider crab (*Hyas araneus*) hemocytes. *Mol Immunol* **2009**, 46, 2604.
- 98- J. Ewer. How the Ecdysozoan Changed Its Coat. *PLoS Biol* **2005**, 3, 1696.
- 99- W.S. Fredrick, S. Ravichandran. Hemolymph proteins in marine crustaceans. *Asian Pac J Trop Biomed*. **2012**, 496, 496.
- 100- N.B. Terwiliger, L. Dangott, M. Ryan. Cryptocyanin, a crustacean molting protein: Evolutionary link with arthropods hemocyanins and insect hexamerins. *Proc. Natl. Acad. Sci. USA*. **1999**, 96, 2013.
- 101- N.B .Terwiliger, M.C. Ryan, D. Towl. Evolution of novel functions: cryptocyanin helps build new exoskeleton in *Cancer magister*. *J. Exp. Biol*. **2005**, 208, 2467.

- 102- J.J. Beintema, W.T. Stam, B. Hazes, M.P. Smidt. Evolution of arthropod hemocyanins and insect storage proteins (hexamerins). *Mol. Biol. Evol.* **1994**, *11*, 493.
- 103- M. Brouwer, R. Syring, T.H. Brouwer. Role of copper-specific metallothionein of the blue crab *Callinectes sapidus*. In copper metabolism associated with degradation and synthesis of hemocyanin. *J. Inorg. Biochem.* **2002**, *88*, 228.
- 104- N.B. Terwiliger. Hemolymph proteins and molting in crustaceans and insects *Am. Zool.* **1999**, *39*, 589.
- 105- N.B. Terwiliger. Gene expression profile, protein production, and functions of cryptocyanin during the crustacean molt cycle. *Invertebr. Reprod. Dev.* **2012**, *56*, 229.
- 106- S.A. Plotkin. Vaccines: correlates of vaccine-induced immunity. *Clin. Infect. Dis.* **2008**, *47*, 401.
- 107- M. Heidelberger, O.T. Avery. The soluble specific substance of pneumococcus. *J. Exp. Med.* **1923**, *38*, 73.
- 108- Vaccine types. [Niaid.nih.gov](http://niaid.nih.gov). (2012-04-03).
- 109- J.K. Sinha, S. Bhattacharya. A text book of immunology (google book preview). academic publishers. p. 318, **2014**.
- 110- J.T. Van Oirschot. Diva vaccines that reduce virus transmission. *J. Biotechnol.* **1999**, *73*, 195.
- 111- I. Capua, C. Terregino, G. Cattoli, F. Mutinelli, J.F. Rodriguez. Development of a DIVA (differentiating infected from vaccinated animals) strategy using a vaccine

containing a heterologous neuraminidase for the control of avian influenza. *Avian Patholo.* **2003**, 32, 47.

112- E. Pupo, A. Aguila, H. Santana, J.F. Núñez, L. Castellanos-Serra, E. Hardy. Mice immunization with gel electrophoresis-micropurified bacterial lipopolysaccharides. *Electrophor.* **1999**, 20, 458.

113- M.R. Davis Jr, J.B. Goldberg. Purification and visualization of lipopolysaccharide from Gram-negative bacteria by hot aqueous-phenol extraction. *J. Vis. Exp.* **2012**, 28, 1.

114- D. Reisser, A. Pance, J.F. Jeannin. Mechanisms of the antitumoral effect of lipid A. *BioEssays* **2002**, 24, 284.

115- G. Nagy, T. Pál. Lipopolysaccharide: a tool and target in enterobacterial vaccine development. *Biol Chem.* **2008**, 389, 513.

116- R.S. Daum, D. Hogerman, M.B. Rennels, K. Bewley, F. Malinoski, E. Rothstein, K. Reisinger, S. Block, H. Keyserling, M. Steinhoff. Infant immunization with pneumococcal CRM<sub>197</sub> vaccines: effect of saccharide size on immunogenicity and interactions with simultaneously administered vaccines. *J. Infect. Dis.* **1997**, 176, 445.

117- D.J. Lefeber, J.P. Kamerling, J.F.G. Vliegthart. Synthesis of Streptococcus pneumoniae type 3 neoglycoproteins varying in oligosaccharide chain length, loading and carrier protein *Chem. Eur. J.* **2001**, 7, 4411.

118- L.C. Paoletti, D.L. Kasper, F. Michon, J. DiFabio, H.J. Jennings, T.D. Tosteson, M.R. Wessels. Effects of chain length on the immunogenicity in rabbits of group B. *Streptococcus* type III oligosaccharide-tetanus toxoid conjugates. *J. Clin. Invest.* **1992**, 89, 203.

- 119- A. Chernyak, S. Kondo, T.K. Wade, M.D. Meeks, P.M. Alzari, J.M. Fournier, R.K. Taylor, P. Kováč, W.F. Wade. Induction of protective immunity by synthetic *Vibrio cholerae* hexasaccharide derived from *V. cholerae* O1 Ogawa lipopolysaccharide bound to a protein carrier. *J. Infect. Dis.* **2002**, *185*, 950.
- 120- R. Roy. New trends in carbohydrate-based vaccines. *Drug Discov. Today: Technol.* **2004**, *1*, 327.
- 121- G .Ada, D. Isaacs. Carbohydrate-protein conjugate vaccines. *Clin. Microbiol. Infect.* **2003**, *9*, 79.
- 122- K. Landsteiner. The specificity of serological reactions. Cambridge: Harvard Univ. Press, **1945**.
- 123- O.T. Avery, W.F. Goebel. Chemo-immunological studies on conjugated carbohydrate-proteins. II. Immunological specificity of synthetic sugar-protein Ags. *J. Exp. Med.* **1929**, *50*, 533.
- 124- W.E. Dick Jr, M. Beurret. In *Glycoconjugates of bacterial carbohydrate Ags*; J. M. Cruse, R. E. Lewis Jr., Eds. Krager: Basel. **1989**, *10*, 48.
- 125- L.F. Tietze, M. Arlt, M. Beller, K.H. Glüsenkamp, E. Jähde, M.F. Rajewsky. Anticancer agents, 15. Squaric acid diethyl ester: a new coupling reagent for the formation of drug biopolymer conjugates. Synthesis of squaric acid ester amides and diamides. *Chem. Ber.* **1991**, *124*, 1215.
- 126- K.H. Glüsenkamp, W. Drosdziok, G. Eberle, E. Ja'hde, M.F.Z. Rajewsky. *Naturforsch. C: Biosci.* **1991**, *46*, 498.

- 127- L.F. Tietze, C. Schröter, S. Gabius, U. Brinck, A. Goerlach-Graw, H.J. Gabius. Conjugation of p-aminophenyl glycosides with squaric acid diesters to a carrier protein and the use of the neoglycoprotein in the histochemical detection of lectines. *Bioconjugate Chem.* **1991**, 2, 148.
- 128- S. Cohen, S.G. Cohen. Preparation and reactions of derivatives of squaric acid. Alkoxy-, hydroxy-, and aminocyclobutenediones<sup>1</sup>. *J. Am. Chem. Soc.* **1966**, 88, 1533.
- 129- J. Grünefeld, G. Bredhauer, G. Zinner. Zur reaktion von quadratsäuredimethylester mit *N,N*-disubstituierten hydrazin-derivaten. *Arch. Pharm. (Weinheim)* **1985**, 318, 984.
- 130- A. Bergh, B.G. Magnusson, J. Ohlsson, U. Wellmar, U.J. Nilsson. Didecyl squarate – A practical amino-reactive cross-linking reagent for neoglycoconjugate synthesis. *Glycoconjugate J.* **2001**, 18, 615.
- 131- V.P. Kamath, P. Diedrich, O. Hindsgaul. Use of diethyl squarate for the coupling of oligosaccharide amines to carrier proteins and characterization of the resulting neoglycoproteins by MALDI-TOF mass spectrometry. *Glycoconjugate J.* **1996**, 13, 315.
- 132- S.J. Hou, R. Saksena, P. Kováč. Preparation of glycoconjugates by dialkyl squarate chemistry revisited. *Carbohydrate Res.* **2008**, 343, 196.
- 133- R. Saksena, R. Adamo, P. Kováč. Immunogens related to the synthetic tetrasaccharide side chain of the *Bacillus anthracis* exosporium. *Bioorg Med Chem.* **2007**, 15, 4283.
- 134- A.F.G. Bongat, R. Saksena, R. Adamo, Y. Fujimoto, Z. Shiokawa, D.C. Peterson, K. Fukase, W.F. Vann, P. Kováč. Multimeric bivalent immunogens from recombinant

tetanus toxin HC fragment, synthetic hexasaccharides and a glycopeptide adjuvant. *Glycoconjugate J.* **2010**, *27*, 69.

135- D.C. Liebler. Protein damage by reactive electrophiles: targets and consequences. *Chem. Res. Toxicol.* **2008**, *21*, 117.

136- P.C. Ohe, R. Kuhne, R.U. Ebert, R. Altenburger. Structural alerts - a new classification model to discriminate excess toxicity from narcotic effect levels of organic compounds in the acute daphnid assay. *Chem. Res. Toxicol.* **2005**, *18*, 536.

137- D.W. Roberts, T.W. Schultz, E.M. Wolf, A.O. Aptula. Experimental reactivity parameters for toxicity modeling: application to the acute aquatic toxicity of S2 electrophiles to *Tetrahymena pyriformis*. *Chem. Res. Toxicol.* **2010**, *23*, 228.

138- I. Chipinda, J.M. Hettick, P.D. Siegel. Haptenation: chemical reactivity and protein binding. *J. Allergy.* **2011**, *2011*, 10.

139- J.F. Janatova, J.K. Fuller, M. Hunter. The heterogeneity of bovine albumin with respect to sulfhydryl and dimer content. *J. J. Biol. Chem.* **1968**, *243*, 3612.

140- K.L. Heredia, D. Bontempo, T. Ly, J.T. Byers, S. Halstenberg, H.D. Maynard. In situ preparation of protein-“smart” polymer conjugates with retention of bioactivity. *J. Am. Chem. Soc.* **2005**, *127*, 16955.

141- A.J. Pollard, K.P. Perrett, P.C. Beverley. Maintaining protection against invasive bacteria with protein-polysaccharide conjugate vaccines. *Nature Rev.* **2009**, *9*, 213.

142- R. Roy, T.C. Shiao. Glycodendrimers as functional antigens and antitumorales Vaccines. *New J. Chem.* **2012**, *36*, 324.

- 143- G.F. Springer. Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy. *J. Mol. Med.* **1997**, *75*, 594.
- 144- B.J. Campbell, I.A. Finnie, E.F. Hounsell, J.M. Rhodes. Direct demonstration of increased expression of Thomsen-Friedenreich (TF) antigen in colonic adenocarcinoma and ulcerative colitis mucin and its concealment in normal mucin. *J. Clin. Invest.* **1995**, *95*, 571.
- 145- G.F. Springer. T and Tn, general carcinoma autoantigens. *Science* **1984**, *224*, 1198.
- 146- W. Dippold, A. Steinborn, K.H.M. Büschenfelde. The role of the thomsen-Friedenreich antigen as a tumor-associated molecule. *Environmental Health Perspectives* **1990**, *88*, 255.
- 147- A.M. Shamsuddin, G.T. Tyner, G.Y. Yang. Common expression of the tumour marker D-Galactose- $\beta$ -[1-3]-N-Acetyl-DGalactosamine by different adenocarcinomas: evidence of field effect phenomenon. *Cancer Res.* **1995**, *55*, 149.
- 148- K. Rittenhouse-Diakun, Z. Xia, D. Pickhardt, S. Morey, M.-G Baek, R. Roy. Development and characterization of monoclonal antibody to T-antigen: (Gal $\beta$ 1-3GalNAc- $\alpha$ -O). *Hybridoma* **1998**, *17*, 165.
- 149- J. Heimbürg-Molinaro, A. Almogren, S. Morey, O.V. Glinskii, R. Roy, G.E. Wilding, R.P. Cheng, V.V. Glinsky, K. Rittenhouse-Olson. Development, characterization, and immunotherapeutic utilization of peptide mimics of the Thomsen-Friedenreich carbohydrate antigen. *Neoplasia* **2009**, *11*, 780.

- 150- Y. Zhang, E.P. Go, H. Desaire. Maximizing coverage of glycosylation heterogeneity in MALDI-MS analysis of glycoproteins with up to 27 glycosylation sites. *Anal. Chem.* **2008**, *80*, 3144.
- 151- M. Laštovičková, J. Chmelik, J. Bobalova. The combination of simple MALDI matrices for the improvement of intact glycoproteins and glycans analysis. *Int. J. Mass Spectrom.* **2009**, *281*, 82.
- 152- V.P. Kamath, P. Diedrich, O. Hindsgaul. Use of diethyl squarate for the coupling of oligosaccharide amines to carrier proteins and characterization of the resulting neoglycoproteins by MALDI-TOF mass spectrometry. *Glycoconj. J.* **1996**, *13*, 315.
- 153- H.J. Issaq, T.P. Conrads, D.A. Prieto, R. Tirumalai, T.D. Veenstra. SELDI-TOF MS for diagnostic proteomics. *Anal Chem.* **2003**, *75*, 148A.
- 154- C. Liu. The application of SELDI-TOF-MS in clinical diagnosis of cancers. *J Biomed Biotechnol.* **2011**, *2011*, 1.
- 155- F. Jahouh, P. Xu, W.F. Vann, P. Kováč, J.H. Banoub. Mapping the glycation sites in the neoglycoconjugate from hexasaccharide antigen of *Vibrio cholerae*, serotype Ogawa and the recombinant tetanus toxin C-fragment carrier. *J. Mass Spectrom.* **2013**, *48*, 1083.

**Chapter II: Differentiation between the crustacean cryptocyanin protein during molting and non-molting processes of the snow crab (*Chionoecetes opilio*) using matrix-assisted laser desorption/ionization mass spectrometry and tandem mass spectrometry**

This chapter has been published: Wael L. L. Demian, Farid Jahouh, Don Stansbury, Edward Randell, Robert J. Brown and Joseph H. Banoub (*Rapid Commun. Mass Spectrom.* **2014**, 28, 355–369)

**Chapter II** was designed by Wael L. L. Demian and Joseph H. Banoub, and Wael L. L. Demian conducted the research, analyzed the data and prepared the manuscript with guidance from Joseph H. Banoub. The final manuscript was read and approved by all authors.

## 2.1 Introduction

The harvesting of snow crab is one of the largest and commercially important fisheries on the Grand Bank of Newfoundland's southeast coast in the northwest Atlantic.<sup>[1]</sup> It is one of the most valuable crustacean invertebrate catches for commercial fishers. It is well known that the immune system of crabs is very simple and it lacks immunological specificity since it does not contain lymphocytes or memory cells; instead, the immune system relies on antimicrobial peptides or proteins.<sup>[2]</sup>

During its molting period, the crab reabsorbs some of the calcium carbonate from the old exoskeleton, and secretes enzymes to separate the old shell from the underlying skin (or epidermis) which secretes a new, soft, paper-like shell beneath the old one.<sup>[2-4]</sup> During this molting phase the snow crabs are susceptible to infection as their outer shells become soft.<sup>[4,5]</sup>

It has been suggested that the high level of cryptocyanin protein in the crab molting stage, reflects its function adaptability as an antimicrobial protein used to kill microbes.<sup>[2-4]</sup> In consequence, cryptocyanin is considered a significant antifungal, antimicrobial and antiviral protein which is present in the hemolymph.

Both cryptocyanin and hemocyanin proteins are synthesized in specific cells of the crab hepatopancreas and are secreted directly in the hemolymph.<sup>[2-4]</sup> In addition, jointly hemocyanin and cryptocyanin constitute the most abundant proteins of the crustacean hemolymph. Unlike hemocyanin which contains copper, cryptocyanin is a copper free protein which does not bind to the oxygen. Consequently, cryptocyanin has been described as a non-respiratory protein.<sup>[3]</sup> Nevertheless, cryptocyanin resembles hemocyanin in its sequence, structure, size and it is produced from the hemocyanin

duplication gene.<sup>[3]</sup> Besides, cryptocyanin is responsible for profound physiological functions such as ion regulation and cardiovascular functions.<sup>[3]</sup>

The expression of cryptocyanin has been reported to be controlled by the crustacean molt inhibiting hormone (MIH) which is synthesized and excreted from the important endocrine organ in the optic ganglia of the crustacean's eyestalk called the X organ-sinus gland system (XO-SG) eyestalk.<sup>[3-11]</sup>

In this manuscript, we report the measurements of the level of cryptocyanin during the molting and non-molting stages and we demonstrate that there are indeed structural differences between the molting and non-molting cryptocyanin proteins of the crustacean snow crab (*Chionoecetes opilio*) hemolymph.

## **2.2 Material and methods**

### **2.2.1 Chemicals and reagents**

Acetonitrile (ACN) and methanol were purchased from Anachemia (Lachine, QC, Canada). Glacial acetic acid and ammonium bicarbonate ( $\text{NH}_4\text{HCO}_3$ ) and were bought from Sigma-Aldrich (Oakville, ON, Canada). Iodoacetamide and Dithiotreitol (DTT) and Bovine Serum Albumin (BSA) were bought from Sigma Aldrich (Saint Louis, MO, USA). All SDS-PAGE chemicals and reagents were purchased from Bio-Rad (Mississauga, ON, Canada). Zip Tips C18 were bought from Millipore (Bedford, MA, USA).

### **2.2.2 Sample preparation**

Juvenile and adult snow crabs were collected from Grand Bank of Newfoundland and kept in running seawater at ambient temperature and salinity at the Department of Fisheries and Oceans Canada (DFO).

### **2.2.3 Extraction of snow crab hemolymph**

Using a sterile syringe, 20-35 ml of hemolymph (blue blood) were withdrawn and collected into sterile propylene centrifuge tube cooled with ice to decrease the rate of blood clotting. The blood was centrifuged immediately at 730 rpm for 10-15 minutes and the supernatant (plasma) is recollected in another sterile tube and stored frozen at -80° C for 24 hours. The frozen sample was lyophilized for three days and kept in 4° C for further use.<sup>[12]</sup>

### **2.2.4 Protein precipitation**

The freeze-dried lyophilized plasma sample was dissolved in distilled water (10ml) and the protein was precipitated using 25 ml of a 50% polyethylene glycol 4000 (PEG 4000) solution, and then centrifuged at 700 rpm for 10 minutes. The supernatant was further treated gradually with 50% PEG 4000 until a final concentration of 16% PEG 4000 was obtained. The protein precipitate was kept at 4° C for further used.<sup>[12]</sup>

### **2.2.5 Electrophoresis**

The protein precipitate was diluted with SDS buffer (50 mg protein sample + 500 µl SDS buffer), and then boiled for 3 minutes. The loading buffer was prepared at pH 6.8 Tris Buffer (65ml glycerol + 6.5g SDS + sodium azide 0.1 g + 3.94g Tris, dissolved in 500 ml distilled H<sub>2</sub>O). Ten microliters (per well) of the diluted sample was run on a 8% polyacrylamide gel at 100 volts for 60 minutes. The gel was stained and fixed using Coomassie Blue stain in 10% glacial acetic acid and followed destaining in 50% methanol solution for 45 minutes.

In order to isolate the intact protein, the protein band corresponding to cryptocyanin was excised from the gel, cut into 1 mm-cubes and then was transferred in

0.5 ml microfuge tubes. The gel pieces were washed with water and a solution of 100  $\mu$ l 50 mM  $\text{NH}_4\text{HCO}_3$ / 50% a ACN 1+1 (v/v) for 15 minutes. Then, all remaining liquid was removed and 100  $\mu$ l excess acetonitrile was added and the gel pieces were shrunk and stuck together. Acetonitrile was removed and the gel pieces were rehydrated by 25  $\mu$ l of 50 mM  $\text{NH}_4\text{HCO}_3$ . The intact protein was extracted with an extraction buffer of 50  $\mu$ l of 50% acetonitrile containing 0.1% trifluoroacetic acid (TFA). The extract was dialyzed in a dialysis tube with cut off of 35 kDa. Single-step desalting, concentration, and purification before MALDI measurements was performed using Zip Tip C18.<sup>[13]</sup>

### **2.2.6 Confirmation of the identity of the cryptocyanin gel band**

Two methods were used to confirm the identity of the cryptocyanin gel band. The first method relied on the detection of the presence or absence of copper metal, measured by atomic absorption spectrometry, on the proteins extracted from the 75 kDa gel band.<sup>[4]</sup> The second method was based on the ablation of the eyestalk during the molting process of three juvenile crabs following collections of hemolymph samples at day 7, 14 and 21.<sup>[4]</sup> Similarly, like the previous steps, the hemolymph proteins were run on the gel with the same percentage of polyacrylamide, which was 8%.<sup>[12]</sup>

### **2.2.7 Tryptic Digestion of Cryptocyanin**

The identification and sequencing of cryptocyanin peptides was achieved by tryptic in-gel digestion of cryptocyanin. The excised gel bands were dehydrated with 50  $\mu$ l acetonitrile and rehydrated with the same volume of 50 mM of  $\text{NH}_4\text{HCO}_3$ . The proteins were then reduced with 50  $\mu$ l of reduction solution (10 mM of dithiotreitol (DTT) in 50 mM  $\text{NH}_4\text{HCO}_3$ ) and the cysteine amino acids were alkylated using 50  $\mu$ l of alkylated solution (55 mM iodoacetamide solution, containing 50 mM of  $\text{NH}_4\text{HCO}_3$ ).

Enzymatic digestion was performed by adding 25  $\mu$ l of a trypsin solution (a 20 ng/ml trypsin in 25 mM solution of  $\text{NH}_4\text{HCO}_3$ ) to the gel pieces, and incubated at 37°C, for 30 minutes, and the excess trypsin solution was removed and replaced with 3  $\mu$ l of a 25 mM solution of  $\text{NH}_4\text{HCO}_3$  to keep the gel wet. The gel pieces were then incubated again at 37°C, overnight. The peptides extraction buffer was composed of 50% of acetonitrile and 0.1% of trifluoroacetic acid (TFA). The extracted peptides were further purified using Zip Tips C18.

### **2.2.8 MALDI-TOF/TOF-MS analysis**

The extracted intact cryptocyanin protein was analyzed using a MALDI-TOF/TOF-MS instrument (Applied Biosystems, 4800 MALDI TOF/TOF Analyzer, International-MDS Sciex, Foster City, CA, USA), with a 200-Hz frequency Nd: YAG laser. The intact protein sample (2mg/ml) was mixed with an equal volume of sinnapic acid (10 mg/ml, in 50% acetonitrile solution) and spotted on the MALDI plate.<sup>[14,15]</sup> The mass calibration of the MALDI-TOF/TOF-MS instrument was performed using a pure Bovine Serum Albumin (BSA) protein solution (2 mg/ml), looking at the following molecular ions:  $[\text{M}+\text{H}]^+$  at  $m/z$  66431.2213,  $[\text{M}+2\text{H}]^{2+}$  at  $m/z$  33216.3776 and  $[\text{M}+3\text{H}]^{3+}$  at  $m/z$  22144.1304. The spotting of the analyte and matrix on the MALDI plate was done as described below.

### **2.2.9 MALDI-QqTOF-MS**

The digested cryptocyanin protein was analyzed by MALDI-QqTOF-MS and low-energy CID-MS/MS on an API-QSTAR XL Quadrupole-orthogonal time-of-flight QTOF-MS/MS hybrid instrument (Applied Biosystems, International-MDS Sciex, Foster City, CA, USA) using a *o*-MALDI source. Briefly, 1  $\mu$ l of a 20 mg/ml solution of  $\alpha$ -

CHCA dissolved in ACN, 0.1% TFA was spotted with 1  $\mu$ l of 2mg/ml of digested sample on the MALDI plate and dried at room temperature. The extracted peptides were further purified using Zip Tips C18. The calibration of the instrument was performed in the range of 600-3200 Da, using standards peptide molecular ions known to be present in the BSA.<sup>[16-26]</sup>

### **2.2.10 MALDI-QqTOF-CID-MS/MS**

The digested peptides were sequenced by MALDI-QqTOF-low-energy CID-MS/MS using the same QqTOF-MS/MS hybrid instrument described above. The sample preparation by in-gel trypsin digestion was carried out as described previously, and a mixture of 50 % ACN and 0.1% TFA was used for the peptides extraction. The extract from the in-gel protein spot digestion was trapped on a Millipore C18 Zip Tip (15-20) sample aspiration. 1  $\mu$ l of the 2mg/ml digested and eluted sample was spotted directly on the MALDI plate with an equal volume of 1  $\mu$ l of a 20 mg/ml solution of  $\alpha$ -CHCA dissolved in ACN, 0.1% TFA, and allowed to dry. The target was inserted into the o-MALDI source and the spectra were acquired with the Analyst QS software operating under control of the o-MALDI server software. The laser was operated at a rate of 20Hz and at a calibrated intensity of 40  $\mu$ J. An initial survey (peptide mass map) scan of singly charged peptide ions was acquired over the course of 1 min with the mass range  $m/z$  600-3200 with Q1 and Q2 set to run at full transmission mode. MS/MS spectra were then acquired with Q1 set to allow the complete isotope spectrum of the selected ions to pass into Q2 where CID-MS/MS was performed in presence of nitrogen. The Analyst QS software (ABI/MDS Sciex) was used to set the lower mass range of the MS/MS to  $m/z$  50 and the upper limit to that of the parent ion. The CID-MS/MS of each precursor ion was

acquired using the Multi-Channel Analysis (MCA) mode generating an average spectrum of different scans monitored at different collision energies. It is important to state that for all the CID-MS/MS sequencing we used the  $[M+H]^+$  monoisotopic  $m/z$  values of the released peptides using MS/MS ions search of the ‘Mascot search engine’.

### **2.2.11 Peptide mass fingerprint and MS/MS ion searches**

Identification of the cryptocyanin protein from the PMF and MS/MS ion searches was carried out using the Mascot search engine on the National Center for Biotechnology Information Non-redundant protein database (NCBIInr). The search parameters used were: one missed cleavage allowed, carbamidomethyl cysteine (Cys CAM) was introduced as a fixed modification, methionine oxidation was set as a variable modification, and the peptide mass tolerance was fixed at  $\pm 0.2$  Daltons (Da). No other constraints were submitted. <sup>[25, 26]</sup>

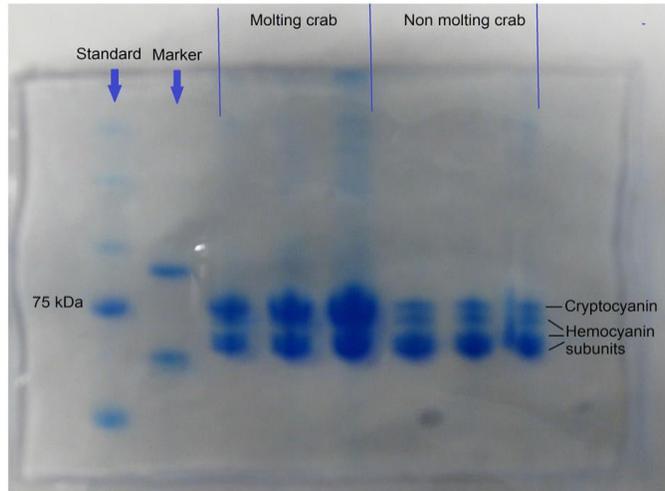
## **2.3 Result and discussion**

### **2.3.1 SDS-PAGE purification of the cryptocynin**

The SDS-PAGE of the purified hemolymph proteins obtained during molting and non-molting showed different protein bands above 75 kDa (**Figure 2.1**). However, we observed that the concentration of the 75kDa protein is higher in the molting crab hemolymph, when compared to the non-molting crab. The elemental analysis of the 75 kDa gel band using the atomic absorption spectrometer indicated the total lack of copper. Therefore, we can propose that this protein corresponds to cryptocyanin.

It is imperative to mention, that we have attempted to improve the SDS-PAGE electrophoresis purification of the cryptocyanin protein by increasing the percentage of

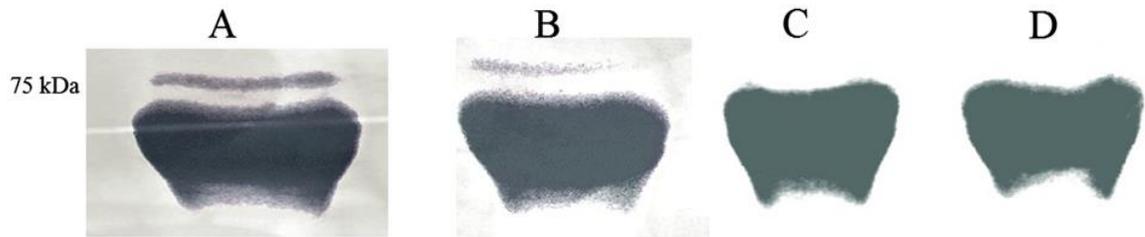
polyacrylamide gel from 8% to 12%. Unfortunately, with the increase in the gel concentration, the separation between the cryptocyanin and hemocyanin bands disappeared and united to form an undistinguishable band.



**Figure 2.1** SDS-PAGE comparison of hemolymph proteins sampled from juvenile crabs during their molting and nonmolting processes.

In order to confirm that the cryptocyanin band did not coalesce with the hemocyanin subunits adjacent heavy band, containing copper, we have experimentally differentiated and distinguished the cryptocyanin and hemocyanin bands. It has been reported in the literature, that ablation of the crab eyestalk affects the crustacean molt inhibiting hormones (MIH) and consequently stops the production of cryptocyanin. It was reported that MIH is synthesized and stored in the X organ/sinus gland complex of the eyestalk of the crab and it has a major effect on the expression and catabolism of cryptocyanin.<sup>[3-11]</sup> For this reason, we carried out the surgical ablation of the eyestalk of juvenile snow crabs, which theoretically stops the cryptocyanin formation during the molting process.<sup>[4]</sup> Therefore, three hemolymph samples were collected, 7 days following the ablation, and analyzed by SDS-PAGE electrophoresis. These analyses showed that the cryptocyanin gel band has disappeared on the samples collected 14 and 21 days after the eyestalk ablation (**Figure 2.2**)<sup>[3-6]</sup>

In the following section, we will trek around the similarities and differences between the molting and non-molting cryptocyanin proteins analyzed by mass spectrometry. We have initially used MALDI-TOF/TOF-MS instrument for the detection of the molecular masses of the intact cryptocyanin proteins and MALDI-QqTOF-MS hybrid instrument for the peptide mass fingerprinting (PMF) protein identification, and finally MALDI-QqTOF-MS/MS for the sequencing of the molting and non-molting cryptocyanin proteins.



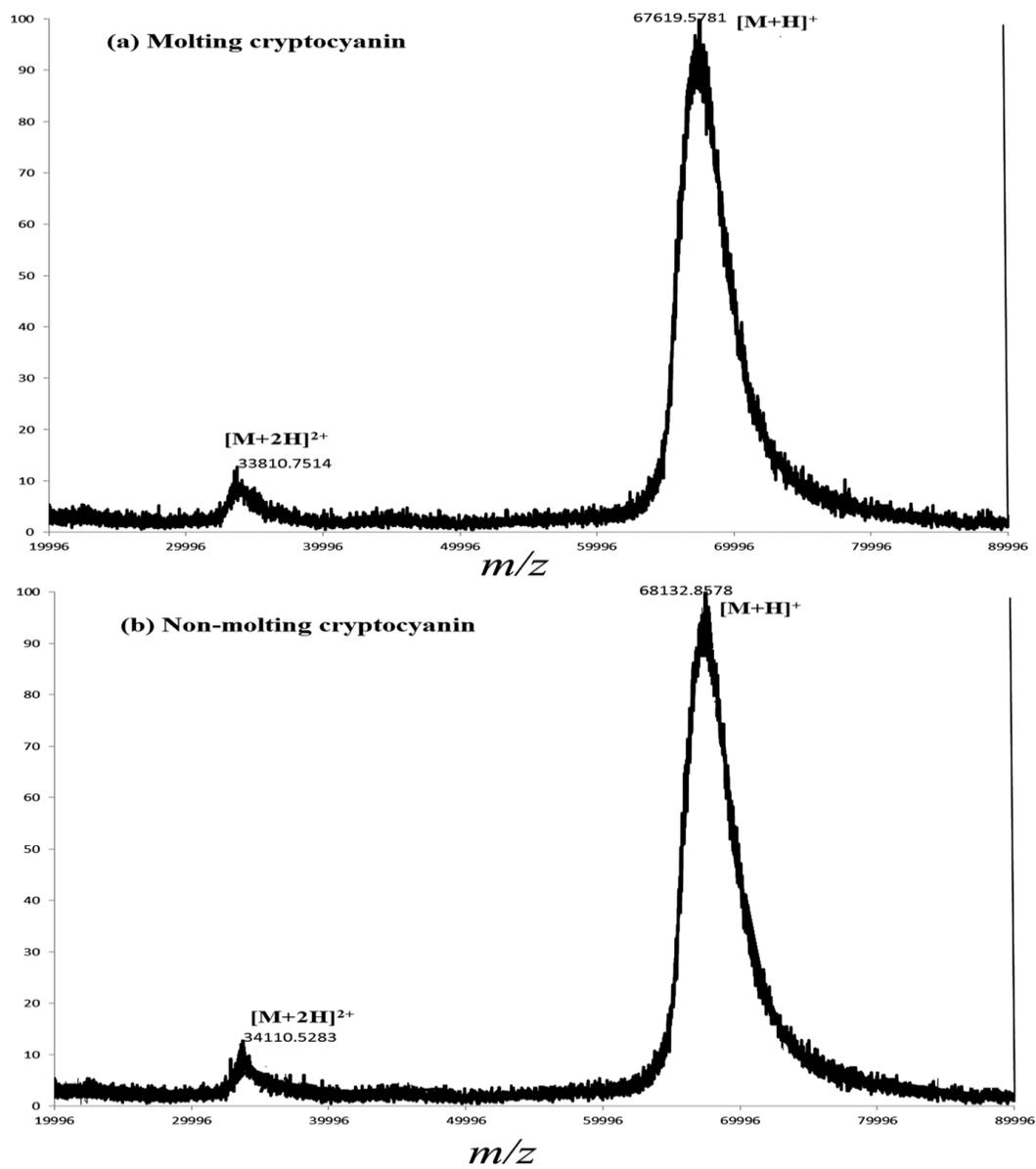
**Figure 2.2** SDS-PAGE separation of molting crab hemolymph proteins stained with Coomassie Blue stain: (A): no eyestalk ablation was performed; (B): after 7 days ablation of eyestalk; (C): after 21 days eyestalk ablation; and (D): after 14 days eyestalk ablation.

### 2.3.2 MALDI-TOF/TOF-MS determination of the molecular mass of the cryptocyanin proteins during the non-molting and molting stages

The 75 kDa SDS-PAGE band of molting cryptocyanin protein was cut and the extracted purified protein was subjected to further analysis. The MALDI-TOF/TOF-MS of the extracted and purified protein showed the presence of the protonated molecule  $[M+H]^+$  at  $m/z$  67619.56 and the bis-protonated molecule  $[M+ 2H]^{2+}$  at  $m/z$  33810.75 (**Figure 2.3a**).

Similarly, the 75kDa gel SDS-PAGE band of the non-molting stage was cut and the extracted purified cryptocyanin protein was subjected to further analysis. The MALDI-TOF/TOF-MS of the extracted and purified protein showed the presence of the protonated molecule  $[M+H]^+$  at  $m/z$  68132.86 and a bis-protonated molecule  $[M+2H]^{2+}$  at  $m/z$  34110.53 (**Figure 2.3b**).

It is noteworthy to mention, that we have noted blatant differences between the proteins molecular masses observed during the SDS-PAGE and those which were measured by MALDI-TOF/TOF-MS. These differences were attributed to the presence of the high number of proline residues (heavy amino acid), which are known to slow protein migrations during SDS-PAGE, thus making it appearing heavier than it actually is.<sup>[27, 3]</sup>

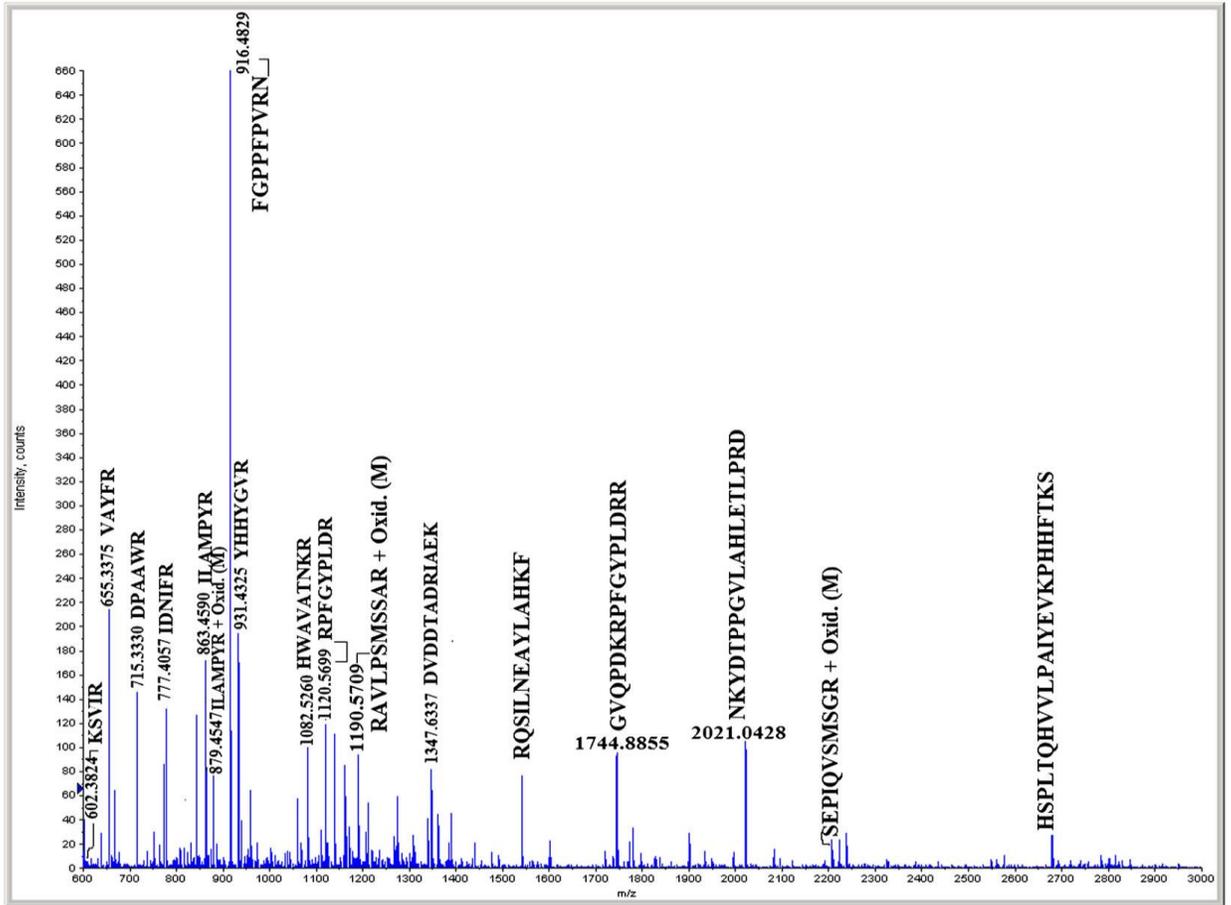


**Figure 2.3** MALDI-TOF-MS of the intact cryptocyanin protein (a) during the molting period and (b) during the non-molting period.

### 2.3.3 Peptide mass fingerprinting of the molting cryptocyanin

The 75 kDa protein band of the SDS PAGE separation of the extracted and purified molting hemolymph proteins was subjected to trypsin digestion for the purpose of PMF identification by MALDI-QqTOF-MS (**Figure 2.4**). Therefore, as a result of using the peptide mass fingerprint database, the sequences of the digested peptides were matched with other types of cryptocyanin in the Mascot library. These were derived from three different species with four accessions, namely: (gi|121484232) for cryptocyanin 1 [*Portunus pelagicus*], (gi|81230850) for cryptocyanin 2 [*Metacarcinus magister*], (gi|24658770) for CG3499 [*Drosophila melanogaster*], and (gi|28557619) for RE25996p [*Drosophila melanogaster*].

The statistically significant scores reported for *Metacarcinus magister* were as follows: a score of 108, an expectation value  $3e^{-05}$  and 28 matching peptides; whereas for the *Portunus pelagicus* had a score of 74, an expectation value 0.075 and 19 matching peptides. Like the *Chionoecetes opilio* both these species (*Portunus pelagicus* and *Metacarcinus magister*) are crustacean crabs and provided us with an additional evidence for assigning the purified band as cryptocyanin. We have presented in this work, only the matching peptides obtained from the digested cryptocyanin protein obtained from the molting stage and these are shown in **Table 2.1**.



**Figure 2.4** MALDI-QqTOF-MS of 'in gel' trypsin digestion of snow crab during the molting process.

**Table 2.1** List of matching peptides obtained from the peptide mass fingerprint (PMF) obtained by MALDI-QqTOF-MS analysis of the digested molting cryptocyanin

Observed ion ( <i>m/z</i> )	Mr (exp)	Calculated Mr	Mass	Sequence* <sup>T</sup>
			Deviation (Da)	
602.3824	601.3751	601.3911	-0.0160	K.KSVIR.I* <sup>2</sup>
655.3375	654.3302	654.3489	-0.0187	R.VAYFR.E <sup>2</sup>
667.2876	666.2803	666.2973	-0.0169	R.YEAER.I <sup>1,2</sup>
715.3330	714.3257	714.3449	-0.0192	R.DPAAWR.L <sup>1,2</sup>
774.4263	773.4190	773.4436	-0.0245	K.IFEPLR.E <sup>1,2</sup>
777.4057	776.3984	776.4181	-0.0197	R.IDNIFR.E <sup>1,2</sup>
863.4590	862.4517	862.4735	-0.0217	R.ILAMPYR.D <sup>2</sup>
879.4547	878.4474	878.4684	-0.0210	R.ILAMPYR.D + Oxid. (M) <sup>2</sup>
916.4829	915.4756	915.4967	-0.0211	K.FGPPFPVR.N <sup>1</sup>
931.4325	930.4252	930.4460	-0.0208	K.YHHYGV.R.G <sup>2</sup>
933.4968	932.4895	932.5192	-0.0297	K.RIDNIFR.E* <sup>1,2</sup>
1037.5076	1036.5003	1036.4938	0.0066	K.GNEKGTEFR.L* <sup>2</sup>
1082.5260	1081.5188	1081.5781	-0.0593	K.HWAVATNKR.H* <sup>1,2</sup>
1120.5699	1119.5626	1119.5825	-0.0199	K.RPFGYPLDR.R <sup>2</sup>
1139.4826	1138.4753	1138.5043	-0.0290	K.DTYGYHIDR.K <sup>1,2</sup>
1170.6035	1169.5962	1169.5717	0.0245	K.QGFTPQTTYK.F <sup>2</sup>
1171.6136	1170.8543	1170.5560	0.0065	K.EGFTPQTTYK.F <sup>1,2</sup>
1179.5418	1178.5345	1178.6084	-0.0739	R.LNHKEFTYK.I* <sup>2</sup>
1219.5817	1218.5744	1218.5450	0.0293	R.HLMDEFTNGR.L <sup>1</sup>
1235.5787	1234.5401	1234.5400	0.0314	R.HLMDEFTNGR.L + Oxid. (M) <sup>1</sup>
1267.5732	1266.5659	1266.5993	-0.0334	K.DTYGYHIDRK.G* <sup>1,2</sup>
1276.6553	1275.6480	1275.6836	-0.0356	K.RPFGYPLDRR.V* <sup>2</sup>
1348.6249	1347.6167	1347.6170	0.0009	R.NDDIHLHDVDR.I <sup>1</sup>
1389.7590	1388.7517	1388.7663	-0.0146	R.ISNHLPPLEELK.L <sup>2</sup>
1543.7277	1542.7204	1542.7390	-0.0184	K.AVSFNPLGDVSMYK.D + Oxid. (M) <sup>1</sup>
1744.8855	1743.8782	1743.9057	-0.0274	R.GVQPKRPFYPLDR.R* <sup>2</sup>
1771.8616	1770.8543	1770.8487	0.0065	K.GENFFYAYHQLLNR.Y <sup>1,2</sup>
1778.9014	1777.8941	1777.9363	-0.0422	K.YDTPPGVLAHLETLP.R.D <sup>1,2</sup>
1779.0025	1777.9952	1777.9363	0.0589	K.YDTPPGVLAHLETLP.R.D <sup>1,2</sup>
1899.9073	1898.9000	1898.9427	-0.0427	R.KGENFFYAYHQLLNR.Y* <sup>1,2</sup>
2021.0428	2020.0355	2020.0742	-0.0386	K.NKYDTPPGVLAHLETLP.R.D* <sup>1,2</sup>
2206.0416	2205.0343	2205.1542	-0.1199	K.ALTFNPVGDVSIYRDGGVAVR.Q* <sup>2</sup>
2239.0934	2238.0861	2238.1182	-0.0320	K.FGAPFPIRNDIQLHDVER.V* <sup>2</sup>
2678.3880	2677.3807	2677.4493	-0.0686	K.HSPLTQHVVLPAIYEVKPHHFTK.S <sup>2</sup>

\* The sequence peptide column marked subscripts <sup>1</sup> and <sup>2</sup> respectively. (\*) indicates a sequence with one missed cleavage.

#### **2.3.4 MALDI-CID-MS/MS sequencing of the digested peptides obtained from the molting cryptocyanin protein**

It is worthy to mention that for exploring the MS/MS sequencing data of the digested peptides, we conducted all the CID-MS/MS analyses on the monoisotopic precursor ion  $[M+H]^+$  of each peptide. The resulting product ion spectra of the digested peptides obtained from the molting cryptocyanin were obtained and submitted to Mascot database for MS/MS ion search.<sup>[28, 29]</sup> Therefore, we have found that the CID-MS/MS peptide sequences matched the same cryptocyanin proteins of the two species with three accessions identified previously. These were the following: (gi|121484232) accession for cryptocyanin 1 [*Portunus pelagicus*], (gi|81230850) accession for cryptocyanin 2 [*Metacarcinus magister*] and (gi|4191390) accession for cryptocyanin [*Metacarcinus magister*] as shown in **Table 2.2**

Each of the identified peptide which was resulted from the CID-MS/MS analysis had a separate statistically significant score and all the matched peptides of cryptocyanin are indicated in **Table 2.2**. The CID-fragmentations of the digested peptides obtained from the molting cryptocyanin peptides followed, without any ambiguity, the fragmentation routes proposed by Roepstorff and Fohlman.<sup>[28]</sup>

**Table 2.2** CID-MS/MS sequences of the peptides obtained by the digested cryptocyanin molting protein.

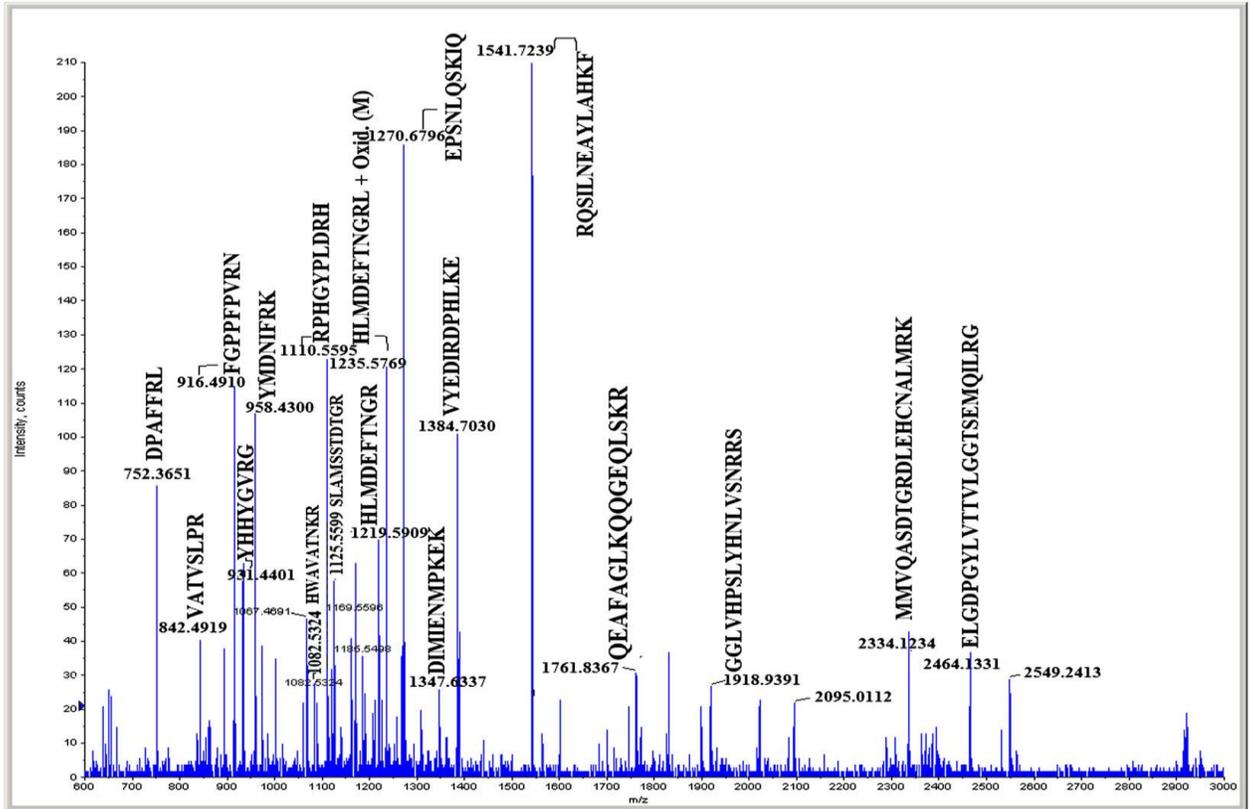
Observed ion ( <i>m/z</i> )	Sequence	Identified Cryptocyanin peptides in the Mascot library		
		gi 121484232	gi 81230850	gi 4191390
602.3824	K.KSVIR.I (U)	N	Y	N
655.3375	R.VAYFR.E	N	Y	N
667.2876	R.YEAER.I	Y	Y	Y
715.333	R.DPAAWR.L	Y	Y	Y
774.4263	K.IFEPLR.E	Y	Y	Y
777.4057	R.IDNIFR.E	Y	Y	Y
863.459	R.ILAMPYR.D (U)	N	Y	N
879.4547	R.ILAMPYR.D+ Oxid. M (U)	N	Y	N
916.4829	K.FGPPFPVR.N (U)	Y	N	Y
931.4325	K.YHHYGV.R	N	Y	N
933.4968	K.RIDNIFR.E*	Y	Y	Y
1082.526	K.HWAVATNKR.H* (U)	Y	Y	Y
1120.5699	K.RPFGYPLDR.R*	N	Y	N
1139.4826	K.DTYGYHIDR.K	Y	Y	Y
1267.5732	K.DTYGYHIDRK.G	Y	Y	Y
1360.6744	R.NDDIHLHDVDR.I (U)	Y	N	N
1744.8855	R.GVQPKRPFYPLDR.R** (U)	N	Y	N
1771.8616	K.GENFFYAYHQLLR.Y	Y	Y	Y
1899.9073	R.KGENFFYAYHQLLR.Y	Y	Y	Y
2021.0428	K.NKYDTPPGVLAHLETLPR.D*	Y	Y	Y
2678.388	K.HSPLTQHVVLPALIEVKKPHHFTK.S** (U)	N	Y	N

(\*) and (\*\*) indicate a sequence with one and two missed cleavages respectively (U) indicates that the identified sequence of the peptide was unique.

### 2.3.5. Peptide mass fingerprinting of the non-molting cryptocyanin

Similarly, the SDS-PAGE 75kDa protein band of the extracted and purified non-molting hemolymph proteins was subjected to trypsin digestion for the purpose of PMF identification by MALDI-QqTOF-MS (**Figure 2.5**).

It is very interesting to note that the PMF analysis of the obtained digested peptides matched four different proteins. Of these four proteins, two accessions were similar to the molting stage, namely: the (gi|121484232) for cryptocyanin 1 [*Portunus pelagicus*] and (gi|81230850) for cryptocyanin 2 [*Metacarcinus magister*]. In addition, two other species-accessions which were not related to the cryptocyanin of two different proteins were identified as: the LP04011p (gi|21429122) protein of the *Drosophila melanogaster* specie and the WD repeat domain 61 (gi|119619579), a CRA protein isoform from the *Homo sapiens* specie. The statistical significant score of *Drosophila melanogaster* was of 68, an expectation value 0.3 and 30 matching peptides, with the WD repeat domain had a score of 79, an expectation value 0.025 and 11 matching peptides (**Table 2.3**).



**Figure 2.5** MALDI-QqTOF-MS of 'in gel' trypsin digestion of snow crab during the non-molting process

**Table 2.3** List of matching peptides obtained from the peptide mass fingerprint (PMF) after MALDI-QqTOF-MS analysis of the digests of the gel band at 75 kDa of non-molting stage

Observed ion ( <i>m/z</i> )	Mr (exp)	Calculated Mr	Mass	Sequence
			deviation (Da)	
863.4706	862.4633	862.4548	0.0085	R.QEAFAGLK.Q <sup>1</sup>
895.4852	894.4779	894.4447	0.0333	R.SLYEIDR.W <sup>1</sup>
916.4910	915.4837	915.4563	0.0275	K.QVHVYDR.V <sup>1</sup>
917.4968	916.4895	916.4614	0.0281	K.QQGEQLSK.R <sup>1</sup>
932.4477	931.4404	931.4399	0.0005	K.KHDEFEK.L* <sup>1</sup>
959.4364	958.4291	958.5083	-0.0792	K.DADLIREK.L* <sup>1</sup>
1067.4691	1066.4618	1066.5229	-0.0611	K.LASHSCPAPK.E <sup>2</sup>
1111.5692	1110.5619	1110.4798	0.0821	K.TSIGLFAKFK.S* <sup>1,2</sup>
1125.5599	1124.5526	1124.5132	0.0394	K.SLAMSTDTGR.D <sup>1</sup>
1126.5632	1125.5559	1125.5415	0.0145	R.DEVSVQGHQK.K <sup>1</sup>
1127.5658	1126.5585	1126.4747	0.0838	-.MMVQASDTGR.D + 2Oxid.(M) <sup>1</sup>
1141.4890	1140.4817	1140.5081	-0.0264	K.SLAMSTDTGR.D + Oxid.(M) <sup>1</sup>
1170.5757	1169.5684	1169.7132	-0.1448	K.VAKLATGLIER.N* <sup>1</sup>
1171.5936	1170.5863	1170.6972	-0.1109	K.DLAAVSNLLKK.H* <sup>1</sup>
1220.5833	1219.5760	1219.6197	-0.0437	K.LQVALDENYR.E <sup>1</sup>
1185.5498	1184.5425	1184.6700	-0.1274	-.MTNQVRPVLK.G <sup>2</sup>
1226.6118	1225.6045	1225.7142	-0.1097	R.DLAAVEALIRR.E* <sup>1,2</sup>
1236.5814	1235.5741	1235.6258	-0.0517	K.RPVLSSSDYGR.D <sup>1</sup>
1258.5769	1258.5696	1257.5594	0.0102	R.DLEHCNALMR.K <sup>1</sup>
1271.6861	1270.6788	1270.6881	-0.0093	R.EPSNLQSKIQK.H* <sup>1</sup>
1274.6666	1273.6593	1273.5543	0.1050	R.DLEHCNALMR.K + Oxid.(M) <sup>1</sup>
1347.6337	1346.6264	1346.6314	-0.0050	R.DVDDTADRIA EK.S* <sup>1</sup>
1386.7182	1385.7109	1385.7303	-0.0193	R.QSILNEAYLAHK.F <sup>1,2</sup>
1389.7643	1388.7570	1388.7412	0.0158	K.LLHADHVDTLQK.F <sup>1</sup>
1542.7315	1541.7242	1541.8314	-0.1072	K.RQSILNEAYLAHK.F* <sup>1</sup>
1745.8901	1744.8828	1744.8744	0.0084	R.LEADVAAHGELADQLK.Q <sup>1</sup>
1761.8367	1760.8294	1760.9057	-0.0763	R.QEAFAGLKQQGEQLSK.R* <sup>1</sup>
1762.8220	1761.8147	1761.9274	-0.1127	R.GGLVHPSLYHNLVSNR.R <sup>2</sup>
1770.8965	1769.8892	1769.9312	-0.0419	K.QKIDQHETAAEFLIK.K* <sup>1</sup>
1918.9391	1917.9318	1918.0285	-0.0967	R.GGLVHPSLYHNLVSNRR.S* <sup>2</sup>
2334.9224	2333.9151	2334.0330	-0.1186	-.MMVQASDTGRDLEHCNALMR.K* <sup>1</sup>
2335.1442	2334.1369	2334.0337	0.1032	-.MMVQASDTGRDLEHCNALMR.K* <sup>1</sup>
2465.1457	2464.1384	2464.2520	-0.1135	K.ELGDPGYLVTTVLGGTSEMQLR.G+ Oxid.(M) <sup>2</sup>

Sequence peptide column with subscripts 1 and 2 respectively. (\*) indicates a sequence with one missed cleavage.

### **2.3.6. MALDI-CID-MS/MS sequencing of the digested peptides obtained from the non-molting cryptocyanin protein**

The product ion spectra of the digested peptides were submitted to the Mascot database for MS/MS ion search. Once more, it is important to state that for the CID-MS/MS sequencing we used the  $[M+H]^+$  monoisotopic  $m/z$  values of the released peptides for the MS/MS ions search of the 'Mascot search engine'.

The obtained MS/MS data matched the respective hemocyanin subunit proteins of three different crustaceans: the hemocyanin subunit 4 (gi|57901145) of the *Metacarcinus magister*, the hemocyanin subunit 3 (gi|57901143) of the *Metacarcinus magister* and hemocyanin alpha-subunit (gi|7105883) of the *Homarus americanus* (**Table 2.4**).

As previously mentioned, we have indicated that the crab cryptocyanin protein was created from the duplication of the hemocyanin gene. Therefore, it was no surprise that we identified non-molting cryptocyanin peptides that matched a number of the peptides of the six hemocyanin subunits. These matching non-molting peptides are indicated in **Table 2.4**.

It is worth mentioning that for the MS/MS of the protonated molecules extracted from the digested molting peptides, described in section 4, none of the sequences of these peptides could be matched to the peptides originating from the six hemocyanin subunits. This means that these molting peptides were modified during the molting process to adapt to their new biological functions.

**Table 2.4** CID-MS/MS sequences of the peptides obtained the digested cryptocyanin non-molting protein.

<b>Observed ion</b>	<b>Sequence</b>	<b>Identified Cryptocyanin peptides in the Mascot library</b>		
		<b>gi 57901145</b>	<b>gi 57901143</b>	<b>gi 7105883</b>
637.2000	R.FDAER.L*	Y	Y	Y
752.3000	R.DPAFFR.L*	Y	Y	N
958.4000	K.YMDNIFR.E*	Y	Y	N
1110.5000	K.RPHGYPLDR.H*(U)	N	Y	N
1166.5000	R.LNHEEFSYK.I*(U)	N	N	Y
1384.7000	K.VYEDIRDPHLK.E*(U)	Y	N	N

The matching (\*) indicate a sequence with one missed cleavages (U) indicates that the identified sequence of the peptide was unique

### 2.3.7 Structural similarities and differences between the identified peptides of the digested cryptocyanin protein in molting and non-molting snow crab

We compared the CID-MS/MS measurements of the digested molting and non-molting cryptocyanins to show the similarities and differences between the protonated peptides. For that reason, once more the product ion scans of identical series of monoisotopic  $m/z$  value, for both the digested molting and non-molting cryptocyanin peptides were recorded. Accordingly, we measured a series of 15 product ion scans from the individual extracted protonated molecule of these peptides, which were respectively assigned with the Mascot library. The structures of these peptides were determined as follows: VAYFRE at  $m/z$  655.3381, VATVSLPR at  $m/z$  842.4919, FGPPFPVRN at  $m/z$  916.4829, QQGEQLSKR at  $m/z$  917.4968, YHHYGVRG at  $m/z$  931.4401, KHDEFEKL at  $m/z$  932.4477, RIDNIFRE  $m/z$  933.4968, VLHPVTNRK  $m/z$  934.5133, HWAVATNKRH at  $m/z$  1082.5324, RPFGYPLDRR at  $m/z$  1120.5604, SFVEANPELR  $m/z$  1161.5891, RAVLPSMSSAR at  $m/z$  1190.5574 and DVDDTADRIA EKS  $m/z$  1347.6337 AQLLAMMDTMMGGR at  $m/z$  1541.7257, and NKYDTPPGVLAHLETLPR.D at  $m/z$  2021.0428.

Actually, seven of these similar  $[M+H]^+$  ions of the peptide sequences were assigned with matched sequences of cryptocyanin of different species (*Portunus pelagicus* and *Metacarcinus magister*); VAYFRE at  $m/z$  655.3381, FGPPFPVRN at  $m/z$  916.4829, YHHYGVR at  $m/z$  931.4401, RIDNIFRE at  $m/z$  933.4968, HWAVATNKRH at  $m/z$  1082.5324, RPFGYPLDRR at  $m/z$  1120.5604, and NKYDTPPGVLAHLETLPR.D at  $m/z$  2021.0428 (full sequence is shown in **Fig. 2.6**). On the other hand, the remaining eight similar  $[M+H]^+$  ions of the peptide sequences were assigned with other matched

sequences of different proteins of different species (*Drosophila melanogaster* and *Homo sapiens*); VATVSLPR at  $m/z$  842.4919, QQGEQLSKR at  $m/z$  917.4968, KHDEFEKL at  $m/z$  932.4477, VLHPVTNRK at  $m/z$  934.5133, SFVEANPELR at  $m/z$  1161.5891, RAVLPSMSSAR at  $m/z$  1190.5574, DVDDTADRIAES at  $m/z$  1347.6337, and AQLLAMMDTMMGGR at  $m/z$  1541.7257.

1 MKVLVVFALV ALSAAAAQAN EPDGIPTYQK QHDVNNALYK **IFEPLREDNL**  
 51 RDK**ALTFNPV** **GDVSIYRDGG** **VAVRQLMYEL** THDRLLEK**KH** **WAVATNKRHL**  
 101 EEAIMLFEV FMQCKDWNCAA SNAAYFRERV NEEEFYAA**Y** **HAVKHSPLTQ**  
 151 **HVVLP**AIYEV **KPHHF**TKSQV INEAYEAKEM KLRNIYFQNN FTGTPNDIEH  
 201 **RVAYF**REDIG VGTHHLMINL ENPCWW**KD**TY **GYHIDR**KGEN **FFYAYHQLLN**  
 251 **RYEAER**ISNH LPPLEELKLD EPL**KQG**FT**PQ** TTYK.FGAPF IRNDDIQLHD  
 301 **VERV**GRIHEI VHLEDRIHDA IA**HGY**VEDEQ GNKINIENDQ GIDIIGDIIQ  
 351 S**SLY**SPNRKY YGNLTTLAYN MLDH**Q**TDPKN **KYD**TPPGVLA **HLET**LPDP  
 401 **AWRL**HKRIDN **I**FREHIDSLP PYTKEQLEFT GISVENVQIQ GNLETYFEEY  
 451 KYDLINGFND NTTQTEFYGI YATMPRL**NH**K **E**FTYKINVQN NNGTPK**S**VI  
 501 RILAMPYRDG NGAIISFDEG RWLAIEMDLF VKTLTEGKNE IIRKSSEASI  
 551 TVPDVPTYKT L**VEM**TEDHQS LERYESATGI PN**RLL**LPKGN **EKG**TEFRLV  
 601 AATNAEEDIN DESIITM**N**KY **H**HYGVR**G**VQ**P** **D**KRPF**G**YPLD **R**RVFDEHIVN  
 651 EVSSIKETMV KVN**H**DVLR LPHY

(a) Full sequence of cryptocyanin 2 [Metacarcinus magister]

1 DEPDGVSKYQ KQHDVYYAFY **KIFEPL**RESN LVEKAATFNP VGDVSMYKDG  
 51 GVAVQNLMYE LTHQR**LLE**KK **HWAVAT**NKRH LEEAIMLFEV FMQCKDWN**C**V  
 101 ASNAAYFRER VNEEFY**Y**AA YHAIKHSPLT QHV**VLP**AIYE VKPRH**F**TKTQ  
 151 IAEEAYEAKE MKLRNVFFQN NFTGTPNDIE HR**VAYF**REDI GVGTHHLM**I**H  
 201 LENPF**W**KDT **YGYHIDR**KGE **NFFYAYHQLL** **NR**YEAE**RIS**N HLP**P**LEELKL  
 251 DKPL**KQ**GFAP QTTYN**F**GGPF ATRNDNVHLQ DVDNVGRIHE IVH**MED**RIHD  
 301 AIAHGYVEDE QGNKINIENE QGIDILGDII QSSLYSPNRK YYGNLTTLAY  
 351 TMLDHQ**T**DPK **NKYD**TPPGVL **AHLET**LPDP **AAWRL**HKRID **NIF**REHIDSL  
 401 PPY**TKE**QLEF TGITVSDVQI QGSLETYFEE YKYDLINGLN DNTTETEFY**G**  
 451 IYATMPRLNH KEFSYKIHVQ N**NG**ASKKAV IRLLAMPYRD **NGAI**IP**FDE**  
 501 **GR**WLAIEMDL FVKTLTSGDN EIDRKSSEAS ITVPDVPTYK TLVEKTENRE  
 551 NLEMYESSTG IPN**RLL**LPKG CEEGTEFRLP VAVTDAENDV NDESIITM**Q**K  
 601 **YH**HYGVR**G**VQ PDNRPF**G**YPL DRRVPDEHIV DEVSNIKATM VKVYNHNVFI  
 651 PLS

(b) Full sequence of cryptocyanin [Metacarcinus magister]

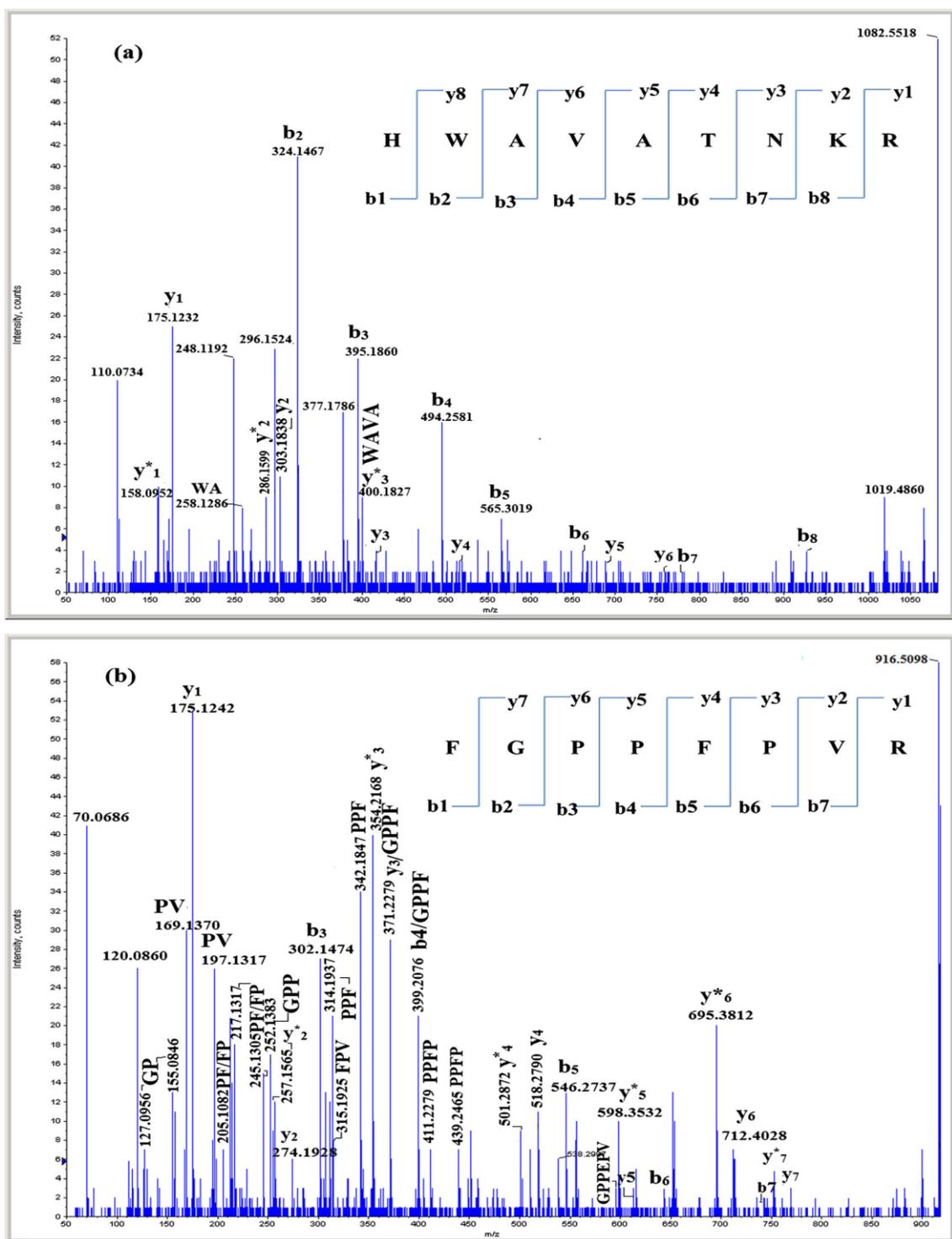
1 MKLLVILALV ALTGASAWPN FMSDEPDGVP AHQ**Q**HDVNY AFY**KIFEPL**R  
 51 DNNLAE**KAVS** **FNPLGDVSMY** **KDGGVAVRHL** **MDEF**TNGRLL **EKKHWAVAT**N  
 101 **KR**HLEEAIML FEVFMQCRDW NCVASNGAYF RERVNEEFY YAA**Y**HAIKHS  
 151 PLTQHVVLP**A** MYEVKPHHFT KTQVIEEAYE AKEMKLRNII FQNNFTGTPN  
 201 DIEQRVAYYR EDVGVGTHHL MIHLENPFWW **KD**TY**GYHIDR** **KGEN**FFYAYH  
 251 QLLNRYEAER ISNYLPPLQE LKLDEPL**KEG** FTPQTTYK.FG PPFVRND**DI**  
 301 **HLHDVDR**IGR IHEVVHMEDR IHDAIAHGYV EDEQGNKINI DNDHGIDILG  
 351 DIIQSSMYSP NRKYGNLTT LAYTLLDHQ**T** DPK**NKYD**TP**P** **GVL**AHLETLP  
 401 **RDPA**AWRLHK **RIDN**IFREHI DTLPPYTKE

(c) Full sequence of cryptocyanin 1 [Portunus pelagicus]

**Figure 2.6** The matched peptides locations of cryptocyanin in different species. The matched peptides during molting process were coloured in red. The green matched peptides were sharing in both molting and non-molting stages. The underlined sequence indicates to isobaric peptides

We have chosen to illustrate in this discussion the MS/MS data of the peptide HAWAVATNKR at  $m/z$  1082.5324 from the aforementioned series, as a show case common peptide example for the digested molting and non-molting cryptocyanin protein. Hence the CID-MS/MS spectrum of the protonated molecule at  $m/z$  1082.5324 is shown in **Figure 2.7a** and the identified product ions are reported in **Table A.1** (supplementary information).

In addition, we also recognized from this series, the presence of numerous isobaric peptides of the cryptocyanin protein from the molting and non-molting stages (**Table 2.1 and 2.3**). As an example, we are only showing one protonated peptide molecule at  $m/z$  916.4829, which could be assigned to two different peptide structures, either FGPPFPVRN and/or QVHVYDRV. It is interesting to note that these two isobaric peptides, matched respectively two different proteins of different species, namely: the cryptocyanin protein of *Portunus pelagicus* and the 'LP04011p protein of *Drosophila melanogaster*. Hence, in this manuscript, we have assigned the peptide 'FGPPFPVRN' at  $m/z$  916.4829 to the cryptocyanin of *Portunus pelagicus*, as it is the closest specie to *Chionoecetes Opilio*. Accordingly, **Figure 2.7b** illustrates the MS/MS of the FGPPFPVR peptide at  $m/z$  916.4829 and the identified product ions are listed in **Table A.2** (supplementary information).



**Figure 2.7** MALDI-CID-MS/MS of the  $[M+H]^+$  ions obtained from the tryptic peptides of the molting and non-molting cryptocyanins: (a) HAWAVATNKR at  $m/z$  1082.5324 and (b) FGPPFPVR at  $m/z$  916.5099.

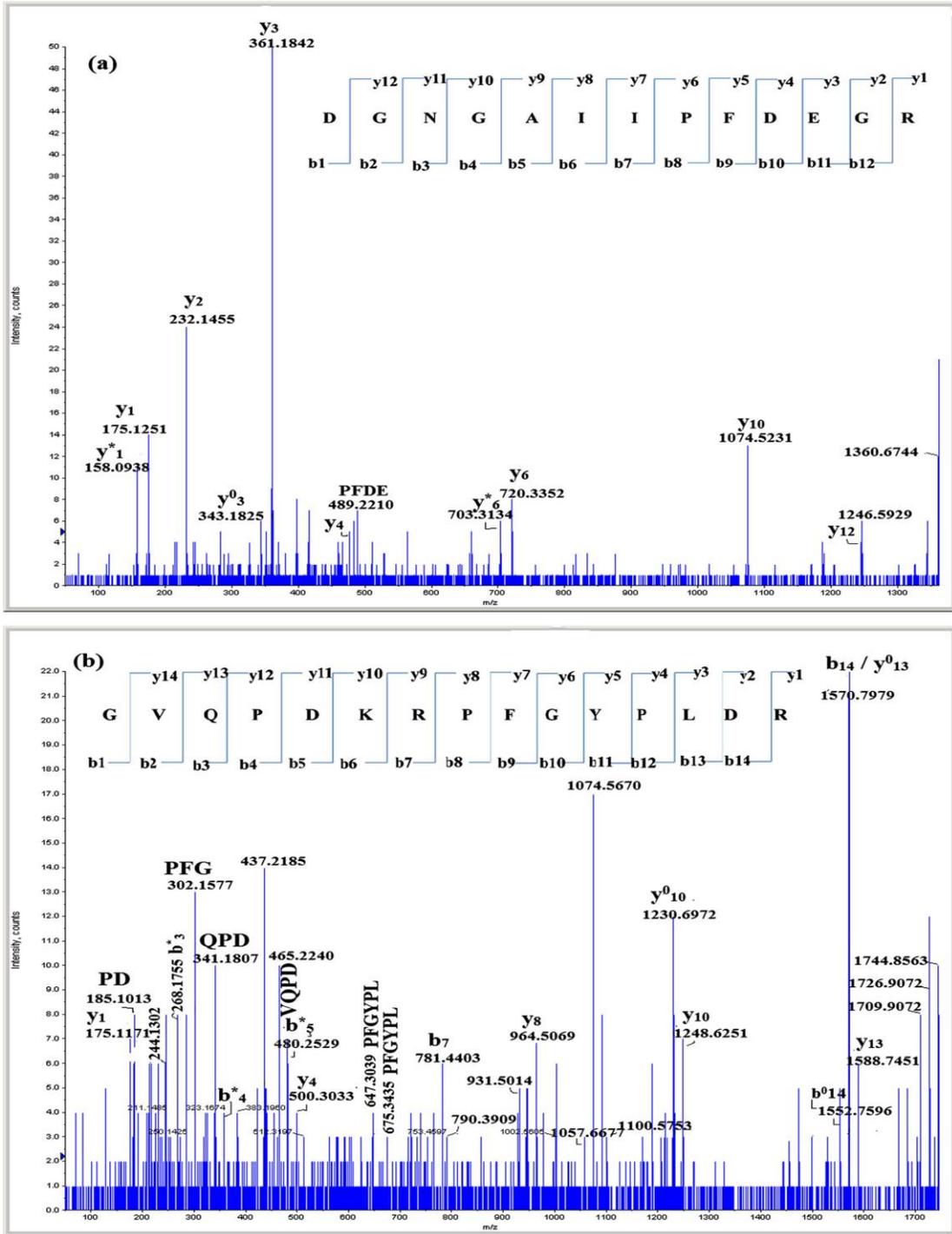
In addition, we concluded that without any doubts the MS/MS data correlate the peptide sequences identified during the PMF analysis. In the several paragraphs discussed above, we have indicated the structural similarities between the digested molting and non-molting cryptocyanin peptides.

Unquestionably, it is fundamental to distinguish the structural differences between the protonated peptides of the molting and non-molting cryptocyanin peptides. In the following section we will discuss their structural differences.

As described earlier, we have anticipated that such differences and/or modifications between the peptides of the molting and non-molting digested cryptocyanins must be inherent of a biosynthetic alteration, which occurs in order to prepare the cryptocyanin to adapt to its new biological function. This biosynthetic alteration of the cryptocyanin is attributed to its new role as a key player in forming the new exoskeleton during the molting process. For these reasons, it is understandable that some of the characterized molting cryptocyanin peptides will result from modification of the original peptides present in the non-molting cryptocyanin protein. On the other hand, it is important to remember that during the non-molting process, the cryptocyanin protein maintains its normal physiological functions such as cardiovascular function and ion regulation.<sup>[3]</sup>

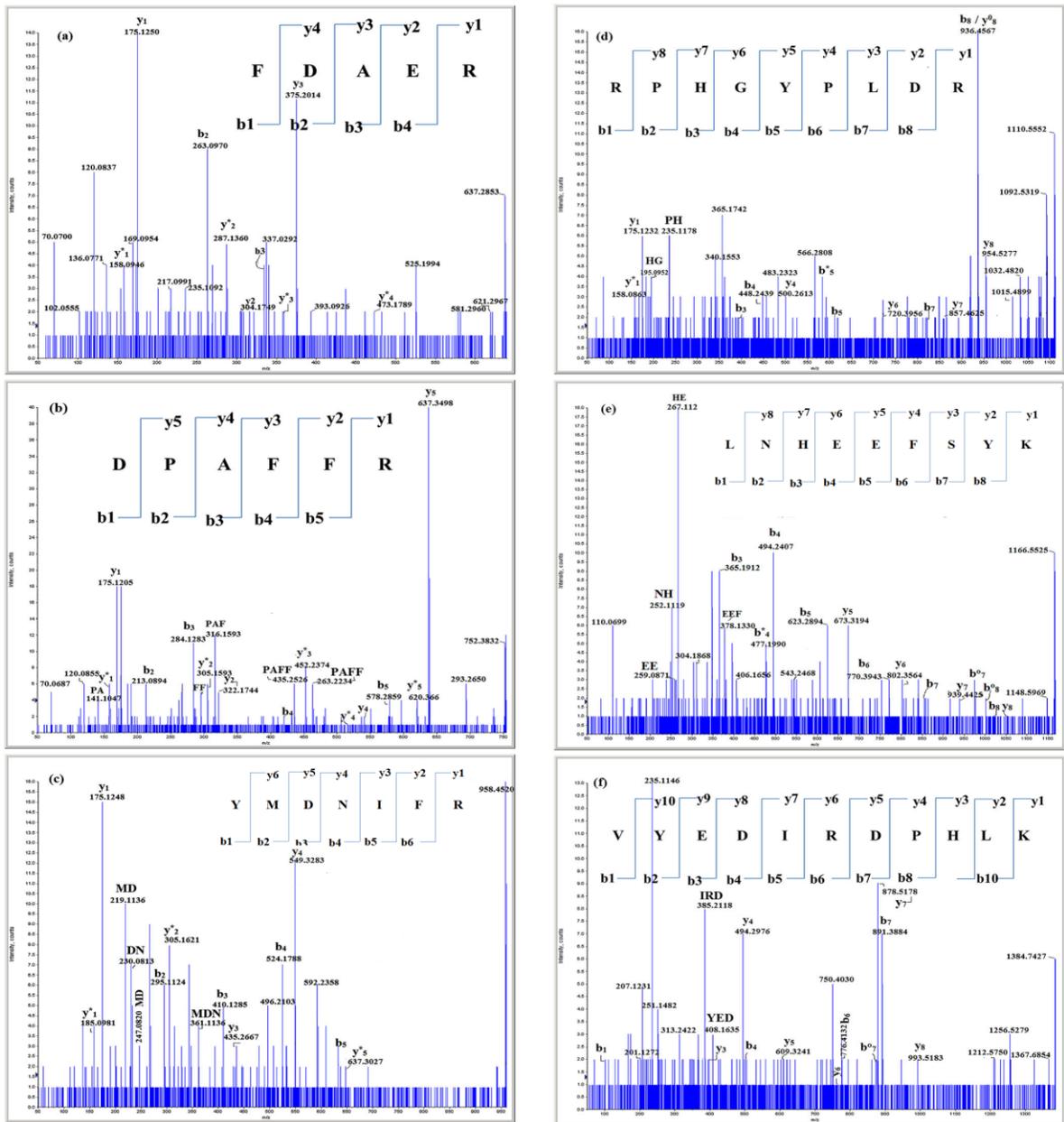
Concurrently, for the characterization of the obtained unique molting peptides we have chosen two discriminating modified peptides for CID-MS/MS sequencing which were considered as unique peptides for the molting cryptocyanin. Consequently, both the product ion scans of the  $[M+H]^+$  precursor protonated peptides DGNGAIIPFDEGR at  $m/z$  1360.6744 and GVQPKRPFPGYPLDRR at  $m/z$  1744.8554 are shown in **Figures 2.8a**

**and 2.8b.** The identified respective product ions are presented in **Tables A.3 and A.4** (supplementary information). These series of product ions are indeed diagnostic for the molting stage cryptocyanin protein.



**Figure 3.8** MALDI-CID-MS/MS of the  $[M+H]^+$  ions of both the exclusive tryptic peptides of molting cryptocyanin: (a) DGNGAIIPFDEGR at  $m/z$  1360.6744 and (b) GVQPKRPFYPLDRR at  $m/z$  1744.8554.

In a similar manner, we conducted six additional CID-MSMS analyses of selected precursor protonated peptide ions, diagnostic of the non-molting stage cryptocyanin protein. As already discussed, the structures of these six chosen peptides are matched to the structure of the six hemocyanin subunits. Henceforth, the product ion scan of the following peptides FDAER at  $m/z$  637.2853, DPAFFR at  $m/z$  752.3832, YMDNIFR at  $m/z$  958.4520, RPHGYPLDR at  $m/z$  1110. 5552, LNHEEFSYK at  $m/z$  1166. 5525 and VYEDIRDPHLK at  $m/z$  1384.7427 are shown in **Figures 2.9a, 2.9b, 2.9c, 2.9d, 2.9e and 2.9f** and the identified product ions are displayed in **Tables A.5, A.6, A.7, A.8, A.9 and A.10** (supplementary information), respectively.



**Figure 2.9** MALDI-CID-MS/MS of the  $[M + H]^+$  ions of the tryptic changed and/or modified peptides obtained from nonmolting cryptocyanin (a) FDAER at  $m/z$  637.2000, (b) DPAFFR at  $m/z$  752.3000, (c) YMDNIFR at  $m/z$  958.4000, (d) RPHGYPLDR at  $m/z$  1110.5000, (e) LNHEEFSYK at  $m/z$  1166.5000, and (f) VYEDIRDPHLK at  $m/z$  1384.7000

It is important to note that from the six selected non-molting precursor protonated peptide ions described previously, we were capable of choosing three characteristic and unique peptide ions. These were assigned as follows: *\_RPHGYPLDR* at  $m/z$  1110.5000, *LNHEEFSYK* at  $m/z$  1166.5000 and *VYEDIRDPHLK* at  $m/z$  1384.7000. In that instance, it is important to mention that the selection criterion for the characterized unique peptide ions was based on very simple criteria. Hence, initially, the chosen peptides have to be free of post-translational modifications. In addition, the choice of these unique peptide ions must respect the different following theoretical, empirical and technical caveats. From the theoretical point of view, the peptide sequences should be free of methionine and cysteine to avoid amino acid modifications, which may occur during the digestion of the protein by either the oxidation of the methionine (Met) and/or carbamidomethylation of the Cys residue. As well, the empirical and technical properties are based on the signal intensity of the peptide ions obtained during the various mass spectrometric analyses.<sup>[31]</sup> In view of that, the strongest peptide ion signals are preferred over the others because of their high sensitivity and their lower limit of detection. Furthermore, the selection of the precursor peptide ions is also based on their high susceptibility to fragment under CID conditions, and to produce abundant high molecular mass product ions.<sup>[31]</sup>

## 2.4 Conclusion

In this report, we have demonstrated for the first time, using MALDI TOF/TOF-MS, that the actual molecular mass of the cryptocyanin protein in the snow crab (*Chionoecetes opilio*) during molting and non-molting phases. The peptide mass fingerprinting (PMF) was conducted using an MALDI-QqTOF-MS hybrid instrument.

The MALDI-CID-MS/MS analyses allowed the sequencing of the cryptocyanin protein after tryptic digestion, during the molting and non-molting stages. These CID-MS/MS analyses showed some similarities and staggering differences between the identified cryptocyanin peptides of the two different stages.

In quest for developing a “signature peptides” method for monitoring the levels of the cryptocyanin protein during the molting and non-molting stages in the hemolymph, we compared the identified peptides of this protein in the two different stages.

We have shown that six peptides of the non-molting cryptocyanin matched the known hemocyanin peptides. These were selected as diagnostic peptides to elucidate the ability of snow crab to change its genetic code for adaptability to new biological functions during molting.

### **Appendix A. Supplementary data**

*Supplementary are data associated with this chapter (P.165)*

## References

- 1- D.R.J. Mallowney, E.M. Hynick, E.G. Dawe, W.A. Coffey. In distribution and habitat of cold water species on the Grand Bank of Newfoundland, (Eds: K. Sasruwatari, M. Nisshimura), NOVA, New York, **2012**.
- 2- W.S. Fredrick, S. Ravichandran. Hemolymph proteins in marine crustaceans. *Asian Pac J Trop Biomed.* **2012**, 496, 496.
- 3- N.B. Terwiliger, L. Dangott, M. Ryan. Cryptocyanin, a crustacean molting protein: Evolutionary link with arthropods hemocyanins and insect hexamerins. *Proc. Natl. Acad. Sci. USA.* **1999**, 96, 2013.
- 4- N.B. Terwiliger, M.C. Ryan, D. Towl. Evolution of novel functions: cryptocyanin helps build new exoskeleton in Cancer magister. *J. Exp. Biol.* **2005**, 208, 2467.
- 5- J.J. Beintema, W.T. Stam, B. Hazes, M.P. Smidt. Evolution of arthropod hemocyanins and insect storage proteins (hexamerins). *Mol. Biol. Evol.* **1994**, 11, 493.
- 6- M. Brouwer, R. Syring, T.H. Brouwer. Role of copper-specific metallothionein of the blue crab. *Callinectes sapidus*. In copper metabolism associated with degradation and synthesis of hemocyanin. *J. Inorg. Biochem.* **2002**, 88, 228.
- 7- N.B. Terwiliger, Hemolymph proteins and molting in crustaceans and insects *Am. Zool.* **1999**, 39, 589.
- 8- C. Zeleny. Compensatory regulation. *J. Rxp. Zool.* **1905**, 2, 1.
- 9- X. Kang, S. Ma, G. Li, Q. Wang, G. Liu, G. Cao. In neurosecretory structure and gonad inhibiting hormone in eyestalk and the physiology and biochemistry of spermatozoa in eriocheir sinesis, (Eds: K. Sasruwatari, M. Nisshimura), NOVA, New York, **2012**.

- 10- D. Skinner. Molting and regeneration in biology of crustacea. Academic Press, Orlando, **1985**.
- 11- E.S. Chang, M. J. Bruce, S. L. Tamone. Regulation of crustacean molting: a multi-hormonal system. *Am. Zool.* **1993**, 33, 324.
- 12- N. Ismail, S.S. Faezah, S.K. Dzulkipli, M.M. Fauzi and A.B. Luqman. In protein profiling of the hemolymph plasma from horseshoe crab, *Tachypleus gigas*. In proceedings of international conference on Life science, UMTAS, Kuala Terengganu, Malaysia **2011**.
- 13- A. Shevchenko, M. Wilm, O. Vorm, M. Mann. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* **1996**, 68, 850.
- 14- A. Dell, H.R. Morris. Glycoprotein structure determination by mass spectrometry. *Science* **2001**, 291, 2351.
- 15- E. Hoffmann, V. Stroobant. Mass spectrometry: principles and applications. John Wiley: New York, **2007**.
- 16- M. Cohen, F. Jahouh, S. Sioud, R.M. Rideout, M.J. Morgan, J.H. Banoub. Quantification of Greenland halibut serum vitellogenin: a trip from the deep sea to the mass spectrometer. *Rapid Commun. Mass Spectrom.* **2009**, 23, 1049.
- 17- M.A. Rahman, A.L. Lopata, R.E. O'Hehir, J.J. Robinson, J.H. Banoub, R.J. Helleur. Characterization and de novo sequencing of snow crab tropomyosin enzymatic peptides by both electrospray ionization and matrix-assisted laser desorption ionization QqTOF tandem mass spectrometry. *J. Mass Spectrom.* **2010**, 45, 372.

- 18- T.J. Cornish, R.J. Cotter. A curved field reflection time-of-flight mass spectrometer for the simultaneous focusing of metastable product ions. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 781.
- 19- P. Roespstroff, J. Foehlman. Letter to the editors *Biol. Mass spectrom.* **1984**, *11*, 601.
- 20- J. Hui, H. Desaire, V. Y. Butnev, G. R. Bousfield. Glycoprotein profiling by electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 750.
- 21- R. Koufmann, D. Kirsch, B. Spengler. Sequencing of peptides in a time-of-flight mass spectrometer: evaluation of post source decay following matrix-assisted laser desorption ionization (MALDI). *Int. J. Mass Spectrom. Ion process.* **1994**, *131*, 355.
- 22- M. Cohen, A.A. Mansour, J.H. Banoub. 'De novo' sequencing of atlantic cod vitellogenin tryptic peptides by matrix-assisted laser desorption/ionization quadrupole time-of-flight tandem mass spectrometry: similarities with haddock vitellogenin. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 2454.
- 23- M. Kinter, N.E. Sherman. Protein sequencing and identification using tandem mass spectrometry. John Wiley: New York, **2000**.
- 24- W.H. Vensel, C.K. Tanaka, N. Cai, J.H. Wong, B.B. Buchanan, W.J. Hurkman. Development changes in metabolic protein profiles of wheat endosperm. *Proteomics* **2005**, *5*, 1594.
- 25- T. Berggard, S. Linse, P. James. Methods for the detection and analysis of protein-protein interactions. *Proteomics* **2007**, *7*, 2833.
- 26- D.J. Pappin, P. Hojrup, A.J. Bleasby. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **1993**, *3*, 487.

- 27- M.A. Ziemer, A. Mason A, D.M. Carlson. Cell-free translations of proline-rich protein mRNAs. *J. Biol. Chem.* **1982**, *18*, 11176.
- 28- P. Roepstorff, J. Fohlman. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **1984**, *11*, 601.
- 29- W.I. Burkitt, A.E. Giannakopoulos, F. Sideeridou, S. Bashir, P.J. Derrick. Discrimination effects in MALDI-MS of mixtures of peptides- analysis of the proteome. *Austr. J. Chem.* **2003**, *56*, 369.
- 30- R. Kartazer, C. Eckerskorn, M. Karas, F. Lottspeich. Suppression effects in enzymatic peptide ladder sequencing using ultraviolet – matrix assisted laser desorption/ionization- mass spectrometry. *Electrophor.* **1998**, *19*, 1910.
- 31- M. Cohen, A.A. Mansour, J.H. Banoub. Absolute quantification of atlantic salmon and rainbow trout vitellogenin by the 'signature peptide' approach using electrospray ionization QqTOF tandem mass spectrometry. *J. Mass Spectrom.* **2006**, *4*, 646.

**Chapter III: Direct targeted glycation of the free sulfhydryl group of cysteine residue (Cys-34) of BSA. Mapping of the glycation sites of the anti-tumor Thomsen-Friedenreich neoglycoconjugate vaccine prepared by Michael addition reaction**

This chapter has been published: Wael L. L. Demian, Naresh Kotari, Tze Chieh Shiao, Edward Randell, René Roy and Joseph H. Banoub (*J. Mass Spectrom.* **2014**, *49*, 1223–1233)

**Chapter III** was designed by Wael L. L. Demian and Joseph H. Banoub, and Wael L. L. Demian conducted the research, analyzed the data and prepared the manuscript with guidance from Joseph H. Banoub. The final manuscript was read and approved by all authors. The carbohydrate hapten-BSA glycoconjugate was synthesized by the group of Professor Dr. René Roy (Department of Chemistry, University of Quebec at Montreal, CP 8888, Montreal, Qc, Canada H3C 3P8).

### 3.1 Introduction

The cytotoxic T cells (CD8+ T cells) play an important role in fighting cancer cells through recognizing the TACAs in a conventional class I MHC-restricted fashion.<sup>[1-9]</sup> Generally; cancer cells are often characterized by changes in their shape and size. These cancer cells attack and invade normal healthy cells and their surrounding environment.<sup>[1-9]</sup> One of the most noticeable changes during progression of cancer is the aberrant glycosylation.<sup>[8-10]</sup> Accordingly, in some forms, the resulting increase in the numbers of glycosylation sites are reflected by the presence of the TACAs such as; Tn (GalNAc $\alpha$ 1 $\rightarrow$ Ser/Thr), sialyl-Tn (Neu5Ac $\alpha$ 2 $\rightarrow$ 6GalNAc) and Thomsen-Friedenreich (TF: Gal $\beta$ 1 $\rightarrow$ 3GalNAc $\alpha$ 1 $\rightarrow$ Ser/Thr) antigens.<sup>[11-14]</sup>

The TF tumor associated antigen (Gal $\beta$ 1 $\rightarrow$ 3GalNAc $\alpha$ 1 $\rightarrow$ Ser/Thr) is present on the core 1 structure of *O*-linked mucin type glycan (MUC-1). This TF tumor associated antigen is formed by a disaccharide composed of one D-Galactose and one D-GalNAc residue attached by a  $\beta$ -D-(1 $\rightarrow$ 3) glycosidic linkage. Moreover, this disaccharide is linked by an  $\alpha$ -D-(1 $\rightarrow$ O) bond to the L-serine or L-threonine residues.<sup>[14,15]</sup>

It was observed that this antigen is heavily expressed in several carcinomas, such as those of breast, colon, bladder, and prostate.<sup>[13,15-18]</sup> Nevertheless, 90% of human cancer cells possess the tumor associated Gal $\beta$ 1 $\rightarrow$ 3GalNAc antigen on their epithelial tissues, in contrast to normal cells which are covered by sialic acids and sulphates groups.<sup>[12]</sup>

Different studies have been carried out on the use of the TF antigen as a target for the design of different cancer vaccines.<sup>[18-23]</sup> For example, Heimburg *et al.* found that the anti-Thomsen-Friedenreich antigen monoclonal antibody JAA-F11 inhibited spontaneous

breast cancer metastasis.<sup>[23]</sup> More recently, Hoffmann-Röder *et al.* prepared several synthetic antitumor vaccines composed of tetanus toxoid conjugates of Mucin 1 (MUC1) glycopeptides with the Thomsen-Friedenreich antigen (TF-MUC1-BSA conjugate) and a fluorine-substituted analogue (difluoro-TF-MUC1-BSA conjugate).<sup>[24]</sup> These authors also conjugated the tandem-repeat sequences of the epithelial mucin MUC1 containing the TF and/or its difluoro-TF saccharide side chains to the tetanus toxoid protein carrier.<sup>[24]</sup>

Different protein carriers such as keyhole limpet hemocyanine (KLH), diphtheria toxoid (DT), cross-reactive material 197 (CRM197) and tetanus toxoid (TT) have been used for the synthesis of various anticancer carbohydrate-based vaccines.<sup>[25-27]</sup> In these syntheses the carbohydrate antigens used were often allyl or pentenyl glycosides convertible into an aldehyde for subsequent attachment to the immunogenic protein carriers via reductive amination.<sup>[20-27]</sup> In addition, other frequently employed chemical ligation involved the 1,4-conjugate addition of multiply thiolated proteins onto maleimido derivatives.<sup>[28-30]</sup>

To the best of our knowledge, the systematic use of the L-lysine  $\epsilon$ -amino groups of the protein carrier, has scarcely been investigated as nucleophilic partners for the chemical ligation by a 1,4-conjugate addition.<sup>[31-33]</sup> Furthermore, the detailed structural identity of the glycoforms which are resulted from the 1,4-conjugate addition (Michael reaction) has never been used for the conjugation of the anticancer TF to BSA as a protein carrier.

In this manuscript, we have characterized two synthetic TF:BSA protein conjugate based vaccines and have revealed the respective C-*N* and C-*S* glycation sites of the synthetic TF to the various amino acid residues of bovine serum albumin protein (BSA).

## **3.2 Material and methods**

### **3.2.1 Preparation of the TF-BSA vaccine conjugate**

The TF was prepared according to a published procedure.<sup>[33]</sup> Briefly, the photocatalyzed reaction of the allyl glycoside of Thomsen–Friedenreich (TF) disaccharide [allyl 3-*O*-( $\beta$ -D-galactopyranosyl)-2-acetamido-2-deoxy- $\alpha$ -D-glucopyranoside] **1** with cysteamine gave the TF glycoside derivative **2** with a terminal primary amino group in 83% yield. Subsequently, compound **2** was treated with acryloyl chloride in methanol to yield the  $\alpha,\beta$ -unsaturated amide derivative **3** via the nucleophilic displacement of the chloride with the amino function. This reaction was carried out in the presence a strong base anion resin to remove the formed HCl. Finally, the acrylamide-ending TF **3** was anchored to the BSA carrier protein by treating the  $\alpha,\beta$ -unsaturated amide with BSA at room temperature and at 40 °C, respectively in a 0.2 M carbonate buffer for 3 days to afford the two desired TF-BSA anti-tumor vaccine conjugates having a low and a high TF-content. A colorimetric phenol-sulfuric acid test was performed onto the BSA-TF conjugates using compound **1** a standard.<sup>[34]</sup> The analysis showed the conjugates to contain  $2\pm 1$  and  $8\pm 2$  TFs, respectively.

### **3.2.2 Trypsin digestion of the glycoconjugate**

The digestion of the hapten-BSA glycoconjugate was carried out with trypsin (Sigma Aldrich, Saint Louis, MO, USA). Thus, 100  $\mu$ g of the glycoconjugate was

dissolved in a mixture of 0.1% RapiGest SF Surfactant (1  $\mu$ g, Waters, Milford, MA, USA) in 50 mM of  $\text{NH}_4\text{HCO}_3$  (100  $\mu$ l) at a pH of 8.0 and reduced by treatment with 2  $\mu$ l of 10mM dithiothreitol (Sigma Aldrich, Saint Louis, MO, USA) for 30 min at room temperature, followed by alkylation with 2  $\mu$ l of 50 mM iodoacetamide (Sigma Aldrich, Saint Louis, MO, USA) for 1 h at room temperature. A portion (50  $\mu$ g) of the glycoconjugate was digested with trypsin using a 20 ng/ml solution of trypsin dissolved in  $\text{NH}_4\text{HCO}_3$  (50 mM, 1 ml) at a trypsin-glycoprotein ratio of 1:25 (w/w) and incubated at 37°C overnight with shaking. The sample was then dried under vacuum, and the residue was dissolved in 20  $\mu$ l of 1% acetic acid (Sigma Aldrich, Saint Louis, MO, USA). An aliquot of each sample (10  $\mu$ l) was then cleaned up using ZipTip C18 (Millipore, Bedford, MA, USA) before mass spectral analysis.

### **3.2.3 MALDI-TOF-MS analysis**

MALDI-TOF-MS analysis was carried out on a 4800 Proteomics analyzer with TOF/TOF optics (Applied Biosystems Foster City, CA) (FWHM 20,000) and a 200-Hz frequency-tripled Nd:YAG laser.  $\alpha$ -Cyano-4-hydroxycinnamic acid ( $\alpha$ -CHCA) was used as matrix for the analysis of BSA and hapten-BSA conjugates with an average of 5000 to 8000 laser shots per spectra. Briefly, 1  $\mu$ L of a 20 mg/mL solution of  $\alpha$ -CHCA (dissolved in ACN, 0.1% trifluoroacetic acid) was spotted on the MALDI plate and dried at room temperature (the use of ACN allows a good homogeneity of the matrix in the spot). Then, an aliquot of 1  $\mu$ L of sample was spotted on the top of the dried matrix and allowed to dry before the MALDI-MS experiments. The analysis was achieved in the linear mode (positive ion mode) and the MALDI-TOF-MS was calibrated using BSA.

### 3.2.4 LC-ESI-QqTOF-CID-MS/MS analysis

The trypsin released glycopeptides and peptides were separated on a DIONEX UltiMate3000 Nano LC System (Germering, Germany). 250 fmol of digested glycoprotein was dissolved in 0.1% TFA and loaded onto a precolumn (300  $\mu\text{m}$  i.d.  $\times$  5 mm, C<sub>18</sub> PepMap100, 5  $\mu\text{m}$  (LC Packing, Sunnyvale, CA)) in order to desalt and concentrate the sample. After their elution from the precolumn, the peptides and glycopeptide mixtures were separated on a nanoflow analytical column (75  $\mu\text{m}$  i.d.  $\times$  15 cm, C<sub>18</sub> PepMap 100, 3  $\mu\text{m}$ , 100, (LC Packing, Sunnyvale, CA)) at a flow rate of 180 nL/min. The elution of the peptides and glycopeptides was achieved using the following mobile phases: 0.1% formic acid (FA)/0.01% TFA/2% ACN (A) and 0.08% FA/0.008% TFA/98% ACN (B). The elution started with 0% B for 10 min, followed by a gradient of 0-60% B in 55 min and 60-90% B in 3 min and was kept at 90% B for 3 min. The tandem mass spectrometry analysis of the eluted peptides and glycopeptides was accomplished using an Applied Biosystems API-QSTAR XL quadrupole orthogonal time-of-flight (QqTOF)-MS/MS hybrid tandem mass spectrometer (Applied Biosystems International-MDS SCIEX, Foster City, CA, USA) (FWHM 15,000) equipped with a nanoelectrospray source (Protana XYZ manipulator) which produces the electrospray through a PicoTip needle (10  $\mu\text{m}$  i.d., New Objectives, Wobum, MA, USA) carrying a voltage of 2400 V. The TOF analyzer was calibrated on the ions at  $m/z$  586.9815 and  $m/z$  879.9723, derived from renin (1 pmol/ $\mu\text{L}$  solution). The collision energies used during the CID-MS/MS analyses were determined automatically using the Information Dependent Acquisition (IDA) method integrated in the Analyst software.

### 3.2.5 MS/MS ion searches

Identification of glycoprotein from the MS/MS ions search was carried out with the Mascot search engine on NCBIInr. The search parameters used were: one missed cleavage was allowed, Cys CAM was introduced as a fixed modification, and the peptide tolerance was fixed at  $\pm 0.2$   $m/z$  unit. No other constraints were submitted.<sup>[35, 36]</sup>

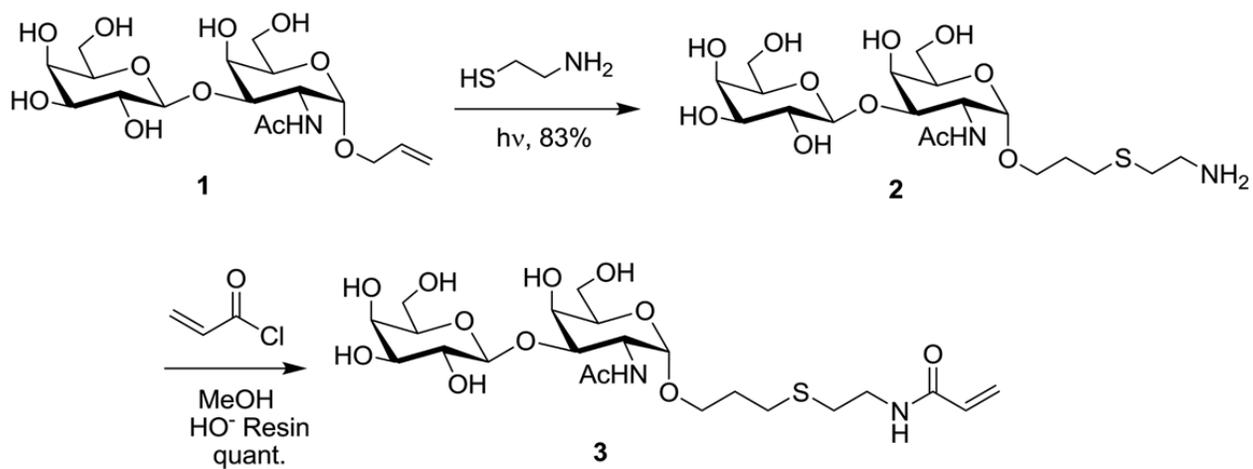
## 3.3 Results and discussion

### 3.3.1 Synthesis of the TF-BSA vaccine conjugate by the Michael addition reaction

BSA is one of the most widely studied proteins used as a carrier protein for the synthesis of neoglycoconjugates.<sup>[27,37-41]</sup> BSA contains 35 cysteine residues linked by S-S disulphide bonds among which, however, there is only a single buried free Cys 34 residue<sup>[42]</sup> that has been systematically used for other conjugation.<sup>[43,44]</sup> It is well known, that the 35 cysteine residues of BSA form 17 disulphide bonds at these positions (1) 77-86; (2) 99-115; (3) 114-125; (4) 147-192; (5) 191-200; (6) 223-269; (7) 268-276; (8) 288-302; (9) 301-312; (10) 339-384; (11) 383-392; (12) 415-461; (13) 460-471; (14) 484-500; (15) 499-510; (16) 537-582; (17) 581-590<sup>[42, 45-51]</sup> As a result, the free remaining reactive single sulfhydryl (Cys 34) group can serve as potential acceptor for the Michael addition reaction with a reactive acceptor.

In this report, the synthesis of the TF acceptor **3** was carried out by an already published strategy using the Michael addition reaction (Scheme 1).<sup>[31-33]</sup> The Michael addition reaction belongs to the larger class of conjugate additions and its mechanism involves the attack by biological nucleophile (cysteine or lysine) at the  $\beta$ -carbon atom of the  $\alpha,\beta$ -unsaturated Michael acceptor (**Scheme 3.1 and Figure 3.1**).<sup>[49-54]</sup> In order to

investigate the possible alternative conjugate addition originating from the lysine  $\epsilon$ -amine residues (or other amino acids), we tested various buffers, pH, and stoichiometric conditions. To our surprise, at pH 10 in a carbonate buffer at 40°C for three days, we found the highly immunogenic conjugated vaccine <sup>[43,44]</sup> to contain several copies of the TFs. <sup>[31,32]</sup>

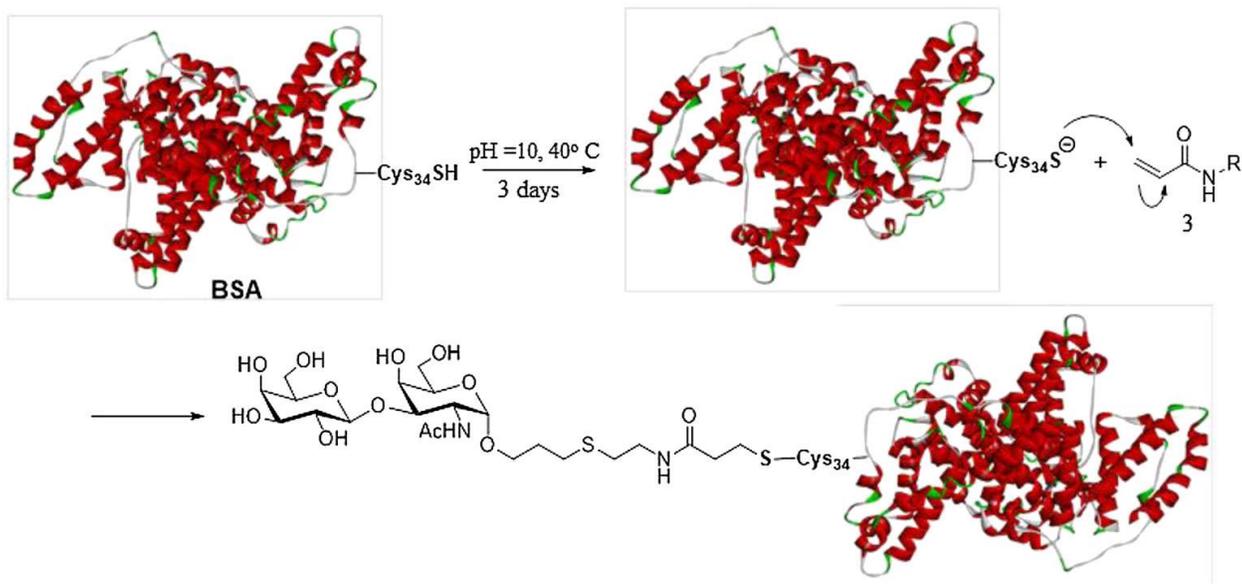


**Scheme 3.1** Structure of the TF antigen and its chemical transformation into a suitable Michael acceptor 3.

The mechanism of the based-catalyzed thio-Michael addition reaction between the TF and BSA is governed by the electrophilic addition of the acrylamide aglycone part of the disaccharide Gal $\beta$ 1 $\rightarrow$ 3GalNAc $\alpha$ , which easily reacts with the free sulphhydryl group of Cys 34 is illustrated in **Figure 3.1**.

In the presented work, we have attempted two different syntheses based on different stoichiometry of the glycosyl donor (TF) and acceptor carrier protein (BSA) of 2:1 and 8:1, respectively. As expected, these conditions resulted in the formation of two different TF neoglycoconjugate vaccines.

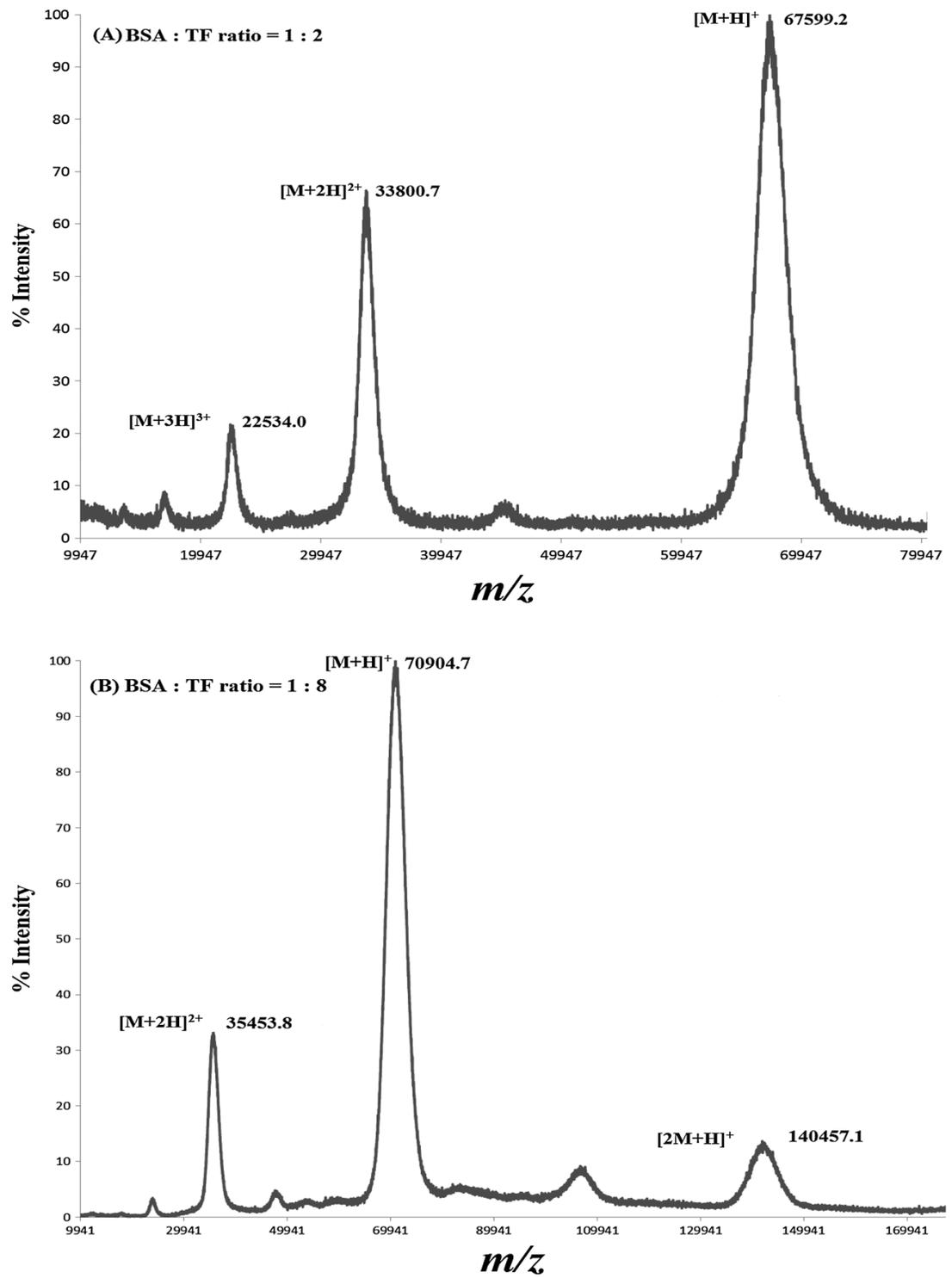
The best advantage of the synthetic tactic used for producing the TF:BSA conjugate over other existing approaches for protein glycosylation, is that it can be applied to cysteine containing proteins (free of disulphide bonds with neighbouring residues) in their native form without any additional chemical modification or activation.



**Figure 3.1** Synthesis of Thomsen–Friedenreich antigen–BSA glycoconjugate vaccines using a Michael addition.

### 3.3.2 MALDI-TOF-MS analysis of the TF-BSA glycoconjugates

For the determination of the average number of carbohydrate-spacer moieties linked to BSA (TF: BSA), the two glycoconjugates were analyzed by matrix assisted laser desorption/ionization- mass spectrometry (**Figure 3.2a and 3.2b and Table 3.1**).<sup>[38-41]</sup> The MALDI-TOF-MS were characterized by the formation of a protonated molecular ions  $[M+H]^+$  at  $m/z$  67599 and 70904 and their TF-BSA ratios were calculated to be individually 2:1 and 8:1.



**Figure 3.2** a) MALDI-TOF-MS analysis of TF:BSA glycoconjugates with ratio= 2:1. b) MALDI-TOF-MS analysis of TF:BSA glycoconjugates with ratio = 8:1.

The molecular mass calculation of the glycoconjugate formed with a hapten:BSA ratio of 2:1, was based on the assumption that one molecule of the TF will covalently link by a Michael addition reaction on the free sulfhydryl (Cys 34) group, forming a C-S bond. The remaining stoichiometric amount (one molecule) of the TF, can react with specific free available  $\epsilon$ -amino group of lysine residue of BSA, forming a C-N bond. For the glycoconjugate designed with the hapten:BSA ratio of 2:8 we expected similar glycation of Cys 34, whereas the seven additional molecules of the TF (hapten) directly targeted the  $\epsilon$ -amino groups of the various lysine residues of BSA, forming a C-N bond, which occurred by loss of ethanol molecules.

It is important to recognize that the TF is composed of the disaccharide Gal $\beta$ 1 $\rightarrow$ 3GalNAc $\alpha$ 1 $\rightarrow$ O-linked to the spacer  $[-(\text{CH}_2)_3\text{S}(\text{CH}_2)_2\text{NH-CO-CH=CH}_2]$  (containing the acrylamide portion) and it has a molecular mass of 555.22 Da (**Table 3.1**).

**Table 3.1.** MALDI-TOF-MS of BSA and TF: BSA glycoconjugate vaccines and the calculated carbohydrate:BSA ratios

	$[M+2H]^{2+}$	$[M+H]^+$	<b>TF: BSA Ratio</b>
<b>Standard BSA</b>	33216	66431	-
<b>TF:BSA 2:1</b>	33800	67599	2.1:1
<b>TF:BSA 8:1</b>	35453	70904	8.1:1

### **3.3.3 LC-ESI-MS and CID-MS/MS analyses of tryptic digest of the glycoconjugate with a TF: BSA ratio of 2:1 and 8:1**

Nano-LC-ESI-QqTOF-MS/MS analysis was carried out on both tryptic digests of glycoconjugates with TF: protein ratio 2: 1 and 8:1, and this was followed by *de novo* peptide sequencing using low-energy collision dissociation tandem mass spectrometry CID-MS/MS to identify the possible S- and N- glycosylated sites.<sup>[55,56]</sup> The obtained low-energy CID-MS/MS were submitted to the Mascot library to identify the glycopeptides and peptides released from BSA.

The Mascot reports of MS/MS data of the tryptic digests of both glycoconjugates identified two serum albumin isoforms from *Bos taurus* species. The serum albumin precursor (gi|1351907) was identified with sequence coverages: 46% and 21% for both glycoconjugates formed with TF: BSA ratio of 2:1 and 8:1 respectively. In addition, serum albumin protein isoform (gi|229552) was also identified for the tryptic digest for both glycoconjugates with sequence coverages of 42% for the TF: BSA ratio of 2:1 and 19% for the glycoconjugated having TF: BSA ratio of 8:1.

Please note that, the identified glycopeptides  $m/z$  values were calculated by adding the molecular mass of the carbohydrate TF (Molecular mass, 555.22 Da), to the  $m/z$  value of the “correct” identified peptides. The “correct” peptide was identified by comparing the enzymatically released peptides of both the glycoconjugate vaccine and BSA. This comparison allows the selection of the missing peptide ions from blueprint of the BSA digestion. Once these peptides are identified, it is possible to add their respective masses to the mass of the carbohydrate TF. This method permits to determine the exact structures and masses of the glycosylated peptides, which sequences are then verified by MS/MS.

Following *de novo* sequencing of the digested glycoconjugate formed with a TF: BSA, ratio of 2:1, we were able to identify of 3 glycated peptides (**Table 3.2**) ; FK\*DLGEEHFK **4** (*C-N* glycation site is Lys 12) at 902.11 (+2), GLVLIAFSQYLQQC\*PFDEHVK **5** (*C-S* glycation site is Cys 34) at 1015.48 (+3) and HKPK\*AEEQLK **6** (*C-N* glycation site is Lys 535) at 620.54 (+3) (**Table 3.2**).

**Table 3.2** CID-MS/MS analysis of the glycated peptides ions obtained by the trypsin digestion of the TF:BSA vaccine with ratio 2:1

Glycated peptide sequence	Missed Cleavage	TF:BSA ratio 2:1			
		Molecular weight	Calculated <i>m/z</i> (charge)	Observed <i>m/z</i> (charge)	Error
FK*DLGEEHFK (Lys 12) <b>4</b>	1	1802.18	902.09(+2)	902.11(+2)	-0.02
GLVLIAFSQYLQQC*PFDEHVK <b>5</b> (Cys 34)	0	3043.61	1015.53(+3)	1015.48(+3)	0.05
HKPK*ATEEQLK (Lys 535) <b>6</b>	2	1858.61	620.53(+3)	620.54(+3)	-0.01

Sequence of the glycated peptides (\* = glycation site)

Subsequently, *de novo* sequencing of the digested glycoconjugate formed with a TF: BSA, ratio of 8:1, allowed us to identify 14 glycosylated peptides. The structures of this series of glycopeptides are specified in **Table 3.3**. This series of glycopeptides contained the three above-mentioned identified peptides **4** (*C-N* glycation site is Lys 12), **5** (*C-S* glycation site is Cys 34) and **6** (*C-N* glycation site is Lys 535); in which all ions had the same charges as previously described. Furthermore, it also contained the following peptides: **7** (*C-N* glycation site is Lys 116) at  $m/z$  1065.58(+2), **8** (*C-N* glycation site is Lys 522) at  $m/z$  848.71(+2), **9** (*C-N* glycation site is Lys 412) at  $m/z$  731.46(+3), **10** (*C-N* glycation site is Lys 127) at  $m/z$  858.28(+3), **11** (*C-N* glycation site is Lys 106) at  $m/z$  818.79(+3), **12** (*C-N* glycation site is Lys 312) at  $m/z$  1004.26(+3), **13** (*C-N* glycation site is Lys 429) at  $m/z$  685.25, **14** (*C-N* glycation site is Lys 186) at  $m/z$  765.85(+2), **15** (*C-N* glycation site is Lys 210) at  $m/z$  777.37(+2), **16** (*C-N* glycation site is Lys 4) at  $m/z$  873.42(+2), **17** (*C-N* glycation site is Lys 179) at  $m/z$  942.63(+2).

**Table 3.3** CID-MS/MS analysis of the glycosylated peptides ions obtained by the trypsin digestion of the TF:BSA vaccine with ratio 8:1

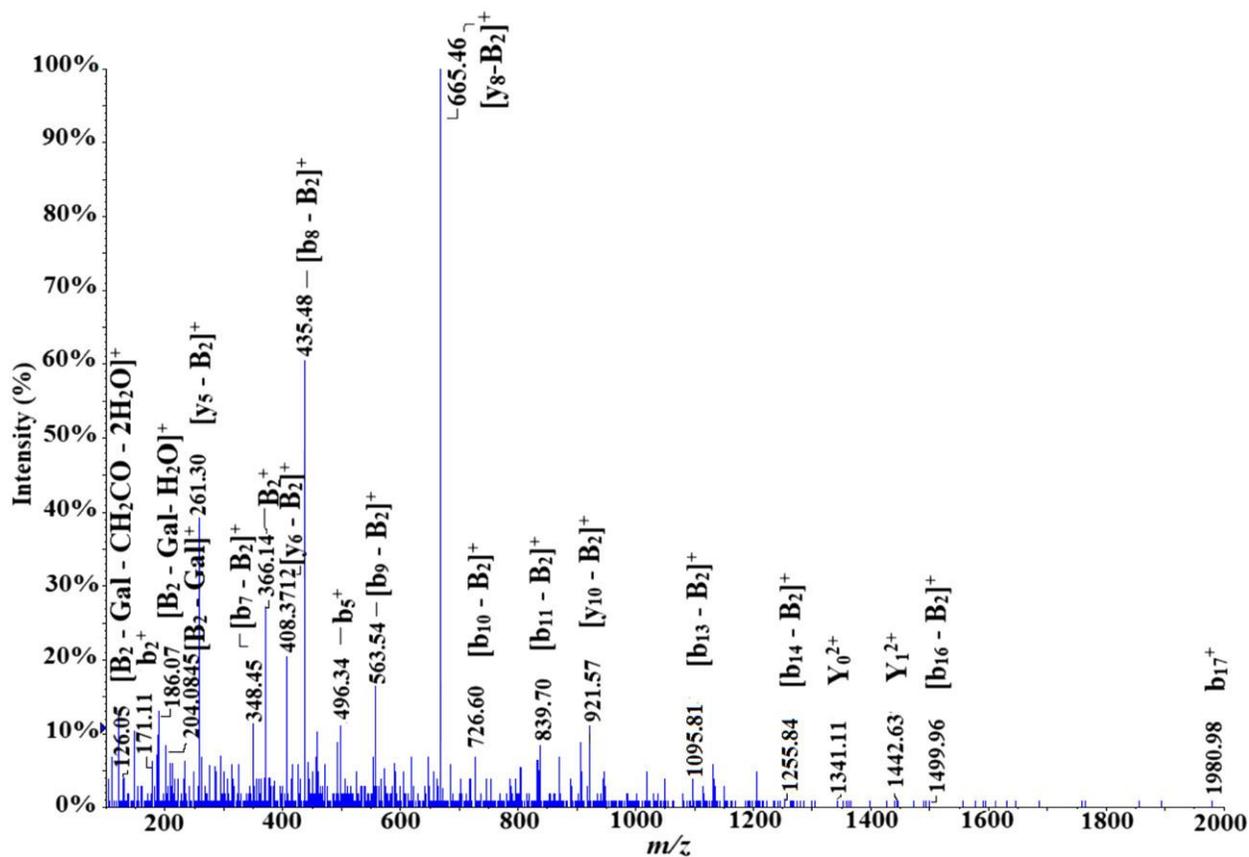
Glycosylated peptide sequence	Missed Cleavage	TF: BSA ratio 8:1			
		Molecular weight	Calculated <i>m/z</i> (charge)	Observed <i>m/z</i> (charge)	Error
FK*DLGEEHFK (Lys 12) <b>4</b>	1	1802.18	902.09 (+2)	902.12(+2)	-0.03
GLVLIAFSQYLQQC*PFDEHVK (Cys 34) <b>5</b>	0	3043.61	1015.53 (+3)	1015.46(+3)	0.07
HKPK*ATEEQLK (Lys 535) <b>6</b>	2	1858.61	620.53 (+3)	620.53 (+3)	0.00
LK*PDPNTLCDEFK (Lys 116) <b>7</b>	1	2129.33	1065.66(+2)	1065.58 (+2)	0.08
K*QTALVELLK (Lys 522) <b>8</b>	1	1695.27	848.63 (+2)	848.71 (+2)	-0.08
K*VPQVSTPTLVEVSR (Lys 412) <b>9</b>	1	2191.50	731.50 (+3)	731.46 (+3)	0.04
LKPDPNTLCDEFK*ADEK (Lys 127) <b>10</b>	2	2571.62	858.20 (+3)	858.28 (+3)	-0.08
NECFLSHK*DDSPDLPK (Lys 106) <b>11</b>	1	2453.43	818.81 (+3)	818.79 (+3)	0.02
DAIPENLPPLTADFAEDK*DVCK (Lys 312) <b>12</b>	1	3009.74	1004.24 (+3)	1004.26 (+3)	-0.02
SLGK*VGTR (Lys 429) <b>13</b>	1	1368.52	685.26 (+2)	685.25 (+2)	0.01

EK*VLTSAR (Lys 186) <b>14</b>	1	1529.69	765.84 (+2)	765.85 (+2)	-0.01
ALK*AWSVAR (Lys 210) <b>15</b>	1	1552.75	777.37 (+2)	777.37 (+2)	0.00
DTHK*SEIAHR (Lys 4) <b>16</b>	1	1744.85	873.42 (+2)	873.42 (+2)	0.00
GACLLPK*IETMR (Lys 179) <b>17</b>	1	1883.26	942.63 (+2)	942.63 (+2)	0.00

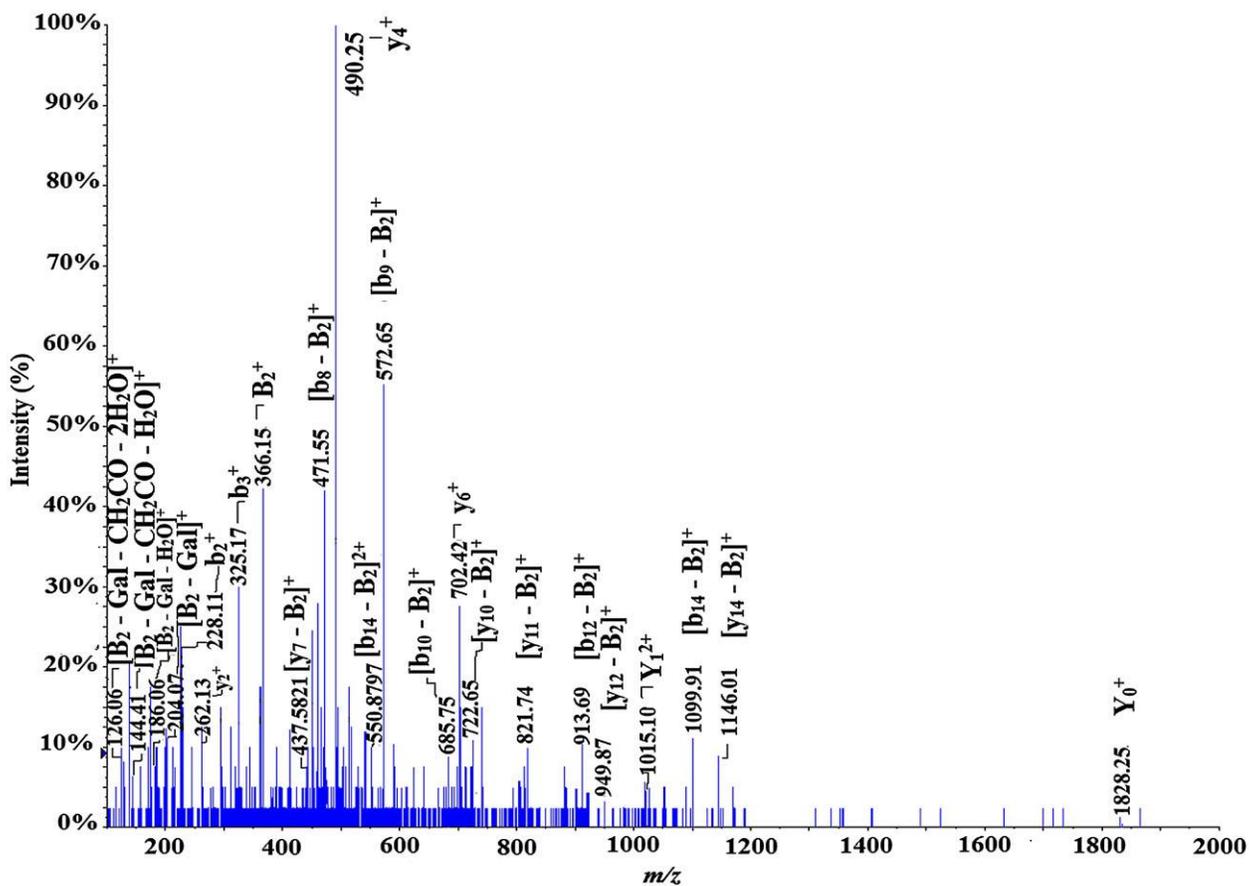
Sequence of the glycated peptides (\* = glycation site)

It is essential to note that we have identified for both synthetic glycoconjugates the expected glycation site on Cys 34 residue, resulting from a Michael addition reaction between the TF antigen and the BSA protein carrier. Whereas, the direct conjugation of the TF ensued on the other  $\epsilon$ -amino groups of the various lysine residues of BSA. Nevertheless, it was noted that for the two previously described trypsin digestions, we have identified a higher number of glycopeptides than one should expect, for a reaction performed with different stoichiometry of the TF:BSA ratio 2:1 and 8:1. This means that the uses of these different synthetic ratios result in neoglycoconjugates possessing different glycation occupancies. The net increase in the occupancy sites is caused by the presence of several numbers of glycoforms, formed during the synthetic carbohydrate carrier protein conjugation.<sup>[38]</sup> This formation of glycoforms is not unforeseen as BSA contains several lysine residues which are free of steric hindrances and are straightforwardly available for other N-glycation reactions.

For the purpose of illustration of the *de novo* sequencing of the formed glycopeptides, we have chosen to describe three examples. Henceforth, the product ion scans of the protonated molecules of the glycated peptides **5** (*C-S* glycation site is Cys 34) at  $m/z$  1015.46(+3) (**Fig. 3.3**); **9** (*C-N* glycation site located at Lys412) at  $m/z$  731.46(+3) (**Fig. 3.4**) and **4** (*C-N* glycation site is Lys12) at  $m/z$  902.12(+2) (**Figure B.1**, supplementary material) are respectively described.



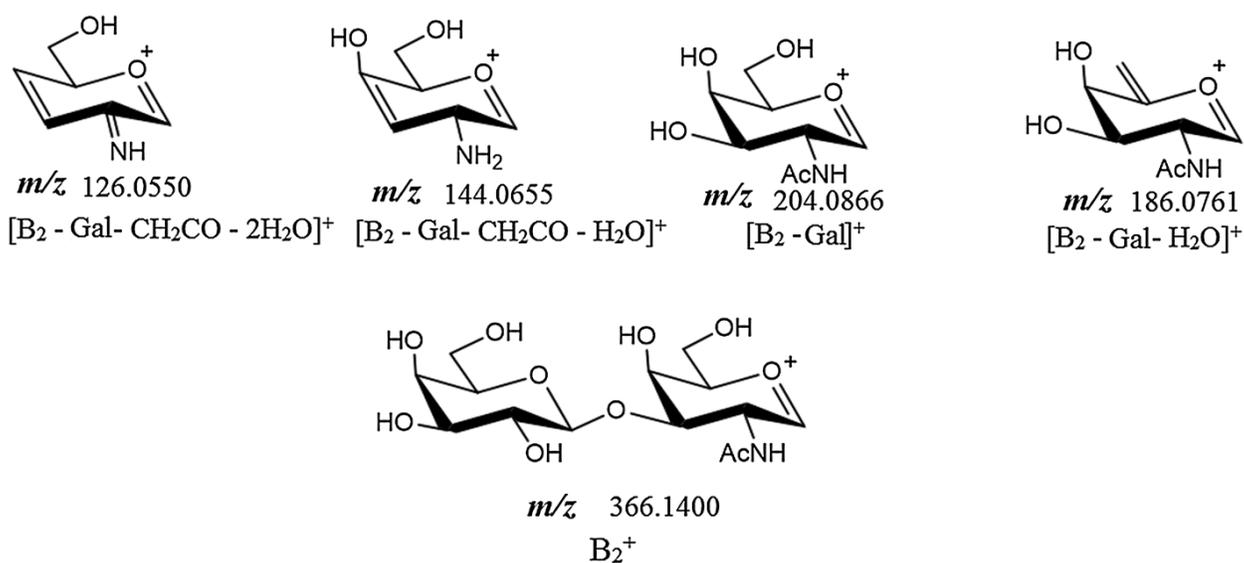
**Figure 3.3** CID-MS/MS spectra of identified glycosylated peptide GLVLIAFSQYLQQC\*PFDEHK 5 (Cys34) at  $m/z$  1015.4614 (+3).



**Figure 3.4** CID-MSMS spectra of identified glycosylated peptide K\*VPQVSTPTLVEVSR 9 (Lys412) at  $m/z$  731.4685 (+3).

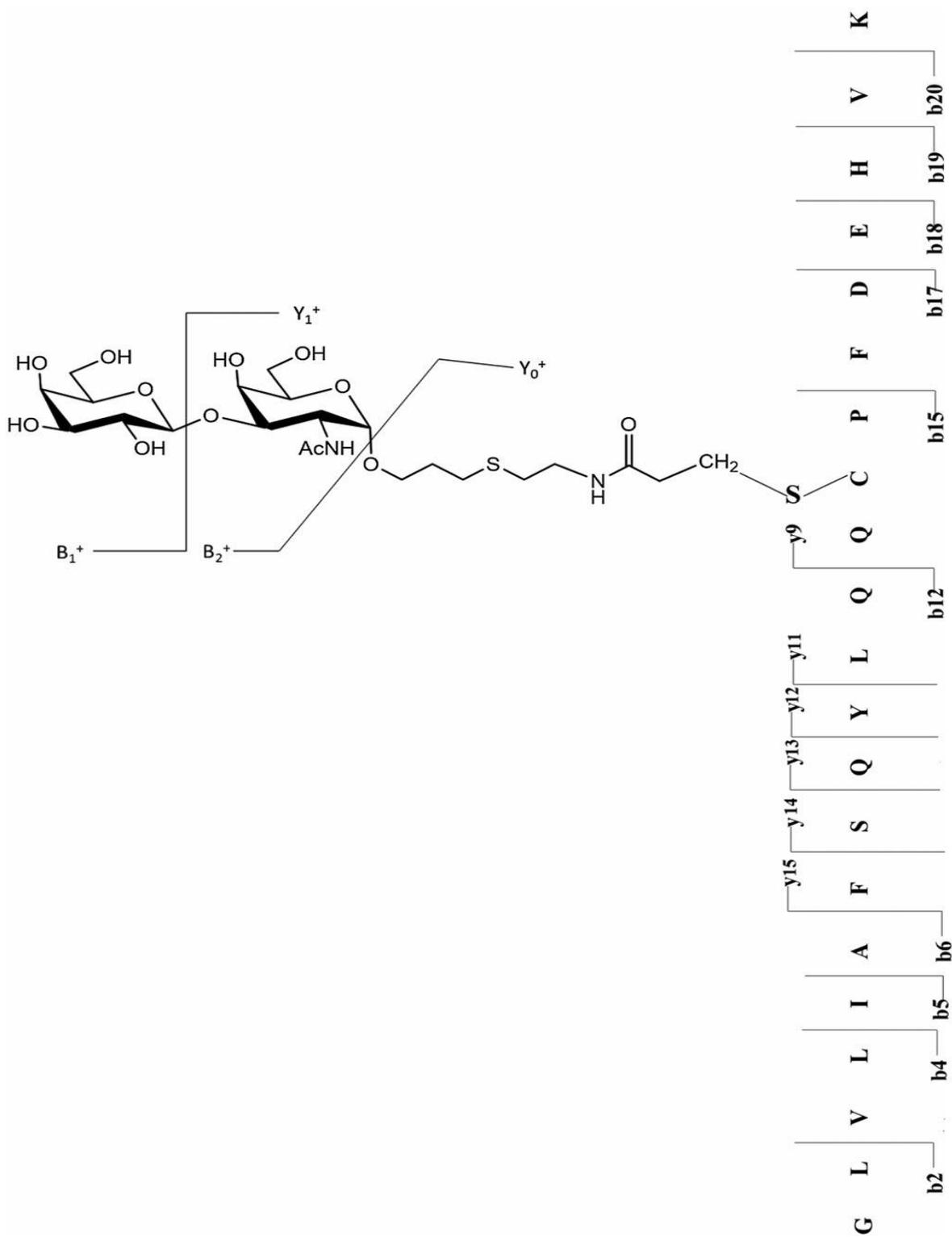
In addition, the formation of the diagnostic product ions obtained by the gas-phase CID-fragmentation of these protonated molecules of these glycosylated peptides **4**, **5** and **9** are presented in **Tables 3.4, 3.5** and (**Table B.1**, supplementary material). In this rationale, we have designated the carbohydrate product ions obtained by the MS/MS fragmentation of the glycopeptide protonated molecules the nomenclature generated for the B- and Y- ions product ions described by Domon and Costello,<sup>[57]</sup> whereas for the peptide product ions, we have used the b- and y- ions nomenclature of Roepstorff and Fohlman,<sup>[58]</sup>

As hypothetically expected, we have noted for both digested glycoconjugates, the diagnostic formation of product ions corresponding to specific fragmentation of TF carbohydrate part, which is displayed in **Figure 3.5**. These product ions were characterized as follows:  $[B_2 - \text{Gal} - \text{CH}_2\text{CO} - 2\text{H}_2\text{O}]^+$  at  $m/z$  126.05,  $[B_2 - \text{Gal} - \text{CH}_2\text{CO} - \text{H}_2\text{O}]^+$  at  $m/z$  144.06,  $[B_2 - \text{Gal} - \text{H}_2\text{O}]^+$  at  $m/z$  186.07,  $[B_2 - \text{Gal}]^+$  at  $m/z$  204.08 and  $B_2^+$  at  $m/z$  366.14.

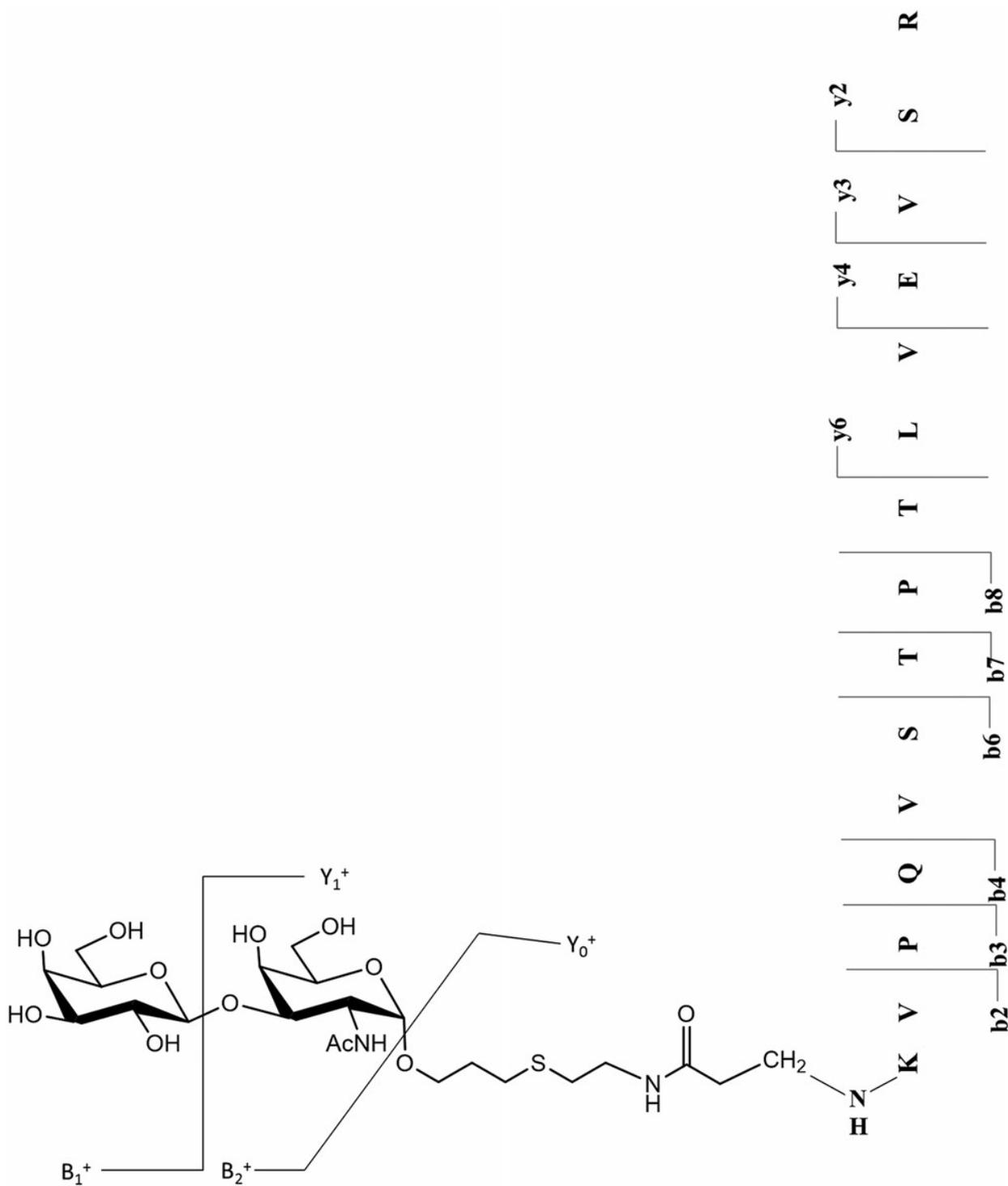


**Figure 3.5** Product ions signature of the carbohydrate antigens, observed during the CID-MS/MS analysis of the glycosylated peptides.

A point of interest worth mentioning, is that we have observed a difference in the CID-fragmentation routes between the product ion scans of the two types of identified glycopeptides **5** (*C-S* glycation site located at Cys 34) at  $m/z$  1015.46(+3) (**Scheme 3.2**), and **9** (*C-N* glycation site located at Lys 412) at  $m/z$  731.46(+3) (**Scheme 3.3**).



**Scheme 3.2** CID-MS/MS fragmentation of the TF-glycopeptide attached with the C—S bond at Cys 34.



**Scheme 3.3** CID-MS/MS fragmentation of the TF-glycopeptide attached with C—N bond at Lys 412.

The CID-fragmentation of the (Cys 34) *S*-glycated peptide: **5** at  $m/z$  1015.46(+3) afforded the series of products ions indicated in Table 4. This MS/MS analysis allowed us to identify the product ions formed by the carbohydrate hapten exclusively:  $[B_2 - Gal - CH_2CO - 2H_2O]^+$  at  $m/z$  126.05,  $[B_2 - Gal - H_2O]^+$  at  $m/z$  186.07,  $[B_2 - Gal]^+$  at  $m/z$  204.08,  $B_2^+$  at  $m/z$  366.14,  $Y_0^{2+}$  at 1341.11 and  $Y_1^{2+}$  1442.63. Additionally, we also noted the peptide product ions resulting from the release of the carbohydrate portion from the glycopeptide protonated molecule  $[-B_2^+]$  and these were identified as:  $[y_5 - B_2]^+$  at  $m/z$  261.30,  $[b_7 - B_2]^+$  at  $m/z$  348.45,  $[y_6 - B_2]^+$  at  $m/z$  408.37,  $[b_8 - B_2]^+$  at  $m/z$  435.48,  $[b_9 - B_2]^+$  at  $m/z$  563.54,  $[y_8 - B_2]^+$  at  $m/z$  665.46,  $[b_{10} - B_2]^+$  at  $m/z$  726.60,  $[b_{11} - B_2]^+$  at  $m/z$  839.70,  $[y_{10} - B_2]^+$  at  $m/z$  921.57,  $[b_{13} - B_2]^+$  at  $m/z$  1095.81,  $[b_{14} - B_2]^+$  at  $m/z$  1255.84 and  $[b_{16} - B_2]^+$  at  $m/z$  1499.96. Finally, we noted the products ions created by the forthright fragmentation of the peptide portion only:  $b_2^+$  at  $m/z$  171.11,  $b_5^+$  at  $m/z$  496.34 and  $b_{17}^+$  at  $m/z$  1980.98.

Finally, it is important to mention that for this above described MS/MS analysis of this glycopeptide ion at  $m/z$  1015.46, no product ion containing the intact spacer linker arm was ever observed.

**Table 3.4** Product ions obtained following the CID-MS/MS analysis of GLVLIAFSQYLQQC\*PFDEHVK 5 (Cys 34) ion at  $m/z$  1015.46(+3).

<b>Product ion</b>	<b>Calculated <math>m/z</math></b>	<b>Observed <math>m/z</math></b>	<b>Error</b>
$b_{17}^+$	1980.98	1980.98	0.00
$[b_{16} - B_2]^+$	1499.96	1499.96	0.00
$Y_1^{2+}$	1442.66	1442.63	0.03
$Y_0^{2+}$	1341.16	1341.11	0.05
$[b_{14} - B_2]^+$	1255.84	1255.84	0.00
$[b_{13} - B_2]^+$	1095.81	1095.81	0.00
$[y_{10} - B_2]^+$	921.57	921.57	0.00
$[b_{11} - B_2]^+$	839.70	839.70	0.00
$[b_{10} - B_2]^+$	726.60	726.60	0.00
$[y_8 - B_2]^+$	665.46	665.46	0.00
$[b_9 - B_2]^+$	563.54	563.54	0.00
$b_5^+$	496.34	496.34	0.00
$[b_8 - B_2]^+$	435.48	435.48	0.00
$[y_6 - B_2]^+$	408.37	408.37	0.00
$B_2^+$	366.14	366.14	0.00
$[b_7 - B_2]^+$	348.45	348.45	0.00
$[y_5 - B_2]^+$	261.30	261.30	0.00
$[B_2 - Gal]^+$	204.08	204.08	0.00
$[B_2 - Gal - H_2O]^+$	186.07	186.07	0.00

$b_2^+$	171.11	171.11	0.00
$[B_2 - Gal - CH_2CO - 2H_2O]^+$	126.05	126.05	0.00

On the other hand, the CID fragmentation for *N*-glycated peptides **9** (*C*-*N* glycation site is Lys 412) at  $m/z$  731.46(+3) and **4** (*C*-*N* glycation site is Lys12) at  $m/z$  902.12(+2) are shown in **Table 3.5** and (**Table S3.1**, supplementary material).

**Table 3.5** Product ions obtained following the CID-MS/MS analysis of K\*VPQVSTPTLVEVSR 6 (Lys 412) ion at  $m/z$  731.46(+3)

<b>Product ion</b>	<b>Calculated <math>m/z</math></b>	<b>Observed <math>m/z</math></b>	<b>Error</b>
$Y_0^+$	1828.35	1828.25	-0.1
$[y_{14} - B_2]^+$	1145.84	1146.01	-0.17
$[b_{14} - B_2]^+$	1099.82	1099.91	-0.09
$Y_1^{2+}$	1015.17	1015.10	-0.07
$[y_{12} - B_2]^+$	949.72	949.87	-0.15
$[b_{12} - B_2]^+$	913.72	913.69	0.03
$[y_{11} - B_2]^+$	821.66	821.74	-0.08
$[y_{10} - B_2]^+$	722.59	722.65	-0.06
$y_6^+$	702.41	702.42	-0.01
$[b_{10} - B_2]^+$	685.61	685.75	-0.14
$[b_9 - B_2]^+$	572.53	572.65	-0.12
$[b_{14} - B_2]^{2+}$	550.91	550.87	0.04
$y_4^+$	490.26	490.25	0.01
$[b_8 - B_2]^+$	471.48	471.55	-0.07
$[y_7 - B_2]^+$	437.46	437.58	-0.12
$B_2^+$	366.14	366.15	-0.01
$b_3^+$	325.22	325.17	0.05
$y_2^+$	262.15	262.13	0.02
$b_2^+$	228.17	228.11	0.06

$[\text{B}_2 - \text{Gal}]^+$	204.08	204.07	0.01
$[\text{B}_2 - \text{Gal} - \text{H}_2\text{O}]^+$	186.07	186.06	0.01
$[\text{B}_2 - \text{Gal} - \text{CH}_2\text{CO} - \text{H}_2\text{O}]^+$	144.06	144.41	-0.35
$[\text{B}_2 - \text{Gal} - \text{CH}_2\text{CO} - 2\text{H}_2\text{O}]^+$	126.05	126.06	-0.01

To visualize the location of the glycation occupancies of the TF: BSA synthetic vaccines, we have indicated the complete sequence of BSA and the various glycation sites in **Figure 3.6**. It is important to note that the three glycation sites of the TF: BSA glycoconjugate vaccine formed with a ratio of 2:1 are underlined. Whereas, the 14 glycation sites are indicated with a red asterisk for the vaccine formed with a ratio of 8:1.

1	DTH <u>K</u> *SEIAHR	<u>F</u> <u>K</u> *DLGEEHFK	GLVLIAFSQY	<u>L</u> <u>Q</u> <u>Q</u> <u>C</u> *PFDEHV	KLVNELTEFA
51	KTCVADESHA	GCEKSLHTLF	GDELCKVASL	RETYGDMADC	CEKEQPERNE
101	CFLSH <u>K</u> *DDSP	DLPKL <u>K</u> *PDPN	TLCDEF <u>K</u> *ADE	KKFWGKYLYE	IARRHPYFYA
151	PELLYANKYN	GVFQECCQAA	DKGACLL <u>P</u> <u>K</u> *I	ETMRE <u>K</u> *VLTS	SARQRLRCAS
201	IQKFGERAL <u>K</u> *	AWSVARLSQK	FPKAEFVEVT	KLVTDLTKVH	KECCHGDLE
251	CADDRADLAK	YICBBZBTIS	SKLKECKDPC	LLEKSHCIAE	VEKDAIPEDL
301	PPLTADFAED	<u>K</u> *DVCKNYQEA	KDAFLGSFLY	EYSRRHPEYA	VSVLLRLAKE
351	YEATLEECCA	KDDPHACYTS	VFDKCLKHLVD	EPQNLIKZBC	BZFEKLGEYX
401	XXALIVRYTR	<u>K</u> *VPQVSTPTL	VEVSRSL <u>G</u> <u>K</u> *V	GTRCCTKPES	ERMPCTEDYL
451	SLILNRLCVL	HEKTPVESKV	TKCCTESLVN	RRPCFSALTP	DETYVPKAFD
501	EKLFTFHADI	CTLDPTEKQI	<u>K</u> <u>K</u> *QTALVELL	KHK <u>P</u> <u>K</u> *ATEEQ	LKTVMENFVA
551	FVDKCCAADD	KEACFAVEGP	KLVVSTQTAL	A	

**Figure 3.6** BSA sequence where the 14 glycation sites are indicated by red asterisk = identified peptides in TF:BSA ratio = 8: 1. The three underlined aminoacids = identified in TF:BSA ratio= 2: 1 and in TF:BSA ratio = 8:1.

The structure generated from the X-ray data of BSA (PDB access No. 3V03) using the software Discovery Studio v 3.1 is shown in Figure 7. It is noteworthy to mention that the labeled Cys 34 and the 14 Lys residues are designated in the yellow and blue color respectively (ball and stick; C-S group of Cys 34 is in yellow, and the  $\epsilon$ -amino groups of the Lys are in blue). Additionally, it is interesting to note that the characterized glycation sites proposed in this manuscript corresponded to the sterically non-crowded lysine and Cys 34 residues which were available for glycation by the hapten (**Figure 3.7**).



**Figure 3.7** 3D structure of BSA generated from the X-ray data using the software Discovery Studio v 3.1.

### 3.4 Conclusion

In this work, we have determined the precise glycation sites of two different TF:BSA (TACAs) vaccine prepared by a 1,4-conjugate addition (Michael reaction) between an acrylamide-ending TF (**3**) with BSA as carrier protein using a 0.2 M carbonate buffer for three days at either room temperature or at 40°C. This rationale reflects the attractiveness of glycoscience in the ability of using the free sulfhydryl group of Cys 34 of BSA as protein carrier for conjugation with TACAs but at the same time emphasize the fact that even the  $\epsilon$ -amino groups of the Lys residues can enter into such conjugate addition at proper pH, temperature, and reaction time.

Two synthetic neoglycoconjugate vaccines were prepared by the Michael addition reaction. Their molecular composition reflected exactly the stoichiometry found by our colorimetric analysis (Dubois test). These ratios were confirmed by MALDI-TOF-MS analysis. *De novo* peptide sequencing using low-energy CID-MS/MS tandem mass spectrometry of the enzymatic digests allowed us to identify 3 glycation sites for the glycoconjugate prepared with a TF: protein ratio 2: 1 and 14 glycation sites for the glycoconjugate prepared with a TF: protein ratio 8: 1. All the identified glycation sites are located on lysine residues were the results of *N*-alkylation. However, this manuscript reports the formation of the unique *C-S* linked glycoconjugate identified exclusively on Cys 34.

This systematic investigation constitutes a versatile example for the identification of accurate quality control necessary in commercial production of glycoconjugate vaccines. The glycopeptides isolated and fully characterized during this work may well

represent useful reference compounds to be used in standardization analyses. Moreover, although BSA usually serves a universal model carrier protein for novel conjugation chemistry, we found it perfectly legitimate as a vaccine in mouse experiments since the monoclonal antibodies isolated from the above conjugates were found to be anti-metastatic.<sup>[23]</sup> Most importantly, the identified glycopeptides may well form the basis for fully synthetic carbohydrate-based vaccines.<sup>[2-5]</sup> This would be particularly true when performed on other more immunogenic protein carriers such as tetanus toxoid and KLH.

### **Appendix B. Supplementary data**

*Supplementary are data associated with this chapter (P.176)*

## References

- 1- G.F. Gao, B.K. Jakobsen. Molecular interactions of coreceptor CD+8 and MHC class I: the molecular basis for functional coordination with the T-cell receptor. *Immunol Today* **2000**, *21*, 630.
- 2- R. Roy. New trends in carbohydrate-based vaccines. *Drug Discov. Today: Technologies*. **2004**, *1*, 327.
- 3- R. Roy, T.C. Shiao, K. Rittenhouse-Olson. Glycodendrimers : versatile tools for nanotechnology. *Braz. J. Pharm. Sci.* **2013**, *49*, 85.
- 4- R. Roy, T.C. Shiao. Glycodendrimers as functional antigens and antitumorales vaccines. *New J. Chem.* **2012**, *36*, 324.
- 5- R. Roy, T.C. Shiao. Organic chemistry and immunochemical strategies in the design of potent carbohydrate-based vaccines. *Chimia*, **2011**, *65*, 24.
- 6- L.P. Icart, V. Fernandez-Santana, R.C. Veloso, T. Carmenate, R. Sirois, R. Roy, V. Verez Bencomo. T-cell immunity of carbohydrates. In carbohydrate-based vaccines. R. Roy, Ed., *ACS Symp. Ser.* **2007**, 989, 1.
- 7- M.A. Daniels, S.C. Jameson. Critical role for CD+8 in T cell receptor binding and activation by peptide/major histocompatibilitycomplex multimers. *J. Exp. Med.* **2000**, *191*, 335.
- 8- Y.S. Kim, J. Gum, I. Brockhausen, Mucin glycoproteins in neoplasia. *Glycoconj. J.* **1996**, *13*, 693.
- 9- Y.J. Kim, A. Varki. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconj. J.* **1997**, *14*, 569.

- 10- M. Ono, S. Hakomori. Glycosylation defining cancer cell motility and invasiveness. *Glycoconj. J.* **2004**, *20*, 71.
- 11- G.F. Springer. Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy. *J. Mol. Med.* **1997**, *75*, 594.
- 12- B.J. Campbell, I.A. Finnie, E.F. Hounsell, J.M. Rhodes. Direct demonstration of increased expression of Thomsen-Friedenreich (TF) antigen in colonic adenocarcinoma and ulcerative colitis mucin and its concealment in normal mucin. *J. Clin. Invest.* **1995**, *95*, 571.
- 13- G.F. Springer. T and Tn, general carcinoma autoantigens. *Science* **1984**, *224*, 1198.
- 14- W. Dippold, A. Steinborn, K.H.M. Büschenfelde. The role of the Thomsen-Friedenreich antigens a tumor-associated molecule. *Environ. Health Persp.* **1990**, *88*, 255.
- 15- A.M. Shamsuddin, G.T. Tyner, G.Y. Yang. Common expression of the tumour marker D-Galactose- $\beta$ -[1-3]-N-Acetyl-DGalactosamine by different adenocarcinomas: evidence of field effect phenomenon. *Cancer Res.* **1995**, *55*, 149.
- 16- J.S. Coon, R.S. Weinstein, J.L. Summers. Blood group precursor T-antigen expression in human urinary bladder carcinoma. *Am. J. Clin. Pathol.* **1982**, *77*, 692.
- 17- T. Janssen, M. Petein, R. Van Velthoven, P. Van Leer, M. Fourmarier, J.P. Vanegas, A. Danguy, C. Schulman, J.L. Pasteels, R. Kiss. Differential histochemical peanut agglutinin stain in benign and malignant human prostate tumors: relationship with prostatic specific antigen immunostain and nuclear DNA content. *Human Pathol.* **1996**, *27*, 1341.

- 18- K. Rittenhouse-Olson, J. Abdulla, J. Heimbürg-Molinaro, P. Sahoo, A. Almogren, R. Roy, R. Chang, S. Morey, J. Barchi. Carbohydrate tumor antigen vaccines using unique strategies. *J. Immunol.* **2010**, *184*, 131.
- 19- K. Rittenhouse-Diakun, K. Xia, D. Pickhardt, M.G. Baek, R. Roy. Development and characterization of monoclonal antibody to T-antigen: (Gal $\beta$ -(1-3)-GalNAc $\alpha$ -O). *Hybridoma* **1998**, *17*, 165.
- 20- J. Heimbürg-Molinaro, A. Almogren, S. Morey, O.V. Glinskii, R. Roy, G. E. Wilding, R. P. Cheng, V. V. Glinsky, K. Rittenhouse-Olson. Development, characterization, and immunotherapeutic utilization of peptide mimics of the Thomsen-Friedenreich carbohydrate antigen. *Neoplasia* **2009**, *11*, 780.
- 21- S.F. Slovin, G. Ragupathi, C. Musselli, C. Fernandez, M. Diani, D. Verbel, S. Danishefsky, P. Livingston, H. I. Scher. Thomsen-Friedenreich (TF) antigen as a target for prostate cancer vaccine: clinical trial results with TF cluster (c)-KLH plus QS21 conjugate vaccine in patients with biochemically relapsed prostate cancer. *Cancer Immunol Immunother.* **2005**, *54*, 694.
- 22- U. Philippe, H. Gemma, T. Kawe, L. Anja, S. Jens, K. Uwe, B. Michael, G. Steffen. Specific humoral immune response to the Thomsen-Friedenreich tumor antigen (CD176) in mice after vaccination with the commensal bacterium *Bacteroides ovatus* D-6. *Cancer Immunol Immunother.* **2013**, *62*, 875.
- 23- J. Heimbürg, J. Yan, S. Morey, O.V Glinskii, V.H Huxley, L. Wild, R. Klick, R. Roy, V.V. Glinsky, K.R. Olson. Inhibition of spontaneous breast cancer metastasis by

- anti-Thomsen-Friedenreich antigen monoclonal antibody JAA-F11. *Neoplasia* **2006**, *8*, 939.
- 24- A. Hoffman-Röder, A. Kaiser, S. Wagner, N. Gaidzik, D. Kowalczyk, U. Westerlind, B. Gerlitzki, E. Schmitt, H. Kunz. Synthetic antitumor vaccines from tetanus toxoid conjugates of MUC1 glycopeptides with the Thomsen–Friedenreich antigen and a fluorine-substituted analogue. *Ang. Chem.Int. Ed.*, **2010**, *49*, 8498.
- 25- Y.L. Huang, J.T. Hung, S.K.C. Cheung, H.Y. Lee, K.C. Chu, S.T. Li, Y.C. Lin, C.T. Ren, T.J.R. Cheng, T.L. Hsu, A.L. Yu, C.Y. Wu, C.H. Wong. Carbohydrate-based vaccines with a glycolipid adjuvant for breast cancer. *PNAS*. **2013**, *110*, 2517.
- 26- J. Zhu, J.D. Warren, S.J. Danishefsky. Synthetic carbohydrate-based anticancer vaccines: the memorial sloan-kettering experience. *Expert Rev Vaccines* **2009**, *8*, 1399.
- 27- S.J. Keding, S.J. Danishefsky. Prospects for total synthesis: a vision for a totally synthetic vaccine targeting epithelial tumors. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 11937.
- 28- V. Verez-Bencomo, V. Fernández-Santana, E. Hardy, M.E. Toledo, M.C. Rodríguez, L. Heynngnezz, A. Rodriguez, A. Baly, L. Herrera, M. Izquierdo, A. Villar, Y. Valdés, K. Cosme, M.L. Deler, M. Montane, E. Garcia, A. Ramos, A. Aguilar, E. Medina, G.Toraño, I. Sosa, Y. Carbonell, I. Hernandez, R. Martínez, A. Muzachio, A. Carmenates, L. Costa, F. Cardoso, C. Campa, M. Diaz, R. Roy. A Synthetic conjugate polysaccharide vaccine against *Haemophilus Influenzae* type b. *Science* **2004**, *305*, 522.
- 29- V.Verez Bencomo, R. Roy, M. C. Rodriguez, A. Villar, V. Fernandez-Santana, E. Garcia, Y. Valdes, L. Heynngnezz, I. Sosa, E. Medina. *ACS Symp. Ser.* **2007**, *989*, 71.

- 30- V. Fernández-Santana, F. Cardoso, A. Rodriguez, T. Carmenate, L. Peña, Y. Valdés, E. Hardy, F. Mawas, L. Heynngnezz, M. C. Rodríguez, I. Figueroa, J. Chang, M. E. Toledo, A. Musacchio, I. Hernández, M. Izquierdo, K. Cosme, R. Roy, V. Verez-Bencomo. Agicity and immunogenicity of a synthetic oligosaccharide protein conjugates against *Haemophilus influenzae* type b. *Infect. Immun.* **2004**, 72, 7115.
- 31- R. Roy, M. G. Baek, K. Rittenhouse-Olson. Synthesis of N,N'-bis(acrylamido)acetic acid base-T antigen glycodendrimers and their mouse monoclonal IgG antibody binding properties. *J. Am. Chem. Soc.* **2001**, 123, 1809.
- 32- M.-G.Baek, K. Rittenhouse-Olson, R. Roy. Synthesis and antibody binding properties of glycodendrimers bearing the tumor related T-antigen. *Chem Comm.* **2001**, 2001, 257.
- 33- M.-G. Baek, R. Roy. Glycodendrimers: novel glycotope isosteres unmasking sugar coding. Case study with T-Ag markers from breast cancer MUC1 glycoprotein. *Rev. Molec. Biotechnol.* **2002**, 90, 291.
- 34- M. Dubois, K. A. Gilles, J. K. Hamilton, P. A. Rebers, F. Smith. Colorimetric method for determination of sugars and related substances. *Anal. Chem.* **1956**, 28, 350.
- 35- T. Berggard, S. Linse, P. James. Methods for the detection and analysis of protein-protein interactions. *Proteomics* **2007**, 7, 2833.
- 36- W.L.L. Demian, F. Jahouh, D. Stansbury, E. Randell, R.J. Brown, J.H. Banoub. Characterizing changes in snow crab (*Chionoecetes opilio*) cryptocyanin protein during molting using matrix-assisted laser desorption/ionization mass spectrometry and tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2014**, 28, 355.

- 37- F. Jahouh, R. Saksena, D. Aiello, A. Napoli, G. Sindona, P. Kováč, J.H. Banoub. Glycation sites in neoglycoconjugates from the terminal monosaccharide antigen of the O-PS of *Vibrio cholerae* O1, serotype Ogawa, and BSA revealed by matrix-assisted laser desorption-ionization tandem mass spectrometry. *J. Mass Spectrom.* **2010**, *10*, 1148.
- 38- F. Jahouh, S. Hou, P. Kováč, J.H. Banoub. Determination of the glycation sites of *Bacillus anthracis* neoglycoconjugate vaccine by MALDI-TOF/TOF-CID-MS/MS and LC-ESI-QqTOF-tandem mass spectrometry. *J. Mass Spectrom.* **2011**, *46*, 993.
- 39- F. Jahouh, S. Hou, P. Kováč, J.H. Banoub. Determination of glycation sites by tandem mass spectrometry in a synthetic lactose-bovine serum albumin conjugate, a vaccine model prepared by dialkyl squarate chemistry. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1.
- 40- F. Jahouh, R. Saksena, P. Kováč, J.H. Banoub. Revealing the glycation sites in neoglycoconjugate models formed by conjugation of the antigenic monosaccharide hapten of *Vibrio cholerae* O1 serotype Ogawa with the BSA protein carrier using LC-ESI-QqTOF-MS/MS. *J. Mass Spectrom.* **2012**, *47*, 890.
- 41- F. Jahouh, P. Xu, W.F. Vann, P. Kováč J.H. Banoub. Mapping the glycation sites in the neoglycoconjugate from hexasaccharide antigen of *Vibrio cholerae*, serotype Ogawa and the recombinant tetanus toxin C-fragment carrier. *J. Mass Spectrom.* **2013**, *48*, 1083.
- 42- R.G. Reed, F.W. Putnam, T. Jr. Peters. Sequence of residues 400-403 of bovine serum albumin. *Biochem. J.* **1980**, *191*, 867.
- 43- J.F. Janatova, J. K.Fuller, M. Hunter. The heterogeneity of bovine albumin with respect to sulfhydryl and dimer content. *J. J. Biol. Chem.* **1968**, *243*, 3612.

- 44- K.L. Heredia, D. Bontempo, T. Ly, J.T. Byers, S. Halstenberg, H.D. Maynard. In situ preparation of protein-“smart” polymer conjugates with retention of bioactivity. *J. Am. Chem. Soc.* **2005**, *127*, 16955
- 45- J.R. Brown. Structure of bovine serum albumin. *Fed. Proc.* **1975**, *34*, 591.
- 46- J.E. Patterson, D. M. Geller. Bovine microsomal albumin: amino terminal sequence of bovine proalbumin. *Biochem. Biophys. Res. Commun.* **1977**, *74*, 1220.
- 47- K. Hirayama, S. Akashi, M. Furuya, K.I. Fukuhara. Rapid confirmation and revision of the primary structure of bovine serum albumin by ESIMS and FRIT-FAB LC/MS. *Biochem. Biophys. Res. Commun.* **1990**, *173*, 639.
- 48- R.T.A. McGillivray, D.W. Chung, E.W. Davie. Biosynthesis of bovine plasma proteins in a cell-free system. Amino-terminal sequence of preproalbumin. *Eur. J. Biochem.* **1979**, *98*, 477.
- 49- P.C. Ohe, R. Kuhne, R.U. Ebert, R. Altenburger. Structural alerts - a new classification model to discriminate excess toxicity from narcotic effect levels of organic compounds in the acute daphnid assay. *Chem. Res. Toxicol.* **2005**, *18*, 536.
- 50- D.C. Liebler. Protein damage by reactive electrophiles: targets and consequences. *Chem. Res. Toxicol.* **2008**, *21*, 117.
- 51- D.W. Roberts, T.W. Schultz, E.M. Wolf, A.O. Aptula. Experimental reactivity parameters for toxicity modeling: application to the acute aquatic toxicity of S2 electrophiles to *Tetrahymena pyriformis*. *Chem. Res. Toxicol.* **2010**, *23*, 228.
- 52- T.W. Schultz, J.W. Yarbrough, E.L. Johnson. Structure-activity relationships for reactivity of carbonyl-containing compounds with glutathione. *SAR and QSAR in Environ Res.* **2005**, *16*, 313.

- 53- T.W. Schultz, R.E. Carlson, M.T. Cronin. A conceptual framework for predicting the toxicity of reactive chemicals: modeling soft electrophilicity. *SAR and QSAR in Environ Res.* **2006**, *17*, 413.
- 54- I. Chipinda, J.M. Hettick, P.D. Siegel. Haptenation: chemical reactivity and protein binding. *J. Allergy.* **2011**, *2011*, 10.
- 55- W.I. Burkitt, A.E. Giannakopoulos, F. Sideeridou, S. Bashir, P.J. Derrick. Discrimination effects in MALDI-MS of mixtures of peptides- analysis of the proteome. *Austr. J. Chem.* **2003**, *56*, 369.
- 56- R. Kartazer, C. Eckerskorn, M. Karas, F. Lottspeich. Suppression effects in enzymatic peptide ladder sequencing using ultraviolet – matrix assisted laser desorption/ionization- mass spectrometry. *Electrophor.* **1998**, *19*, 1910.
- 57- B. Domon, C. Costello. Asystematic nomenclature of carbohydrate fragmentation in FAB-MS/MS spectra of glycoconjugates. *Glycoconj. J.* **1988**, *5*, 397.
- 58- P. Roepstorff, J. Fohlman. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **1984**, *11*, 601.

## Chapter IV: General Discussion and Conclusion

In the first part of this thesis, MALDI-MS was used for characterizing the changes of cryptocyanin protein from juvenile snow crabs during their molting and non-molting phases. It was demonstrated that the actual molecular masses of the non-molting and molting cryptocyanin proteins were different; these were 67.6 kDa and 68.1 kDa, respectively. In the second part of the thesis, the characterization of the exact glycation sites of a TF:BSA antigen was presented.

We have demonstrated for the first time in the literature that the actual molecular masses of the cryptocyanin protein of snow crab during the molting and non-molting phases were different. It was noticed that the molecular masses of cryptocyanin in other species were dissimilar. Accordingly, the molecular mass of blue crab (*Portunus pelagicus*) cryptocyanin was reported during the molting stage by using cDNA sequencing, and it was 50.1 kDa.<sup>[1]</sup> In addition, Terwilliger reported the molecular mass of Dungeness crab (*Metacarcinus magister*) cryptocyanin during its pre-molting stage by using cDNA sequencing, and it was 78.3 kDa.<sup>[2]</sup> The molecular mass of cryptocyanin from *Metacarcinus magister* was also previously assessed during development, molting, and physiological perturbations;<sup>[3]</sup> the cryptocyanin molecular mass varied from 72.6 to 77.3 kDa.<sup>[3]</sup> Unfortunately, no information about the cryptocyanin molecular mass from these other species of crab is available for a comparison of masses between molting and non-molting stages, thus it is difficult at this time to make a comparison of the differential findings with snow crab.

Our results showed that the MALDI-CID-MS/MS analyses of the molting and non-molting cryptocyanin after tryptic digestion had some similarities and staggering differences between the identified cryptocyanin peptides. Our explanation is that a part of non-molting cryptocyanin, which was determined to be 513 Da, was lost during the molting process, which allowed the cryptocyanin to adopt new functions that are associated with the hardening the exoskeleton. It would be of future interest to compare the structure of the modified cryptocyanin with unmodified cryptocyanin to understand how the new functions arise.

In the second study, it was observed that the majority of the glycation sites in the studied vaccine were located in the outer surface of the BSA. The importance of this discovery relies on the fact that the carbohydrate antigen is thus exposed and accessible to interact with specific receptors located on the surface of B cells. Three glycation sites - Lys 12, Cys 34, and Lys 535 - which were identified with TF:BSA ratio of 2:1, were completely accessible on to the surface of BSA. Nevertheless, only 11 glycation sites (Lys 4, Lys 12, Cys 34, Lys 106, Lys 116, Lys 127, Lys 179, Lys 186, Lys 210 Lys 522, and Lys 535) of 14 glycation sites with a TF:BSA ratio of 8:1 were accessible on the surface of BSA. Then, the three remaining glycation sites of the 14 glycation sites (Lys 312, Lys 412 and Lys 429) were not accessible, but they were glycated because the BSA structure became less compact by an increasing in the temperature to 40°C during the glycoconjugate synthesis with TF:BSA ratio of 8:1.

In both of the aforementioned studies, bottom up proteomics was important for identifying large biological molecules, namely crab cryptocyanin and the TF:BSA

antigen. The application of MS/MS sequencing to the large biomolecules is considered a bright innovation in the field of protein mass spectrometry.<sup>[4-6]</sup>

The two stages of bottom up proteomics, PMF and MS/MS ion searches, are done by the submission of the MS and MS/MS data of the digested protein to the Mascot library. While useful, the PMF method does have disadvantages, such that proteins with translational modifications resulting in mass shifts, partial proteolysis of proteins, and non-specific protein cleavages result in peptide sequences that cannot be identified using databases.<sup>[4-9]</sup> Both PMF and MS/MS ion searches were used to identify the change to cryptocyanin during molting and non-molting stages. Cryptocyanin of snow crab was a previously unknown protein, thus the data obtained with PMF was vital to depict the different forms of cryptocyanin during molting and non-molting. As a result of using the peptide mass fingerprint database, the sequences of the digested peptides during molting were matched with other types of cryptocyanin in the Mascot library, thus leading to the discovery of snow crab cryptocyanin.

For the data presented in Chapter III, PMF was not achieved in the determination of glycation sites of the glycoconjugate antigen. For determining the glycation sites of BSA in the glycoconjugate vaccine, PMF is unworkable because the BSA structure is altered by the conjugation of the carbohydrate antigen. However, a database search of ions obtained from the MS/MS spectra of glycated and non-glycated BSA was performed to elucidate the peptide sequences. The total sequence coverage for the free BSA, which is an important parameter to be considered in the proteomics field, was found to be in the range of higher than 90 %. However, the total sequence coverage for the hapten conjugated BSA was decreased to less than 50 %. The decrease in the total sequence

coverage reflects that some of the amino acid residues were mutated by the hapten conjugation, thus their peptides were recorded as non-matched peptides. Ultimately, the glycation sites were determined from the non-matched peptides using the Mascot library.

The proper choice of ionization source and analyzer type will depend on the nature of the sample.<sup>[10-14]</sup> Accordingly, MALDI-QqTOF-MS was selected for the determination of changes in snow crab cryptocyanin protein because MALDI is well-suited and highly sensitive for large biomolecules, and it additionally has a high tolerance for the salts and contaminants.<sup>[12-18]</sup> Moreover, the MALDI spectra are easy to interpret because they contain singly charged ions.<sup>[12,15]</sup> It was reported that MALDI has a suppression effect on the number of glycation sites during their determination in the glycoconjugates.<sup>[8,19]</sup> For example, a recent study showed that the number of glycation sites in a glycoconjugate vaccine was 5 when determined by MALDI-CID-MS/MS, but using ESI-CID-MS/MS led to the identity of 30 glycation sites in the same vaccine. Accordingly, ESI was nominated as the ionization method for the determination of glycation sites in the glycoconjugate vaccine.

*De novo* sequencing of the glycated peptides of the glycoconjugate vaccine was achieved by using LC coupled with ESI-MS, and low-energy ESI-CID-MS/MS analysis with a QqTOF-MS/MS hybrid instrument. LC-ESI is a qualified technique for determining glycation sites, since the use of LC offers a separation method for digested peptides in a solution phase.<sup>[12]</sup> ESI-QqTOF-MS was extensively used in previous studies for determining glycation sites of glycoproteins.<sup>[9-12]</sup> Additionally, ESI-FTICR-MS and

ESI-QqTOF-MS were also reported for mapping and sequencing the *O*-glycosylated peptides.<sup>[11]</sup>

Ultimately, the bottom up proteomics approach is based on database homology searches, so only matched sequences of proteins are recognized in the databases. Accordingly, a complete sequence of cryptocyanin was not achieved. Differentiation between two isomers, such as the amino acids leucine and isoleucine, can also be problematic in determining a peptide sequence, especially in the case of peptides with non-matched sequences in the homology database. Furthermore, the MS analysis of glycoproteins is a complex task because the resulted spectra contain a structural mixture of carbohydrate and protein. A spectrum includes peptide sequences (b and y ions), carbohydrate fragments (B and Y ions), and spacer linked to peptide or carbohydrate fragments.

#### **4.1 Future aspects**

Cryptocyanin has been sequenced by MS/MS and a signature peptide can be chosen to allow for the quantification of the protein. From the theoretical point of view, the signature peptide sequence should be free of Met and Cys, as these may be modified during the digestion of the protein by either the oxidation of the Met and/or carbamidomethylation of the Cys residue.<sup>[20]</sup> As well, the empirical and technical properties of a signature peptide are based on the signal intensity of the peptide ions during the various mass spectrometric analyses,<sup>[20]</sup> thus peptides with strong signal intensities would be preferred.

Future work should include the study of other crab species, such as *Portunus pelagicus* and *Metacarcinus magister*, to compare cryptocyanin protein and RNA sequences between molting and non-molting stages. Furthermore, because of the 513 Da difference in snow crab cryptocyanin mass between molting and non-molting, additional work with snow crab will be necessary, to identify whether the difference exists at the RNA level or just the protein level.

The use of the methods in Chapter III to identify glycation sites could be applied to future studies of other synthetic carbohydrate-based antigens that use more immunogenic protein carriers (instead of BSA), such as tetanus toxoid and KLH. These different forms of antigens could be administrated *in vivo* to test their effectiveness. Additionally, different ion fragmentation approaches could be used for mapping the glycation sites of glycoconjugates, such as ECD or ETD (electron transfer dissociation), which confer the fragmentation of the peptide backbone and allow top down proteomics to be performed on the intact proteins.

## **4.2 Conclusion**

In this thesis, snow crab cryptocyanin protein was isolated by SDS-PAGE and analyzed using a MALDI-TOF/TOF-MS instrument. The purified cryptocyanin protein was sequenced using a ‘bottom up’ approach, and it was found that two different sequences for cryptocyanin exist during molting and non-molting. Furthermore, the exact glycation sites of the TF-BSA antigen were determined by using an ESI-QqTOF-MS/MS hybrid instrument. The work of my thesis clearly shows the power of MS and how it can be used to sequence large biomolecules.

## References

1. A.V. Kuballa, D.J. Merritt, A. Elizur. Gene expression profiling of cuticular proteins across the moult cycle of the crab *Portunus pelagicus*. *BMC Biol.* **2007**, *5*, 45.
2. N.B. Terwilliger, M.C. Ryan. Functional and phylogenetic analyses of phenoloxidases from Brachyuran (*Cancer magister*) and Branchiopod (*Artemia franciscana*, *Triops longicaudatus*) Crustaceans. *Biol. Bull.* **2006**, *210*, 38.
3. N.B. Terwilliger, M. Ryan, M.R Phillips. Crustacean hemocyanin gene family and microarray studies of expression change during eco-physiological stress. *Integr. Comp. Biol.* **2006**, *46*, 991.
4. R. Aebersold, M. Mann. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198.
5. D.A. Wolters, M.P. Washburn, J.R. Yates. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **2001**, *73*, 5683.
6. J.R. Yates, C.I. Ruse, A. Nakorchevsky. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **2009**, *11*, 49.
7. F. Jahouh, R. Saksena, D. Aiello, A. Napoli, G. Sindona, P. Kovác, J.H. Banoub. Glycation sites in neoglycoconjugates from the terminal monosaccharide antigen of the O-PS of *Vibrio cholerae* O1, serotype Ogawa, and BSA revealed by matrix-assisted laser desorption-ionization tandem mass spectrometry. *J. Mass Spectrom.* **2010**, *10*, 1148.

8. F. Jahouh, S. Hou, P. Kováč, J.H. Banoub. Determination of the glycation sites of *Bacillus anthracis* neoglycoconjugate vaccine by MALDI-TOF/TOF-CID-MS/MS and LC-ESI-QqTOF-tandem mass spectrometry. *J. Mass Spectrom.* **2011**, *46*, 993.
9. F. Jahouh, S. Hou, P. Kováč, J.H. Banoub. Determination of glycation sites by tandem mass spectrometry in a synthetic lactose-bovine serum albumin conjugate, a vaccine model prepared by dialkyl squarate chemistry. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1.
10. F. Jahouh, P. Xu, W.F. Vann, P. Kováč, J.H. Banoub. Mapping the glycation sites in the neoglycoconjugate from hexasaccharide antigen of *Vibrio cholerae*, serotype Ogawa and the recombinant tetanus toxin C-fragment carrier. *J. Mass Spectrom.* **2013**, *48*, 1083.
11. J. Peter-Katalinić. Methods in enzymology: *O*-glycosylation of proteins. *Methods Enzymol.* **2005**, *405*, 139
12. K. Downard. Mass spectrometry: a foundation course, TJ International Ltd, Cornwall, **2004**.
13. M.A. Baldwin. Mass spectrometers for the analysis of biomolecules. *Methods Enzymol.* **2005**, *402*, 3.
14. F.W. Aston. Mass Spectra and isotopes, Longmans, Green and Co, New York **1942**.
15. R.C. Beavis, B.T. Chait. Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Commun. Mass Spectrom.* **1989**, *3*, 432.

16. R.C. Beavis, B.T. Chait.  $\alpha$ -cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry. *Org. Mass Spectrom.* **1992**, 27, 156.
17. R.C. Beavis, B.T. Chait. Velocity distributions of intact high mass polypeptide molecule ions produced by matrix assisted laser desorption. *Chem. Phys. Lett.* **1991**, 181, 479.
18. R. Knochenmuss, R. Zenobi. MALDI ionization: the role of in-plume processes. *Chem. Rev.* **2003**, 103, 441.
19. R. Kartazer, C. Eckerskorn, M. Karas, F. Lottspeich. Suppression effects in enzymatic peptide ladder sequencing using ultraviolet – matrix assisted laser desorption/ionization- mass spectrometry. *Electrophor.* **1998**, 19, 1910.
20. M. Cohen, A.A. Mansour, J.H. Banoub. Absolute quantification of Atlantic salmon and rainbow trout vitellogenin by the 'signature peptide' approach using electrospray ionization QqTOF tandem mass spectrometry. *J. Mass Spectrom.* **2006**, 4, 646.

## **APPENDIX A**

### **Additional tables for**

**CHAPTER II:** Differentiation between the crustacean cryptocyanin protein during molting and non-molting processes of the snow crab (*Chionoecetes opilio*) using matrix-assisted laser desorption/ionization mass spectrometry and tandem mass spectrometry

**Table A.1** Identified product ions obtained by CID-MS/MS of the  $[M+H]^+$  ion of the peptide HAWAVATNKR at  $m/z$  1082.5324

<b>Product ion</b>	<b>Observed <i>m/z</i></b>	<b>Calculated <i>m/z</i></b>	<b>Deviation Da</b>
$b_8^+$	908.4719	908.4743	-0.0024
$b_7^+$	780.3789	780.3793	-0.0004
$y_7^+$	759.8805	759.8735	0.0070
$b_6^+$	666.3303	666.3364	-0.0061
$b_5^+$	565.3019	565.2887	0.0132
$b_4^+$	494.2581	494.2516	0.0065
$b_3^+$	395.1860	395.1832	0.0028
$b_2^+$	324.1467	324.1461	0.0006
$y_1^+$	175.1232	175.1184	0.0048

**Table A.2** Identified product ions obtained by CID-MS/MS of the  $[M+H]^+$  ion obtained from the peptide **FGPPFPVRN** at  $m/z$  916.5099

<b>Product ion</b>	<b>Observed <i>m/z</i></b>	<b>Calculated <i>m/z</i></b>	<b>Deviation Da</b>
$y_7^+$	769.4282	769.4350	-0.0068
$y_6^+$	712.4028	712.4135	-0.0107
$y_5^+$	615.3532	615.3908	-0.0376
$b_5^+$	546.2737	546.2716	0.0021
$y_4^+$	518.2790	518.3080	-0.0290
$b_4^+$	399.2076	399.2032	0.0044
$y_3^+$	371.2279	371.2396	-0.0117
$b_3^+$	302.1474	302.1505	-0.0031
$y_1^+$	175.1242	175.1184	0.0058

**Table A.3** Identified product ions obtained by CID-MS/MS of the  $[M+H]^+$  ion obtained from the peptide **DGNGAIIPFDEGR** at  $m/z$  1360.6744

<b>Product ion</b>	<b>Observed <i>m/z</i></b>	<b>Calculated <i>m/z</i></b>	<b>Deviation Da</b>
$y_{12}^+$	1245.6289	1245.6217	0.0072
$y_{10}^+$	1074.5231	1074.5573	-0.0342
$y_9^+$	1017.5388	1017.5358	0.0030
$y_7^+$	833.4111	833.4146	-0.0035
$y_6^+$	720.3352	720.3306	0.0046
$y_5^+$	623.2765	623.2778	-0.0013
$y_4^+$	476.2148	476.2094	0.0054
$y_3^+$	361.1842	361.1825	0.0017
$y_2^+$	232.1455	232.1399	0.0056
$y_1^+$	175.1251	175.1184	0.0067

**Table A.4** Identified product ions obtained by CID-MS/MS of the  $[M+H]^+$  ion obtained from the peptide **GVQPKRPF $\text{G}$ YPLDRR** at  $m/z$  1744.8554

<b>Product ion</b>	<b>Observed <i>m/z</i></b>	<b>Calculated <i>m/z</i></b>	<b>Deviation <b>Da</b></b>
$y_{13}^+$	1588.7451	1588.8225	-0.0774
$b_{14}^+$	1570.7979	1570.8018	-0.0039
$b_{13}^+$	1455.7718	1455.7749	-0.0031
$y_{10}^+$	1248.6251	1248.6842	-0.0591
$b_{11}^+$	1245.6323	1245.6380	-0.0057
$b_{10}^+$	1082.5711	1082.5747	-0.0036
$b_9^+$	1025.5579	1025.5532	0.0047
$y_8^+$	964.2069	964.4881	-0.2812
$b_7^+$	781.4403	781.4321	0.0082
$y_5^+$	663.3499	663.3455	0.0044
$y_4^+$	500.3033	500.2822	0.0211
$b_4^+$	382.2101	382.2090	0.0011
$b_3^+$	285.1502	285.1563	-0.0061
$y_1^+$	175.1171	175.1184	-0.0013
$b_2^+$	157.1012	157.0977	0.0035

**Table A.5** Identified product ions obtained by CID-MS/MS of the  $[M+H]^+$  ion obtained from the peptide **FDAER** at  $m/z$  637.2853

<b>Product ion</b>	<b>Observed <i>m/z</i></b>	<b>Calculated <i>m/z</i></b>	<b>Deviation Da</b>
$y_3^+$	375.2041	375.1981	0.0060
$b_3^+$	334.1609	334.1403	0.0206
$b_2^+$	263.0970	263.1032	-0.0062
$y_1^+$	175.1250	175.1184	0.0066

**Table A.6** Identified product ions obtained by CID-MS/MS of the  $[M+H]^+$  ion obtained from the peptide **DPAFFR** at  $m/z$  752.3832

<b>Product ion</b>	<b>Observed <i>m/z</i></b>	<b>Calculated <i>m/z</i></b>	<b>Deviation Da</b>
$y_5^+$	637.3498	637.3451	0.0047
$b_4^+$	431.1894	431.1931	-0.0037
$b_3^+$	284.1283	284.1246	0.0037
$b_2^+$	213.0894	213.0875	0.0019
$y_1^+$	175.1205	175.1184	0.0021

**Table A.7** Identified product ions obtained by CID-MS/MS of the  $[M+H]^+$  ion obtained from the peptide **YMDNIFR** at  $m/z$  958.4520

<b>Product ion</b>	<b>Observed <i>m/z</i></b>	<b>Calculated <i>m/z</i></b>	<b>Deviation <b>Da</b></b>
$b_5^+$	637.3027	637.2656	0.0371
$y_4^+$	549.3283	549.3138	0.0145
$b_4^+$	524.1718	524.1816	-0.0098
$b_3^+$	410.1285	410.1386	-0.0101
$b_2^+$	295.1124	295.1116	0.0008
$y_1^+$	175.1248	175.1184	0.0064

**Table A.8** Identified product ions obtained by CID-MS/MS of the  $[M+H]^+$  ion obtained from the peptide **RPHGYPLDR** at  $m/z$  1110.5552

<b>Product ion</b>	<b>Observed <i>m/z</i></b>	<b>Calculated <i>m/z</i></b>	<b>Deviation Da</b>
$y_8^+$	954.5277	954.4786	0.0491
$b_8^+$	936.4587	936.4692	-0.0105
$b_7^+$	821.4485	821.4422	0.0063
$y_6^+$	720.3956	720.3670	0.0286
$b_5^+$	611.3155	611.3054	0.0101
$y_4^+$	500.2613	500.2822	-0.0209
$b_4^+$	448.2439	448.2421	0.0018
$y_1^+$	175.1232	175.1184	0.0048

**Table A.9** Identified product ions obtained by CID-MS/MS of the  $[M+H]^+$  ion obtained from the peptide **LNHEEFSYK** at  $m/z$  1166.5525

<b>Product ion</b>	<b>Observed <i>m/z</i></b>	<b>Calculated <i>m/z</i></b>	<b>Deviation <b>Da</b></b>
$y_8^+$	1053.4593	1053.4631	-0.0038
$b_8^+$	1020.4508	1020.4427	0.0081
$y_7^+$	939.4425	939.4201	0.0224
$b_7^+$	857.3710	857.3794	-0.0084
$y_6^+$	802.3564	802.3612	-0.0048
$b_6^+$	770.3943	770.3473	0.0470
$y_5^+$	673.3194	673.3186	0.0008
$b_5^+$	623.2894	623.2789	0.0105
$b_4^+$	494.2407	494.2363	0.0044
$b_3^+$	365.1912	365.1937	-0.0025

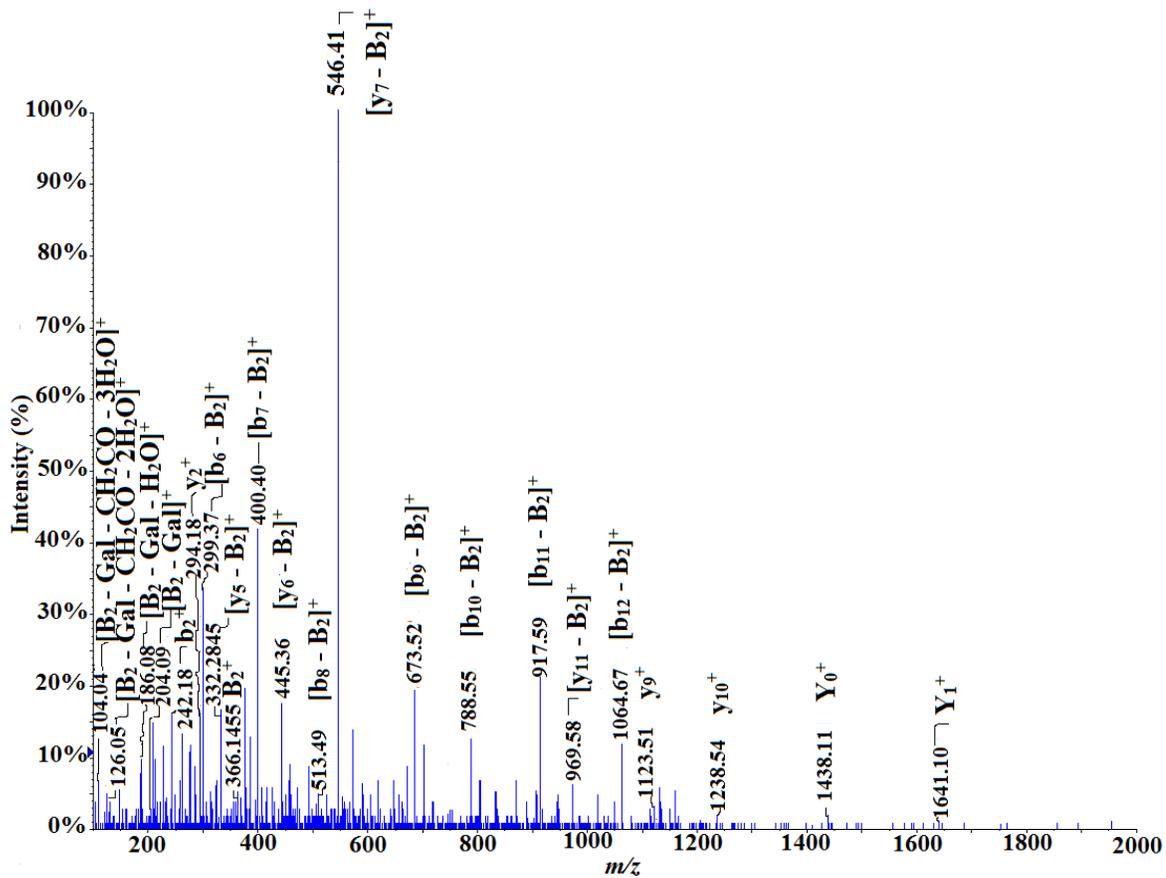
**Table A.10** Identified product ions obtained by CID-MS/MS of the  $[M+H]^+$  ion obtained from the peptide **VYEDIRDPHLK** at  $m/z$  1384.7427

<b>Product ion</b>	<b>Observed <i>m/z</i></b>	<b>Calculated <i>m/z</i></b>	<b>Deviation <b>Da</b></b>
$y_8^+$	993.5187	993.5471	-0.0284
$b_7^+$	891.3884	891.4212	-0.0328
$y_7^+$	878.5179	878.5201	-0.0022
$b_6^+$	776.4132	776.3943	0.0189
$y_6^+$	765.4312	765.4360	-0.0048
$y_5^+$	609.3241	609.3349	-0.0108
$b_4^+$	507.2185	507.2091	0.0094
$y_4^+$	494.2976	494.3080	-0.0104
$y_3^+$	397.2605	397.2552	0.0053

## **APPENDIX B**

### **Additional Figure and table for**

**CHAPTER III:** Direct targeted glycation of the free sulfhydryl group of cysteine residue (Cys 34) of BSA. Mapping of the glycation sites of the anti-tumor Thomsen-Friedenreich neoglycoconjugate vaccine prepared by Michael addition reaction.



**Figure B.1** CID-MS/MS spectrum of identified glycosylated peptide FK\*DLGEEHFK 4 (Lys12) at  $m/z$  902.12(+2)

**Table B.1** Product ions obtained following the CID-MS/MS analysis of FK\*DLGEEHFK 4 (Lys12) ion at  $m/z$  902.12(+2)

<b>Product ion</b>	<b>Calculated <math>m/z</math></b>	<b>Observed <math>m/z</math></b>	<b>Error</b>
$Y_1^+$	1641.16	1641.10	0.06
$Y_0^+$	1438.17	1438.11	-0.06
$y_{10}^+$	1238.53	1238.54	-0.01
$y_9^+$	1123.50	1123.51	-0.01
$[b_{12} - B_2]^+$	1064.67	1064.67	0.00
$[y_{11} - B_2]^+$	969.58	969.58	0.00
$[b_{11} - B_2]^+$	917.59	917.59	0.00
$[b_{10} - B_2]^+$	788.55	788.55	0.00
$[b_9 - B_2]^+$	673.52	673.52	0.00
$[y_7 - B_2]^+$	546.41	546.41	0.00
$[b_8 - B_2]^+$	513.49	513.49	0.00
$[y_6 - B_2]^+$	445.36	445.36	0.00
$[b_7 - B_2]^+$	400.40	400.40	0.00
$B_2^+$	366.14	366.14	0.00
$[y_5 - B_2]^+$	332.28	332.28	0.00
$[b_6 - B_2]^+$	299.36	299.37	-0.01
$y_2^+$	294.18	294.18	0.00
$b_2^+$	242.18	242.18	0.00
$[B_2 - Gal]^+$	204.08	204.09	-0.01
$[B_2 - Gal - H_2O]^+$	186.07	186.08	-0.01
$[B_2 - Gal - CH_2CO - 2H_2O]^+$	126.05	126.05	0.00
$[B_2 - Gal - CH_2CO - 3H_2O]^+$	104.04	104.04	0.00